Advances in Multimedia

Edited by:

Jovan Pehcevski

ARCLER PRESS

Advances in Multimedia

Advances in Multimedia

Edited by:

Jovan Pehcevski

ARCLER
PRESS

www.arclerpress.com

Advances in Multimedia

Jovan Pehcevski

Arcler Press
224 Shoreacres Road
Burlington, ON L7L 2H2
Canada
www.arclerpress.com
Email: orders@arclereducation.com

e-book Edition 2022

ISBN: 978-1-77469-353-7 (e-book)

This book contains information obtained from highly regarded resources. Reprinted material sources are indicated. Copyright for individual articles remains with the authors as indicated and published under Creative Commons License. A Wide variety of references are listed. Reasonable efforts have been made to publish reliable data and views articulated in the chapters are those of the individual contributors, and not necessarily those of the editors or publishers. Editors or publishers are not responsible for the accuracy of the information in the published chapters or consequences of their use. The publisher assumes no responsibility for any damage or grievance to the persons or property arising out of the use of any materials, instructions, methods or thoughts in the book. The editors and the publisher have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission has not been obtained. If any copyright holder has not been acknowledged, please write to us so we may rectify.

Notice: Registered trademark of products or corporate names are used only for explanation and identification without intent of infringement.

© 2022 Arcler Press

ISBN: 978-1-77469-174-8 (Hardcover)

Arcler Press publishes wide variety of books and eBooks. For more information about Arcler Press and its products, visit our website at www.arclerpress.com

DECLARATION

Some content or chapters in this book are open access copyright free published research work, which is published under Creative Commons License and are indicated with the citation. We are thankful to the publishers and authors of the content and chapters as without them this book wouldn’t have been possible.

ABOUT THE EDITOR

Jovan obtained his PhD in Computer Science from RMIT University in Melbourne, Australia in 2007. His research interests include big data, business intelligence and predictive analytics, data and information science, information retrieval, XML, web services and service-oriented architectures, and relational and NoSQL database systems. He has published over 30 journal and conference papers and he also serves as a journal and conference reviewer. He is currently working as a Dean and Associate Professor at European University in Skopje, Macedonia.

TABLE OF CONTENTS

Section 2: Multimedia applications in Health and Medicine

Section 3: Multimedia transmission in Wireless Networks

Section 4: Multimedia applications in Education

LIST OF CONTRIBUTORS

Xiaodong Liu and Miao Wang
School of Computing Henan University of Engineering, Zhengzhou, China
Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, China

Chen Zhang, Bin Hu, Yucong Suo, Zhiqiang Zou and Yimu Ji
College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China
College of Geographic Science, Nanjing Normal University, Nanjing, China
Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Ministry of Education, Nanjing, China
Bell Honor School, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China
Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing, Jiangsu, China

Fei Yang
School of Computer Science and Technology, Shandong University, Jinan 250101, China
School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, 264209, China

Xiangxu Meng
School of Computer Science and Technology, Shandong University, Jinan 250101, China

JiYing Lang
School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, 264209, China

Weigang Lu
Department of Educational Technology, Ocean University of China, Qingdao, 266100, China

Lei Liu
The Institute of Acoustics of the Chinese Academy of Sciences, Beijing, 100190, China

Yunfei Han
The Xinjiang Technical Institute of Physics & Chemistry, Urumqi 830011, China
Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
University of Chinese Academy of Sciences, Beijing 100049, China

Tonghai Jiang
The Xinjiang Technical Institute of Physics & Chemistry, Urumqi 830011, China
Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
University of Chinese Academy of Sciences, Beijing 100049, China

Yupeng Ma
The Xinjiang Technical Institute of Physics & Chemistry, Urumqi 830011, China
Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
University of Chinese Academy of Sciences, Beijing 100049, China

Chunxiang Xu
The Xinjiang Technical Institute of Physics & Chemistry, Urumqi 830011, China
Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
University of Chinese Academy of Sciences, Beijing 100049, China

Chih-Chen Chen
Department of Management Information System, Hwa Hsia Institute of Technology, Taipei, Chinese Taipei

De-Jou Hong
Department of Computer Science, National Taipei University of Education, Taipei, Chinese Taipei

Shih-Ching Chen
Department of Physical Medicine & Rehabilitation, Taipei Medical University Hospital, Taipei, Chinese Taipei

Ying-Ying Shih
Department Physical & Rehabilitation Medicine, Chang Gung Memorial Hospital, Tao-Yuan, Chinese Taipei

Yu-Luen Chen
Department of Computer Science, National Taipei University of Education, Taipei, Chinese Taipei
Department of Information Management, St. Mary’s Medicine, Nursing and Management College, Yilan, Chinese Taipei

Mai Xin
Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Zhifeng Ye
Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Changhua Chen
Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Yuhan Cai
Sichuan Hangbiao Electric Power Construction Co., Ltd., Meishan, China

Haili Ding
Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Ran Wang
Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Shiqian Wang
Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Gali Dar
Physical Therapy Department, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel
Ribstein Center for Research and Sports Medicine, Wingate Institute, Netanya, Israel

Yaron Marx
Maccabi Health Care Services, Afula and Tiberias, Israel

Emma Ioffe
Maccabi Health Care Services, Grand Canyon, Haifa, Israel

Einat Kodesh
Physical Therapy Department, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel

Christiane Eichenberg and
University of Cologne, Germany

Carolin Wolters
University of Cologne, Germany

Pushpender Kumar and
Department of Computer Science and Engineering, National Institute of Technology, Hamirpur, India

Narottam Chand
Department of Computer Science and Engineering, National Institute of Technology, Hamirpur, India

Ahmed Riadh Rebai
Wireless Research Group, Texas A&M University at Qatar, Qatar

Mariam Fliss
Wireless Research Group, Texas A&M University at Qatar, Qatar

Ahmed Riadh Rebai
Texas A&M University at Qatar – Doha, Qatar

Mariam Fliss
Texas A&M University at Qatar – Doha, Qatar

Saïd Hanafi
University of Valenciennes et du Hainaut-Cambrésis, France

K. L. Eddie Law
Kirin Cloud Solutions, Ltd Hong Kong

Jacek Ilow
Dalhousie University Canada

Atef F. Mashagbh
Al al-Bayt University, Mafraq, Jordan

Rosseni Din
UniversitiKebangsaan Malaysia, Selangor, Malaysia

M. Khalid M. Nasir
UniversitiKebangsaan Malaysia, Selangor, Malaysia

Lilia Halim
UniversitiKebangsaan Malaysia, Selangor, Malaysia

Rania Ahmad Al-Batainah
UniversitiKebangsaan Malaysia, Selangor, Malaysia

André Koscianski
UTFPR, Ponta Grossa, Brazil

Denise do Carmo Farago Zanotto
UTFPR, Ponta Grossa, Brazil

Prince Hycy Bull
North Carolina Central University, Durham, USA

Jorge Montalvo
University of Lima, Scientific Research Institute –IDIC Peru

Marina Milovanović
Faculty of Real Estate Management, Union University, Belgrade, Serbia

Đurđica Takači
Faculty of Natural Sciences, University of Novi Sad, Novi Sad, Serbia

Aleksandar Milajić
Faculty of Management in Civil Engeneering, Union University, Serbia

LIST OF ABBREVIATIONS

APs	Access Points
AQ	Acrophobia Questionnaire
APFH	Adaptive Preemptive Fast Handoff
APIs	Application programming interfaces
ATHQ	Attitudes Towards Heights Questionnaire
AR	Augmented Reality
BSS	Basic service set
BAT	Behavioral Avoidance Test
BEB	Binary Exponential Backoff
BER	Bit Error Rate
CCTV	Closed-circuit television
CH	Cluster head
CTML	Cognitive Theory of Multimedia Learning
CC/PP	Composite Capability/Preference Profile
CAVE	Computer animated virtual environment
CT	Computer tomography
CW	Congestion Window
CBIR	Content-based image retrieval
CAAN	Context-aware attention network
CNNs	Convolutional neural networks
CLAA	Cross-Layer Adaptive Approach
CLLA	Cross-layer link adaptation
DFSE	Deep Feature Spatial Encoding
DCT	Discrete Cosine Transform
DHTs	Distributed Hash Tables
EA	Egress Agent
ESS	Extended service set
XML	Extensible Markup Language

FABQ	Fear-Avoidance Beliefs Questionnaire
FOTO	Focus on Therapeutic Outcomes
HMD	Head-mounted display
HEIV	Human emotion in the video
ITU	International Telecommunication Union
IETF	Internet Engineering Task Force
LBTT	Local Binary Temporal Tracking
MOS	Mean Opinion Score
MAC	Medium Access Control
MR	Mixed Realities
MSs	Mobile Stations
MLP	Multilayer perceptron
MAF	Multimedia Agent Framework
MI	Multimedia Instruction
MMS	Multimedia messaging services
NSFC	National Natural Science Foundation of China
OCD	Obsessive-compulsive disorder
OIMI	Online Individualized Multimedia Instruction
PSNR	Peak signal-to-noise ratio
PESQ	Perceptual Evaluation of Speech Quality
PTSD	Posttraumatic stress disorder
PSM	Power-saving mode
PSHP	Prevent Scan Handoff Procedure
PCA	Principal component analysis
QoE	Quality of Experience
QoS	Quality of Service
RCT	Randomized controlled trial
RTT	Round Trip Time
SLA	Service Level Agreement
SMS	Short message service
SNR	Signal to Noise Ratio
SGP	Spectral graph partitioning
SGP	Spectral Graph Partitioning
SVM	Support vector machine

SOAQ	Symmetry, Ordering and Arrangement Questionnaire
TML	Theory of Meaningful Learning
TLI	Tucker-Lewis coefficient
VQM	Video Quality Measurement
VR	Virtual reality
VDP	Visual Differences Predictor
VCA	Voting-based clustering algorithm
WAMI	Wide Area Motion Imagery
WLANs	Wireless local area networks

PREFACE

Multimedia uses multiple forms of information content and processing to inform or entertain the user, or the audience. Generally speaking, we can say that multimedia refers to the use of electronic media for storage and processing of various data types. We experience the multimedia content presented by text, images, videos, animation, audio, and other media types, which stimulate our senses.

There are different definitions of the term multimedia, depending on the context of use. One of the accepted definitions of multimedia is as follows: Multimedia is an area that integrates text, graphics, drawings, images and videos, animation, audio and other media, where any type of information can be displayed, stored, transmitted and processed in digital form. Despite this definition of multimedia as a scientific field, the word multimedia is also used to denote plain multimedia content, which represents the convergence of text, images, video, animation and sound (elements of multimedia) into a single form. The power of multimedia lies in the way this information is interconnected.

Multimedia applications in addition to multimedia content also include the processes that execute inside them (during playback). The classification of multimedia content can be done in different ways. One of them classifies multimedia content into static and time-dependent. Text and images form static multimedia content, while audio, video and animations change over time, and the multimedia content in which they are included is called time-dependent.

By multimedia structure we mean the form in which multimedia content appears. Multimedia structures are divided in two basic categories: linear and nonlinear. An example of a linear multimedia structure is a film without navigation controls. Nonlinear structures often use interactivity to control the flow, e.g. in video games or computer-assisted learning. Interactivity in multimedia means the ability to display or broadcast multimedia content according to user instructions.

This edition covers several recent advances in multimedia, including: machine learning and AI methods in multimedia research, multimedia applications in health and medicine, multimedia transmission in wireless networks and multimedia applications in education.

Section 1 focuses on machine learning and AI methods in multimedia research, describing context-aware attention network for human emotion recognition in video, large-scale video retrieval via deep local convolutional features, region space guided transfer function design for nonlinear neural network augmented image visualization, data-driven methods for image and video understanding, and pretraining convolutional neural networks for image-based vehicle classification.

Section 2 focuses on multimedia applications in health and medicine, describing study of multimedia technology in posture training for the elderly, rapid extraction of target information in messy multimedia medical data, effectiveness of a multimedia messaging service reminder system in the management of knee osteoarthritis, and virtual realities in the treatment of mental disorders: a review of the current state of research.

Section 3 focuses on multimedia transmission in wireless networks, describing clustering in wireless multimedia sensor networks, a dynamic link adaptation for multimedia quality-based communications in IEEE_802.11 wireless networks, multimedia and VoIP-oriented cell search technique for the IEEE 802.11 WLANS, and ubiquitous control framework for delivering perceptual satisfaction of multimedia traffic.

Section 4 focuses on multimedia applications in education, describing online individualized multimedia instruction instrument for engineering communication skills, design model for educational multimedia software, cognitive constructivist theory of multimedia, design of videogame based on inca abacus, and multimedia approach in teaching mathematics through examples of interactive lessons from mathematical analysis and geometry.

SECTION 1:

MACHINE LEARNING AND AI METHODS IN MULTIMEDIA

CHAPTER 1

CONTEXT-AWARE ATTENTION NETWORK FOR HUMAN EMOTION RECOGNITION IN VIDEO

Xiaodong Liu[1,2] and Miao Wang[1]

[1]School of Computing Henan University of Engineering, Zhengzhou, China
[2]Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, China

ABSTRACT

Recognition of human emotion from facial expression is affected by distortions of pictorial quality and facial pose, which is often ignored by traditional video emotion recognition methods. On the other hand, context information can also provide different degrees of extra clues, which can further improve the recognition accuracy. In this paper, we first build a video dataset with seven categories of human emotion, named human emotion in the video (HEIV). With the HEIV dataset, we trained a Context-aware

Citation: Xiaodong Liu, Miao Wang, "Context-Aware Attention Network for Human Emotion Recognition in Video", Advances in Multimedia, vol. 2020, Article ID 8843413, 10 pages, 2020. https://doi.org/10.1155/2020/8843413.

Copyright: © 2020 by Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

attention network (CAAN) to recognize human emotion. The network consists of two subnetworks to process both face and context information. Features from facial expression and context clues are fused to represent the emotion of video frames, which will be then passed through an attention network and generate emotion scores. Then, the emotion features of all frames will be aggregated according to their emotional score. Experimental results show that our proposed method is effective on HEIV dataset.

INTRODUCTION

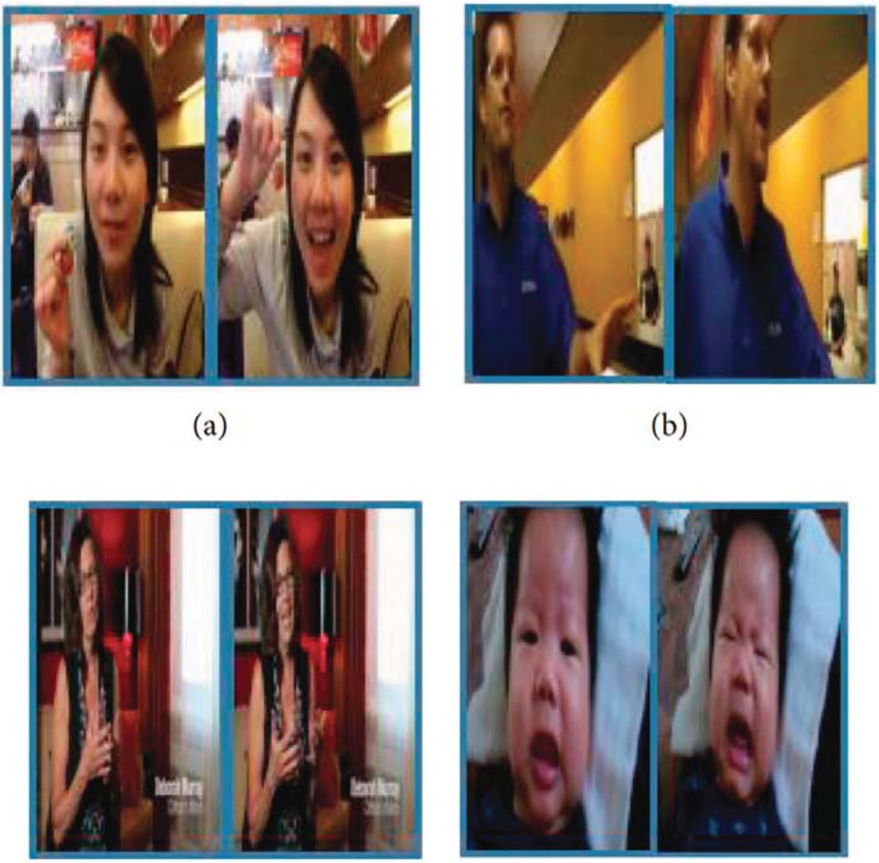
Estimating a person's emotional state is essential in our everyday life. This capacity is necessary to perceive and anticipate people's reactions [1]. Particularly, this emotion recognition challenge has a wide range of applications. For example, the emotional recognition platform can be used to recognize a potential suspicious person on intelligent security. Video recommendation services can match users' interest with video emotion, and the government sector can better understand people's response to hot events or new policies. Thus, human emotion recognition has attracted more and more attention as a new research field.

The human face contains rich emotional clues. Chu et al. [2] proposed a human emotion recognition method based on facial action coding system, which encodes facial expressions through a series of specific location movements of the face (action units). The action units can be identified by geometric features and appearance features extracted from face images [3]. Recently, along with the development of convolutional neural networks (CNNs), researchers attempt to further improve the performance of emotion recognition via CNNs [4]. Barrett et al. used CNNs to recognize action units and facial emotion. These studies were mainly focused on facial emotion recognition. However, the context information can also provide extra clues to recognize emotion. For example, persons are usually happy at a wedding and are usually sad at a funeral. When the context is incorporated, the recognition accuracy can be further improved. Previous researches have shown the importance of context in the perception of emotions [5]. In some scenarios, when we analyze a wider view rather than focus on the face of the person, we can more easily judge one's feeling. Kosti et al. [6] built an emotions-in-context database and showed the emotion recognition accuracy is improved when the person and the whole scene are jointly analyzed. Chen et al. [7] exploited context clues, including events, objects, and scenes for video emotion recognition, to improve performance. However, these methods treat the features of different frames equally and the difference of emotional information contained in these frames is not considered. Although the research of context-aware video emotion recognition has made great progress, it still has two major challenges:(1) Combination of face and context information. Face feature is associated with its context information. However, traditional video emotion recognition often computes the maximum or average value of images' feature of face and context separately and then fuses the features of these two modes, lack of organic fusion of face features, and context clues of the same image. Face feature and its context feature on the same image cannot be effectively integrated. As shown in Figure 1(a), when we try to estimate the emotion of the people in image sequences, context information is difficult to provide effective emotional features. For example, it is difficult to determine whether a person in the image sequences is teasing or being attacked by a dog through context information. However, when we combine face and context information in an image, it is easier to judge people's emotions as angry than using face information alone. Similarly detailed estimations can be made in Figure 1(b).(2)Emotional differences in different images. Each frame in video contains a certain amount of emotional information, and these pieces of information can be complementary to each other. The most frequently used method is simply max/average pooling emotional features of all frames. However, different images of one video may contain different emotional information because of the difference of face size, pose, perspective, and context information. As an example, let us try to estimate the emotion of these people in Figure 2. In Figure 2(a), we can recognize that the emotion of the right image is joy with a greater probability. That is to say, the right image contains more emotional clues. Similar detailed estimations method can be made in the other images of Figure 2. Similarly, context information including surrounding environment and human body can also provide different emotional information. Therefore, how to solve the problem of emotional differences between different images is an important challenge for video emotional recognition.

(a)

(b)

Figure 1. Illustration of information combination.



(a)

(b)

(c)

(d)

Figure 2. Illustration of emotion difference.

To overcome the above two challenges, inspired by the attention mechanism [8, 9], we propose a context-aware attention network (CAAN), which is robust to frames containing less emotional information and simultaneously uses the rich emotional clues provided by the other frames. Firstly, CAAN uses two subnetworks to extract both face and context features, respectively, and these two features on the same image are fused to represent the emotion of the image. Similar to literature [6], we take as input the entire image and extract global features for providing the necessary contextual support. Then, an attention network takes as input the image feature and generates emotional score of the image. Finally, the emotion features of all images in one video will be aggregated according to their emotional score, and the final emotion representation of the video is produced.

In addition, existing video emotion recognition datasets, such as video emotion dataset [10] and Ekman Emotion Dataset [11], mainly focus on the psychological feelings of viewers brought by video content and there are no humans in many videos, which cannot effectively evaluate the human emotion in the videos. Therefore, this paper builds a human emotion in video (HEIV) dataset, which is based on video emotion dataset [10, 11] and downloads some videos from the network. The HEIV dataset contains 1012 videos and the human emotions in the videos are annotated according to the emotion category defined by psychologists Ekman and Friesen, as well as the neutral emotional categories. Besides, some videos also are annotated by neutral. We will describe it in detail in Section 3. The performance of the CAAN network is evaluated on the HEIV dataset. It improves top-1 matching rates over the state of the art by 2.22%.

The main contributions of the paper are summarized as follows. We constructed a HEIV dataset consisting of 1012 annotated videos, which mainly focuses on human emotion in the video rather than the psychological feelings of viewers brought by video content in existing video emotion recognition datasets. It is important for the design of good video emotion recognition model. CAAN can automatically generate emotion scores for each frame and lead to better representation for the difference of emotional information in different video frames. The effect of different weight function of attention mechanisms is evaluated which is helpful for the design of attention-based computational model.

The remainder of this paper is organized as follows. In the next section, we discuss related work on video-based emotion recognition. Section 3 describes the proposed dataset. Section 4 introduces the proposed CAAN. Section 5 gives experimental results. Section 6 concludes the paper and gives our future work.

RELATED WORK

Facial Emotion Recognition

Faces are the most commonly used stimuli to recognize the emotional states of people by researchers in computer vision. facial action coding system uses a set of specific localized movements of the face to encode the facial expression [2]. It deals with images in a nearly frontal pose [3]. However, facial images can be taken from multiple views or people may change their posture while being recorded. Some works that deal with multi-view emotion recognition have been proposed. Tariq et al. [12] learned a single classifier using data from multiple views. CSGPR [13] model performed the view normalization, where the features from different poses are combined. However, these approaches are not model relationships among different views. Eleftheriadis et al. proposed a discriminative shared Gaussian process latent variable model for learning a discriminative shared manifold of facial expressions from multiple views [3]. Different from the existing multiple-view facial emotion recognition, this paper mainly solves the problem of different facial poses in emotion recognition in video. There are also a few emotion recognition works using other clues apart from the face. For example, Nicolaou et al. [14] considered the location of shoulders as additional information to the face features to recognize emotions.

Recognizing Emotion from Videos

There is some early work that recognized emotion through audio-visual features (e.g., [15–18]). Wang et al. [15] used audio-visual features to recognize emotion in 36 Hollywood movies. Irie et al. [16] extracted audio-visual features and combined them with a Hidden-Markov-like dynamic model. The audio-visual features fusion is evaluated by decision level fusion and feature level fusion [17]. However, they only use simple multimodal feature fusion without considering the potential relation of multimodal features, and the appearance features are low-level features. Singh et al. [19] proposed an improved technique for order preference by similarity to ideal solution (TOPSIS) method, which is based on the co-occurrence behavior of facial action coding system in the visual sequence to select key frames. Wang et al. [20] proposed two-level attention with two-stage multi-task learning framework. Firstly, the features of corresponding region are extracted and enhanced automatically. Secondly, the relationship features of different layers are fully utilized by bi-directional RNN with self-attention. Wang et al. [21] defined a multimodal domain adaptive method to obtain the interaction between modes.

The performance of emotion recognition is evaluated by using different architectures CNN and different CNN feature layers in paper [11]. Nicolaou et al. [14] fused facial expression, shoulder posture, and audio clues for emotion recognition. Vielzeuf et al. [22] proposed a hierarchical approach, where scores and features are fused at different levels. It can retain the information of different levels, but the potential connection between multi-modal data in video has not yet been considered. Xue et al. [23] proposed a Bayesian nonparametric multimodal data modeling framework to learn emotions in videos, but it does not reflect the time evolution of emotional expression in videos. Kahou et al. [24] used CNN and RNN to model dynamic expression of videos, and the results show that the performance is better than the features fusion of frames. The temporal evolution of facial features is modeled through RNN in paper [25]. Zhang et al. [26] constructed kernel functions to convert CNN features into kernelized features. Xu et al. [27] conducted concept selection to investigate the relations between high-level concept features and emotions. This paper considers not only the emotional fusion of different facial features but also the difference of the amount of emotional information of video frames.

HUMAN EMOTION DATASET

We constructed a human emotion dataset based on video emotion dataset [10, 11] and videos downloaded from the web. Each video in the video emotion dataset is longer and contains multiple human clips in each video. It mainly focuses on the psychological feelings of viewers brought on by video content. We clipped human clips from videos of video emotion dataset and annotate the emotions of humans in the video. We also downloaded short video clips from YouTube. The database contains a total number of 1012 videos, and it uses a training set of 607 videos and a testing set of 405 videos. Figure 3 shows example frames of each emotion category from the HEIV dataset.

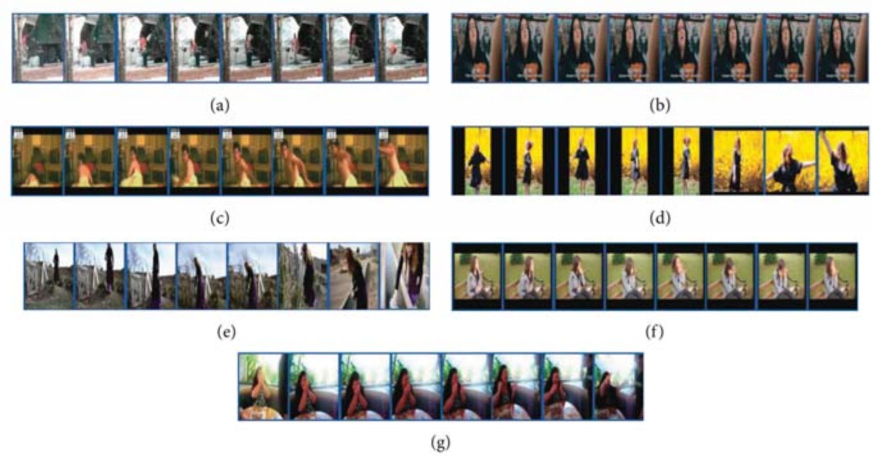


Figure 3. Example frames of each emotion category from the HEIV dataset. (a) Anger. (b) Disgust. (c) Fear. (d) Joy. (e) Neutral. (f) Sadness. (g) Surprise.

Video Annotation

The HEIV dataset was manually annotated by 10 annotators: 5 males and 5 females. Neutral and six emotion categories, including "anger," "disgust," "fear," "joy," "sadness," and "surprise," defined by psychologists Ekman and Friesen [28] are considered. In order to ensure the quality of the annotations, some videos clips with emotion labels coming from existing video emotion recognition dataset are exercised by annotators. After learning and practicing, annotators are asked to annotate our HEIV dataset. When we show a video with a person marked, we ask the annotators to select one of the emotion categories that suit that video. Each annotator independently annotates emotions, and emotion catalogue of a video marked by the most annotators is selected as the emotion label of the video. Furthermore, the gender (male/female) and the age range (child, teenager, adult) of persons in the video are also annotated.

Database Statistics

Of the 1012 annotated videos, 64% are males and 36% are females. Their ages are distributed as follows: 10% children, 11% teenagers, and 79% adults. Table 1 shows the number of videos for each of the categories.

Table 1. The number of videos per emotion category in HEIV dataset.

Category	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Number	103	105	121	207	125	158	197

CONTEXT-AWARE ATTENTION NETWORK

In our work, we focus on improving the accuracy of emotion recognition. The primary challenge in human emotion recognition is the difference of facial scale, poses, perspective, and different degrees of contextual information. We aim to tackle this by context-aware attention network (CAAN), where facial and context emotion features are fused and the emotional scores of the fusion feature are generated by attention network. The fusion features of all images and their emotion scores are aggregated to make a human emotion prediction in video.

Our proposed framework is shown in Figure 4. The architecture consists of three main modules: two emotion feature extractors and an attention fusion module. The face feature extraction module takes as input the region of the human face and extracts its facial emotion features. The context extraction module takes as input the entire frame, which extracts global features for providing the necessary contextual information. Finally, the third module is an attention fusion network which takes as input the fusion features of face and context information. It is composed of two branches. The first branch is a tiny CNN network which takes as input the fusion feature and generates emotion features of frames. The other branch is also a tiny CNN network and is used to generate an emotion score for each frame. Then, the emotion features of frames and their emotion scores will be aggregated, and the final emotion representation of the human in the video will be produced.

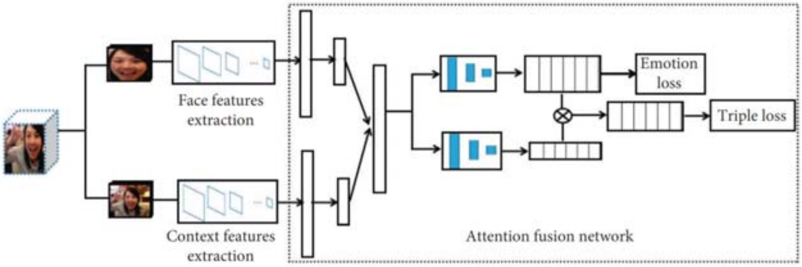


Figure 4. CAAN structure.

Emotion Features Extraction

The face is the main part of a human to express emotion. Previous research on emotion recognition mainly focused on facial expression. However, the context plays an important role in emotion recognition, and when context information is incorporated, recognition accuracy can be further improved. To jointly analyze the human face and context features to recognize rich information about human emotion in the video, facial and contextual emotion features are extracted separately and then are fused. Meanwhile, the importance of fusion feature is judged by the attention mechanism. This section describes facial and contextual emotion features extraction.

In the emotion feature extraction stage, face features extraction module and context features extraction module are used to extract face and context information, respectively. Given a video V = (v_1, v_2, . . . , v_K) with emotion labels where K is the total number of frames of the video V and vi is the i-th video frame, human faces are first extracted from video frames by faster-rcnn [29] trained on WIDER face dataset [30]. Then, the faces detected in videos are resized to 224 × 224. Let n be the number of frames with faces in the video. The human faces in video V can be denoted as F = (f_1, f_2, . . . , f_n), where fi is the human face extracted from vi . VGG-Face model trained on the VGG-Face dataset [31] as initialization is used to extract facial emotional features. It is trained on the emotion recognition dataset and obtains the facial emotion feature extractor. In this paper, HEIV dataset is used to train face feature extractor and context feature extractor. HEIV is mainly used for human emotion recognition in video, and most images in video contain human face. VGG-Face uses face images as training samples and is supervised by emotional category to which the image belongs. The trained VGG-Face can extract the emotional features of different images. Therefore, given an image sequence F = (f_1, f_2, . . . , f_n), the fc6 feature is extracted as the facial emotion feature of each image through the forward propagation operation of face feature extractor. Let X = {x_i |i = 1, 2, . . . , n} denote the fc6 layer features of F, where xi is the fc6 layer feature of fi . In order to fuse face and context information on the same image, only image sequences containing faces S = (I_1, I_2, . . . , I_n) are selected in this paper. VGG network [32] is selected as context feature extractor. It is pre-trained on ImageNet dataset [33] and then trained on HEIV emotion recognition dataset. The whole image containing face is taken as training sample, and the emotional category of the video is taken as supervisory signal. The VGG trained can extract the contextual emotional features about the scenes, environments, and backgrounds of different images. Therefore, given an image sequence S = (I_1, I_2, . . . , I_n), the fc6 feature is extracted as the context emotion feature of each image through the forward propagation operation of context feature extractor. Let C = (c_1, c_2, . . . , c_n) denote the fc6 layer features of V, where ci is the context information of vi. The obtained facial features and context features are fed to the attention fusion network for effective fusion, so as to further improve the accuracy of video emotion recognition.

Attention Fusion Network.

For a video clip v, we now have two high-level semantic features (X, C). These two features characterize the human of video from different perspectives and there are also differences in the amount of emotional information contained in different video frames. In order to fuse the features of face image sequence and context image sequence into a unified feature representation, the face feature sequence and context feature sequence can be fused separately, and then the face fusion feature and context fusion feature are fused. However, this will isolate the face feature from its context in the same image. Faces are closely related to the context of the same image, and their emotional features can complement each other to reflect a person's emotions in the image more comprehensively. In addition, traditional average pooling or maximum pooling feature fusion methods are difficult to effectively mine the complementarity between different image features and cannot reflect the emotional differences of different images. Therefore, this paper proposes an attention fusion network to effectively fuse image feature sequences of face and context. It can quantify the emotion difference of different video frames and fuse the features of face image sequence and context image sequence according to their importance and derive a unified feature representation

More precisely, inspired by [7], we first transform face and context features of all images to a high-level space (1024 neurons for face and context features) and then face and context feature of each image are fused. Therefore, their distinct properties can be preserved and discriminative ability will be increased. Since these features are extracted from different video frames and have different discrimination, we then use an attention fusion mechanism to fuse these features, which can be able to be robust to frames with poor emotion information and simultaneously use the rich emotion information provided by the other video frames. Our basic idea is that each emotion feature can have an emotion score in aggregation, and the emotion features are aggregated according to their emotion scores. For that, emotion features are passed through two branches and then aggregated together. the first branch named fusion feature generation part extracts higher-level fusion emotion feature, and the other branch named emotion score generation part predicts an emotion score for each fusion feature. Features of video frames are then aggregated according to emotion scores.

Since the face and context features are extracted from the same frames demonstrating the same emotion in different forms, fusion feature generation part first fuses them using a fusion layer with 2048 neurons for absorbing all the information to obtain a shared representation, which can be expressed by the following:

$$r_i = \Phi\left(x_i, c_i\right), \tag{1}$$

where r_i is the fusion feature of the i-th video frame and Φ(,) is a fusion function

The fusion feature r_i will be fed to two branch networks. the first branch named fusion generates subnetworks for generating higher-level fusion features, and it can be expressed by a fully connected layer

$$g_i = W_1^f \times r_i + b_1^f. \tag{2}$$

The obtained higher-level fusion features are passed through a fully connected layer and generate emotion prediction vector. this branch is supervised by softmax-loss, which optimizes the probability of each image feature.

The other branch is the emotion score generation subnetwork, which is used to generate an emotion score for fusion feature of each image. We rely on an attention mechanism to obtain an emotion score. Its responsibility is first to analyze the amount of emotional information contained in video frames and then generate an emotion score which is used to bestow the feature with as much emotion information as possible. We use higher-level fusion features gi to represent fusion features of each image, and its corresponding emotion score can be calculated using a fully connected layer that has only one cell, which is signed as 1C

$$s_i = W_1^s \times g_i + b_1^s, \tag{3}$$

where W_1^s and b_1^s are parameters to be learned for the emotion score generation part. Similarly, emotion score can also be generated by two or three successive fully connected layers, which is signed as 2CF and 3CF separately. In the experiments in Section 5, we will compare the effects of these different weighting functions

The emotion feature representation of video V can be obtained by aggregating the fusion features gi and emotion scores s_i of all images. It can be expressed as follows

$$R_V = \frac{\sum_{i=1}^{n} (g_i \times s_i)}{\sum_{i=1}^{n} s_i}, \tag{4}$$

where R_V is the emotion feature of the video V. It is supervised by triplet loss [34], which minimizes variances of intra-class samples and discrimination of the emotion representation of the vide

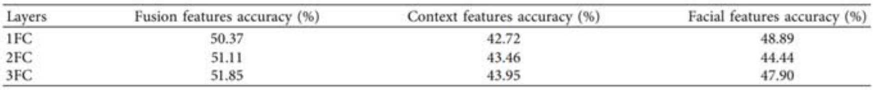
EXPERIMENTS

Effect of Weighting Function

In this subsection, we analyze the effect of different weighting function of emotion score generation part on the emotion recognition performance. First of all, we extract the fc6 layer feature of the face and context features by face feature extraction part and context feature contraction part. The fc6 features of face and context information are first passed through a fully connected layer with 1024 neurons and then fused. Fusion features are fed to two branches: one is used to generate higher-level fusion features, and the other is used to generate emotion scores. These two branches will be aggregated to generate the final emotion representation of the video. We consider three different weight functions of attention network, 1CF, 2CF, and 3CF, as described in Section 4.2.

We also give the evaluation results by the attention network which takes as inputs face and context information separately. For the face or context information, the network is divided into two branches beginning with pool5 layer features. The first branch is used to extract facial or context features through pre-training vgg face or vgg16 model, and the other branch takes as input the middle features of face or context information and generates emotion score for each face or context feature. Then, the facial or context emotion features and their emotion scores will be aggregated, and the final emotion representation of the face or context will be produced. Similar to the attention fusion network, emotion score can be calculated by one or two or three convolution layers and a fully connected layer that has only one cell, which is also signed as 1CF and 2CF and 3CF separately. Table 2 shows the accuracy of emotion recognition using different weight functions in attention networks on HEIV datasets.

Table 2. Accuracy of emotion recognition on HEIV dataset.



Layers	Fusion features accuracy (%)	Context features accuracy (%)	Facial features accuracy (%)
1FC	50.37	42.72	48.89
2FC	51.11	43.46	44.44
3FC	51.85	43.95	47.90

As shown in Table 2, we observe that the recognition accuracy is different with different weighting functions in emotion score generation part, which means that attention mechanism can play an effective role in this situation. We also observe that 3CF is slightly better than 2FC and 1FC for fusion emotion features and context information, but 1CF is slightly better than 2FC and 3FC for the face features. From these three emotion feature accuracies, we can infer that deeper attention networks can get better emotion score, but when the attention network exceeds a certain degree, we cannot get better emotion score. We rely on the 3CF weighting function for fusion features and context features emotion score generation part and 1CF weighting function for facial emotion score generation part as the default in all subsequent experiments.

Effect of Attention Mechanism and Feature Fusion

In this subsection, we evaluate the performance of attention mechanism and feature fusion. In order to validate the effectiveness of our attention mechanism and feature fusion, we implement the following three average aggregate baseline approaches:

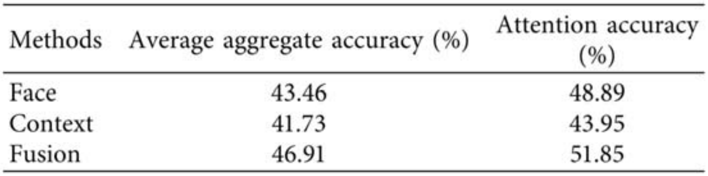
Face Average Aggregate (FAA). The fc6 layer features of all faces are extracted by VGG-Face. These features are aggregated by average pool and then passed through two successive fully connected layers and are supervised by softmax-loss.

Context Average Aggregate (CAA). The fc6 layer features of all context images are extracted by VGG16. These features are aggregated by the average pool and then passed through two successive fully connected layers and are supervised by softmax-loss.

Fusion Feature Average Aggregate (FFAA). The fc6 layer features of all faces and context images are extracted by VGG-Face and vgg16 separately. These two features are first passed through a fully connected layer with 1024 neurons and then fused. The fusion feature is passed through two successive fully connected layers and is supervised by softmax-loss.

Table 3 shows the accuracy of emotion recognition using attention mechanism and the above three average aggregation methods: FAA, CAA, and FFAA. As shown in Table 3, on HEIV dataset, attention mechanism increases top-1 emotion recognition accuracy by 5.43%, 2.22%, and 4.94%, respectively, compared with FAA, CAA, and FFAA. We also notice that feature fusion increases top-1 emotion recognition accuracy by 3.45% and 5.18%, respectively, compared with face features and context features on average aggregation and feature fusion increases top-1 emotion recognition accuracy by 2.96% and 7.9%, respectively, compared with face features and context features on attention mechanism. Based on these experiments, we can infer that attention fusion network outperforms the average aggregate method on HEIV dataset, and feature fusion outperforms single face or context feature on HEIV dataset.

Table 3. Performance evaluation of attention and feature fusion.



Methods	Average aggregate accuracy (%)	Attention accuracy (%)
Face	43.46	48.89
Context	41.73	43.95
Fusion	46.91	51.85

VISUALIZATION OF CAAN

In order to visualize the CAAN, some image sequences in the test set and their corresponding emotional scores are shown in Figure 5. As shown in Figure 5, the emotional scores of different images are different because of the difference of their facial posture and context information. Some images contain abundant emotion clues on human face and the context information, such as the 3rd image in Figure 5(b) and the 5th image in Figure 5(f); thus, CAAN gives these images higher emotional scores. On the contrary, some images contain little emotion clues on human face and the context information, such as the 1st image in Figure 5(a) and the 7th image in Figure 5(e), and CAAN gives these images lower emotional scores.

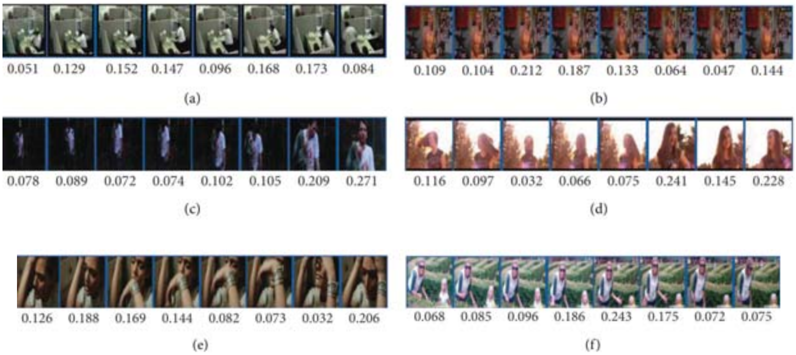


Figure 5. Samples with their emotion scores predicted by CAAN.

Comparison with State of the Art

We also compare state-of-the-art performance in recent literature. To validate the effectiveness of our CAAN method, we compare with the following state-of-the-art approaches on HEIV dataset.

Attention-Based Network

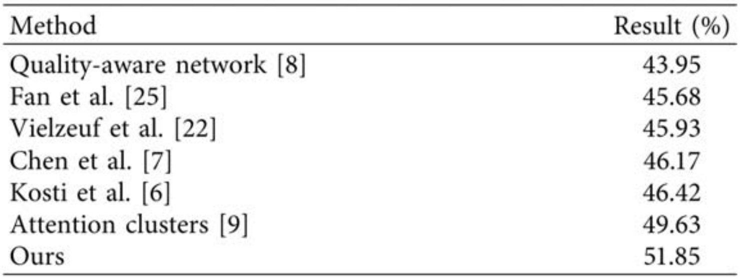
QAN [8] and attention clusters [9] are two attention-based networks. QAN is a quality network which takes as input images of video on HEIV dataset, and attention clusters are a multimodal attention network which takes as input fc6 layer features of face and context on HEIV dataset.

Feature Fusion Network

Recent literature [6, 7, 24, 29] used multimodal feature fusion network. It implemented two modes of face and context on HEIV dataset.

Table 4 gives top-1 accuracy (%) of different methods on HEIV. As shown in Table 4, our context-aware attention fusion network achieves 2.22% performance gain on HEIV dataset. We also noticed that the performance of QAN only taking as input video frames is lower than fusion feature. By the attention mechanism, the performance of attention clusters [9] taking as input two modes of face and context is higher than feature fusion without attention mechanism. Note that our CAAN attains superior performance for two reasons: firstly, attention mechanism is robust to frames containing less emotional information and simultaneously uses the rich emotional clues provided by the other frames. Secondly, our feature fusion not only jointly exploits the face features and context information but also preserves their distinct properties. Based on these experiments, CAAN outperforms state-of-the-art results on HEIV datasets. The improvement of CAAN network proves CAAN's ability to deal with videos with different emotion information.

Table 4. Top-1 accuracy (%) compared with state-of-the-art methods on HEIV

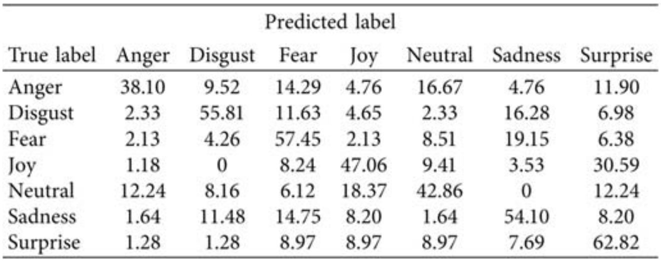


Method	Result (%)
Quality-aware network [8]	43.95
Fan et al. [25]	45.68
Vielzeuf et al. [22]	45.93
Chen et al. [7]	46.17
Kosti et al. [6]	46.42
Attention clusters [9]	49.63
Ours	51.85

Confusion Matrix

To analyze the recognition accuracy of different emotion categories, we gave the confusion matrix of the recognition accuracy using CAAN on HEIV which is shown in Table 5. The vertical is true label, and the horizontal is the recognition accuracy of each emotion category. We observed that the surprise, fear, and disgust are well recognized, and the anger, neutral, and joy have a greater number of false positives. We inferred that it is because anger and joy add more emphasis on psychological activities, and their behavior expression is relatively low. We also noticed that 30.59% of joy is recognized as a surprise. We inferred that it is because some humans have both feelings of joy and surprise, and it is hard to determine which emotion dominates. We also observed that happiness is not recognized as disgust and neutral is not recognized as sadness. It is because the expressions of these two emotion categories are quite different.

Table 5. Confusion matrix.



	Predicted label						
True label	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
Anger	38.10	9.52	14.29	4.76	16.67	4.76	11.90
Disgust	2.33	55.81	11.63	4.65	2.33	16.28	6.98
Fear	2.13	4.26	57.45	2.13	8.51	19.15	6.38
Joy	1.18	0	8.24	47.06	9.41	3.53	30.59
Neutral	12.24	8.16	6.12	18.37	42.86	0	12.24
Sadness	1.64	11.48	14.75	8.20	1.64	54.10	8.20
Surprise	1.28	1.28	8.97	8.97	8.97	7.69	62.82

Result on Ekman-6 and VideoEmotion-8

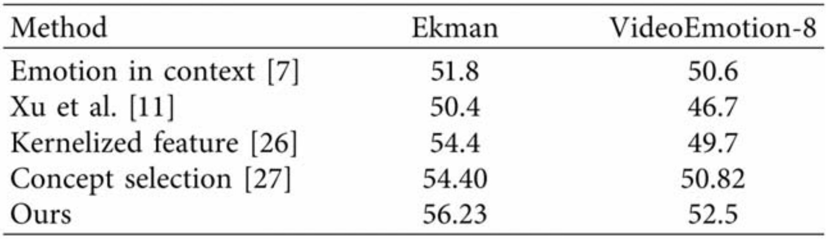
In this section, we conduct experiments on Ekman-6 [35] and VideoEmotion-8 [10] datasets to further evaluate the effectiveness of our method.

Ekman-6 dataset contains 1637 videos, and it uses a training set of 819 videos and a testing set of 818 videos. It was manually annotated by 10 annotators according to Ekman's theory [28] on six basic human emotion categories, with a minimum of 221 videos per category.

VideoEmotion-8 dataset contains 1101 videos collected from YouTube and Flickr. The average duration of videos is 107 seconds. The experiments were conducted 10 times according to train/test splits provided by [10].

Table 6 gives top-1 accuracy (%) of different methods on Ekman-6 and VideoEmotion-8 datasets. As shown in Table 6, our context-aware attention fusion network achieves 1.83% and 1.68 performance gain on Ekman-6 and VideoEmotion-8 dataset, respectively. The results show that our methods achieve the state-of-the-art results on both Ekman-6 and VideoEmotion-8 datasets.

Table 6. Top-1 accuracy (%) compared with state-of-the-art methods on Ekman-6 and VideoEmotion-8.



Method	Ekman	VideoEmotion-8
Emotion in context [7]	51.8	50.6
Xu et al. [11]	50.4	46.7
Kernelized feature [26]	54.4	49.7
Concept selection [27]	54.40	50.82
Ours	56.23	52.5

CONCLUSION AND FUTURE WORK

In this paper, we first built a video dataset with 7 categories of human emotion, named human emotion in the video (HEIV). With the HEIV dataset, we trained a context-aware attention network (CAAN) to recognize human emotion. CAAN consists of three modules. Two emotion feature extraction modules are used to extract face and context features, respectively. Attention fusion network fuses these two features and generates an emotion score for each fusion feature. Then, the fused emotion features will be aggregated according to their emotion score, and the final emotion representation of the video is produced. The performance of the CAAN network is evaluated and it can achieve excellent results on the HEIV dataset.

Although our approach obtains a promising performance in video emotion recognition, however, because of the diversity of human emotion expression, human emotion can be expressed through multiple body parts. In future work, we will further combine human part semantics for better recognition performance.

ACKNOWLEDGMENTS

This work was supported by the scientific and technological research project of Henan Provincial Science and Technology Department (182102310919) and the Foundation of He’nan Educational Committee (21A520006).

REFERENCES

1. K. Byoung, “A brief review of facial emotion recognition based on visual information,” *Sensors*, vol. 18, no. 2, pp. 401–420, 2018.
2. W.-S. Chu, F. De la Torre, and J. F. Cohn, “Selective transfer machine for personalized facial expression analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 529–545, 2017.
3. S. Eleftheriadis, O. Rudovic, and M. Pantic, “Discriminative shared Gaussian processes for multiview and view-invariant facial expression recognition,” *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 189–204, 2015.
4. C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.
5. L. F. Barrett, B. Mesquita, and M. Gendron, “Context in emotion perception,” *Current Directions in Psychological Science*, vol. 20, no. 5, pp. 286–290, 2011.
6. R. Kosti, J. M. Alvarez, A. Recasens et al., “Emotion recognition in context,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1960–1968, Honolulu, HI, USA, July 2017.
7. C. Chen, Z. Wu, and Y. G. Jiang, “Emotion in context: deep semantic feature fusion for video emotion recognition,” *ACM on Multimedia Conference*, vol. 16, pp. 127–131, 2016.
8. Y. Liu, J. Yan, and W. Ouyang, “Quality aware network for set to set recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4694–4703, Honolulu, HI, USA, July 2017.
9. X. Long, C. Gan, G. D. Melo et al., “Attention clusters: purely attention based local feature integration for video classification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7834–7843, Honolulu, HI, USA, July 2017.
10. Y.-G. Jiang, B. Xu, and X. Xue, “Predicting emotions in user-generated videos,” in *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, pp. 73–79, Québec, Canada, July 2014.

11. B. Xu, Y. Fu, Y.-G. Jiang, B. Li, and L. Sigal, “Video emotion recognition with transferred deep feature encodings,” in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pp. 15–22, New York, NY, USA, June 2016.
12. U. Tariq, J. Yang, and T. S. Huang, “Multi-view facial expression recognition analysis with generic sparse coding feature,” in *Proceedings of the Computer Vision - ECCV 2012. Workshops and Demonstrations European Conference on Computer Vision*, pp. 578–588, Firenze, Italy, October2012.
13. O. Rudovic, M. Pantic, and I. Patras, “Coupled Gaussian processes for pose-invariant facial expression recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1357–1369, 2013.
14. M. A. Nicolaou, H. Gunes, and M. Pantic, “Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space,” *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 92–105, 2011.
15. H. L. Wang and L. F. Cheong, “Affective understanding in film,” *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 16, no. 6, pp. 689–704, 2006.
16. G. Irie, T. Satou, A. Kojima, T. Yamasaki, and K. Aizawa, “Affective audio-visual words and latent topic driving model for realizing movie affective scene classification,” *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 523–535, 2010.
17. M. Xu, C. Xu, X. He, J. S. Jin, S. Luo, and Y. Rui, “Hierarchical affective content analysis in arousal and valence dimensions,” *Signal Processing*, vol. 93, no. 8, pp. 2140–2150, 2014.
18. R. M. A. Teixeira, T. Yamasaki, and K. Aizawa, “Determination of emotional content of video clips by low-level audiovisual features,” *Multimedia Tools and Applications*, vol. 61, no. 1, pp. 21–49, 2012.
19. L. Singh, S. Singh, and N. Aggarwal, “Improved TOPSIS method for peak frame selection in audio-video human emotion recognition,” *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 6277–6308, 2019.

20. X. Wang, M. Peng, L. Pan, H. Min, J. Chunhua, and R. Fuji, “Two-level attention with two-stage multi-task learning for facial emotion recognition,” *Journal of Visual Communication and Image Representation*, vol. 62, no. 7, pp. 217–225, 2019.
21. Y. Wang, J. Wu, and H. Keiichiro, “Multi-attention fusion network for video-based emotion recognition,” in *Proceedings of the 2019 International Conference on Multimodal Interaction*, pp. 595–601, Suzhou, China, October 2019.
22. V. Vielzeuf, S. Pateux, and F. Jurie, “Temporal multimodal fusion for video emotion classification in the wild,” in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pp. 569–576, New York, NY, USA, November 2017.
23. J. Xue, Z. Luo, K. Eguchi, T. Takiguchi, and T. Omoto, “A Bayesian nonparametric multimodal data modeling framework for video emotion recognition,” in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 601–606, Hong Kong, China, July 2017.
24. S. E. Kahou, V. Michalski, K. Konda et al., “Recurrent neural networks for emotion recognition in video,” in *Proceedings of the ACM International Conference on Multimodal Interaction*, pp. 467–474, Seattle, WA, USA, November 2015.
25. L. Fan and K. Yunjie, “Spatiotemporal Networks for Video Emotion Recognition,” 2017, http://arxiv.org/abs/1704.00570.
26. H. Zhang and M. Xu, “Recognition of emotions in user-generated videos with kernelized features,” *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2824–2835, 2018.
27. B. Xu, Y. Zheng, H. Ye et al., “Video motion recognition with concept selection,” in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 406–411, Shanghai, China, July 2019.
28. P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.
29. S. Ren, K. He, R. Girshick et al., “Faster R-CNN: towards real-time object detection with region proposal networks,” in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 91–99, Montreal, Canada, December 2015.
30. S. Yang, P. Luo, C. L. Chen, and X. Tang, “Wider face: a face detection benchmark,” in *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, pp. 5525–5533, Las Vegas, NV, USA, June 2016.

31. O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proceedings of the British Machine Vision Conference*, pp. 1–12, Swansea, UK, September 2015.
32. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, http://arxiv.org/abs/1409.1556.
33. O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2014.
34. F. Schroff, D. Kalenichenko, and P. James, "Facenet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, Boston, MA, USA, June 2015.
35. B. Xu, Y. Fu, Y.-G. Jiang, B. Li, and L. Sigal, "Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 255–270, 2018.

CHAPTER 2

LARGE-SCALE VIDEO RETRIEVAL VIA DEEP LOCAL CONVOLUTIONAL FEATURES

Chen Zhang[1], Bin Hu[2,3], Yucong Suo[4], Zhiqiang Zou[1,5] and Yimu Ji[1]

[1]College of Computer, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China
[2]College of Geographic Science, Nanjing Normal University, Nanjing, China
[3]Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Ministry of Education, Nanjing, China
[4]Bell Honor School, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China
[5]Jiangsu Key Laboratory of Big Data Security & Intelligent Processing, Nanjing, Jiangsu, China

ABSTRACT

In this paper, we study the challenge of image-to-video retrieval, which uses the query image to search relevant frames from a large collection of

Citation: Chen Zhang, Bin Hu, Yucong Suo, Zhiqiang Zou, Yimu Ji, “Large-Scale Video Retrieval via Deep Local Convolutional Features”, Advances in Multimedia, vol. 2020, Article ID 7862894, 8 pages, 2020. https://doi.org/10.1155/2020/7862894

Copyright: © 2020 by Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

videos. A novel framework based on convolutional neural networks (CNNs) is proposed to perform large-scale video retrieval with low storage cost and high search efficiency. Our framework consists of the key-frame extraction algorithm and the feature aggregation strategy. Specifically, the key-frame extraction algorithm takes advantage of the clustering idea so that redundant information is removed in video data and storage cost is greatly reduced. The feature aggregation strategy adopts average pooling to encode deep local convolutional features followed by coarse-to-fine retrieval, which allows rapid retrieval in the large-scale video database. The results from extensive experiments on two publicly available datasets demonstrate that the proposed method achieves superior efficiency as well as accuracy over other state-of-the-art visual search methods.

INTRODUCTION

Enormous images and videos are generated and uploaded onto the Internet. With a large amount of publicly available data, visual search has become an important frontier topic in the field of information retrieval. There exist several kinds of visual search tasks, including image-to-image (I2I) search [1, 2], video-to-video (V2V) search [3, 4], and image-to-video (I2V) search [5, 6]. Specifically, the well-known I2I visual search can be used for product search, in which relevant images are retrieved by the query image. The V2V search is commonly used for copyright protection, in which video clips are found via a relevant video. The I2V search addresses the problem of retrieving relevant video frames or specific timestamps from a large database via the query image. This technology is relevant for numerous applications, such as brand monitoring, searching film using slides, and searching lecture videos using screenshots.

In this work, we study the specific task of I2V search, which is especially challenging because of the asymmetry between the query image and the video data. Video data can be divided into four hierarchical structures: video, scene, shot, and frame. When considering only the visual content, a video is a sequence of frames displayed at a certain rate (as shown in Figure 1). For example, a video with a frame rate of 30 fps is equivalent to 30 images in one second. The structure of a video means that adjacent frames are highly correlated with each other. To perform large-scale retrieval, we should select representative frames of a video frame sequence to reduce redundant information for further processes. Key-frame extraction, which could represent the salient content and information of the video, is the technique employed to remove redundant or duplicate frames. In this work, we propose a cluster-based key-frame extraction algorithm to summarize the video sequences.

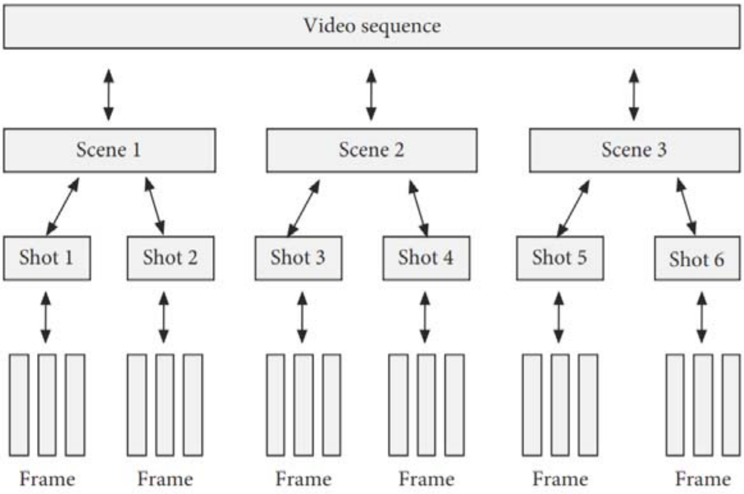


Figure 1. The structure of video data.

Inspired by the advances in content-based image retrieval (CBIR), we propose to take advantage of the image retrieval techniques to image-to-video search. In CBIR, one of the most challenging issues is the association of pixel-level information with human-perceived semantics. Although some hand-crafted features have been proposed to represent images, the performance of these descriptors is not satisfactory. Recently, the CNN-based descriptors have shown excellent performance on various computer vision tasks, such as image classification, instant search, and target tracking. Encouraged by the advances in the deep convolutional neural network, our works share similarities with other CNN-based methods extracting features of the frame via pretrained CNNs.

In visual search tasks, search efficiency plays an essential role. Due to the high computational cost, high-dimensional CNN features are not appropriate for large-scale I2V retrieval. To aggregate these high-dimensional features into a lower-dimensional space, we propose a mechanism with two pooling layers for coarse-to-fine search. Specifically, the low-dimensional frame index generated from the second pooling layer is used for the coarse-level search, which could quickly narrow down the matches. And, the high-dimensional frame descriptor generated from the first pooling layer is used for the fine-level search to improve the retrieval accuracy.

This work presents three contributions:(i) We proposed a cluster-based key-frame extraction algorithm to remove a large amount of redundant information in the video, which could greatly reduce storage cost. (ii) We took advantage of an aggregation method based on average pooling to encode deep local convolutional features, which allows rapid retrieval in the large-scale video database. To further improve efficiency, we introduced a coarse-to-fine strategy performing the search in two steps. (iii) An extensive set of experiments on two publicly available datasets demonstrated that the proposed method outperforms several state-of-the-art visual search methods.

RELATED WORK

Key-frame extraction is an essential part in video analysis and management, providing a suitable video summarization for video indexing, browsing, and retrieval. The existing key-frame extraction methods are roughly divided into three categories. Early works [7, 8] focused on sampling video sequences uniformly or randomly to obtain key frames, which is easy to implement. However, it ignores the contents of the frames and may result in repeated frames or missing of the important frames. A second generation of works [9, 10] reported significant gains in key-frame extraction based on shot segmentation which selects the key frames from shot fragments. The extracted key frames via this method are representative. However, the neglected correlation between different shots may result in information redundancy. In response to the above problems, cluster-based key-frame extraction [11, 12] has emerged. This method divides the video frame into clusters based on frame contents and then extracts several representative frames from each cluster. The key frames extracted by this method faithfully reflect the original video content. In this paper, we propose a key-frame extraction method based on the *k*-means clustering algorithm for further processes.

In the image-to-video task, frame representation plays a critical role. In the early 1990s, images were indexed by the hand-crafted features, like color, texture, and spatial. A straightforward strategy for image representation is to extract global descriptors. However, global signatures may fail the invariance expectation to image changes such as illumination, occlusion, and translation. The performance of these visual descriptors was still limited until the breakthrough of local descriptors. In 2003, with the introduction of the Bag-of-Words (BoW) model in the image retrieval community, the majority of the traditional methods were not used any further. For more than a decade, the retrieval community has witnessed the superiority of the BoW model, and many improvements [13, 14] were proposed. In 2012, Krizhevsky et al. [15] proposed AlexNet, which achieved the state-of-the-art recognition accuracy in ILSRVC 2012. Inspired by the advances of deep convolutional neural networks, many works have focused on deep learning-based methods, especially the CNNs. Early works [16, 17] elaborated that features from fully connected layers of a pretrained CNN network perform much better than traditional hand-crafted descriptors. However, several works [18, 19] reported that local features from the last convolutional layer usually yield superior accuracy compared to the global features from the fully connected layer. Our works share similarities with the former methods that extract convolutional features from pretrained CNNs.

However, to perform large-scale retrieval, it is necessary to compress the high-dimensional features to reduce the storage cost and speed up the retrieval. Several works have tried to encode features from CNNs via BoW [20], VLAD [21], and FV [22], which are commonly used to generate hand-crafted descriptors. Although these methods perform well in some visual search tasks, they require a large code book trained offline, which is difficult to achieve in the large-scale database. Additionally, some information will be lost in the feature encoding stage using these methods. Apart from the aggregation strategies mentioned above, average pooling mechanism was able to generate discriminative descriptors. Lin et al. [23] elaborated the reasons why pooling is effective in encoding deep local convolutional features. Firstly, the mean pooling strategy could largely prevent overfitting. Secondly, it sums up the spatial information, resulting in a more robust spatial transformation of the query image. Inspired by the excellent performance of average pooling, we propose a simple aggregation method to generate compact and discriminative frame representations.

APPROACH

Our method includes three main components: key-frame extraction, frame representation, and coarse-to-fine retrieval, as shown in Figure 2. The first component is a key preprocess to summarize the video data. Subsequently, the feature representation of the key frame is learned by the pretrained deep convolutional neural networks. Ultimately, relevant frames to the query image are retrieved after feature aggregation.

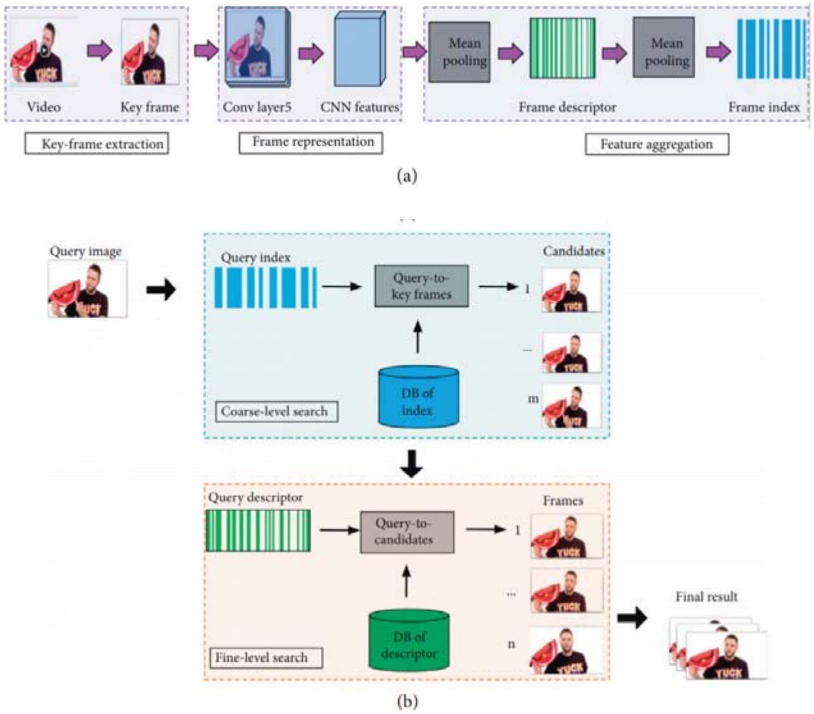


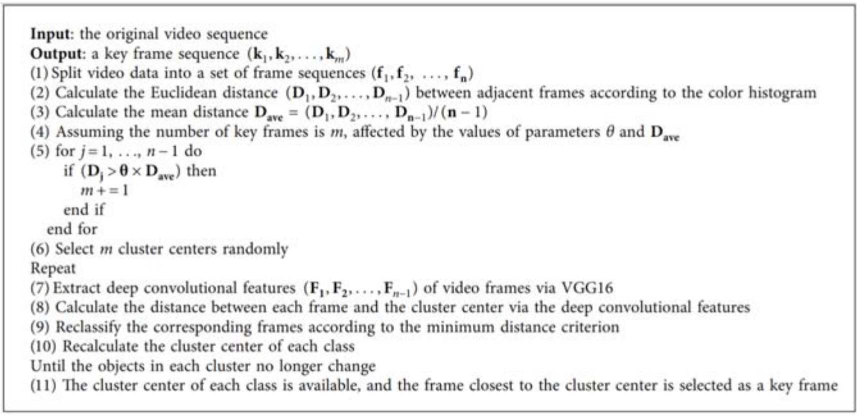
Figure 2. Block diagram of our proposed retrieval approach which searches video databases by images. (a) The process of indexing and extracting the descriptors for an image. (b) The process of coarse-to-fine search.

The focus of our work is shown in Figure 2. Figure 2(a) shows the process of indexing and extracting the descriptors for an image, and note that the length of the index is much smaller than that of the descriptor. For large-scale retrieval tasks, it is very important to quickly narrow down the search using the image index. Figure 2(b) shows the process of coarse-to-fine search. In the coarse-level search, the query image's index is compared to the indices of key frames (DB of the index) which are extracted from video frames to generate m candidates. Then, the descriptor of the query image, which contains more information than the index, is compared to the descriptors (DB of the descriptor) of m candidates in the fine-level search using Euclidean distance. The smaller the Euclidean distance is, the higher the level of similarity of the two images is. Each candidate is ranked in an ascending order by similarity; hence, top n ranked frames are selected as the final result.

Key-Frame Extraction

Key-frame extraction is the basis of video analysis and content-based video retrieval. As mentioned in the previous section, a video is a sequence of frames displayed at a certain rate, and adjacent frames are highly correlated with each other. Key-frame extraction chooses frames to summarize the video while removing redundant information. In this work, we adopt the cluster-based algorithm to extract representative frames.

The main idea of the cluster-based algorithm is to divide the frame sequences into several clusters according to the frame features, and then the frame closest to the cluster center would be selected as a key frame. However, this algorithm requires a prespecified experimental parameter, the number of clusters, which directly affects the result of key-frame extraction. It is very difficult to compute the number of clusters in the case where the video content is uncertain. To address this issue, we propose an improved key-frame extraction algorithm. The specific steps are represented in Algorithm 1, in which steps from (1) to (5) are responsible for computing the number of clusters, while steps from (6) to (9) perform the task of dividing the frame sequences into several clusters and selecting a key-frame sequence.



```
Input: the original video sequence
Output: a key frame sequence $(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_m)$
(1) Split video data into a set of frame sequences $(\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n)$
(2) Calculate the Euclidean distance $(\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{n-1})$ between adjacent frames according to the color histogram
(3) Calculate the mean distance $\mathbf{D}_{\mathbf{ave}} = (\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{\mathbf{n}-1})/(\mathbf{n}-1)$
(4) Assuming the number of key frames is $m$, affected by the values of parameters $\theta$ and $\mathbf{D}_{\mathbf{ave}}$
(5) for $j = 1, \ldots, n-1$ do
      if $(\mathbf{D}_\mathbf{j} > \boldsymbol{\theta} \times \mathbf{D}_{\mathbf{ave}})$ then
        $m += 1$
      end if
   end for
(6) Select $m$ cluster centers randomly
Repeat
(7) Extract deep convolutional features $(\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_{n-1})$ of video frames via VGG16
(8) Calculate the distance between each frame and the cluster center via the deep convolutional features
(9) Reclassify the corresponding frames according to the minimum distance criterion
(10) Recalculate the cluster center of each class
Until the objects in each cluster no longer change
(11) The cluster center of each class is available, and the frame closest to the cluster center is selected as a key frame
```

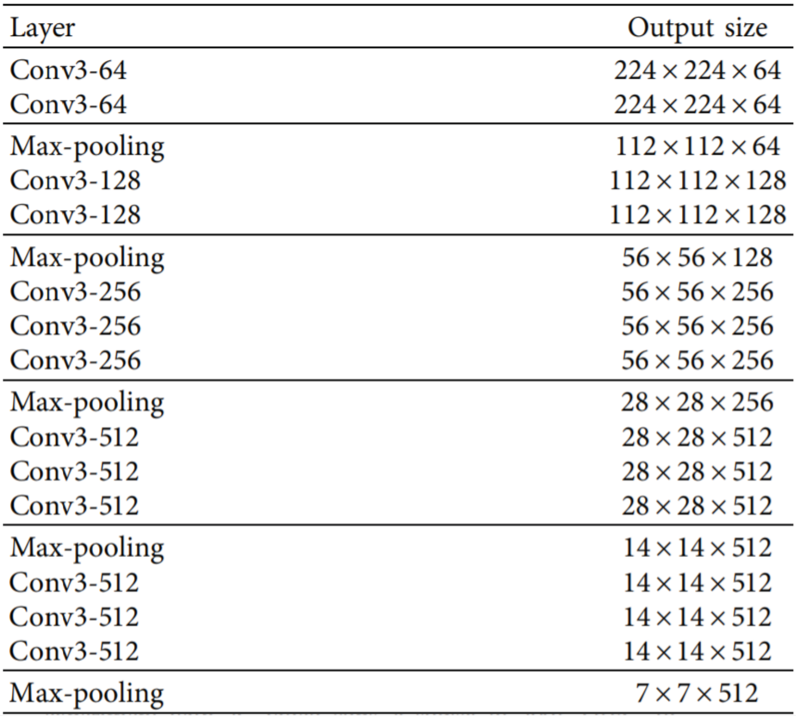
Algorithm 1. Key-frame extraction based on the *K*-means cluster.

Frame Representation

Our approach is similar to former works which extracted convolutional features from pretrained CNNs. However, we discard the softmax and fully connected layers of the original network while keeping convolutional layers to obtain local features. Our work focuses on local features due to the problem that global descriptors may fail the invariance expectation to image changes [24].

In this work, we choose the popular deep neural network named VGG16 to extract frame features, which was trained on the ILSVRC dataset. The network consists of a stacked 3×3 convolutional kernel and max-pooling layers, followed by three fully connected and softmax layers. Table 1 shows the output size of convolutional layers in VGG16. Given a pretrained VGG16 network, an input frame is first rescaled to a predefined image side and then is passed through the network in a forward pass. Finally, we obtain features with size 7×7×512 from the last max-pooling layer.

Table 1. Structure of VGG16.



Layer	Output size
Conv3-64	$224 \times 224 \times 64$
Conv3-64	$224 \times 224 \times 64$
Max-pooling	$112 \times 112 \times 64$
Conv3-128	$112 \times 112 \times 128$
Conv3-128	$112 \times 112 \times 128$
Max-pooling	$56 \times 56 \times 128$
Conv3-256	$56 \times 56 \times 256$
Conv3-256	$56 \times 56 \times 256$
Conv3-256	$56 \times 56 \times 256$
Max-pooling	$28 \times 28 \times 256$
Conv3-512	$28 \times 28 \times 512$
Conv3-512	$28 \times 28 \times 512$
Conv3-512	$28 \times 28 \times 512$
Max-pooling	$14 \times 14 \times 512$
Conv3-512	$14 \times 14 \times 512$
Conv3-512	$14 \times 14 \times 512$
Conv3-512	$14 \times 14 \times 512$
Max-pooling	$7 \times 7 \times 512$

Coarse-to-Fine Retrieval with Aggregated Features

Deep convolutional neural networks have shown their promise as a universal representation for recognition. However, the signatures are high-dimensional vectors that are inefficient in large-scale video retrieval. To facilitate efficient video retrieval, a practical way to reduce the computational cost is to aggregate the CNN features.

Given a frame, we denote the feature map from the last max-pooling layer as f. Assume that f takes the size of k × w × h, where k denotes the number of channels and w and h are the width and height of each channel. Assume that p represents the output of mean pooling and $s \times t(s \leq w, t \leq h)$ is the pooling window size. ,en, we exert mean pooling steps on the local CNN features:

$$p = \frac{1}{s \times t} \sum f(i), \quad i = 1, 2, \ldots, k. \tag{1}$$

Figure 3 depicts the example of encoding the features extracted from the last max-pooling layer before the fully connected layer of the VGG16 network. The given features sized 512 × 7 × 7, and after the first mean pooling process with pooling window sized 7 × 7, we get the feature descriptor sized 512 × 1 × 1. ,en, after the second mean pooling process with pooling window sized 8 × 1, the feature descriptor is resized to 64 × 1 × 1.

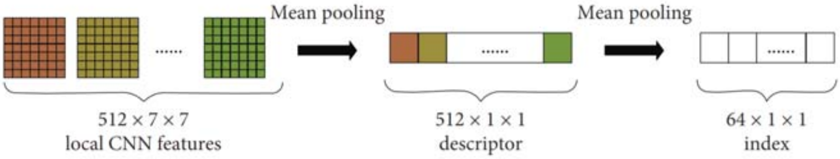


Figure 3. The process of feature encoding.

For large-scale retrieval tasks, it is very important to quickly narrow down the search using the feature index. The initial search is computed using the Euclidean distance of the feature index between the query image and the key frames in the database. After that, the top *m* frames are selected as candidates based on the distance score. Then, to ensure search accuracy, the fine-level search is performed by calculating the distance of the feature descriptor between the query image and the candidates. Finally, top *n* key frames, a subset of candidates, are picked out.

EXPERIMENT

In this section, we demonstrate the benefits of our method. We start with introducing the datasets, evaluation metrics, and parameter setting. Then, we present our experimental results with performance comparison with several existing visual search approaches.

Experimental Preparation

Datasets

We consider 2 datasets. The NTU video object instance dataset (NTU) [25] and the 2001 TREC video retrieval test collection (2001 TREC) [26]. The NTU consists of 146 video clips from YouTube or mobile cameras. The total size of these clips is 274 MB, and the average duration is 10.54 seconds. The second dataset consists of 11 hours of the publicly available MPEG-1 video provided by the TREC conference series. We experiment with 2G video clips, a subset of 2001 TREC, to evaluate the performance of our approach.

Evaluation Metric

Query images for retrieval are captured by OpenCV, an open-source library for computer vision. For evaluation, it is considered a visual match on condition that the query image and the retrieved frame are from the same video clip. Performance is measured in terms of accuracy:

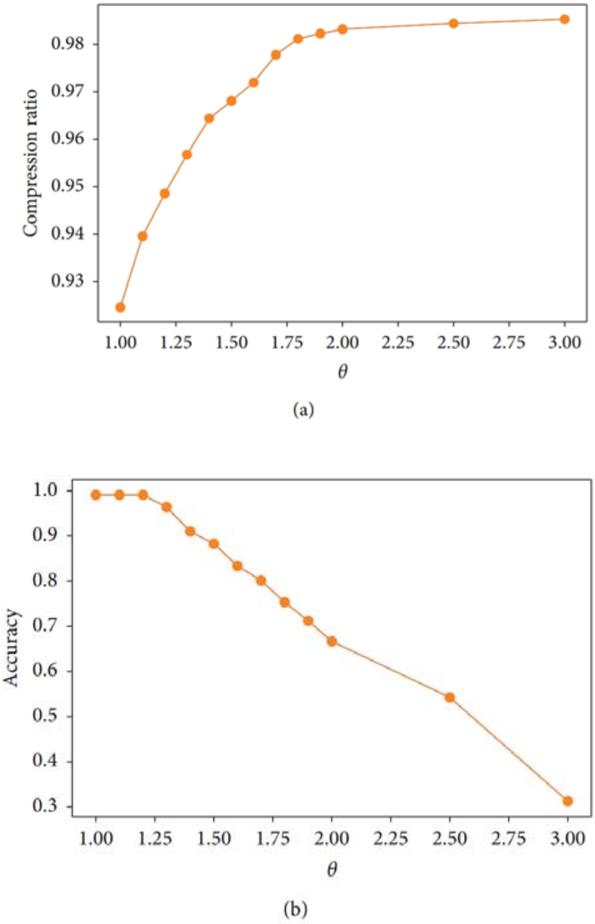
$$\mathrm{Acc} = \frac{\mathrm{no.\,(visual\ matches)}}{\mathrm{no.\,(retrieved\ frames)}}. \tag{2}$$

In order to show the performance variation, we test different parameter settings for our key-frame extraction algorithm. There is one parameter to be tuned in our proposed model: θ. The compression ratio is used to measure the compactness of the extracted key-frame sequence, which is defined as

$$\mathrm{compression\ ratio} = 1 - \frac{\mathrm{no.\,(key\ frames)}}{\mathrm{no.\,(frames)}}. \tag{3}$$

Parameter Setting

Figure 4 shows the compression ratio and retrieval accuracy variations with varying θ. When the value of θ is less than 2, the compression ratio improves dramatically with the increase in θ. After that, the compression ratio keeps steady and infinitely close but no more than 1. The higher the compression ratio, the more the redundant frames are lost and the more the storage space is saved. The accuracy keeps steady when the value of θ is less than 1.4. After that, the accuracy drops dramatically with the increase in θ.



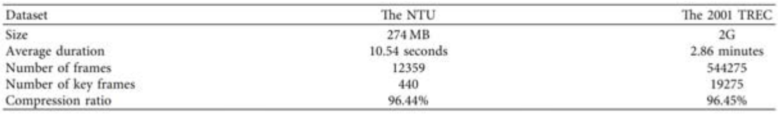
(a)

(b)

Figure 4. The compression ratio variations of different θ.

The accuracy is based on smaller θ. However, it also leads to a lower compression ratio, which will decrease the memory efficiency. We set the final value of θ to 1.4 by making a tradeoff between accuracy and efficiency. The summary of the information for the two datasets is shown in Table 2.

Table 2. The information of two datasets used for experiments.



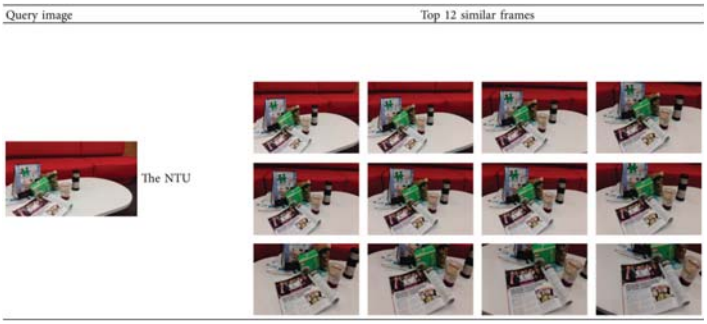
Dataset	The NTU	The 2001 TREC
Size	274 MB	2G
Average duration	10.54 seconds	2.86 minutes
Number of frames	12359	544275
Number of key frames	440	19275
Compression ratio	96.44%	96.45%

Experimental Results

To evaluate the performance of our proposed coarse-to-fine search method, we compare with several existing visual search approaches, which are briefly described as follows:(i) *Deep Feature-Based Method (DF)*. Babenko et al. [16] introduced features of pretrained CNN for image classification to replace traditional hand-crafted descriptors. We use the deep convolutional features from the last convolutional layer of VGG16 as a baseline method. (ii) Deep Feature Spatial Encoding (DFSE). Perronnin et al. [27] focused on encoding the deep convolutional features of CNN using the FV to generate frame descriptors. (iii)*Deep Feature Temporal Aggregation (DFTA)*. Noa et al. [28] proposed to aggregate the deep convolutional features of all frames within one shot via max-pooling. In DFTA, features in the same shot are aggregated into a single feature to reduce redundant information between adjacent frames. (iv) Local Binary Temporal Tracking (LBTT). LBTT [28] is based on the summarization of hand-crafted local binary features, which encode the pixel intensity value of frames into 256-dimensional binary vectors. (v)*Deep Feature Spatial Pooling (DFSP)*. To evaluate the performance of the pooling strategy, we used the 64-dimensional index of the frame for retrieval, which is generated after two pooling layers.

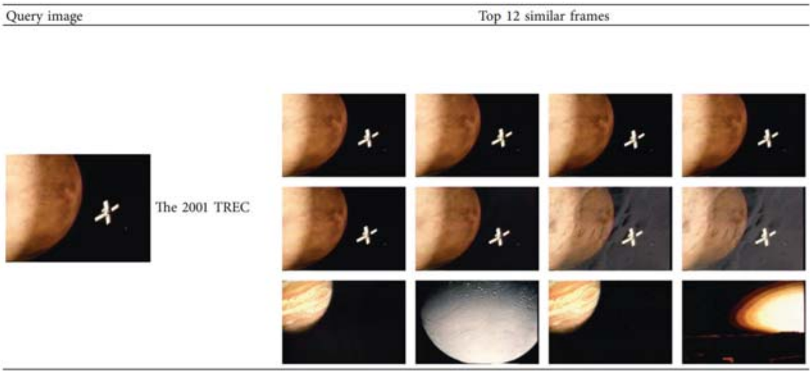
All the experiments are implemented on a computer which has Inter Core i5 2.3 GHz 2 processors, 8 GB RAM, and macOS 10. Tables 3 and 4 show the examples of our retrieval results on the two datasets.

Table 3. Example of the top 12 similar frames for the query image on the NTU.



Query image	Top 12 similar frames
The NTU	

Table 4. Example of the top 12 similar frames for the query image on the 2001 TREC.



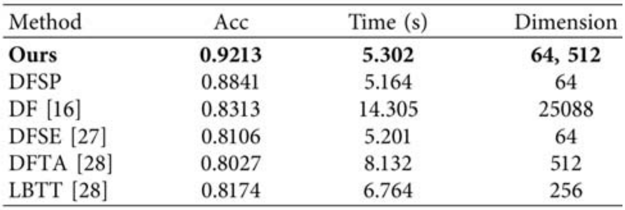
Results on the NTU Dataset

We first test different methods on the NTU. The accuracy, search time, and frame descriptors' dimension of different methods are presented in Table 5. Our method involves a coarse-to-fine retrieval process. In the coarse search, the dimension is 64, and in the fine search, the dimension is 512, which are described in the first line in Table 5. The proposed method achieves the best results in terms of accuracy, improving the performance by 0.05 compared to DF. DFSP and DFSE consume the shortest time without taking into account the time spent in offline training. This is probably because these frame descriptors are 64-dimensional, lower than that of the other methods. To further test the impact of the frame descriptors' dimension on the retrieval speed, we experiment on the large-scale dataset, 2001 TREC. The results of different methods are shown in Table 6.

Table 5. Comparison with existing visual search approaches on the NTU.

Method	Acc	Time (s)	Dimension
Ours	**0.9691**	**1.67**	**64, 512**
DFSP	0.9516	1.613	64
DF [16]	0.9198	1.892	25088
DFSE [27]	0.8441	1.606	64
DFTA [28]	0.8254	1.739	512
LBTT [28]	0.9096	1.693	256

Table 6. Comparison with existing visual search approaches on the 2001 TREC dataset.



Method	Acc	Time (s)	Dimension
Ours	**0.9213**	**5.302**	**64, 512**
DFSP	0.8841	5.164	64
DF [16]	0.8313	14.305	25088
DFSE [27]	0.8106	5.201	64
DFTA [28]	0.8027	8.132	512
LBTT [28]	0.8174	6.764	256

Results on the 2001 TREC Dataset

From Table 6, we can see that the accuracy of all methods is slightly reduced and search time is much longer compared to Table 5. The meaning of the dimension in Table 6 is similar to that in Table 5. For example, the dimensions in our method are 64 and 512, respectively. Our proposed method achieves the best results in terms of accuracy and outperforms other methods by large margins. Note that the accuracy for the proposed method is 0.9153 while that for DFTA is 0.7856. The search time of our method is slightly longer than that of DFSP because it takes time for fine-level search. Although the retrieval speed is slightly reduced, the retrieval accuracy is greatly improved. Therefore, we believe that our proposed coarse-to-fine search is effective. The accuracy of DF is worse than that of DF and DFSP. Furthermore, its search time is about 2-3 times longer than DFSP. This shows that the pooling strategy is effective in encoding deep local convolutional features. However, the accuracy of DFSE and DFTA is worse than DF although the search time is shorter. This indicates that although high-dimensional descriptors could be encoded into a lower-dimensional space via these two methods, they could lose a lot of feature information during the encoding process.

CONCLUSION AND FUTURE WORK

In this paper, we proposed a method based on deep local features to solve the problem of image-to-video retrieval. The models presented in this work are based on key-frame extraction and feature representation. The experimental results demonstrated that our method achieved competitive performance with respect to other CNN-based representations, as well as performed excellent in the cost of indexing and search time.

However, the proposed method appears to be more appropriate for tasks in which query images are from the original video frames. The quality problem of the query image caused by geometric transformations and occlusion might affect the search accuracy. In future work, we aim to explore an effective method to reduce the impact of image quality issues.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of PR China (no. 41571389), the National Key R&D Program of China (2017YFB1401302 and 2017YFB0202200), the Key Laboratory of Spatial Data Mining & Information Sharing of the Ministry of Education, Fuzhou University (no. 2016LSDMIS07), and the Technology Innovation and Application Demonstration Major Program of Chongqing (cstc2018jszx-cyztzxX0015).

REFERENCES

1. K. Lin, Y. Huei-Fang, H. Jen-Hao, and C. Chu-Song, “Deep learning of binary hash codes for fast image retrieval,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, MA, USA, June 2015.
2. M. A. Marzouk, “Improving web based image retrieval with fuzzy descriptors relevance feedback technique,” *Journal of Computers*, vol. 28, no. 3, pp. 11–26, 2017.
3. S. Poullot, T. Shunsuke, N. Anh Phuong, and J. Hervé, “Temporal matching kernel with explicit feature maps,” in *Proceedings of the 23rd ACM International Conference on Multimedia*, ACM, Brisbane, Australia, October 2015.
4. S. R. Shinde and G. G. Chiddarwar, “Recent advances in content based video copy detection,” in *Proceedings of the International Conference on Pervasive Computing (ICPC)*, IEEE, Pune, India, January 2015.
5. A. Araujo, “Large-scale query-by-image video retrieval using bloom filters,” 2016, http://arxiv.org/abs/1604.07939.
6. A. Araujo, “Temporal aggregation for large-scale query-by-image video retrieval,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, IEEE, Quebec City, QC, Canada, September 2015.
7. H. J. Zhang, D. Zhong, and S. W. Smoliar, “An integrated system for content-based video retrieval and browsing,” *Pattern Recognition*, vol. 30, no. 4, pp. 643–658, 1997.
8. Xu-D. Wu, L. Tie-Yan, L. Kwok-Tung, and F. Jian, “Dynamic selection and effective compression of key frames for video abstraction,” *Pattern Recognition Letters*, vol. 24, pp. 9-10, 2003.
9. Li-J. Qin, Z. Yue-Ting, W. Fei, and P. Yun-He, “An integrated framework for shot boundary detection with multi-level features similarity,” in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826)*, IEEE, Shanghai, China, China, August 2004.
10. H. Aoki, S. Shimotsuji, and O. Hori, “A shot classification method of selecting effective key-frames for video browsing,” in *Proceedings of the Fourth ACM International Conference on Multimedia*, ACM, Boston, MA, USA, Febuary 1997.

11. S. E. F. De Avila, A. P. B. Lopes, and A. de Albuquerque Araújo, "VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 56–68, 2011.
12. R. da Luz, B. Qin, and T. Liu, "A novel approach to update summarization using evolutionary manifold-ranking and spectral clustering," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2375–2384, 2012.
13. E. Mohedano, S. Amaia, M. Kevin, M. Ferran, N. E. O'Connor, and N. Xavier Giro-i, "Bags of local convolutional features for scalable instance search," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ACM, New York City, NY, USA, April 2016.
14. G. Csurka, R. D. Christopher, F. Lixin, W. Jutta, and B. Cédric, "Visual categorization with bags of keypoints," in *Proceedings of the Workshop on Statistical Learning in Computer Vision*, Meylan, France, 2004.
15. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 1, pp. 1097–1105, 2012.
16. A. Babenko, "Neural codes for image retrieval," in *Proceedings of the European Conference on Computer Vision*, Springer, Munich, Germany, September 2014.
17. A. Sharif Razavian, A. Hossein, S. Josephine, and C. Stefan, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, USA, June 2014.
18. H. Noh, "Large-scale image retrieval with attentive deep local features," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, October 2017.
19. Y. Kalantidis, C. Mellina, and O. Simon, "Cross-dimensional weighting for aggregated deep convolutional features," in *Proceedings of the European Conference on Computer Vision*, Springer, Munich, Germany, September 2016.
20. P. Kulkarni, Z. Joaquin, J. Frederic, P. Patrick, and C. Louis, "Hybrid multi-layer deep CNN/aggregator feature for image classification," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Brisbane, QLD, Australia, April 2015.

21. Y. Gong, W. Liwei, G. Ruiqi, and L. Svetlana, "Multi-scale orderless pooling of deep convolutional activation features," in *Proceedings of the European Conference on Computer Vision*, Springer, Munich, Germany, September 2014.
22. H. Jégou, D. Matthijs, S. Cordelia, and P. Patrick, "Aggregating local descriptors into a compact image representation," in *Proceedings of the CVPR 2010-23rd IEEE Conference on Computer Vision & Pattern Recognition*, San Francisco, CA, USA, June 2010.
23. M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, http://arxiv.org/abs/1312.4400.
24. J. Richiardi, H. Ketabdar, and A. Drygajlo, "Local and global feature selection for on-line signature verification," in *Proceedings of the Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, IEEE, Seoul, Korea, September 2005.
25. J. Meng, "Object instance search in videos via spatio-temporal trajectory discovery," *IEEE Transactions on Multimedia*, vol. 18, no. 1, pp. 116–127, 2015.
26. E. M. Voorhees, "Overview of the TREC 2001 question answering track," TREC, 2001.
27. F. Perronnin, L. Yan, S. Jorge, and P. Hervé, "Large-scale image retrieval with compressed Fisher vectors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, San Francisco, CA, USA, June 2010.
28. N. Garcia, "Temporal aggregation of visual features for large-scale image-to-video retrieval," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, ACM, Yokohama, Japan, June 2018.

CHAPTER 3

REGION SPACE GUIDED TRANSFER FUNCTION DESIGN FOR NONLINEAR NEURAL NETWORK AUGMENTED IMAGE VISUALIZATION

Fei Yang[1,2], Xiangxu Meng[1], JiYing Lang[2], Weigang Lu[3], and Lei Liu[4]

[1]School of Computer Science and Technology, Shandong University, Jinan 250101, China
[2]School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, 264209, China
[3]Department of Educational Technology, Ocean University of China, Qingdao, 266100, China
[4]The Institute of Acoustics of the Chinese Academy of Sciences, Beijing, 100190, China

ABSTRACT

Visualization provides an interactive investigation of details of interest and improves understanding the implicit information. There is a strong need today for the acquisition of high-quality visualization result for various fields, such as biomedical or other scientific field. Quality of biomedical

Citation: Fei Yang, Xiangxu Meng, JiYing Lang, Weigang Lu, Lei Liu, "Region Space Guided Transfer Function Design for Nonlinear Neural Network Augmented Image Visualization", Advances in Multimedia, vol. 2018, Article ID 7479316, 8 pages, 2018. https://doi.org/10.1155/2018/7479316 .

Copyright: © 2018 by Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

volume data is often impacted by partial effect, noisy, and bias seriously due to the CT (Computed Tomography) or MRI (Magnetic Resonance Imaging) devices, which may give rise to an extremely difficult task of specifying transfer function and thus generate poor visualized image. In this paper, firstly a nonlinear neural network based denoising in the preprocessing stage is provided to improve the quality of 3D volume data. Based on the improved data, a novel region space with depth based 2D histogram construction method is then proposed to identify boundaries between materials, which is helpful for designing the proper semiautomated transfer function. Finally, the volume rendering pipeline with ray-casting algorithm is implemented to visualize several biomedical datasets. The noise in the volume data is suppressed effectively and the boundary between materials can be differentiated clearly by the transfer function designed via the modified 2D histogram.

INTRODUCTION

Since there are the two characteristics of visibility of object and clear detail revealing, visualization has been proven to be of paramount important for exploring meaningful properties of volume data [1]. Because of the ability of obtaining the two-dimensional rendering results on the screen directly from the data field without building a network model in advance, volume rendering is thus certified to be an effective visualization method of extracting underlying information of interest from volumetric data using interactive graphics and imaging [2, 3]. Kniss et al. [4] visualized the muscle, soft tissues, and the bone from the visible male head data using the volume rendering method and produced a set of direct manipulation widgets to make exploring such features convenient. Ching and Chang [5] rendered the feature of interest in CT (Computed Tomography) images and generated a large wide-angle perspective projection view in an endoscopy to help a physician in diagnosis. Zhang et al. [1] synchronized the dual-modality of cardiac MRI (Magnetic Resonance Imaging) and 3D ultrasound volumes and visualized the dynamic heart by 4D cardiac image rendering. Based on volume rendering. Zhang and Wang et al. developed a platform integrating multivolume visualization method for both heart anatomical data and electrophysiological data visualization [6]. Hsieh et al. [7] visualized the three-dimensional (3D) geometry of the ear ossicle with the segmented ossicle computer tomography (CT) slices, which presented the spatial relation with the temporal bone to diagnose middle ear disease. To visualize brain activity conveniently, Holub and Winer [8] performed 3D and 4D volume ray casting on a tablet device in real-time.

Transfer function plays a fundamental role in visualization for its capability of classifying and segmenting features of volume data, which may affect the quality of rendering image and the perception of users to volume data. To measure through-plane MR flow. Thunberg et al. [9] presented a visualization method which combined the magnitude and velocity images into one single image. By using the transfer function, the velocities are color-coded and set to a predefined opacity. How the measured blood flow was related to the underlying anatomy can thus be understood. Zhang presented a statistics-based method to visualize 3D cardiac volume data set [10] and further proposed a novel transfer function design approach for revealing detailed structures in the human heart anatomy via perception-based lighting enhancement [11]. Yang presented a fusion visualization framework to combine the cardiac electrophysiology pattern with the anatomy pattern through a novel multidimensional fusion transfer function [12].

Ebert et al. [13] studied the accuracy of volume rendering for arterial stenosis measurement and the results suggested that the choice of transfer function parameters greatly affects the accuracy of volume rendering, while accurate transfer function parameters selection is still a challenge due to the lack of meaningful guidance information and intuitive user interface. The present methods for this problem mainly are object-centric, image-centric, and data-centric [14]. Object-centric approach first classifies or segments the volume data through clustering, probability, and machine learning which covers artificial neural network, support vector machine, and hidden Markov model [15–18]. Then the optical parameters are specified based on the classification result.

Different from the object-centric method, the image centric transfer function is designed on the rendered images. Through the evaluation of the projective images, parameters of transfer function are automatically adjusted and reapplied to the original data recursively until the satisfied rendering result is achieved. Based on a set of rendered images, He et al. [19] presented the stochastic method to search the satisfied transfer function. Marks et al. [20] proposed the Design Gallery method to provide the user varieties of ordered graphics or animations with different perceptions, which are generated automatically by a series of transfer functions given input parameter vector. Users then explore these images space to search for the satisfactory transfer function.

In data-centric approach, parameters of transfer function are specified by analyzing the volume data. Generally, collecting additional information related to the data prior to confirming transfer function makes the design more convenient. Scalar value of volume data is commonly considered for deriving 1D transfer function. The gradient [21, 22] and curve [23, 24] are introduced as the second variable for the two-dimensional transfer function. Roettger et al. [25] extended the variable of transfer function to spatial information. Spatial regions connected with each other were grouped and thus classified. Huang et al. [26] added spatial information into the transfer function domain and extended the number of dimensions to three. Material boundaries are accurately revealed by taking advantage of a designed cost function in three dimensions to imply regional growth algorithm. Some approaches were proposed to classify the topological structure of volume data for the transfer function design [27, 28]. Through the continuous scale-space analysis and detection filters, Correa et al. [29] obtained the 3D scale fields which represent the scale of every voxel. The size-based transfer function is then proposed using the scale fields and maps the scale of local features to color and opacity. Thus the features with similar scalar values in the complex data can be classified based on the relative size. With the increasing dimensions of transfer function, specifying its parameters properly becomes a more difficult and tedious task.

When the dimension is more than 2, it is difficult to specify the parameters for the higher dimensional transfer functions. The histograms are often used to find satisfied transfer functions. Based on the first and second derivatives in the volume, Kindlmann and Durkin [22] built a 2D histogram and the object boundaries appear as arcs in the histogram. A feature-sensitive transfer function can then be semiautomatically generated according to the arcs to reveal features of interest. Lum et al. [30] employed gradient-aligned samples instead of first derivatives as the first property for creating a variant of 2D histogram. The transfer function designed through the histogram classifies the voxels with different degree of homogeneity by mapping them to different optic parameters. However, with an increasing number of boundaries, their separation based on above methods becomes more difficult due to intersections and overlaps.

In this paper, firstly a neural network volume data preprocessing approach for slice denoising is implemented to improve the quality of 3D biomedical data. Then a two-dimensional transfer function with the preprocessed data is designed based on a modified 2D histogram, which is created using a novel region space based method with depth information. The features of interest in the data are thus exactly explored. In Section 2 of this paper, the method for denoising is described and a two-dimensional transfer function based on 2D histogram is designed. Efficiency and practicability of the presented method are further shown in Section 3. Finally, conclusions are discussed in Section 4.

TRANSFER FUNCTION DESIGN

Augmentation on Slice Data

Biomedical volume data produced by current noninvasive devices such as CT and MRI scanners are usually accompanied by noisy, partial effect, and bias. Data with serious noise or error message which causes low SNR (Signal Noise Ratio) will directly affect transfer function specification and cause the objects obscuring in the resulting image.

Spatial mean low-pass filtering such as General Median filtering and Gaussian smoothing has the advantage of reducing the amplitude of noise fluctuations. While the filtering blurs the details in the data such as the line or edge and does not focus on processing regional boundary or tiny structures, which makes the resulting image too fuzzy. This is a hamper to effectively enhance boundary for those noisy data containing lots of details.

Although nonlinear filtering has the achievement of reserving the edge, it produces the loss of resolution due to the suppression of details. To solve this problem, the non-linear enhancement algorithm uses information of boundary and the neighbor of a pixel to preprocess the image data [14, 31], which effectively removes the noise region with homogeneous physical properties and significantly improves the image quality. Loss of information can thus be minimized by reserving the object boundary and the detailed structure and the shape is enhanced by discontinuous sharpening. Since the anisotropic diffusion filtering smoothes the image along the edge direction rather than in the orthogonal direction to the edge, location and intensity of the edge can be retained. Unlike traditional methods, neural network can learn more features beneficial to the task through hierarchical structure. Based on a multilayer perceptron neural network, Burger et al. [32] presented a denoising algorithm that is learned on a large dataset for image denoising. For the multilayer perceptron (MLP), it can be represented as a nonlinear function which maps a noisy image to a noise-free image:

$$h(\mathrm{u}) = \alpha_3 + W_3 \cdot \Theta\left(\alpha_2 + W_2 \cdot \Theta\left(\alpha_1 + W_1 \cdot \mathrm{u}\right)\right) \tag{1}$$

Here the network has three hidden layers. $\alpha_1, \alpha_2, \alpha_3$ are vector-valued biases. The weight matrices of the structure are W_1, W_2, W_3. The function Θ operates component-wise. In order to realize image denoising, the clean images are selected from the image dataset and the input noise level is employed to produce the corresponding noisy images. The MLP parameters are then estimated by the back propagation algorithm satisfying:

$$\underset{\alpha,W}{\arg\min} \|h(\mathrm{u}) - \mathrm{v}\|^2 \tag{2}$$

where u is the vector-valued noisy images input and h(u) is the mapped vector-valued denoised images output. v is the clean images. The architecture of the network is as Figure 1.

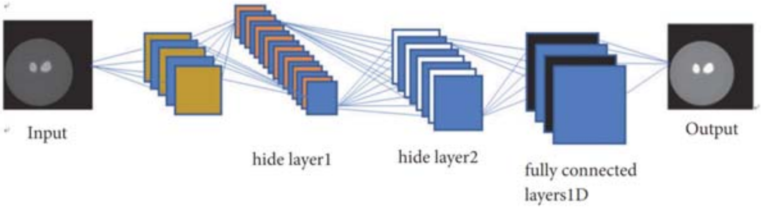


Figure 1. The architecture of the MLP network.

During application for image denoising, MLP uses fully connected neural network to process image fragments and then splits and combines all processed image segments to form a denoising image. First the noisy image is split into overlapping patches and each patch u is denoised separately. Then the denoised patches h(u) are placed at the locations of their noisy counterparts. The denoised image is thus obtained by averaging on the overlapping regions.

Region Space Guided Transfer Function Design

Kindlmann et al. [22] added higher-order derivatives of the voxel to transfer function domain in his presented approach. Those extracted boundaries appeared as arches in the derived histogram with axes representing scalar value and gradient magnitude. Although using this histogram can improve selection of boundaries, intersection or overlapping of two arches caused by different feature voxels sharing the same scalar value and gradient magnitude may result in ambiguities in classification of boundaries. Sereda et al. [33] proposed a multidimensional transfer function based on the LH histogram to facilitate separation of features which are represented by the arches.

LH Histogram based method computes low and high values of each sample voxels which are labeled as FL and FH respectively. For each sample voxel, if the gradient is less than the threshold, the voxel is identified to be internal sample. Otherwise the voxel is considered to be the boundary element. FL and FH of the internal voxel are equal and the value is the scalar of the voxel. For the voxels which are supposed to belong to boundaries, integration is implemented along the gradient and reverse direction in gradient field until the gradient is less than the threshold. FL and FH can then be found. The FL and FH values of all voxels are expressed in the same coordinate system, and the LH histogram is thus obtained. Since the value of FL is not more than FH, points in the LH histogram are only located above the diagonal line. The points on the diagonal indicate the internal voxels; that is, the FL and FH values are equal. The rest represent the boundary voxels and the corresponding FL and FH values are the scalar value of the two materials of the boundary respectively.

In volume rendering, using LH method to design transfer function can not only reduce the dependence on image segmentation, but also include voxel gradient information and boundary gray information. Due to the characteristics of medical data and clinical application, centralization of the voxel is required.

A proper region space Ω is selected and a voxel is compared with the voxels in Ω. Let V(p) be the scalar or intensity value of a voxel p. The intensity mean m and variance V of all voxels within Ω are given the following, respectively:

$$m = \frac{1}{n}\left(\sum_{p_i \in \Omega} V(p_i) + V(p)\right) \tag{3}$$

$$v = \frac{1}{n}\left(\sum_{p_i \in \Omega} \|V(p_i) - m\| + \|V(p) - m\|\right) \tag{4}$$

where p_i represents the adjacent voxel in the region space of voxel p and n is the number of voxels in Ω. The criteria for identifcation boundary voxels can then be formulated as in (5):

$$p \in S_{inner} \quad v < r$$

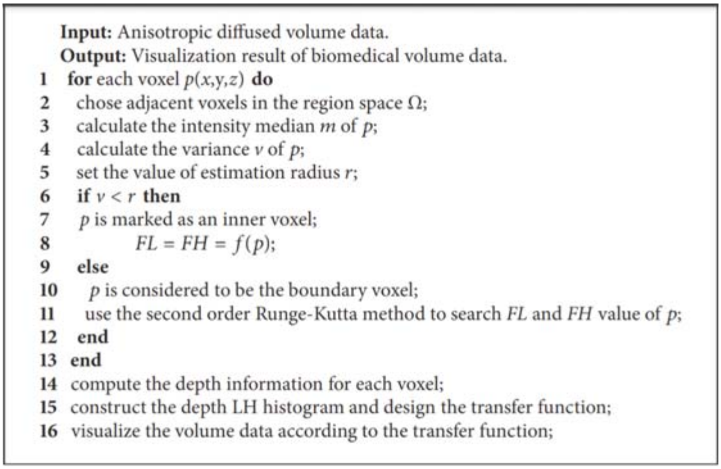
$$p \in S_{boundary} \quad otherwise \qquad (5)$$

where S_{inner} is the set of voxels inside the materials and $S_{boundary}$ is the set of voxels on the boundaries. Tus the voxel that diference between it and the voxels in Ω falls out of the range of r is considered to the boundary voxel.

In the region where the complex boundary exists, using the single criterion will result in boundary determination error. Since some boundaries only appear at a certain depth and then disappear when they reach a certain depth, the complex boundaries can be further differentiated according to the depth information. Then a modified 2D histogram is created using the region space-based method with depth information. In this paper the points in the original histogram are further grouped according to the corresponding depth.

A 2D transfer function can then be specified based on the created LH histogram by selecting relevant areas and by assigning them color and opacity. The corresponding features in the volume data can thus be explored. The details of the proposed method are given in Algorithm 1.

Algorithm 1. Region space guided visualization.

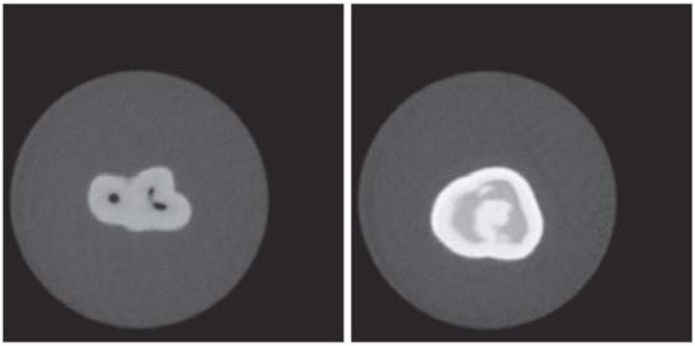


```
Input: Anisotropic diffused volume data.
Output: Visualization result of biomedical volume data.
for each voxel $p(x,y,z)$ do
    chose adjacent voxels in the region space $\Omega$;
    calculate the intensity median $m$ of $p$;
    calculate the variance $v$ of $p$;
    set the value of estimation radius $r$;
    if $v < r$ then
        $p$ is marked as an inner voxel;
            $FL = FH = f(p)$;
    else
        $p$ is considered to be the boundary voxel;
        use the second order Runge-Kutta method to search $FL$ and $FH$ value of $p$;
    end
end
compute the depth information for each voxel;
construct the depth LH histogram and design the transfer function;
visualize the volume data according to the transfer function;
```

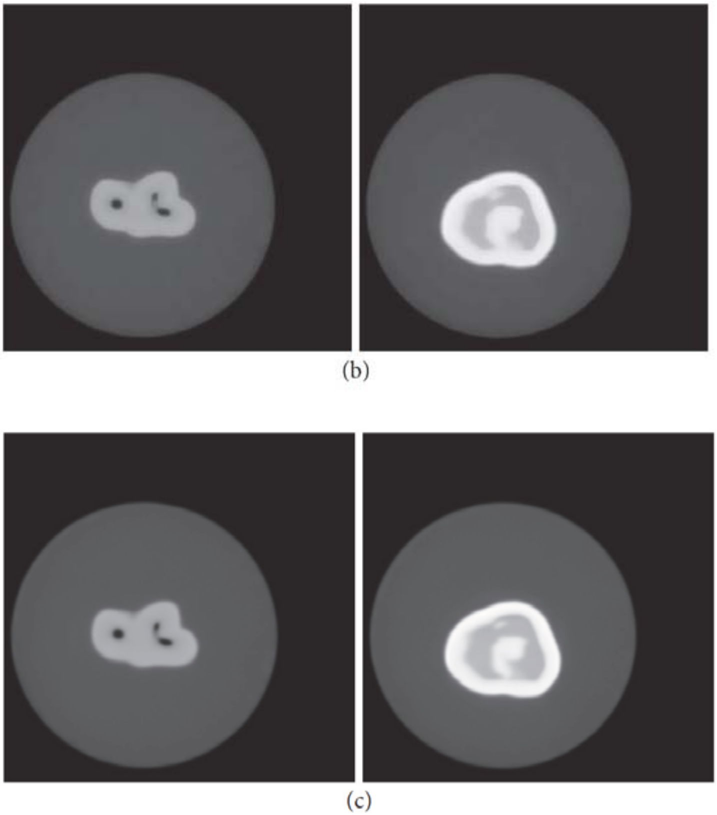
RESULTS AND DISCUSSION

In this section, some data sets are used as the test data, including tooth data and sheep heart data, to evaluate the performance of the proposed transfer function. The size of data set is 256×256×161 and 352×352×256, respectively. All the experiments are carried out on the computer with Intel Core i5 2.66G, 4.00G RAM and graphics card of NVIDIA GeForce GT 650.

Biomedical volume data produced by current noninvasive devices such as CT and MRI scanners are usually accompanied by serious noise, which will generate poor visualized image and cause the blur objects in the resulting image. Thus the MLP neural network is implemented to denoise the volume data. In the experiments, the union of the LabelMe dataset is used to train MLP which contains approximately 150,000 images. Before training, the data are filled with padding operation and each pixel is filled with 6 pixel sizes. The noise level σ is set to 10. We use a patch of size 39×39 to generate the predicted patch and then adopt a filter of size 9×9 to average the output patches, thus its effective patch size is 47×47. In the experiment the learning rate r in each layer is equal to r/N and N is the number of input units of current layer. The basic learning rate was set to 0.1. To improve results slightly we use the sliding window method with stride size of 3 which weights denoised patches with a Gaussian window instead of using all possible overlapping patches. Figure 2 compares the image denoising results of nonlinear enhancement and MLP neural network on the tooth data. The Peak Signal to Noise Ratio (PSNR) with the nonlinear enhancement filtering is 41.4294, and the Structural Similarity Index (SSIM) is 0.9325. The PSNR with the MLP method is 41.4294, and SSIM is 0.9325. As one can see, the MLP network produces more visually pleasant results. Compared with original data, noise in the homogeneous region is suppressed and the quality of the image is improved.



(a)

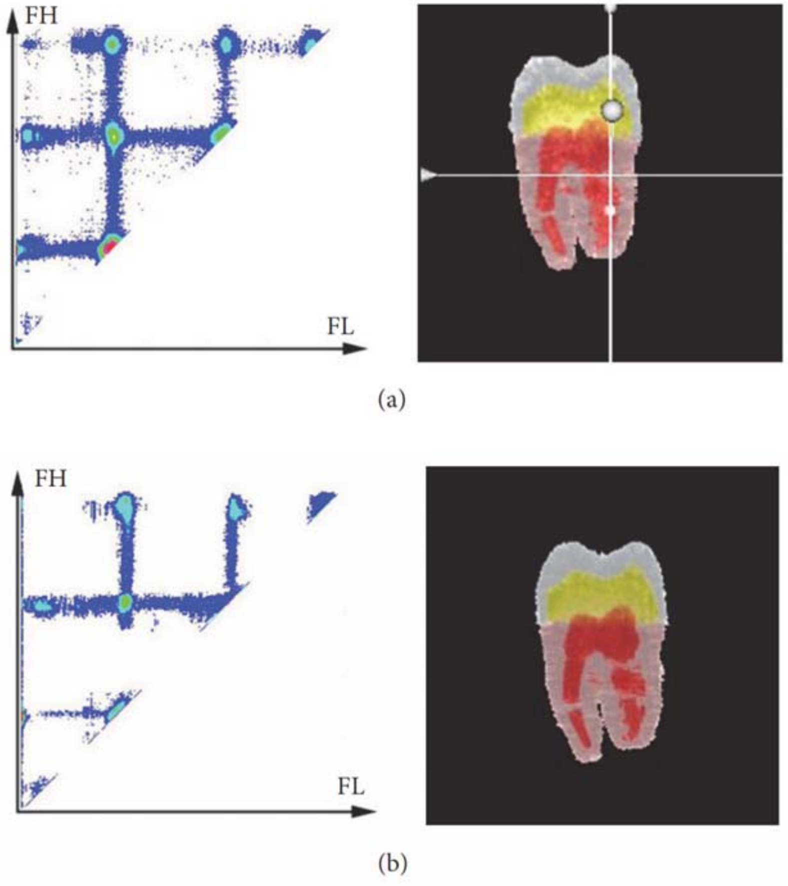


(b)

(c)

Figure 2. Preprocessing results of tooth dataset. From top to bottom: original slice data, denoised images by the nonlinear enhancement algorithm, and denoised images by MLP. (a) Original slice data; (b) nonlinear enhancement denoised result; (c) MLP denoised result.

In LH histogram, points around diagonal represent interior of materials. Regions including those points are thus assigned to lower opacity in the transfer function to fade unimportant information out. Remainder regions are the accumulation of boundary voxels which contain features of interest. Figure 3 shows LH histogram and rendering result of the original tooth data and MLP denoised data. Figure 3(a) describes the created LH histogram (left) and the corresponding rendering result with original tooth data set (right). Figure 3(b) shows the LH histogram (left) based on the denoised data. From Figure 3(b), we can see the more compact separation of points in the LH histogram, which is on account of the suppression of noise in the homogeneous region and enhancement of the boundaries between materials. Since the noise is suppressed effectively, boundaries between different materials can be visualized more clearly. As shown in Figure 3(b), the fact which is specifically manifested through the experiment result is that dentine-enamel (yellow) is explored exactly and noises around the root of the dentine are considerably removed.

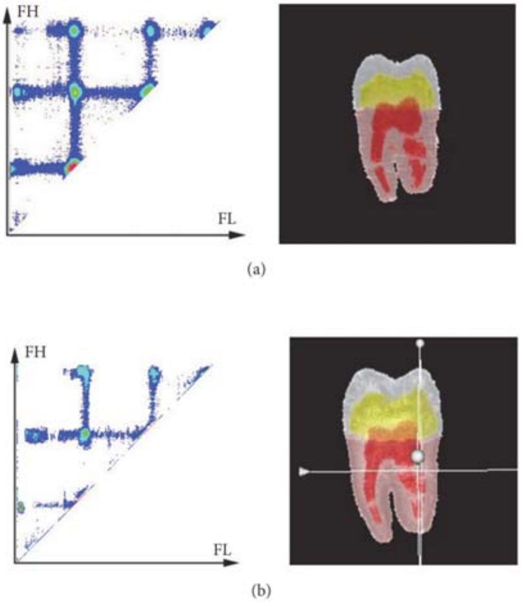


(a)

(b)

Figure 3. LH histogram and rendering result of the original tooth data and MLP denoised data: (a) LH histogram and the corresponding rendering result of original data and (b) LH histogram and corresponding rendering result of denoised data.

Figure 4 shows the created LH histogram and rendering result on the anisotropic diffusion enhanced tooth volume data using the conventional and regional criteria based method. The number of iteration of the nonlinear filtering is the process ordering parameter. Figure 4(a) presents the visualization result of enhanced data using the conventional method. The gradient magnitude threshold for investigating the *FL* and *FH* intensity profile is set to 10. Figure 4(b) shows the LH histogram constructed and corresponding rendering images through regional criteria based method. Here the region range r is set to 1.6. As shown in Figure 4(b), since the noise is suppressed and the path tracing for *FL* and *FH* value starts with regional criteria, the distribution of separated parts of points that correspond to different features is more concentrate in LH histogram, which thus ensures a more accurate identification of boundary voxels. The fact which is specifically manifested through two experiment results is that noises around the root of the dentine are considerably removed, and the phenomenon of blurred boundary is removed and various boundaries of the tooth, i.e., enamel-air (white), dentine-enamel (yellow), pulp-dentine (red), and dentine-air (pink) boundary, are revealed clearly in the final image.



(a)

(b)

Figure 4. LH histogram and corresponding rendering result with nonlinear enhanced tooth dataset through the conventional and region criteria-based method: (a) LH histogram based on conventional method with gradient threshold of 5 and the corresponding rendering result and (b) LH histogram based on region criteria based method and the rendering result.

Figure 5 shows the rendering result after introducing depth information into region space with the MLP augmented data. Classifying boundary voxels through conventional LH histogram will result in the confused boundary exploration. From Figure 4, we can see that there exists a visible discontinuity in the pulp-dentine boundary. This discontinuity is due to classifying those boundary voxels of the dentine tissue falsely. Because the boundary appears at a specific depth, for example, the tooth enamel is in the region near the human eye, while the medulla is at the depth far away from the human eye, thus we add depth information for the histogram construction and can obtain the distinct boundary. In our experiment the depth of enamel is about 120 and the depth of medulla is about 80. As shown in the result image in Figure 5(a), since more voxels are classified into boundary voxels via the transfer function which is designed based on LH histogram with depth information in region space properly, the discontinuity of the pulp-dentine boundary (red) in Figure 4 is corrected. And the exact pulp-dentine boundary is revealed in the rendering result image. The rendering time of the proposed method is 1.1s, which enriches real-time interaction property of visualization.

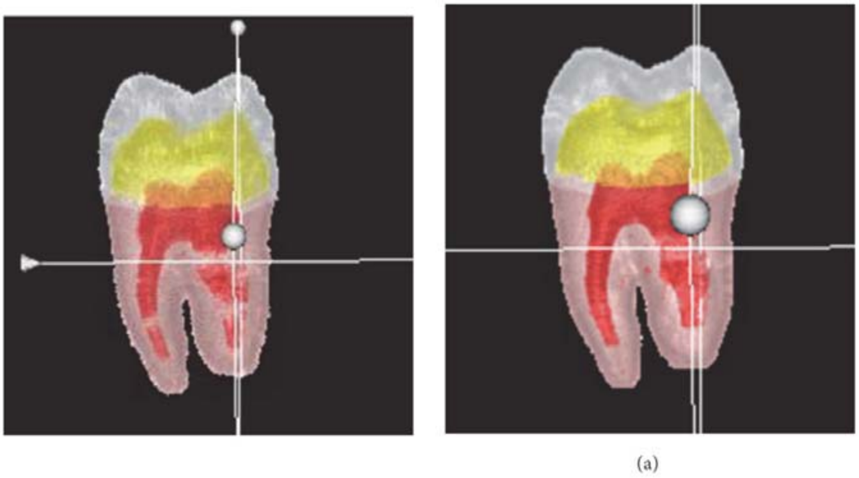
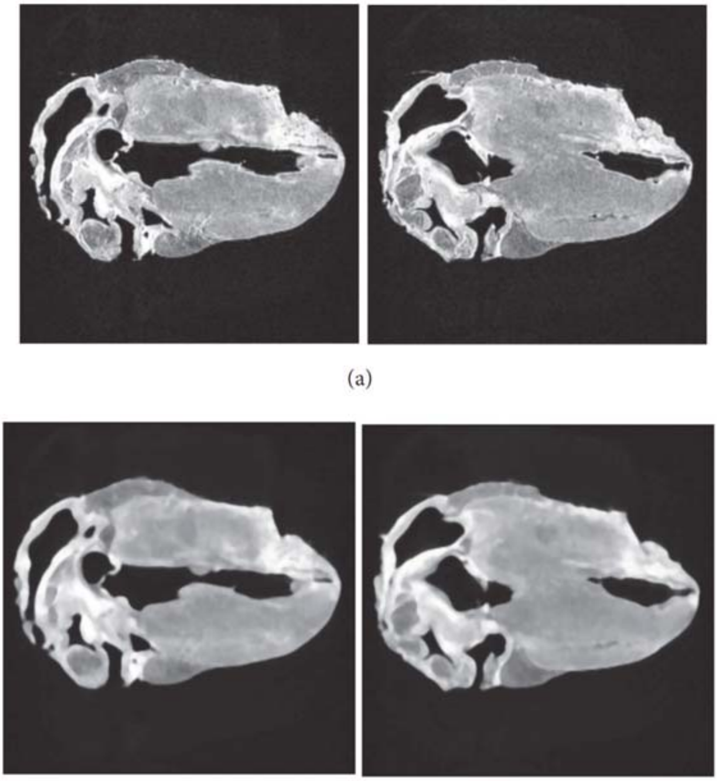


Figure 5. Rendering result of MLP augmented tooth dataset with two methods: (a) rendering result based on region criteria-based method and (b) rendering result based on depth enhanced method.

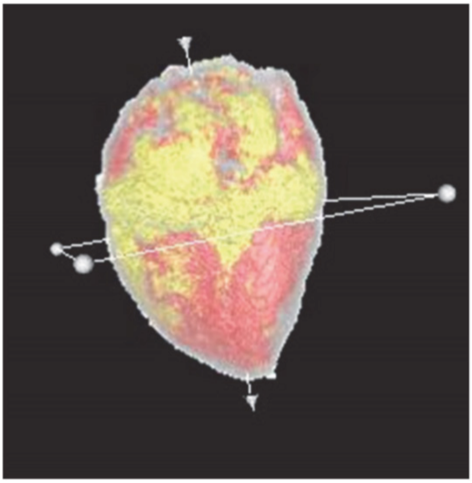
Figure 6 shows the denoising result of sheep heart slice data and gives the rendering result via the proposed region space guided transfer function.

Figure 6(a) shows the original slices. In Figure 6(b), the two corresponding denoised results are presented. It is obvious that preprocessing for tissues of the sheep heart such as the muscle and the fat is effective in decreasing noise and the boundary details are consequently remained. Fine structural features of interesting in the sheep heart are thus clearly visualized through the region space based transfer function with the augmented data and, as in Figure 6(c), the fat of sheet heart is colored in yellow and the muscle is red. The profile of sheep heart is colored by white. From the result, structures of the sheep heart are thus exactly explored and the shape, spatial position, and relationship of the tissues can be observed without ambiguity. The rendering time of the sheep heart data set is 2.6s.



(a)

(b)



(c)

Figure 6. Visualization result of augmented sheep heart dataset with region space guided transfer function: (a) the original sheep heart data; (b) the denoised data; (c) rendering result of sheep heart with the denoised data.

CONCLUSION

Transfer function in performance of volume rendering plays a crucial role for exploring directly detail information hiding in data as well as enhancing important boundaries. In this work we first implement the MLP neural network on volume data to denoise while preserve the boundary. This method can considerably improve quality of volume data acquired by devices. Then we improve the LH method by combining the regional depth information to achieve the transfer function semiautomatic generation. This method can avoid the influence of noise and make the voxels more centralized. In the LH histogram the voxel distribution at the diagonal line is more concentrated, and the boundary of important objects are effectively emphasized. The features of interest in the data can thus be found exactly by mapping scalar value of boundary voxels which correspond to the points in LH histogram to appropriate opacity and color.

ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China (NSFC) under Grant no. 61502275 and the Postdoctoral Science Foundation of China (no. 2017M622210). This work was also supported in part by the Natural Science Foundation of Shandong of Grant no. ZR2017MF051, the National Natural Science Foundation of China (NSFC) of Grant no. 61501450, and the MOE (Ministry of Education in China) Project of Humanities and Social Sciences of Grant no. 16YJC880057.

REFERENCES

1. Q. Zhang, R. Eagleson, and T. M. Peters, “GPU-based visualization and synchronization of 4-D cardiac MR and ultrasound images,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 5, pp. 878–890, 2012.
2. H. Pfister, B. Lorensen, C. Bajaj et al., “The transfer function bake-off,” *IEEE Computer Graphics and Applications*, vol. 21, no. 3, pp. 16–22, 2001.
3. Q. Zhang, R. Eagleson, and T. M. Peters, “Volume visualization: A technical overview with a focus on medical applications,” *Journal of Digital Imaging*, vol. 24, no. 4, pp. 640–664, 2011.
4. J. Kniss, G. Kindlmann, and C. Hansen, “Multidimensional transfer functions for interactive volume rendering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, no. 3, pp. 270–285, 2002.
5. Y.-T. Ching and C.-L. Chang, “A volume rendering technique to generate a very large wide-angle endoscopeic view,” *Journal of Medical and Biological Engineering*, vol. 22, no. 2, pp. 109–112, 2002.
6. L. Zhang, C. Gai, K. Wang, W. Lu, and W. Zuo, “GPU-based high performance wave propagation simulation of ischemia in anatomically detailed ventricle,” in *Proceedings of the Computing in Cardiology Conference (CinC ‘11)*, pp. 469–472, Hangzhou, China, September 2011.
7. M. S. Hsieh, F. P. Lee, and M. D. Tsai, “A virtual reality ear ossicle surgery simulator using three-dimensional computer tomography,” *Journal of Medical and Biological Engineering*, vol. 30, no. 1, pp. 57–63, 2010.
8. J. Holub and E. Winer, “Enabling Real-Time Volume Rendering of Functional Magnetic Resonance Imaging on an iOS Device,” *Journal of Digital Imaging*, vol. 30, no. 6, pp. 738–750, 2017.
9. P. Thunberg and A. Kähäri, “Visualization of Through-Plane Blood Flow Measurements Obtained from Phase-Contrast MRI,” *Journal of Digital Imaging*, vol. 24, no. 3, pp. 470–477, 2011.
10. L. Zhang, K. Wang, F. Yang et al., “A Visualization System for Interactive Exploration of the Cardiac Anatomy,” *Journal of Medical Systems*, vol. 40, no. 6, 2016.
11. L. Zhang, K. Wang, H. Zhang, W. Zuo, X. Liang, and J. Shi, “Illustrative cardiac visualization via perception-based lighting

enhancement," *Journal of Medical Imaging and Health Informatics*, vol. 4, no. 2, pp. 312–316, 2014.

12. F. Yang, W. G. Lu, L. Zhang, W. M. Zuo, K. Q. Wang, and H. G. Zhang, "Fusion visualization for cardiac anatomical and ischemic models with depth weighted optic radiation function," in *Proceedings of the Computing in Cardiology Conference (CinC '15)*, pp. 937–940, IEEE, Nice, France, September 2015.
13. D. S. Ebert, D. G. Heath, B. S. Kuszyk et al., "Evaluating the potential and problems of three-dimensional computed tomography measurements of arterial stenosis," *Journal of Digital Imaging*, vol. 11, no. 3, pp. 151–157, 1998.
14. G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Transactions on Medical Imaging*, vol. 11, no. 2, pp. 221–232, 1992.
15. F.-Y. Tzeng and K.-L. Ma, "A cluster-space visual interface for arbitrary dimensional classification of volume data," in *Proceedings of the in Proceedings of the 6th Joint Eurographics-IEEE TCVG Symposium on Visualization*, pp. 17–24, Konstanz, Germany, 2004.
16. P. Šereda, A. Vilanova, and F. A. Gerritsen, "Automating transfer function design for volume rendering using hierarchical clustering of material boundaries," in *Proceedings of the in Proceedings of the 8th Joint Eurographics-IEEE TCVG Symposium on Visualization*, pp. 243–250, Lisbon, Portugal, 2006.
17. F.-Y. Tzeng, E. B. Lum, and K-L. Ma, "A novel interface for higher-dimensional classification of volume data," in *Proceedings of the IEEE Visualization Conference (VIS '03)*, pp. 505–512, Seattle, WA, USA, 2003.
18. F.-Y. Tzeng, E. B. Lum, and K.-L. Ma, "An intelligent system approach to higher-dimensional classification of volume data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 3, pp. 273–283, 2005.
19. T. S He, L. C. Hong, A. Kaufman, and et al, "Generation of transfer functions with stochastic search techniques," in *Proceedings of the Seventh Annual IEEE Visualization '96*, pp. 227–234, San Francisco, CA, USA.
20. J. Marks, B. Mirtich, B. Andalman et al., "Design Galleries: A general approach to setting parameters for computer graphics and animation,"

in *Proceedings of the 1997 Conference on Computer Graphics, SIGGRAPH*, pp. 389–400, August 1997.

21. M. Levoy, “Display of surfaces from volume data,” *IEEE Computer Graphics and Applications*, vol. 8, no. 3, pp. 29–37, 1988.
22. G. Kindlmann and J. W. Durkin, “Semi-automatic generation of transfer functions for direct volume rendering,” in *Proceedings of the 1998 IEEE Symposium on Volume Visualization, VVS 1998*, pp. 79–86, USA, October 1998.
23. J. Hladůvka, A. König, and E. Gröller, “Curvature-based transfer functions for direct volume rendering,” in *Proceedings of the In Spring Conference on Computer Graphics 2000*, vol. 16, pp. 58–65, 2000.
24. G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Möller, “Curvature-Based Transfer Functions for Direct Volume Rendering: Methods and Applications,” in *Proceedings of the VIS 2003 PROCEEDINGS*, pp. 513–520, USA, October 2003.
25. S. Roettger, M. Bauer, and M. Stamminger, “Spatialized transfer functions,” in *Proceedings of the In Eurographics, IEEE VGTC Symposium on Visualization*, pp. 271–278, 2005.
26. R. Huang, . Kwan-Liu Ma, P. McCormick, and W. Ward, “Visualizing industrial CT volume data for nondestructive testing applications,” in *Proceedings of the IEEE Visualization 2003*, pp. 547–554, Seattle, WA, USA.
27. I. Fujishiro, T. Azuma, and Y. Takeshima, “Automating transfer function design for comprehensible volume rendering based on 3D field topology analysis,” in *Proceedings of the IEEE Visualization ‘99*, pp. 467–470, October 1999.
28. S. Takahashi, Y. Takeshima, and I. Fujishiro, “Topological volume skeletonization and its application to transfer function design,” *Graphical Models*, vol. 66, no. 1, pp. 24–49, 2004.
29. C. D. Correa and K.-L. Ma, “Size-based transfer functions: a new volume exploration technique,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1380–1387, 2008.
30. E. B. Lum and K.-L. Ma, “Lighting transfer functions using gradient aligned sampling,” in *Proceedings of the IEEE Visualization 2004 - Proceedings, VIS 2004*, pp. 289–296, USA, October 2004.

31. P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
32. H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising: Can plain neural networks compete with BM3D?” in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2012*, pp. 2392–2399, USA, June 2012.
33. P. Šereda, A. V. Bartrolí, I. W. O. Serlie, and F. A. Gerritsen, “Visualization of boundaries in volumetric data sets using lh histograms,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 2, pp. 208–217, 2006.

CHAPTER 4

STUDY ON DATA-DRIVEN METHODS FOR IMAGE AND VIDEO UNDERSTANDING

Tatsuya Yamazaki

National Institute of Information and Communications Technology Japan

INTRODUCTION

Owing to progress of broadband Internet communication infrastructure, people can enjoy multimedia services. Among the services, images and videos are making an appeal and people desire to find out what they want to see. But they do not always make a success of searching and it sometimes takes plenty of time to reach their targets. A solution is to attach information tags according to the image or video content. Since image and video submission into Internet is increasing day by day, manual tag attachment is almost impossible. Development of automatic tag attachment is an urgent

Citation: Tatsuya Yamazaki (February 1st 2010). Study on Data-driven Methods for Image and Video Understanding, Multimedia, Kazuki Nishi, IntechOpen, DOI: 10.5772/7650.

Copyright: © 2010 by authors and IntechOpen. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

theme for future Internet service. Another imaging technology in the real world is environmental cameras. The environmental cameras mean the cameras set in the environment such as a ceiling or a street comer. One can easily imagine surveillance cameras as an example of the environmental cameras and sometimes they are considered to be identical. In this paper, I want to segregate them because the purpose of the environment cameras is not only to keep a watch on accidents or crimes but to understand peoples' situations and behaviours. Namely the surveillance cameras are a subset of the environmental cameras.

For both of the above cases, the key point is understanding image and video. It is one of the issues that have been discussed for a long time and related with the artificial intelligence technology. Therein image clustering and object extraction are most essential technologies. Image clustering means to divide an image into several segmented regions in a way of unsupervised, which was used for object extraction (Meier & Ngan, 1998), image compression (Kunt, 1988), or image categorization. Since it is hard to say how many regions are included in an image in generate there are a few studies that estimated the number of regions. Usually the number of regions has been assumed to be known. As a previous research in which the number of region was estimated, Won and Derin (Won & Derin, 1992) proposed to use the Akaike Information Criterion to determine the number of regions. But these are model-based approaches, that is, how to select a suitable model is important. Regarding with the environmental cameras, there have been previous works to extract moving object in order to capture human behaviours. Satake and Shakunaga (Satake & Shakunaga, 2004) proposed an appearance-based condensation tracker, which is composed of a condensation tracker and a sparse template matching method to detect the movement of people with a camera. The template-based condensation tracker is stabilized for tracking even in the case of object occlusion. Thonnat and Rota (Thonnat & Rota, 1999) used low-level image processing techniques to detect and track mobile objects. Their aim was rather to understand images, namely, to generate alarms automatically for operators when interesting scenarios had been recognized by the system. In both of the above works, they used only one camera. On the contrary, Matsuyama (Matsuyama, 1999) proposed a protocol for negotiation among multiple environmental cameras. Because of the protocol they could synchronize several cameras in real time and grab the human behaviours more smoothly and more seamlessly.

As I described in the above paragraphs, automatic processing of image and video media should be deployed more in future Internet services for the information explosion era. Technically adaptation of processing based on the observed data, which is called data driven, is necessary. Therefore, in this paper, data-driven image clustering and object detection methods are proposed.

Data-Driven Image Clustering

In this section, a data-driven clustering algorithm for colour images is proposed based on a multi-dimensional histogram. A statistical model is introduced and the image data is assumed to be derived from a mixture of multi-variant distributions.

Multi-Dimensional Histogram

Although the R-G-B colour space is assumed to make a discussion simple, the proposed method can be applied to other colour spaces. The R-G-B colour image data is represented as $y=(y_1, \dots , y_{Np})$, where $y_i=(y_i^R, y_i^G, y_i^B)$ $(i=1, \dots , N_p)$ is the observed element at the *i-th* pixel *Np* is the total number of pixels, y_i^X is a scalar value observed on the X plane at the *i-th* pixet and y_i^X is assumed to range from G_{min} to G_{max1} where X is K G, or *B*.

The multi-dimensional histogram is formed easily. First, distribute the observed data into the R-G-B color space. Second, construct a complete set of nonoverlapping intervals, called bins, by dividing the cube ([G_{min}, G_{max}]) equally with a width *h*. Finally, count the number of elements in each bin. Fig. 1 shows a construction of a multi-dimensional histogram with $G_{min}=O$. The histogram width *h* is an important parameter, and relates to the data distribution. When the data comes from a single density, several rules have been proposed to determine *h*. When the data are obtained a mixture of densities and the number of densities, *NCJ* is unknown, it is difficult to determine *h*. The multi-dimensional histogram of the color image data deal with in this study corresponds to the latter case.

Data-Driven Clustering Algorithm

It is assumed that there are a set of candidates for the histogram width, that is, $\{H=h_1, h_2, \dots\dots\dots\dots, h_c\}$. Here, a novel algorithm is proposed to determine hand *Nc*.

Step 1) Select a candidate *hj (j=1, ...* , C) from *H.* Construct a histogram with a width *hj.* Select the bins that have at least one element in the histogram, and sort them in the order of the number of elements in each bin. The sorted bins are numbered in the order of the number of elements as $b_1^j, b_2^j, ..., b_{nz}^j$, where b_{nz}^j is the number of bins having at least one element. The *k-th* b_k^j, has n_k^j elements (Fig. 2).

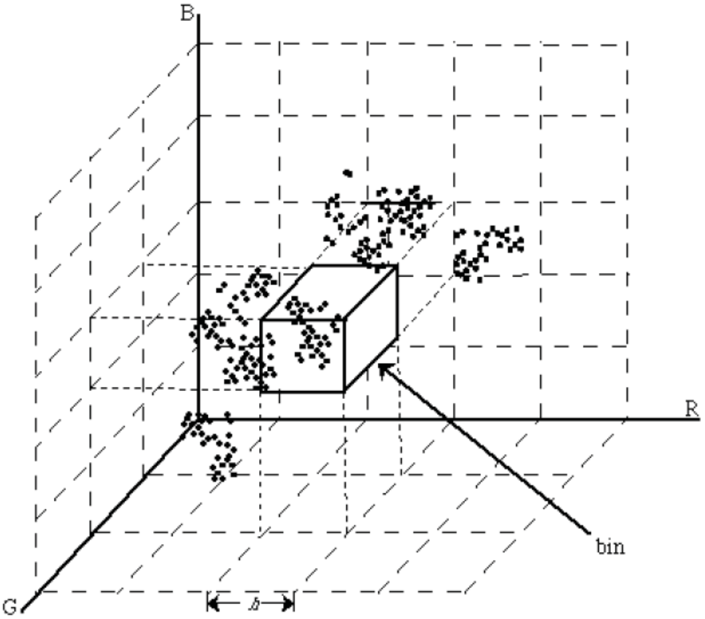


Figure 1. Multi-dimensional histogram

Step 2) If $n_k^j < n_{cut}^j$, then b_k^j is removed. n_{cut}^j is a threshold calculated as

$$n_{cut} = \alpha_{h_j} \frac{b_{nz}}{N_p} \tag{1}$$

where α_{h_j} is a control parameter to be determined for each l_j heuristically. n_{cut} is the threshold that divides explicitly insignificant b_{ins}. Eventually b_{cut}^J bins remain.

Step 3) Extract the top b_{sig}^j bins as significant bins from b_{cut}^J bins, where b_{sig}^j is determined as follows.

$$b_{sig}^j = arg \max_{k=1,\dots,b_{cut}^j} Cr_{sig}(k) \tag{2}$$

$$\mathrm{Cr_{sig}}(k) = \sum_{l=1}^{k} \frac{\sum_{m=1}^{b_{cut}^{j}} n_m^j}{n_l^j} - \frac{b_{cut}^{j}}{k} \qquad (3)$$

$Cr_{sig}{}^{(k)}$ is a criterion to select significant bins that include as many elements as possible considering suppression of the number of selected bins.

Step 4) If $b_{sig}{}^{j}$ for every histogram-width candidate h_j is calculated, then go to Step 5. Otherwise, go to Step 1and pick up another candidate whose $b_{sig}{}^{j}$ has not been calculated.

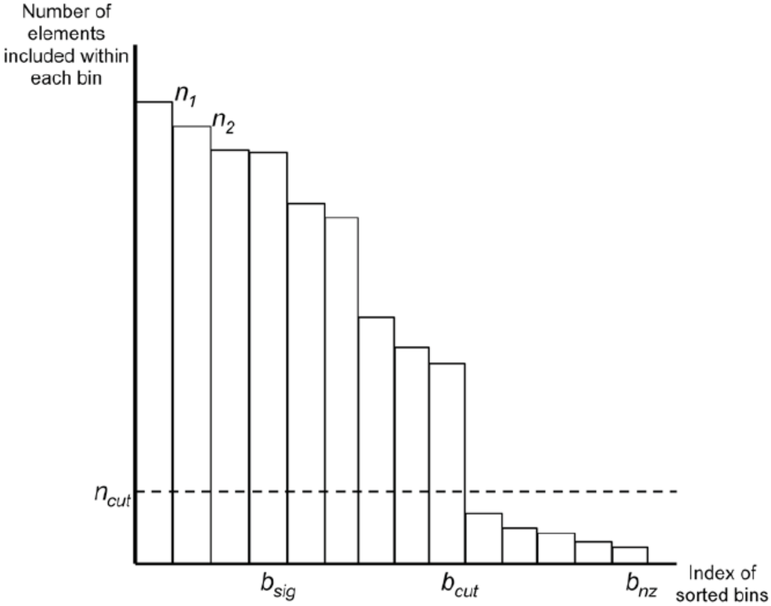


Figure. 2. Sorted frequency distribution of elements within each bin

Step 5) Calculate the optimal histogram width h^* as

$$\mathrm{h^*} = arg \max_{\mathrm{h_j \in H}} \sum_{l=1}^{b_{sig}^{j}} n_l^j \qquad (4)$$

where $b_{sig}{}^{j}$ corresponds to h_j, and $b_{sig}{}^{j}$ corresponding to h^* is denoted as $b_{sig}{}^{*}$

Consequently h^* is selected from Has the optimal width from the perspective of extracting as many significant elements as possible to compute the distribution statistics.

The significant $b_{sig}*$ bins are determined beforehand without considering the mutual relationships. Therefore, the adjacent bins that are supposed to belong to the same density must be merged to determine the final cluster count *Nc.* The clustering criterion is as follows.

Calculate the average $(\mu_k^R, \mu_k^G, \mu_k^B)$ and the standard deviation $(\sigma_k^R, \sigma_k^G, \sigma_k^B)$ for the *k-th* bin *(k=1, ..., b_{sig}*).*

Selectanytwobins,say b_l and b_m, from b_{sig} *bins. If $|\mu_l^R - \mu_m^R| < \beta\overline{\sigma^R}$, $|\mu_l^G - \mu_m^G| < \beta\overline{\sigma^G}$ and $|\mu_l^B - \mu_m^B| < \beta\overline{\sigma^B}$, then the two bins belong to the same cluster, where $\overline{\sigma^X} = \sum_{i=1}^{b_{sig}^*} \sigma_i^X / b_{sig}^*$ for X=R, G, or, Band β is a parameter. Finally, cluster the b_{sig}* bins into *Nc* clusters.

Proposed Algorithm Application to Real Data

The proposed algorithm was applied to real image data. Fig. 3 shows one of the original images, which is called the "Lady with a rose" image in this paper. The image size is **480 X 480.** Although it is shown in grey scale, the original color space is the RGB color space. The histogram width candidates set used in the algorithm is H={4, 8, 16, 32}. The values of α_{hi} corresponding to each histogram width are α_4 =1.0, α_8=0.7, α_{16}=0.35 and α_{32} =0.05. In this experiment, β is set to 0.0; this means that the merging process is skipped. Applied the proposed algorithm, the histogram width was determined to be 32 and the number of clusters was 6 finally. These values were determined in a data-driven way.

Then, the statistics of each cluster were computed by the minimum distance method, and the conventional maximum likelihood method with the estimated statistics, under the normal distribution assumption, was performed to obtain the segmented image. The final result followed by the 3X 3 mode filtering operation is shown in Fig. 4.

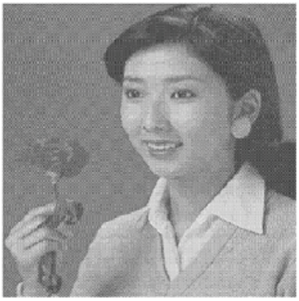


Figure 3. Original image "Lady with a rose"

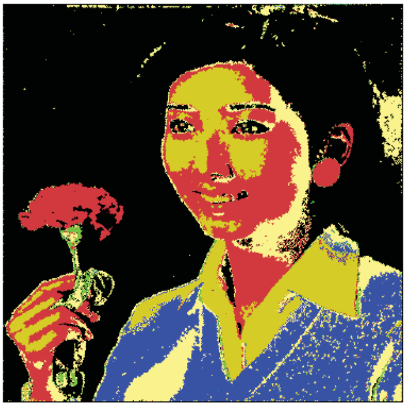
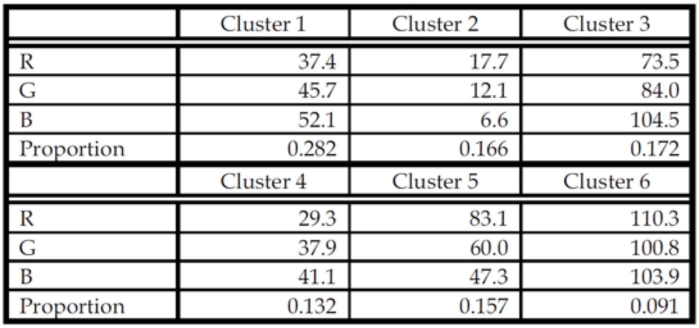


Figure 4. Segmentation result of "Lady with a rose"

The final estimates of the means and the proportions are shown in Table 1.

Table 1. Final parameter estimates of the means and the proportions for "Lady with a rose"



	Cluster 1	Cluster 2	Cluster 3
R	37.4	17.7	73.5
G	45.7	12.1	84.0
B	52.1	6.6	104.5
Proportion	0.282	0.166	0.172
	Cluster 4	Cluster 5	Cluster 6
R	29.3	83.1	110.3
G	37.9	60.0	100.8
B	41.1	47.3	103.9
Proportion	0.132	0.157	0.091

DATA-DRIVEN OBJECT DETECTION

In the real world, cameras are becoming popular to detect and track moving objects not only for surveillance or security but also for digital signage or spoken dialogue systems. The background subtraction method, which can be used to detect objects moving in the foreground by determining the difference between the current frame and an image of the scene's static background, is still one of the useful methods to detect moving objects in video sequences. Although the background subtraction method is a simple and effective method to detect moving objects, it occasionally suffers from illumination changes and unexpected background changes such as shadow. To improve the original background subtraction method, we propose that knowledge application and data-driven parameter adaptation techniques be adopted.

Detection of People by Background Subtraction Method

In order to cope with the illumination changes, we adopt the normalized distance method. Here, a unit vector is defined as the projection onto the unit sphere of a vector whose elments are the intensity values of pixels in a target region. The normalized distance is defined as a distance between two unit vectors. Let $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$ vectors consiting of intensity values of pixels in an observed image and in a background image respectively. Then the distance $\boldsymbol{\delta}$ and normalized distance $\boldsymbol{\delta}$' are shown as follows,

$$\boldsymbol{\delta} = |\boldsymbol{\tau} - \boldsymbol{\beta}| \qquad (5)$$

$$\boldsymbol{\delta}' = \left|\frac{\boldsymbol{\tau}}{|\boldsymbol{\tau}|} - \frac{\boldsymbol{\beta}}{|\boldsymbol{\beta}|}\right| \qquad (6)$$

Supposing that we process image frames during a period of T_{int}- We calculate $\boldsymbol{\delta}$ for each frame in T_{int}, then *Max(6), Min(6),* and *Ave(6)* are calculated as maximum value, minimum value, and averaged value of $\boldsymbol{\delta}$. In the same way, *Max*$(\boldsymbol{\delta}')$ and *Min*$(\boldsymbol{\delta}')$ are calculated as maximum value and minimum value of the normalized distance $(\boldsymbol{\delta}')$. Consequently, the following three discriminant functions are defined,

- discriminant function for scene change

  $\max \boldsymbol{\delta} - \min \boldsymbol{\delta} > Th_{scene}$

- discriminant function for background change

  $\text{Ave } \boldsymbol{\delta} > Th_{bs}$

- discriminant function whether environment change or illumination change

  $\max \quad \boldsymbol{\delta}' - \min \quad \boldsymbol{\delta}' > Th_{ill}$

where Th_{scene}, Th_{bs}, and Th_{ill} are thresholds to be determined.

Using these discriminant functions, whether there is a moving object detection or a background updating in T_{int} can be judged as follows:

- If both (3) and (5) are true, there is a scene change by a moving object.
- If (3) is true and (5) is false, there is a scene change by an illumination change.
- If (3) is false and (4) is true, there is a background change.
- If both (3) and (4) are false, there is nothing. It is a normal background.

A challenging point in this method is adaptively setting the threshold value to differentiate foreground objects from the background image in spite of environmental changes.

To determine the threshold value, Wren et al. (Wren et al., 1997) modeled the background using a Gaussian distribution and estimated the parameters adaptively. Crimson et al. (Crimson et aL 1998) also set up parameters according to the statistical analysis of training samples of the background images. Stauffer and Crimson used a mixture model of Gaussian distributions of the images to cope with multimodal background distributions (Stauffer & Crimson, 1999).

Knowledge Application and Parameter Adaptation

We apply two techniques to the original background subtraction method in order to cope with unexpected moving objects and adaptive threshold parameter setting. Our aim is to detect and track people as moving objects. There are, however, other unexpected moving objects in the scene, such as an automatic door. To avoid detection of such an unexpected moving object, we introduce knowledge about special spots as the first technique. The positions of special spots are assumed to be known and masking is applied not to detect the unexpected moving object. This is a simple but effective technique.

To set the threshold values adaptively, we introduce a kind of steepest descent method as the second technique. The algorithm is depicted in Fig. 5.

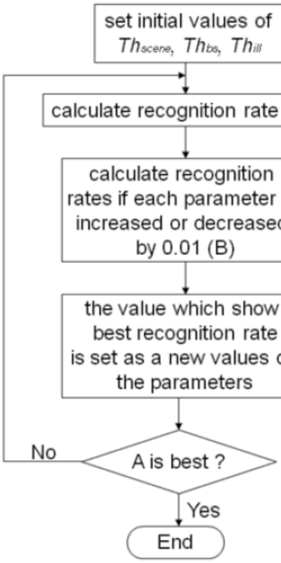

set initial values of Th_{scene}, Th_{bs}, Th_{ill}
calculate recognition rate (A)
calculate recognition rates if each parameter is increased or decreased by 0.01 (B)
the value which show best recognition rate is set as a new values of the parameters
A is best ?
No
Yes
End

Figure 5. Flowchart of threshold adaptation by steepest descent

In the first step of the algorithm shown in Fig. 5, initial values of Th_{scene}, Th_{bs} and $\mathrm{Th}_{\mathrm{ill}}$ are set. Then the recognition rate *A* with the current threshold values that are the same as the initial values at the very first stage of the algorithm. After the threshold values are increased or decreased by a small value, the new recognition rates Bare calculated. 0.01 is set as the small value in Fig. 5. *B* is a set of the recognition rates because increasing and decreasing of each threshold value are tried. *A* is compared with the best recognition rate in *B*. If *A* is superior to the best recognition rate, the algorithm is terminated. Otherwise, the best threshold values that correspond to the best recognition rate are substituted to the current threshold values and the same steps are carried out.

Experimental Results at the Real-World Test Bed

We constructed such a test bed in the entrance of our research laboratory. In the ceiling of the test bed, there are five cameras, and the area covered by the cameras is the meshed region in Fig. 6. The camera can take a 768 X 494-pixel image and has remote pan, tilt and zoom functions.

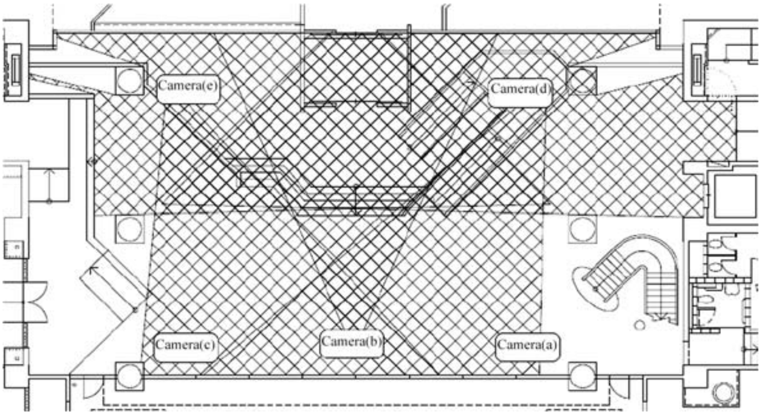


Figure 6. Area monitored by cameras

In the test bed, we tried to detect moving objects using camera images and the background subtraction method. In the background subtraction method, first of alt we selected the initial image without any moving objects as the background image to be subtracted. Then, the difference between the current and the background images is calculated and pixels that have a difference larger than the threshold are registered as candidate pixels of an image of moving objects. This difference calculation is operated for each small image block of 80 X 60 pixels. The judgment described in the previous subsections is applied. The adjacent small image blocks of candidate pixels are merged into a larger image block. To track the moving objects, a two-dimensional histogram with hue and saturation values in an HSV color space is constructed to calculate the correspondence between objects in the current and previously captured images. When the difference between two images in the two-dimensional histogram is smaller, the probability that objects belong to the same object is higher.

We applied the above background subtraction method to a ten-second video captured in an actual situation. The number of frames captured from each camera was different because the time consumed for image compression was different.

Two experiments were carried out. The first experiment was moving object detection by the background subtraction method with knowledge application only. The second one was moving object detection by the background subtraction method with parameter adaptation as well as knowledge application. The first and the second are referred as Experiment I and Experiment II respectively. The applied knowledge is that the position of the automatic door is known.

In Experiment I, the threshold parameters were fixed as $Th_{scene}=0.10$, $Th_{bs}=0.25$ and $Th_{ill}=0.05$.

One result of detecting a moving object is presented in Fig. 7. The people in the image should be recognized as moving objects, and rectangles are drawn as a result. Two larger rectangles (shown in red) are hand-made markings that indicate correct answers. Several smaller rectangles in the left larger rectangle (shown in yellow) indicate detected results obtained by the background subtraction method. In the right larger rectangle, no object was detected by the background subtraction method. We call the larger rectangles ground truth rectangles and the smaller rectangles are called detected rectangles.

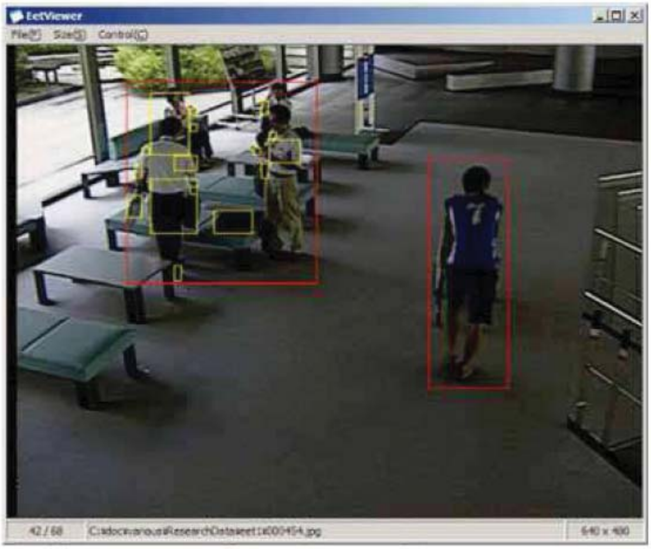

EctViewer
File(F) Size(S) Control(C)
42 / 68
C:\doc\various\ResearchDataset1\c00454.jpg
640 x 480

Figure 7. Result of detecting moving objects in Experiment I

Although an identical match between ground truth and detected rectangles is desirable, detected rectangles are almost always included in or overlapped on ground truth rectangles. Here, we define two kinds of error: the type one error and the type two error. The type one error is that in which no detected rectangle is drawn where there was a ground truth rectangle. The type two error is that in which detected rectangles appeared where there was no ground truth rectangle. The total error rate can be calculated by averaging these two types of error. The rates of occurrence of two types of error and the total error rate are shown in Table 2 for cameras (a)- (e). The positions of the cameras are shown in Fig. 6.

Table 2. Error rates with knowledge application and fixed parameters

	Type one error rate (%)	Type two error rate (%)	Total error rate (%)
Camera(a)	29.38	0.32	14.85
Camera(b)	54.35	27.94	41.15
Camera(c)	28.39	0.31	14.35
Camera(d)	10.23	16.19	13.21
Camera(e)	10.06	8.38	9.22

Next, we applied threshold parameter adaptation presented in Fig. 5 as well as the knowledge application. This is Experiment II. The initial threshold parameters were set as *Thscene=0.10, Thbs=0.25* and Thw=0.05. One moving object detection result with the parameter adaptation is presented in Fig. 8, that corresponds to Fig. 7. By comparing two images, it is found that the person in the right side was detected in Experiment II, who was missed in Experiment I. The rates of occurrence of two types of error and the total error rate are shown in Table 3.

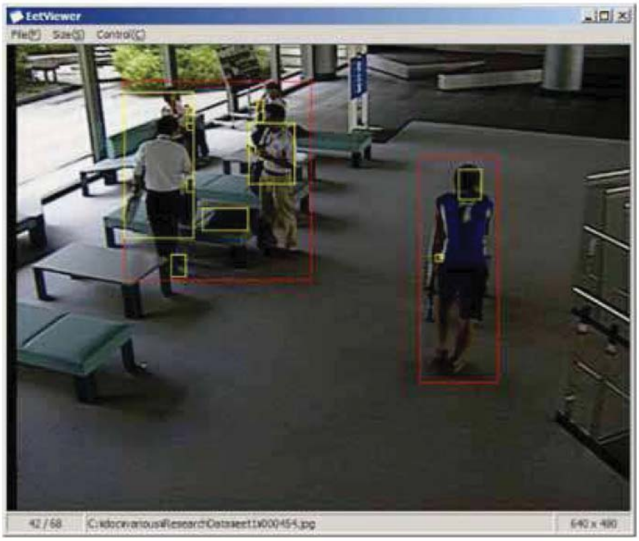
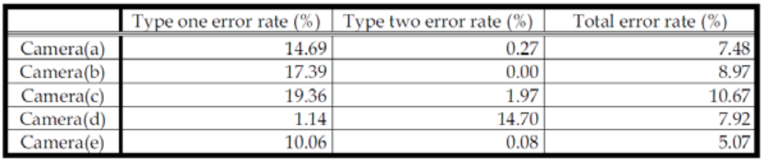

LetViewer
File(F) Size(S) Control(C)
42 / 68
C:\docs\various\Research\Dataset1\000454.jpg
640 x 480

Figure 8. Result of detecting moving objects in Experiment II

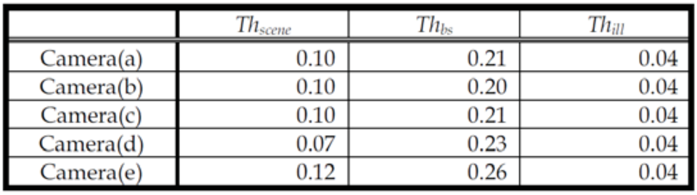
Table 3. Error rates with knowledge application and parameter adaptation



	Type one error rate (%)	Type two error rate (%)	Total error rate (%)
Camera(a)	14.69	0.27	7.48
Camera(b)	17.39	0.00	8.97
Camera(c)	19.36	1.97	10.67
Camera(d)	1.14	14.70	7.92
Camera(e)	10.06	0.08	5.07

Almost all error rates were improved. Especially improvement for camera (b) was splendid. It can be considered that the reason is owing to knowledge application. As the result of parameter adaptation, the final obtained parameters, *Thscer e, Thbs* and Thm, are shown in Table 4.

Table 4. The final threshold parameters in Experiment II



	Th_{scene}	Th_{bs}	Th_{ill}
Camera(a)	0.10	0.21	0.04
Camera(b)	0.10	0.20	0.04
Camera(c)	0.10	0.21	0.04
Camera(d)	0.07	0.23	0.04
Camera(e)	0.12	0.26	0.04

CONCLUSION

Among multimedia, the roles of image and video media are becoming more important both in the cyber and real worlds. The cyber world means the information world structured by computer networks such as Intemet. Videos and images are accumulated in the cyber world and the users are wandering to search for what they want. Also in the real world, cameras are becoming popular to collect the users' and environmental information. How to analyse these more efficiently is one of the issues to be solved urgently.

Image and video understanding has been studied and several approaches have been developed. In the parametric approaches, how to set the parameters is a challenging problem. In this paper, data-driven parameter adaptation was applied to image clustering and object extraction for still and video images. Although the experimental results were limited, definite improvement has been attained.

In the future, it is desired to extract contextual information from image and video media by applying image and video understanding techniques including the methods proposed in this paper. It must contribute to realize a personalized, adaptive, situation-aware service in a ubiquitous network society.

REFERENCES

1. Meier, T. & Ngan, K.N. (1998). Automatic Segmentation of Moving Objects for Video Object Plane Generation. *IEEE Trans. on Circuits Syst., Video Techno!.,* Vol. 8, No. 5, (Sept. 1998) pp. 525-538

2. Kunt, M. (1988). Progress in High Compression Image Coding. *Int.]. Pattern Recognition and Artificial Intelligence,* Vol. 2, No.3, (1988) pp. 387-405

3. Won, C.S. & Derin, H. (1992). Unsupervised Segmentation of Noisy and Textured Images Using Markov Random Fields. *CVGIP: Graphical Models and Image Processing,* Vol. 54, No.4, (July 1992) pp. 308-328

4. Satake, J. & Shakunaga, T. (2004). Multiple target tracking by appearance-based

5. condensation tracker using structure information, *Proceedings of the 17th*

6. *International Conference on Pattern Recognition (ICPR2004),* Vol. lap, No. We-ii, pp. 53*7-5401* 2004

7. Thonnat, M. & Rota, N. (1999). Image Understanding for Visual Surveillance Applications, *Proceedings of Third International Workshop on Cooperative Distributed Vision (CDV VVD'99),* No.3, pp. 51-82, Nov. 2004

8. Matsuyama, T. (1999). Dynamic Memory: Architecture for Real Time Integration of Visual Perception, Camera Action, and Network Communication, *Proceedings of Third International Workshop on Cooperative Distributed Vision (CDV-WD'99),* No.1, pp. 1- 30, Nov. 2004

9. Wren, C.; Azarbayejani, A; Darrelt T. & Pentland, A (1997). Real-time Tracking of the Human Body, *IEEE Trans. on Patt. Anal. and Machine Intell"J* Vol. 19, No.7, (1997) pp.780-785

10. Crimson, W.E.L.; Stauffer, C.; Romano, R. & Lee, L. (1998). Using adaptive tracking to classify and monitor activities in a site, *Proceedings of 1998 Conference on Computer Vision and Pattern Recognition (CVPR '98),*pp. 22-29, 1998

11. Stauffer, C. & Crimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking, *Proceedings of 1999 Conference on Computer Vision and Pattern Recognition (CVPR '99),*pp. 246-

CHAPTER 5

PRETRAINING CONVOLUTIONAL NEURAL NETWORKS FOR IMAGE-BASED VEHICLE CLASSIFICATION

Yunfei Han[1,2,3], Tonghai Jiang[1,2,3], Yupeng Ma[1,2,3] and Chunxiang Xu[1,2,3]

[1]The Xinjiang Technical Institute of Physics & Chemistry, Urumqi 830011, China
[2]Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi 830011, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

Vehicle detection and classification are very important for analysis of vehicle behavior in intelligent transportation system, urban computing, etc. In this paper, an approach based on convolutional neural networks (CNNs) has been applied for vehicle classification. In order to achieve a more accurate classification, we removed the unrelated background as much as possible based on a trained object detection model. In addition, an

Citation: Yunfei Han, Tonghai Jiang, Yupeng Ma, Chunxiang Xu, "Pretraining Convolutional Neural Networks for Image-Based Vehicle Classification", Advances in Multimedia, vol. 2018, Article ID 3138278, 10 pages, 2018. https://doi.org/10.1155/2018/3138278.

Copyright: © 2018 by Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

unsupervised pretraining approach has been introduced to better initialize CNNs parameters to enhance the classification performance. Through the data enhancement on manual labeled images, we got 2000 labeled images in each category of motorcycle, transporter, passenger, and others, with 1400 samples for training and 600 samples for testing. Then, we got 17395 unlabeled images for layer-wise unsupervised pretraining convolutional layers. A remarkable accuracy of 93.50% is obtained, demonstrating the high classification potential of our approach.

INTRODUCTION

Vehicle is one of the greatest inventions in human history. The vehicle has become an indispensable part of modern people's life. The use of a huge large number of vehicles can reflect the population's mobility, intimacy, economic, and so on, and the analysis of vehicle behavior is very meaningful for urban development and government decision-making. In order to collect refueling vehicle information, such as license plate, picture, time, location, volume, type, and so on, we have deployed data collecting equipment in many of refueling stations in Xinjiang, which is mainly responsible for safety supervision and analysis of refueling behavior. Till now, many vehicle profile information such as vehicle color and vehicle type is entered into the system by hand; this is inefficient and not uniform. The accurate, various, and volume of data are the key to dig the value of refueling data. Hence, it has become a problem to be solved urgently that how to obtain the vehicle profile information through the vehicle picture automatically. In this paper, we focus on how to get the vehicle type from the picture. This problem is regarded as image classification, which means we should classify the images containing vehicles into the right type by image processing. Due to the environment in which images are taken is quite varied and complex and the impact of irrelevant background, the vehicles in images are very difficult to recognize.

Thanks to the success of deep learning, we present a combination of approaches for vehicle detection and classification based on convolutional neural networks in this paper. To detect the vehicle in the image more efficiently, a successful object detection approach is used to detect the objects in an image, then the target vehicle waiting for entering refueling station is filtered out. Next, we designed a convolutional neural networks which contains 4 convolutional layers, 3 max pooling layers, and 2 full connected layers for vehicle classification. We trained our model on labeled vehicles images dataset. Comparing it with other five state-of-the-art approaches verified our approach achieves the highest accuracy than others. In order to pursue better classification performance, we taken advantage of unsupervised pretraining to better initialize classification model parameters under the circumstance of a shortage of labeled images. The unsupervised pretraining method was implemented based on deconvolution. After pretraining, the convolutional layers were initialized by the pretrained parameters and trained the model on our labeled images data set; thus, we got a better classification performance than the previous without pretraining.

This paper is organized as follows. The related works are introduced in Section 2. Vehicle detection and classification based on CNNs and a pretraining approach are described in Section 3. In Section 4, the vehicles data set is presented, and we evaluated the presented approaches on our data, and the experimental results and a performance evaluation are given. Finally, Section 5 **concludes the paper.**

RELATED WORKS

Existing methods use various types of signal for vehicle detection and classification, including acoustic signal [2–5], radar signal [6, 7], ultrasonic signal [8], infrared thermal signal [9], magnetic signal [10], 3D lidar signal [11] and image/video signal [12–16]. Furthermore, some of the methods can be combined with a variety of signals, such as radar&vision signal [7] and audio&vision signal [17]. Usually, the detection and classification performance is excellent in these methods because of the precise signal data, but there are many hardware devices involved in these methods, resulting in larger deployment cost and even higher failure rate.

The evolution of image processing techniques, together with wide deployment of surveillance cameras, facilitates image-based vehicle detection and classification. Various approaches to image-based vehicle detection and classification have been proposed over the last few years. Kazemi et al. [13] used 3 different kinds of feature extractors, Fourier transform, Wavelet transform, and Curvelet transform, to recognize and classify 5 models of vehicles; k-nearest neighbor is used as classifier. They compare the 3 proposed approaches and find that the Curvelet transform can extract better features. Chen et al. [18] presented a system for vehicle detection, tracking, and classification from roadside closed-circuit television (CCTV). First, a Kalman filter tracked a vehicle to enable classification by majority voting over several consecutive frames, then they trained a support vector machine (SVM) using a combination of a vehicle silhouette and intensity-based pyramid histogram of oriented gradient (HOG) features extracted following background subtraction, classifying foreground blobs with majority voting. Wen et al. [19] used Haar-like feature pool on a 3232 grayscale image patch to represent a vehicle's appearance and then proposed a rapid incremental learning algorithm of AdaBoost to improve the performance of AdaBoost. Arrospide and Salgado [16] analyzed the individual performance of popular techniques for vehicle verification and found that classifiers based on Gabor and HOG features achieve the best results and outperform principal component analysis (PCA) and other classifiers based on features as symmetry and gradient. Mishra and Banerjee [20] detected vehicle using background, extracted Haar, pyramidal histogram of oriented gradients, shape and scale-invariant feature transform features, designed a multiple kernel classifier based on k-nearest neighbor to divide the vehicles into 4 categories. Tourani and Shahbahrami [21] combined different image/video processing methods including object detection, edge detection, frame differentiation, and Kalman filter to propose a method which resulted in about 95 percent accuracy for classification and about 4 percent error in vehicle detection targets. In these methods, the classification results are very good; however, there are still some problems. First, the image features are limited by the hand-crafted features algorithms to represent rich information. Second, the hand-crafted features algorithms require a lot of calculation, so they are not suitable for real-time applications, especially for embedding in front-end camera devices. Third, most of them are used in fixed scenes and background environments; it is difficult for them to deal with complex environment.

More recently, deep learning has become a hot topic in object detection and object classification area. Wang et al. [22] proposed a novel deep learning based vehicle detection algorithm with 2D deep belief network; the 2D-DBN architecture uses second-order planes instead of first-order vector as input and uses bilinear projection for retaining discriminative information so as to determine the size of the deep architecture which enhances the success rate of vehicle detection. Their algorithm preforms very good in their datasets. He et al. [1] proposed a new efficient vehicle detection and classification approach based on convolutional neural network, the features extracted by this method outperform those generated by traditional approaches. Yi et al. [23] proposed a deep convolution network based on pretrained AlexNet model for deciding whether a certain image patch contains a vehicle or not in Wide Area Motion Imagery (WAMI) imagery analysis. Li et al. [24]

presented the 3D range scan data in a 2D point map and used a single 2D end-to-end fully convolutional network to predict the vehicle confidence and the bounding boxes simultaneously, and they got the state-of-the-art performance on the KITTI dataset.

Meantime, object detection and classification based on Convolutional neural networks (CNNs) [25–27] are very successful in the field of computer vision recently. The first work about object detection and classification based on deep learning has been done in 2013; Sermanet et al. [28] present an integrated framework for using deep learning for object detection, localization, and classification; this framework obtains very competitive results for the detection and classifications tasks. Up to now, excellent object detection and classification models based on deep learning include R-CNN [29], Fast R-CNN [30], YOLO [31], Faster R-CNN [32], SSD [33], and R-FCN [34]; these models achieve state-of-the-art results on several data sets. Before the YOLO, many approaches on object detection, for example, R-CNN and Faster R-CNN, repurpose classifiers to perform detection. Instead, YOLO frame object detection is a regression problem to separated bounding boxes and associated class probabilities. The YOLO framework uses a custom network based on the Googlenet architecture, using 8.52 billion operations for a forward pass. However, a more recent improved model called YOLOv2 [35] achieves comparable results on standard tasks like PASCAL VOC and COCO. In YOLOv2 network, it uses a new model, called Darknet-19, and has 19 convolutional layers and 5 maxpooling layers; the model only requires 5.58 billion operations. In conclusion, YOLOv2 is a state-of-the-art detection system, it is better, faster, and stronger than others and applied for object detection tasks in this work.

At last, unsupervised pretraining initializes the model to a point in the parameter space that somehow renders the optimization process more effective, in the sense of achieving a lower minimum of the empirical loss function [36]. Much recent research has been devoted to learning algorithms for deep architectures such as Deep Belief Networks [37, 38] and stacks of autoencoder variants [39]. After vehicle detection, we can easily get a lot of unlabeled images of vehicles and optimize the classification model parameters initialization by unsupervised pretraining.

METHODOLOGY

In this section, we will present the details of the method based on CNNs for image-based vehicle detection and vehicle classification. This section contains three parts: vehicle detection, vehicle classification, and pretraining approach. The relations between each part and the overall framework of the entire idea is shown in Figure 1.

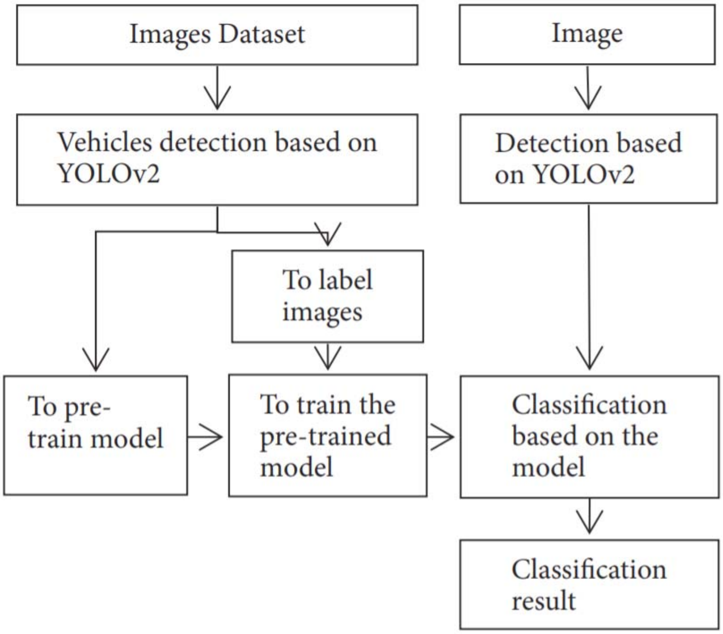

Images Dataset
Vehicles detection based on YOLOv2
To label images
To pre-train model
To train the pre-trained model
Image
Detection based on YOLOv2
Classification based on the model
Classification result

Figure 1. Flowchart of vehicle classification.

Vehicle Detection

Our images taken from static cameras in different refueling station contain front views of vehicles or side views of vehicles at any point. The vehicles in images are very indeterminacy; this makes the vehicle detection more difficult in traditional methods based on hand-crafted features.

The YOLOv2 model is trained on COCO data sets, it can detect 80 common objects in life, such as person, bicycle, car, bus, train, truck, boat, bird, cat, etc., and therefore we can perform vehicle detection based on YOLOv2. In a picture of the vehicle which is waiting for entering the refueling station shown in Figure 2, YOLOv2 can well detect many objects, for example, the security guards, drivers, the vehicle, and queued vehicles, and even vehicles on the side of road. Here, our goal is to pick up the vehicle which is waiting for entering in picture.

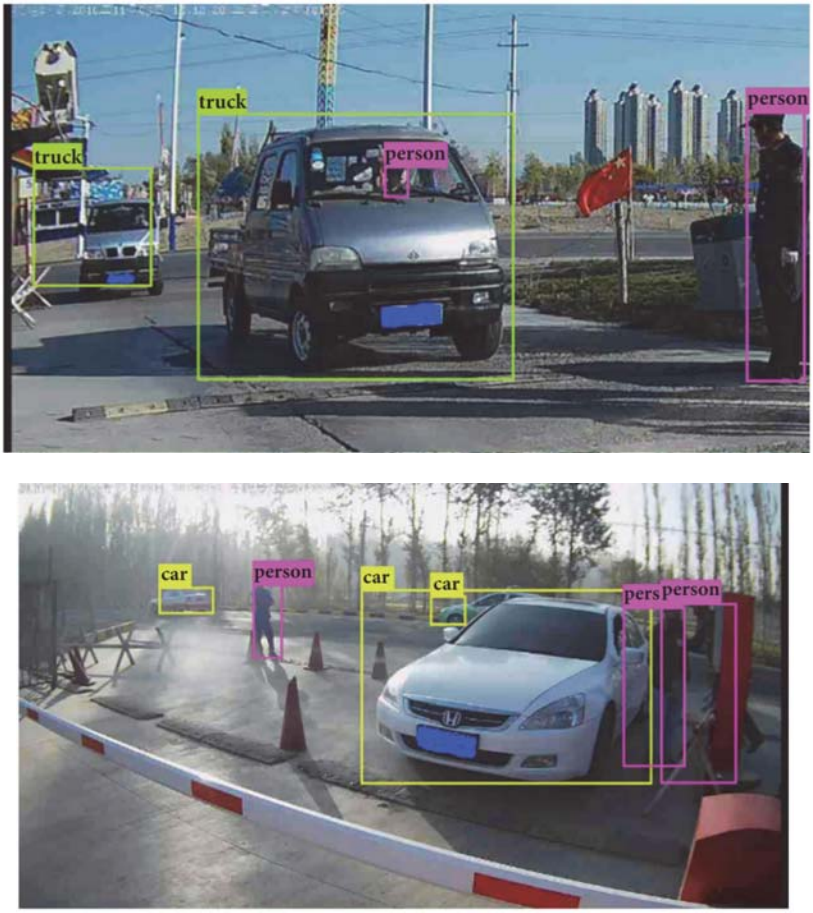


Figure 2. YOLOv2 detection. Top: truck. Down: car.

Although trained YOLOv2 can detect the vehicle and divide it into bicycle, car, motorbike, bus, and truck, it does not meet our classification categories. In order to solve this problem, to fine-tune the YOLOv2 on our data could be a solution, but this method needs amount of manual labeled data and massive computation, it is not a preferable method for us, and then, we presented a rule-based method to detect four categories vehicles more accurately. First, from the YOLOv2 detection results, we select the objects which are very similar to our targets, such as car, bus, truck, and motorbike; second, according to the distance between the vehicle and the camera, the closer the vehicle is, the bigger the target is, we choose the most similar vehicle in picture as the target vehicle entering the refueling station for further vehicle behavior analysis.

Vehicle Classification

According to the function and size of vehicles, vehicles will be divided into four categories of motorcycle, transporter, passenger, and others. Motorcycle includes motorcycle and motor tricycle; transporter includes truck and container car; passenger includes sedan, hatchback, coupe, van, SUV, and MPV; others include vehicles used in agricultural production and infrastructure, such as tractor and crane, and other types of vehicles. Figure 3 shows the sheared samples in four categories in each column. As we can see, samples images are very different in shape, color, size, and angle of camera, even the samples images in the same category. And Figure 3(b) bottom and Figure 3(c) bottom are not in the same category, but they are very similar, especially on front face, shape, and color, which makes the classification between transporter and passenger more difficult.

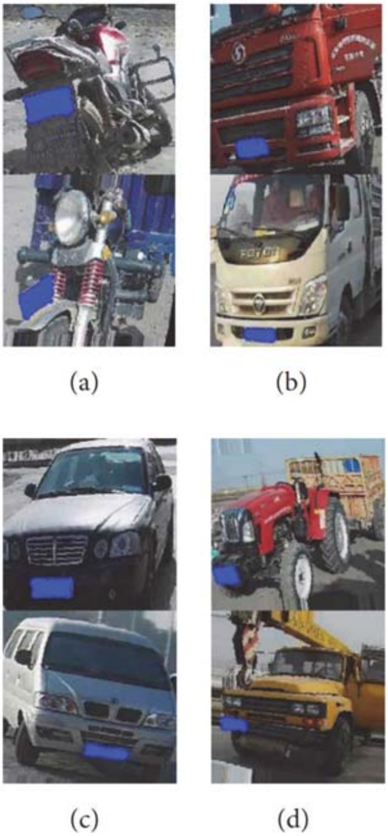
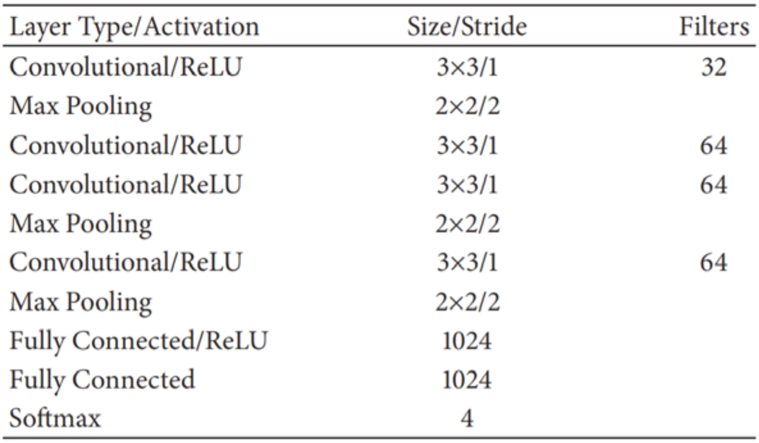


Figure 3. Sheared vehicle image examples for four categories. (a) Motorcycle; (b) transporter; (c) passenger; (d) others.

To solve this difficult problem, we presented a convolutional classification model which is effective and requires little amount of operations. Our model, called C4M3F2, has 4 convolutional layers, 3 max pooling layers, and 2 fully connected layers.

Each convolutional layer contains multiple (32 or 64) 33 kernels, and each kernel represents a filter connected to the outputs of the previous layer. Each max pooling layer contains multiple max pooling with 2×2 filters and stride 2; it effectively reduces the feature dimension and avoids overfitting. For fully connected layers, each layer contains 1024 neurons, each neuron makes prediction from its all input, and it is connected to all the neurons in previous layers. For each sheared vehicle image detected from YOLOv2, it has been resized to 48×48 and then passed into C4M3F2. Eventually, all the features are passed to softmax layer, and what we need to do is just minimizing the cross entropy loss between softmax outputs and the input labels. Table 1 shows the structure of our model C4M3F2.

Table 1. The structure of C4M3F2.



Layer Type/Activation	Size/Stride	Filters
Convolutional/ReLU	3×3/1	32
Max Pooling	2×2/2	
Convolutional/ReLU	3×3/1	64
Convolutional/ReLU	3×3/1	64
Max Pooling	2×2/2	
Convolutional/ReLU	3×3/1	64
Max Pooling	2×2/2	
Fully Connected/ReLU	1024	
Fully Connected	1024	
Softmax	4	

Pretraining Approach

With the purpose of achieving a satisfactory classification result, we need more labeled images for training our model, but there is a shortage of labeled images; however, there are plenty of images collected easily, and how to use the plenty of unlabeled images for optimization of our classification model has become the main content in this subsection.

The motivation of this unsupervised pretraining approach is to optimize the parameters in convolutional kernel parameters. Kernel training in

C4M3F2 starts from a random initialization, and we hope that the kernel training procedure can be optimized and accelerated using a good initial value obtained by unsupervised pretraining. In addition, pretraining initializes the model to a point in the parameter space that somehow renders the optimization process more effective, in the sense of achieving a lower minimum of the loss function [36]. Next, we will explain how to greedy layer-wise pretrain the convolution layers and the fully connected layers in the C4M3F2 model. Max pooling layer function is subsampling; it is not included in layer-wise pretraining process.

An autoencoder [40] neural network is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs. It uses a set of recognition weights to convert the input into code and then uses a set of generative weights to convert the code into an approximate reconstruction of the input. Autoencoder must try to reconstruct the input, which aims to minimize the reconstruction error as shown in Figure 4.

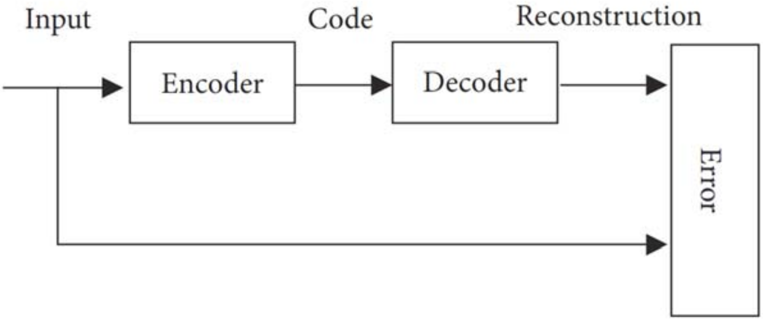


Figure 4. Autoencoder.

According to the aim of autoencoder, our approach for unsupervised pretraining will be explained. In every convolutional layer, convolution can be regarded as the encoder, and deconvolution [41] is taken as the decoder, which is a very unfortunate name and also called transposed convolution. An input image is passed into the encoder, then the output code getting from encoder is passed into the decoder to reconstruct the input image. Here, Euclidean distance, which means the reconstruct error, is used to measure the similarity between the input image and reconstructed image, so the aim of our approach is minimizing the Euclidean paradigm. For pretraining the next layer, first, we should drop the decoder and freeze the weights in the encoder of previous layers and then take the output code of previous layer as the input in this layer and do the same things as in previous layer. Next how to use transposed convolution to construct and minimize the loss function in one convolutional layer will be described in detail as follows.

The convolution of a feature maps and an image can be defined as

$$y_{ij} = w_i \oplus x_j \tag{1}$$

where ⊕ denotes the 2D convolution, $y_{ij} \in R^{r_x \times c_x}$ is the convolution result and the padding is set to keep the input and output dimensions consistent. $w_i \in R^{r_w \times c_w}$ is the i^th kernel, and $x_j \in R^{r_x \times c_x}$ denotes the j^th training image.

Ten, transform the convolution based on circulant matrix in linear system. A circulant matrix is a special kind of Toeplitz matrix where each row vector is rotated one element to the right relative to the preceding row vector. An n × n circulant matrix C takes the form in

$$C = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & & c_3 & c_2 \\ \vdots & & \ddots & & \vdots \\ c_{n-2} & c_{n-3} & \cdots & c_0 & c_{n-1} \\ c_{n-1} & c_{n-2} & & c_1 & c_0 \end{bmatrix} \tag{2}$$

Let $f_j, h_i \in R^{r_e \times c_e}$ be the extension of x_j, w_i, where $r_e = r_x + r_w - 1, c_e = c_x + c_w - 1$. And the method is as follows (3) and (4), where O is zero matrix:

$$f_j = \begin{bmatrix} x_j & O \\ O & O \end{bmatrix} \tag{3}$$

$$h_i = \begin{bmatrix} w_i & O \\ O & O \end{bmatrix} \tag{4}$$

Let $f_j^v \in R^{r_e c_e \times 1}$ be f_j in vectored form, row_a be a row of h_i, and $row_a = [n_0, n_1, n_2, \cdots, n_{c_e-1}]$. To build circulant matrices $H_0, H_1, H_2, \cdots, H_{c_e-1}$ by row_a and by these circulant matrices, a block circulant matrix is defined as shown in formula (5).

$$H = \begin{bmatrix} H_0 & H_{c_e-1} & \cdots & H_2 & H_1 \\ H_1 & H_0 & & H_3 & H_2 \\ & \vdots & \ddots & \vdots & \\ H_{c_e-2} & H_{c_e-3} & \cdots & H_0 & H_{c_e-1} \\ H_{c_e-1} & H_{c_e-2} & & H_1 & H_0 \end{bmatrix} \quad (5)$$

Here, we can transform the convolution into (6).

$$Q = Hf_j^v \quad (6)$$

Q is the vector form of result of convolution calculation and then to reshape Q into $Q' \in R^{r_e \times c_e}$. In this convolution process, the padding is dealing with filling 0, but in actual implementation of our approach, we keep the convolutional input and output dimensions consistent. So we need to prune the extra values to keep the input and output dimensions consistent. So we intercept the matrix Q', then $y_{ij} = Q'[r_w - 1 : r_e - r_w + 1, c_w - 1 : c_e - c_w + 1]$. To simplify the calculation, we extract the effective rows of H according to the efective row indexes which indicates the position of the elements of y_{ij} in Q and denote these rows by $W_i \in R^{(r_e-2r_w+2)(c_e-2c_w+2)\times 2c_e}$. Now, $Y_{ij} \in R^{(r_e-2r_w+2)(c_e-2c_w+2)\times 1}$ is vector form of y_{ij}, so the convolution can be rewritten as

$$Y_{ij} = W_i f_j^v \quad (7)$$

There are J training vehicle images and K kernels. Let $X = [f_1^v, f_2^v, \ldots, f_j^v]$, and $W = [W_1; W_2; \ldots; W_K]$.

The convolution can be calculated as

$$Y = WX \quad (8)$$

And the deconvolution can be calculated

$$X' = W^T Y \quad (9)$$

So X' is the X reconstruction. Ten loss function based on Euclidean paradigm is defined in formula (10), being

$$\text{loss}(W, X) = \|X - X'\|_2 = \sqrt{\sum_{j=1}^{J} (X_j - X_j')^2} \quad (10)$$

Ten, we used adam optimizer, which is an algorithm for first-order gradient-based optimization of stochastic objective functions based on adaptive estimates of lower-order moments, to solve the minimum optimization problem in formula (10).

After the greedy layer-wise unsupervised pretraining, we initiate the parameters in every convolutional layer with the pretrained values and run the supervised training for classification according to the method in previous subsection.

EXPERIMENTS AND DISCUSSIONS

We evaluated the presented algorithm on our data and compared it with other four state-of-the-art methods.

Datasets and Experiment Environment

The vehicles images are taken by the static cameras in different refuel stations; after being compressed, they are sent to servers. The quality of images on servers is lower than that selected by random and classified into four categories of motorcycle, transporter, passenger, and others by hand. We got 498 motorcycle images, 1109 transporter images, 1238 passenger images, and 328 other images. Due to the time-consuming and labor-intensive of manual labeling, there is a shortage of labeled images. Image augmentation has been used to enrich the data. Keras, the excellent high level neural network API, provides the ImageDataGenerator for image data preparation and augmentation. Shear range is set to -0.2 to 0.2, zoom range is set to -0.2 to 0.2, rotation range is set to -7 to 7, size is set to 256×256, and the points outside the boundaries are filled according to the nearest mode. After configuration and taking into account the balance of data, we fitted it on our data and got 1400 samples for every categories on the training set and 600 samples for every categories on the testing set to assess our classification model.

For vehicle classification, the CNNs are under Tensorflow framework, the SIFT is under OpenCV (https://opencv.org/), and other feature embedding methods are under scikit-image (http://scikit-image.org/). All the experiments were conducted on a regular notebook PC (2.5-GHz 8-core CPU, 12-G RAM, and Ubuntu 64-bit OS).

Vehicle Detection Experiment with YOLOv2

For the experiment using original images, the original training set and test set are used for training and testing, for the experiment using sheared images, we used the approach based on trained YOLOv2 to detect the original training set and test set to get the sheared training set and sheared testing set for training and testing.

To verify the importance of vehicle detection for vehicle classification, we designed two groups of vehicle classification experiments, one using original images and the other one using sheared images after vehicle detection, then, the C4M3F2 model is used for vehicle classification experiments.

We initialized the C4M3F2 model by truncated normal distribution, fitted the model on original training set and sheared training set for 2000 epochs, respectively, and recorded the accuracy of our C4M3F2 model on different testing set; the results are shown in Figure 5. As we expected, the sheared images of vehicles more accurately represent the characteristics of vehicles, while pruning more useless information and facilitating the feature extraction and vehicle classification. As we can see in Figure 5, the accuracy of C4M3F2 model using sheared data set is much better than the one using original data set; in the previous training, the characteristics of vehicles were extracted more accurately, so that, the model quickly achieved a better classification results and a stable status.

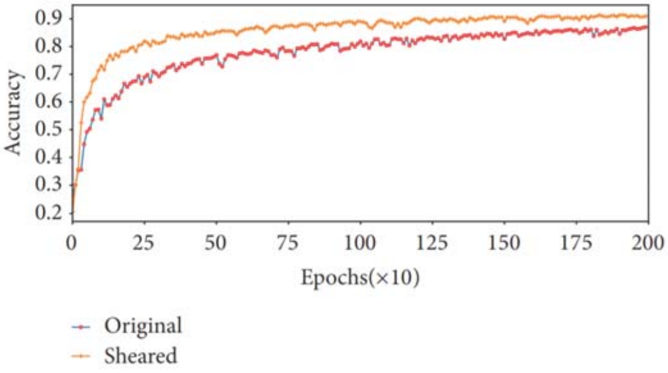


Figure 5. Comparison of training process between original and sheared.

Finally, the accuracy of C4M3F2 using sheared data set is 91.42%, which is 4.53% higher than the 86.89% of C4M3F2 using original data set. It can be concluded that the results of vehicle classification using sheared data set after vehicle detection based on YOLOv2 can be improved effectively.

Compare Our Approach with Others

There are many other image classification methods. To assess our classification model, we compared our approach with other five methods.

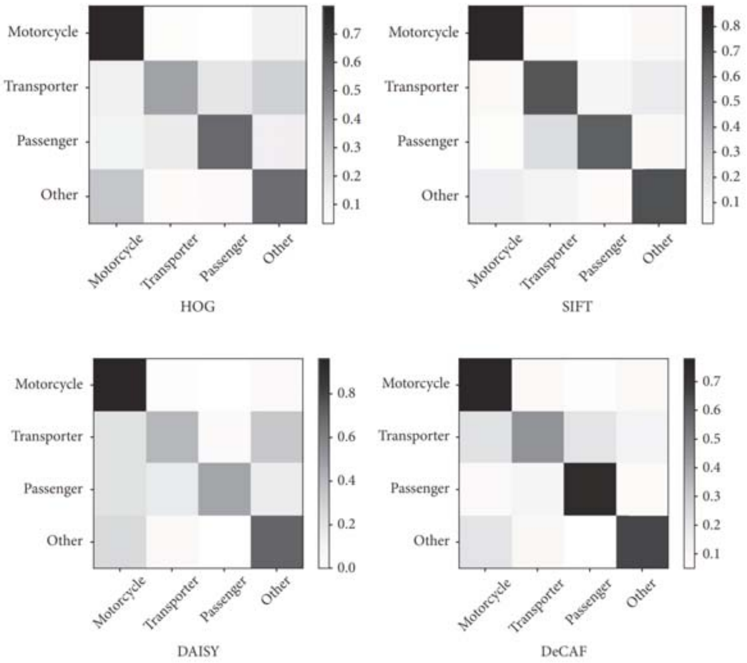
The five methods are based on the image features defined by the scholars in computer image processing. Considering the comprehensive factors, four kinds of image features and a convolutional method are selected, they are histograms of oriented gradient (HOG) [42], DAISY [43], oriented FAST, and rotated BRIEF (ORB)[44], scale-invariant feature transform (SIFT) [45], and DeCAF[1] respectively. These methods are excellent in target object detection in [1, 42–45]. HOG is based on computing and counting the gradient direction histogram of local regions. DAISY is a fast computing local image feature descriptor for dense feature extraction, and it is based on gradient orientation histograms similar to the SIFT descriptor. ORB uses an oriented FAST detection method and the rotated BRIEF descriptors; unlike BRIEF, ORB is comparatively scale and rotation invariant while still employing the very efficient Hamming distance metric for matching. SIFT is the most widely used algorithm of key point detection and description. It takes full advantage of image local information. SIFT feature has a good effect in rotation, scale, and translation, and it is robust to changes in angle of view and illumination; these features are beneficial to the effective expression of targets information. For HOG and DAISY, image features regions are designed; features are computed and sent into SVM classifier to be classified. For ORB and SIFT, they do not have acquisition features regions and specified number of features; we get the image features based on Bag-of-Words (BoW) model by treating image features as words. In the first instance, all features points of all training build the visual vocabulary; in the next place, a feature vector of occurrence counts of the vocabulary is constructed from an image; in the end, the feature vector is sent into SVM classifier to be classified. DeCAF uses five convolutional layers and two fully connected layers to extract features and a SVM to classify the image into the right group [1].

Here, we performed vehicle classification experiments on sheared data set. Table 2 shows the accuracy and FPS of CNNs and other state-of-the-art methods in test; other methods are very slow because they take a lot of time to extract features. It can be observed that the results show the effectiveness of CNNs on vehicle classification problem.

Table 2. Accuracy and FPS of different methods.

Method	Accuracy	FPS
HOG+SVM	60.12%	4
DAISY+SVM	69.04%	2
ORB+BoW+SVM	64.07%	7
SIFT+BoW+SVM	74.49%	5
DeCAF[1]	66.20%	13
CNNs	91.42%	800

From another point of view, we demonstrated the classification ability of each method by confusion matrix of the classification process from the five methods in Figure 6. The main diagonal displays the high recognition accuracy. As shown in Figure 6, the top five comparative methods are better in recognition of motorcycle than other categories. Generally speaking, ORB or SIFT combined with BoW and SVM method is a little better than the other two methods. All taken into account, our CNNs method is the best. But, the performance of CNNs turns out not so satisfactory in view than the confusion of transporter and passenger.



HOG

SIFT

DAISY

DeCAF

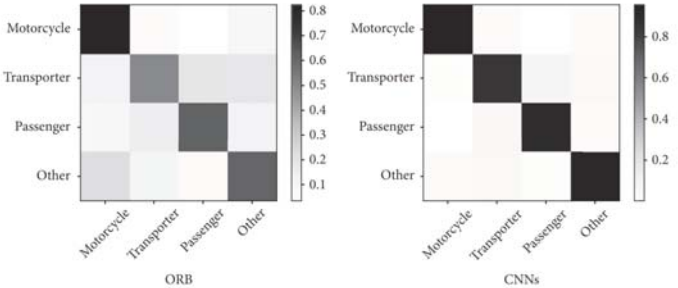


Figure 6. Confusion matrix of classification of different methods.

Next, we will focus on the reason of confusion in CNNs. According to the precision, recall, and f1-score of classification in Table 3, it shows that the identification of motorcycle is very good enough, and the identification of transporter and passenger is relatively poor.

Table 3. Classification precision, recall, and f1-score of CNNs.

Type	Precision	Recall	F1-score
Motorcycle	0.97	0.95	0.96
Transporter	0.87	0.85	0.86
Passenger	0.90	0.91	0.90
Other	0.93	0.93	0.93

As shown in the examples in Figure 7, it can be seen that the wrongly recognized transporter and passenger images include vehicle face information mainly and very little vehicle body information, as far as the main vehicle face is concerned; these vehicles images are so similar in profile that it is still a challenge to recognize same images in Figure 7 manually.

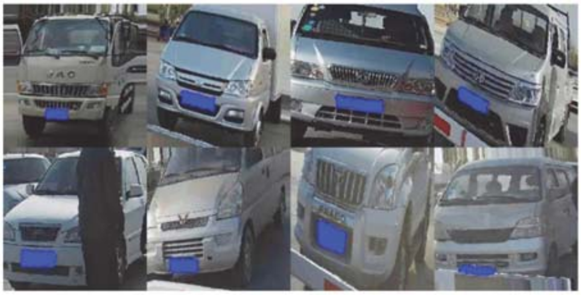


Figure 7. Examples of wrongly recognized vehicles images. First row: wrongly recognized as passenger which should be transporter. Second row: wrongly recognized as transporter which should be passenger.

Pretraining Approach Experiment

We are eager for better performance of our classification model C4M3F2. Here, unsupervised pretraining has been used for optimization of our classification model. 17395 sheared vehicles images are obtained by shearing the unlabeled vehicles images.

We pretrained the parameters of every convolutional layer for 2000 epochs and then supervised training the model on our sheared training set and tested it on our sheared testing set; the results is shown in Figure 8. In the process of training, the conclusion is that the effect is more obvious in the previous epochs, and the overall training process is stable relatively. Ultimately, the accuracy of pretrained CNNs is 93.50%, which is 2.08% higher than the 91.42% of CNNs without pretraining.

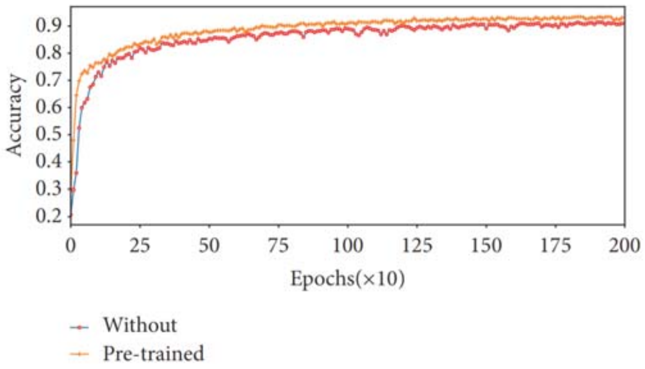
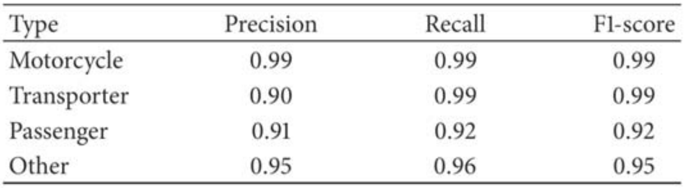


Figure 8. Comparison of training process between pretrained and without pretraining.

By analyzing the classification performance of pretrained CNNs, shown in Table 4, we can draw a conclusion that its performance is better than the one of CNNs without pretraining shown in Table 3, especially for the classification of transporter and passenger. In summary, pretrained CNN is more effective in recognizing the vehicles categories, and it is a state-of-the-art approach for vehicle classification.

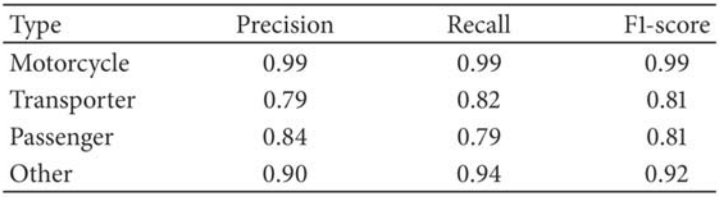
Table 4. Classification precision, recall, and f1-score of pretrained CNNs based on sheared dataset.



Type	Precision	Recall	F1-score
Motorcycle	0.99	0.99	0.99
Transporter	0.90	0.99	0.99
Passenger	0.91	0.92	0.92
Other	0.95	0.96	0.95

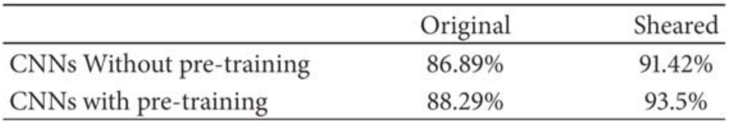
In the end, to verify the effect of detection in the entire system, we conducted ablation study by pretraining and testing our model on the original dataset which has not been sheared by YOLOv2 and contains a large quantity of irrelevant background. Ultimately, the accuracy of pretrained CNNs on original dataset is 88.29%, which is 5.21% lower than the 93.5% of pretrained CNNs on sheared dataset and even lower than the 91.42% of CNNs without pretraining on sheared dataset; the classification performance is shown in Table 5. And according the classification accuracy in Table 6, we can conclude that this ablation study confirms the essentiality of detection virtually in the whole vehicle classification system.

Table 5. Classification precision, recall, and f1-score of pretrained CNNs based on original dataset.



Type	Precision	Recall	F1-score
Motorcycle	0.99	0.99	0.99
Transporter	0.79	0.82	0.81
Passenger	0.84	0.79	0.81
Other	0.90	0.94	0.92

Table 6. The classification accuracy under different strategies.



	Original	Sheared
CNNs Without pre-training	86.89%	91.42%
CNNs with pre-training	88.29%	93.5%

CONCLUSIONS

A classification method based on CNNs has been detailed in this paper. To improve the accuracy, we used vehicle detection to removing the unrelated background for facilitating the feature extraction and vehicle classification. Then, an autoencoder-based layer-wise unsupervised pretraining is introduced to improve the CNNs model by enhancing the classification performance. Several state-of-the-art methods have been evaluated on our labeled data set containing four categories of motorcycle, transporter, passenger, and others. Experimental results have demonstrated that the pretrained CNNs method based on vehicle detection is the most effective for vehicle classification.

In addition, the success of our vehicle classification makes a vehicle color or logo recognition system possible in our refueling behavior analysis; meanwhile, it is a great help to urban computing, intelligent transportation system, etc.

ACKNOWLEDGMENTS

This research is supported by Youth Innovation Promotion Association CAS (2015355). The authors gratefully acknowledge the invaluable contribution of Yupeng Ma and the members of his laboratory during this collaboration.

REFERENCES

1. D. He, C. Lang, S. Feng, X. Du, and C. Zhang, "Vehicle detection and classification based on convolutional neural network," in *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*, 2015.
2. J. F. Forren and D. Jaarsma, "Traffic monitoring by tire noise," *Computer Standards & Interfaces*, vol. 20, pp. 466-467, 1999.
3. J. George, A. Cyril, B. I. Koshy, and L. Mary, "Exploring Sound Signature for Vehicle Detection and Classification Using ANN," *International Journal on Soft Computing*, vol. 4, no. 2, pp. 29–36, 2013.
4. J. George, L. Mary, and K. S. Riyas, "Vehicle detection and classification from acoustic signal using ANN and KNN," in *Proceedings of the 2013 International Conference on Control Communication and Computing, (ICCC '13)*, pp. 436–439, Thiruvananthapuram, India, 2013.
5. K. Wang, R. Wang, Y. Feng et al., "Vehicle recognition in acoustic sensor networks via sparse representation," in *Proceedings of the 2014 IEEE International Conference on Multimedia and Expo Workshops, (ICMEW '14)*, pp. 1–4, Chengdu, China, 2014.
6. A. Duzdar and G. Kompa, "Applications using a low-cost baseband pulsed microwave radar sensor," in *Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, (IMTC '01)*, vol. 1 of *Rediscovering Measurement in the Age of Informatics*, pp. 239–243, IEEE, Budapest, Hungary, 2001.
7. H.-T. Kim and B. Song, "Vehicle recognition based on radar and vision sensor fusion for automatic emergency braking," in *Proceedings of the 13th International Conference on Control, Automation and Systems, (ICCAS '13)*, pp. 1342–1346, Gwangju, Republic of Korea, 2013.
8. Y. Jo and I. Jung, "Analysis of vehicle detection with wsn-based ultrasonic sensors," *Sensors*, vol. 14, no. 8, pp. 14050–14069, 2014.
9. Y. Iwasaki, M. Misumi, and T. Nakamiya, "Robust vehicle detection under various environments to realize road traffic flow surveillance using an infrared thermal camera," *The Scientific World Journal*, vol. 2015, Article ID 947272, 2015.
10. J. Lan, Y. Xiang, L. Wang, and Y. Shi, "Vehicle detection and classification by measuring and processing magnetic signal," *Measurement*, vol. 44, no. 1, pp. 174–180, 2011.

11. B. Li, T. Zhang, and T. Xia, "Vehicle Detection from 3D Lidar Using Fully Convolutional Network," https://arxiv.org/abs/1608.07916, 2016.
12. V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image and Vision Computing*, vol. 21, no. 4, pp. 359–381, 2003.
13. F. M. Kazemi, S. Samadi, H. R. Poorreza, and M.-R. Akbarzadeh-T, "Vehicle recognition based on fourier, wavelet and curvelet transforms - A comparative study," in *Proceedings of the 4th International Conference on Information Technology-New Generations, ITNG '07*, pp. 939-940, Las Vegas, Nev, USA, 2007.
14. J. Y. Ng and Y. H. Tay, "Image-based Vehicle Classification System," https://arxiv.org/abs/1204.2114, 2012.
15. R. A. Hadi, G. Sulong, and L. E. George, "Vehicle Detection and Tracking Techniques: A Concise Review," *Signal & Image Processing: An International Journal*, vol. 5, no. 1, pp. 1–12, 2014.
16. J. Arróspide and L. Salgado, "A study of feature combination for vehicle detection based on image processing," *The Scientific World Journal*, vol. 2014, Article ID 196251, 13 pages, 2014.
17. P. Piyush, R. Rajan, L. Mary, and B. I. Koshy, "Vehicle detection and classification using audio-visual cues," in *Proceedings of the 3rd International Conference on Signal Processing and Integrated Networks, (SPIN '16)*, pp. 726–730, Noida, India, 2016.
18. Z. Chen, T. Ellis, and S. A. Velastin, "Vehicle detection, tracking and classification in urban traffic," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems, ITSC '12*, pp. 951–956, Anchorage, Alaska, USA, 2012.
19. X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Information Sciences*, vol. 295, pp. 395–406, 2015.
20. P. Mishra and B. Banerjee, "Multiple Kernel based KNN Classifiers for Vehicle Classification," *International Journal of Computer Applications*, vol. 71, no. 6, pp. 1–7, 2013.
21. A. Tourani and A. Shahbahrami, "Vehicle counting method based on digital image processing algorithms," in *Proceedings of the 2nd International Conference on Pattern Recognition and Image Analysis, IPRIA '15*, pp. 1–6, Rasht, Iran, 2015.

22. H. Wang, Y. Cai, and L. Chen, "A vehicle detection algorithm based on deep belief network," *The Scientific World Journal*, vol. 2014, Article ID 647380, 7 pages, 2014.

23. M. Yi, F. Yang, E. Blashch et al., "Vehicle Classification in WAMI Imagery using Deep Network," in *Proceedings of the SPIE 9838: Sensors and Systems for Space Applicatioins IX*, 2016.

24. B. Li, T. Zhang, and T. Xia, "Vehicle detection from 3D lidar using fully convolutional network," in *Proceedings of the Robotics: Science and Systems*, 2016.

25. Y. Lecun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

26. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

27. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Neural Information Processing Systems*, pp. 1097–1105, 2012.

28. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," https://arxiv.org/abs/1312.6229, 2013.

29. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 580–587, Columbus, Ohio, USA, 2014.

30. R. Girshick, "Fast R-CNN," in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV '15)*, pp. 1440–1448, Santiago, Chile, 2015.

31. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR '16)*, pp. 779–788, Las Vegas, Nev, USA, 2016.

32. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

33. W. Liu, D. Anguelov, D. Erhan et al., “SSD: single shot multibox detector,” in *Proceedings of the Computer Vision – ECCV 2016*, vol. 9905 of *Lecture Notes in Computer Science*, pp. 21–37, 2016.
34. J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object Detection via Region-based Fully Convolutional Networks,” https://arxiv.org/abs/1605.06409, 2016.
35. J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, (CVPR ‘17)*, pp. 6517–6525, Honolulu, Hawaii, USA, 2017.
36. D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, “Why does unsupervised pre-training help deep learning?” *Journal of Machine Learning Research*, vol. 11, pp. 625–660, 2010.
37. G. E. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
38. Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” *Neural Information Processing Systems*, pp. 153–160, 2007.
39. J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, “Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction,” in *Proceedings of the Artificial Neural Networks and Machine Learning - ICANN 2011*, vol. 6791 of *Lecture Notes in Computer Science*, pp. 52–59, Springer, Berlin, Heidelberg, Germany, 2011.
40. G. E. Hinton and R. S. Zemel, “Autoencoders, minimum description length and helmholtz free energy,” *Neural Information Processing Systems*, pp. 3–10, 1994.
41. M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, “Deconvolutional networks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’ 10)*, pp. 2528–2535, San Francisco, Calif, USA, 2010.
42. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ‘05)*, vol. 1, pp. 886–893, 2005.

43. E. Tola, V. Lepetit, and P. Fua, "DAISY: an efficient dense descriptor applied to wide-baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, 2010.

44. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2564–2571, Barcelona, Spain, 2011.

45. D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1150–1157, Kerkyra, Greece, 1999.

SECTION 2:
MULTIMEDIA APPLICATIONS IN HEALTH AND MEDICINE

CHAPTER 6

STUDY OF MULTIMEDIA TECHNOLOGY IN POSTURE TRAINING FOR THE ELDERLY

Chih-Chen Chen[1], De-Jou Hong[2], Shih-Ching Chen[3], Ying-Ying Shih[4], Yu-Luen Chen[2,5]

[1]Department of Management Information System, Hwa Hsia Institute of Technology, Taipei, Chinese Taipei

[2] Department of Computer Science, National Taipei University of Education, Taipei, Chinese Taipei

[3] Department of Physical Medicine & Rehabilitation, Taipei Medical University Hospital, Taipei, Chinese Taipei

[4] Department Physical & Rehabilitation Medicine, Chang Gung Memorial Hospital, Tao-Yuan, Chinese Taipei

[5] Department of Information Management, St. Mary's Medicine, Nursing and Management College, Yilan, Chinese Taipei

Citation: Chen, C., Hong, D., Chen, S., Shih, Y. and Chen, Y. (2013). Study of Multimedia Technology in Posture Training for the Elderly. Engineering, vol. 5, 47-52. doi: 10.4236/eng.2013.510B010.

Copyright: © 2013 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

ABSTRACT

Wrist diseases, also known as “mama hand”, “mouse hand” or “keyboard hand”, are commonly seen and easily over-looked symptoms in daily life. The diseases are mostly related to incorrect exercise or excessive force imposed on hands, leading to tenosynovitis or carpal tunnel syndrome. To alleviate the symptoms or even to recover, besides drug treatment, supplementary rehabilitation training is necessary. The rehabilitation process forces the affected wrist area to continually exercise in order to stimulate the self-repair signals sent to the affected area for partial restoration, if not all, of the supposed functions. In modern world, using information technology to improve the rehabilitation environment with better approaches has become a trend. In this study, software programs coded in C# and Flash are developed to work on the Bluetooth ball hardware to facilitate the rehabilitation. The basic principle is to use the reaction from the Bluetooth ball’s swaying or swinging to drive the movement of the objects on the computer screen, making the supposedly boring rehabilitation much more fun and vivid in the interactive multimedia environment, thus gaining better treatment effect. Also, RFID, Internet and databases can be integrated into this facility to provide patient identification and data storage for references in making rehabilitation training programs.

Keywords: Rehabilitation; Bluetooth Ball; RFID; Databases; Interactive Multimedia

PREFACE

The interactive multimedia controller has evolved rapidly, from the early knob and joystick to the Xbox360 controller, the PS3 controller’s direction button and press button, and then the Wii remote controller. The controller design has changed from the two-hand gripping to the single handheld rod, coupled with the triaxial acceleration detection technology to determine the action of the hand swaying and infrared optical positioning technology to detect the direction of the controller’s front end, for the interface control.

This study makes an insight into the rehabilitation training aided with the Bluetooth ball for signals transmission. The numeric values generated by the Bluetooth ball actions drive the movement of the objects on the computer screen, making the supposedly boring rehabilitation much more fun and vivid, for a better medical result. And the rehabilitation data can be integrated into the RFID patient identification, Internet access connection, computer server storage, and database queries, to constitute a comprehensive management mechanism for medical references.

Of all the wrist diseases, the “mama hand”, whose scientific name is stenosing tenosynovitis, occurs at age 30 to 50, and female’s incidence is about 10 times that of male. It’s the stenotic tendon bursitis of the extended thumb brevis tendon and longus tendon of the thumb abduction near the side of wrist, which brings tendon glide difficulties and causes wrist pain. However, the “mouse hand” and “keyboard hand” are of the carpal tunnel syndrome, which is caused by the wrist transverse ligament pressing the wrist median nerve, occurring at age 30 to 60 to those who repeat their wrist exercises over and over, and statistics shows the male and female incidence ratio is 1:4. Also, using mobile phones continuously for a long time can bring great pressure to the palm, causing wrist pains in many cases. To alleviate the symptom or even recover from the disease, besides drug treatment, complimentary rehabilitation is necessary.

LITERATURE REVIEW

Bluetooth ball was created by Finland Ball-It [1], since it has been massively used in entertainment. Following are some of the literatures on its applications. In his article titled “Interactive multimedia applications—illustrated with the campus navigation system”, Chau-Cheng Cheung argues that the merit of Bluetooth ball wireless transmission is to get rid of the keyboard and mouse for control, and instead use the handheld Bluetooth ball to do pinch, press, toss, and rotation to send instruction signals to Quest3D for corresponding actions to control the objects in the scene, making the campus navigation system more flexible and fun [2].

Yu-Jie Lin, in his article titled “A hand rehabilitation system based on Bluetooth ball: case studies of stroke disabled adults,” points out that the rehabilitation system built on Bluetooth ball games can increase stoke patients’ interest in rehabilitation. The Bluetooth ball games presented on the computer screen can increase the visual stimulation, which appears very helpful in the interference phase and helps the participants understand their own rehabilitation statues. Besides, the visual and audio combination gets patients’ attention and makes them more interested in rehabilitation [3].

In his paper titled “The development of traceable intelligent weighing system” released for an industrial-university cooperative research project, Bai-Sheng Chen suggests a mechanism that can automatically integrate consumers' health information (e.g. height, weight, BMI, eating habit, exercise habit), calories of consumer food (e.g. beverages, light food, fruit) and its traceability system, health maintenance recommendations, and shopping recommendations, which are processed by the triaxial six-directional wireless Bluetooth ball to calculate the calorie consumption in exercise, so as to fulfill the goals of food traceability and health management [4].

Chin-Jui Chang et al. in the article "Applications for medical recovery using wireless control of a Bluetooth ball with a hybrid G-sensor and human-computer interface technology" point out that extension of the Bluetooth applications helps the development of human-machine interfaced interactive games which allow players to follow the rhythm of songs to change the hand pressing actions or the speed of swaying the Bluetooth ball, so as to improve the flexibility of the hand movement and enhance the body balancing [5].

THE RESEARCH METHODS AND PURPOSES

Wrist Diseases

The commonly seen wrist diseases are mama hand, mouse hand and keyboard hand. When housewives cook, they use hands to grip the pot handle and turning shovel, and move their wrists up and down, and such a persistent movement can hurt the wrist and cause the mama hand disease. When working on a computer, one will use fingers to click the mouse or two hands to strike the keyboard, with the wrist posed in a fixed position for a long duration, and such a persistent movement day after day can finally cause serious damage to the wrist. Also, riding motorcycle with a fixed hand position to turn the throttle can also cause keyboard hand disease after a long-term operation.

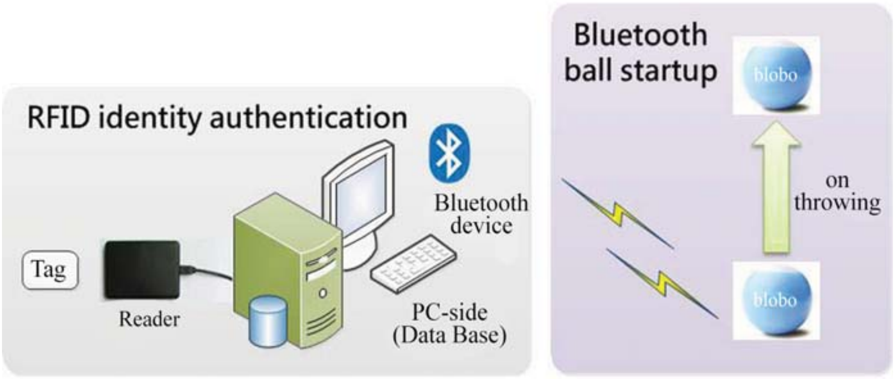
System Planning

Figure 1 shows the schematic overview of the study. Before using the Bluetooth system, following preprocesses are required:

- The hardware: This includes an RFID reader and tag, a PC with a Bluetooth device, and a Bluetooth ball (make sure it is adequately charged and can communicate with the PC via the Bluetooth). Since Bluetooth ball stays in sleep mode under constant state, we need to wake it up by tossing it, shaking is or swaying it. On the system startup, the ball's blue light will flash.

- The software: Microsoft Visual Studio C# 2008, Microsoft SQL Server 2008, Adobe Flash CS4. Visual Studio C# constitutes the main program codes, and SQL Server is responsible for managing the database that stores the RFID card numbers, user information and the Bluetooth ball-related data. As for Flash CS4, it supports the Action Script 3.0 syntax and the Flash file is played under Visual Studio C#.

After all the preparation is done, we can start to execute the program. To begin with, it is the RFID authentication, during which the RFID tag is detected by the reader which then displays the tag number and sent it to the SQL Server to search for a matched tag number and its holder information stored in the database; if the RFID holder is an administrator, the administrator's screen will be displayed, and if the holder is a general user, the user training screen will be displayed. Before doing the exercise, toss up the Bluetooth ball to start it, then connect the PC to the Bluetooth and test the communication by moving the ball left and right in the hand for the ball's built-in electronic gyroscope, pressure sensor and GSensor to detect the X-axis acceleration values. When the exercise is done, the wireless communication between the PC and the Bluetooth ball will terminate, and X-axis acceleration values resulted from the exercise will be averaged and stored in the SQL Server database for retrieval later on when one wants to resume previous exercise so the averaged X-axis acceleration value can be applied directly to the gaming screen.



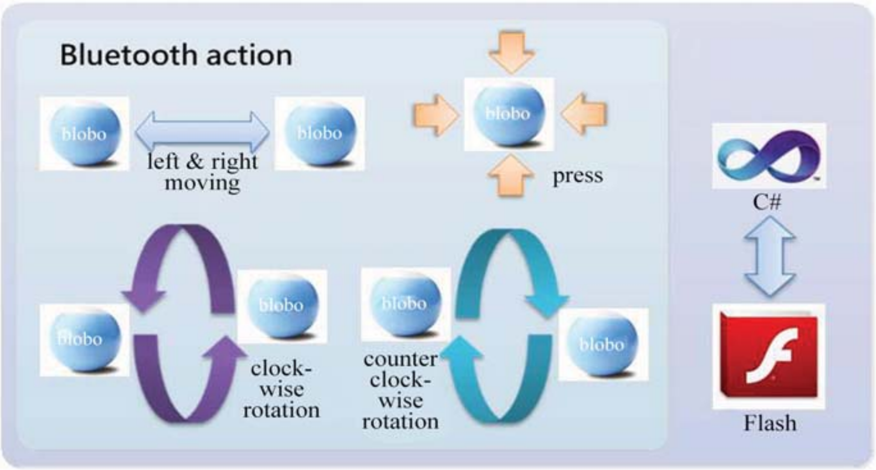


Figure 1. Overall schematic diagram.

The gaming mechanism is schemed into two parts: the Bluetooth ball movement and the PC end reactions. The former part includes moving left and right, press, clockwise and counter clockwise rotations which are respectively reacted in the latter part PC-end screen display of moving left, right, up and down, and rotation. Before the game is started, play the Flash file under the Visual Studio C# execution first, and then proceed to connect the PC with the Bluetooth ball so that the ball can do its work, which is correspondingly reacted on the PC screen. After the game is done, the communication between the PC and the Bluetooth ball will terminate, and the data collected during the gaming period will be stored in the SQL Server database and can be used for analysis on the training effect.

Process Flow Design

- The hardware: connection between the PC and Bluetooth ball: Figure 2 demonstrates the flow of connection between the PC and Bluetooth ball, and the connection is based on the Bluetooth communication protocol. Bluetooth ball is in sleep mode when it stays still. So, toss the ball up or heavily shake it to wake it up before using it. Once the ball is initiated, it is in on status, ready for subsequent connection to the PC. The ball will seek available connection devices, and if the connection fails, it could be insufficient charge of the ball; if so, recharge the ball and have it seek connection spots again.

- The software: This is categorized into preprocessing of database creation, identity authentication and training activities.

1) Identity authentication

The flow chart of identity authentication shown in Figure 3 is to determine the identity of an administrator or a user. The authentication is done by matching the RFID tag number which is unique to an individual user, and is required for entry into the program. If the identity is determined as an administrator, the maintenance mechanism for managing the electronic tag numbers and user identities in the database is enabled; if the tag number is absent in the database, an authentication failure message will be displayed and adding a new tag number to the database is required so that the new user can proceed with the operations.

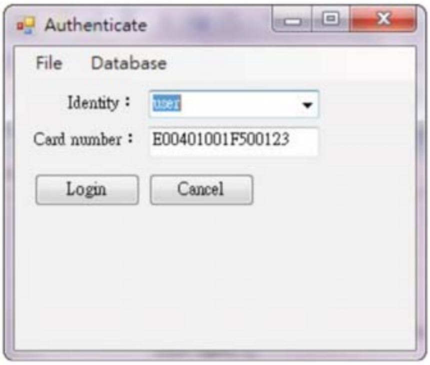
2) Training activities

The overall flow chart of the training activities shown in Figure 4 can be divided into the practice screen and gaming screen. After being authenticated, the user enters the practice screen. Since every user has his/her unique hand acceleration force, before getting to the formal exercise, the user is required to do the acceleration force practice and test. This is done by holding the Bluetooth ball for a specified duration and continuously swinging it left and right for it to capture the acceleration values in the right and left directions (X-axis); the values are then averaged and stored in the database. One of the games designed in this study is Tetris, which has graphic boxes in different shapes on the screen for the player to rearrange them into appropriate positions, and the player scores when a bunch of boxes are seamlessly put together into a row, and the row is then erased. Later on, when the patient in training re-enters the gaming screen, the patient's acceleration force data is first retrieved from the database and passed to the gaming screen. Before the game begins, the Flash needs to be initiated and then the PC and the Bluetooth ball be connected for communication. The Bluetooth movements are left and right moves, press and rotation, which are correspondingly reacted on the screen as the manipulated object's left and right, up and down moves and rotation. When the gaming is finished, the connection between the PC and Bluetooth ball terminates, and the time used as well as the scores gained during the gaming session are stored in the database and can be used later on for rehabilitation analysis and program planning.

RESULTS AND DISCUSSION

The purpose of this study is to introduce the multimedia interactive technology to the wrist rehabilitation training programs, in the hope of presenting an interesting and convenient wrist rehabilitation approach.

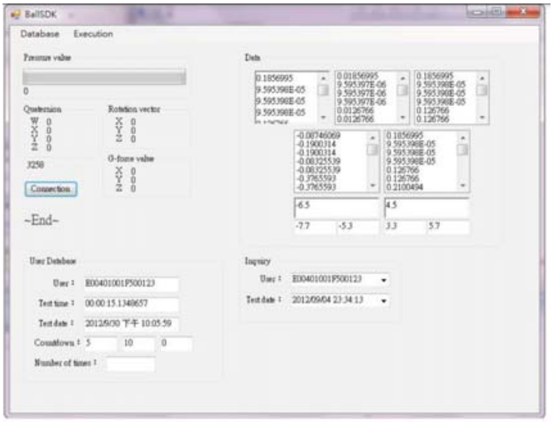
- *Identity authentication*: Figure 5 shows the RFID authentication screen. The RFID reader detects the tag number and sends it to the database for identity matching authentication
- *The practice screen*: Figure 6 shows the practice screen. The user holds the Bluetooth ball in his hand and sways it left and right, and the system fetches the ball's X-axis acceleration values in a fixed time interval. When the practice is done, the user's averaged acceleration values in the left and right directions are calculated for later use in the formal gaming exercise.
- *The gaming screen*: One of the games developed in this study is Tetris as shown in Figure 7. Originally, Tetris was played with the keyboard. The new Tetris developed in this study is manipulated with the electronic gyroscope, pressure sensor and G-Sensor inside the Bluetooth ball to make wrist rehabilitation much more effective and fun as well
- *The empirical test:* In this study, a middle-aged woman suffered from minor mama hand is taken as an empirical test subject (see Figure 8). The training is given 3 times a week, with each time taking 10 minutes, for a total training period of 3 weeks. In this training, the Bluetooth ball's left and right moves, rotation and press are given to manipulate the Tetris game. Of the manipulative moves, the left and right moves command horizontal shifts of the square to the left and right; the rotation flips the square over; the press action drops the square quickly. It is a 25-point gain when a bunch of boxes are put together into a seamless row and the row will then be erased. The score will be accumulated as more rows are completed. During the gaming, if any of the squares is stacked up to touch the top of the screen, the game is over. If the training time is not up yet, click the game again to re-enter the gaming screen and continue the training.



Identity : user

Card number : E00401001F500123

Figure 5. Identity authentication.



Pressure value
0

Quaternion
W 0
X 0
Y 0
Z 0

Rotation vector
X 0
Y 0
Z 0

G-force value
X 0
Y 0
Z 0

3250

Data

```
0.1856995        0.01856995       0.1856995
9.595398E-05     9.595397E-06     9.595398E-05
9.595398E-05     9.595397E-06     9.595398E-05
9.595398E-05     9.595397E-06     9.595398E-05
                 0.0126766        0.126766
                 0.0126766        0.126766

-0.08746069      0.1856995
-0.1900314       9.595398E-05
-0.1900314       9.595398E-05
-0.08325539      9.595398E-05
-0.08325539      0.126766
-0.3765593       0.126766
-0.3765593       0.2100494

-6.5             4.5
-7.7   -5.3      3.3   5.7
```

User Database

User : E00401001F500123

Test time : 00:00:15.1348657

Test date : 2012/9/30 下午 10:05:59

Countdown : 5 10 0

Number of times :

Inquiry

User : E00401001F500123

Test date : 2012/09/04 23:34:13

Figure 6. Practice screen.

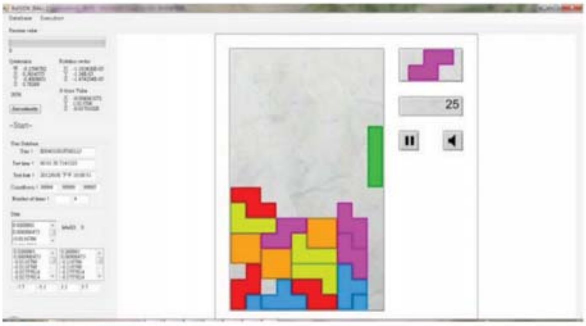


Figure 7. Game screen.

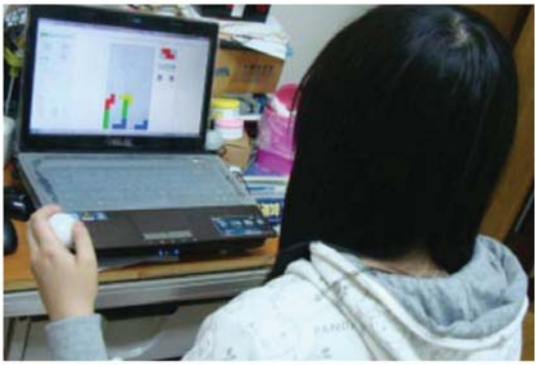


Figure 8. Empirical test.

CONCLUSION

In this study, the Bluetooth wireless communication technology is applied to the medical rehabilitation practices. In contrast with the traditional rehabilitation approach that is compulsive exercise-centric, monotonous and unable to automatically record and analyze the rehabilitation data, the interactive multimedia-aided rehabilitation developed in this study offers versatility, vividness and mobility, bringing lots of fun to the supposedly boring and long-lasting rehabilitation process. Furthermore, the RFID authentication technology makes it possible to store the rehabilitation data, such as the rehabilitation dates and times used in the exercises, which can then be used for analyses. This makes breakthrough for the traditional approach, thus increases the rehabilitation effects.

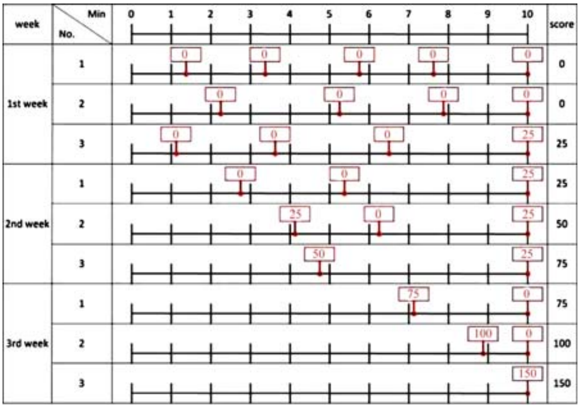


Figure 9. Training results.

ACKNOWLEDGEMENTS

The authors wish to thank the National Science Council (NSC) in Taiwan (Grant Number: NSC 101-2221-E-146- 001-, NSC101-2221-E-152-001-, NSC101-2627-E-002- 006-) for support this research.

REFERENCES

1. Ball-It. http://www.ball-it.com/index.html
2. C.-C. Cheung, “Interactive Multimedia Applications— Illustrated with the Campus Navigation System,” Master Thesis, Nan Jeon Institute of Technology, 2011.
3. Y.-J. Lin, “A Hand Rehabilitation System Based on Bluetooth Ball: Case Studies of Stroke Disabled Adults,” 2012.
4. B.-S. Chen, “The Development of Traceable Intelligent Weighing System,” A Concise Report on the Outcomes of the Industry, University Cooperative Research Program Subsidied by the National Science Council of the Executive Yuan, The Technology and Knowledge Application Type, 2010.
5. C.-J. Chang, C.-Y. Chen and C.-W. Huang, “Applications for Medical Recovery Using Wireless Control of a Bluetooth Ball with a Hybrid G-Sensor and Human-Computer Interface Technology,” Journal of Vibration and Control, Vol. 19, No. 8, 2013, pp. 1139-1151.

CHAPTER 7

RAPID EXTRACTION OF TARGET INFORMATION IN MESSY MULTIMEDIA MEDICAL DATA

Mai Xin[1], Zhifeng Ye[1], Changhua Chen[1], Yuhan Cai[2], Haili Ding[1], Ran Wang[1], Shiqian Wang[1]

[1]Library, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[2]Sichuan Hangbiao Electric Power Construction Co., Ltd., Meishan, China

ABSTRACT

The information work in the hospital is very complicated. The amount of data is huge, and the forms are various. It is very troublesome to organize and find the data you need. In order to solve the problem, which often appears in the process of analyzing medical case and data, we use Labview2017 to design the corresponding multimedia data calling system for the purpose of rapid searching and effective extraction. This program, which greatly improves work efficiency, can realize fast searching multimedia data in the

Citation: Xin, M., Ye, Z., Chen, C. , Cai, Y., Ding, H., Wang, R. and Wang, S. (2019). Rapid Extraction of Target Information in Messy Multimedia Medical Data. Intelligent Information Management, 11, 38-41. doi: 10.4236/iim.2019.112004.

Copyright: © 2019 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

folder. Since we use of this type of system, the work efficiency has been greatly improved, and the burden on the staff has been greatly reduced.

Keywords: Multimedia, Information Retrieval, LabVIEW2017, Fast Searching

INTRODUCTION

With the advancement of times, people's requirements for office efficiency are getting higher and higher [1]., which makes the level of office automation systems correspondingly higher [2]. In the daily information statistics of large hospitals, there are often huge folders of messy storage. It is very troublesome to organize such folders and obtain specific information in need [3]. When the data are stored in multimedia style, the file retrieval function of the operating system alone is difficult to meet the demand [4] [5] [6]. Using the search system that comes with the computer system, you can only retrieve files simply with inefficient [7].

If the corresponding data is damaged or cannot be used, it requires manual discrimination. Designing a kind of software which can quickly retrieve and access the contents of multimedia files to improve the retrieval efficiency before manual dis crimination is very necessary in office practice. Through further research work, we fully understood the needs of our staff. After that, work progress has been made and each task node has been controlled. On the basis of active communication and collaboration among all personnel, we conscientiously implemented relevant work.

CONSTRUCTION OF THIS SYSTEM

The process is shown in Figure 1: By setting the retrieved folder as a file path, we use the software written to do the following work. The files in the folder are matched according to the requirements of the retrieved fields, and then the files in various formats containing the retrieved fields are called for review. After that, they are directly opened by the default matching software for the operator to filter and decide the trade-off. It saves the manual verification process, which needs to open one by one. Using this method, we greatly improve the work efficiency.

Considering that the designed system should have the requirements of cross-platform, simplicity and rapidity, Labview2017 with good portability and operability is used as the development platform to design the software system. The designed operation interface and program interface are shown in Figure 2 and Figure 3.

EXPERIMENTAL ARGUMENTATION

The following is a demonstration by experiment. The searched folder is input in accordance with the search path shown in Figure 2, and the types of files corresponding to "Type 2" in the folder shown in Figure 4 which are matched and searched. We do it in this way to determine which information about type 2 diabetes is needed and which is useless, then get useful information and organize messy folders.

The results of various formats, which meet the search conditions, are all listed quickly and opened by their default software for the searcher to view, as shown in Figure 5.

CONCLUSIONS

It can be seen from the results shown in Figure 5 that the system can smoothly achieve the desired target, and also can quickly call data in various formats to realize the function of searching and retrieving data.

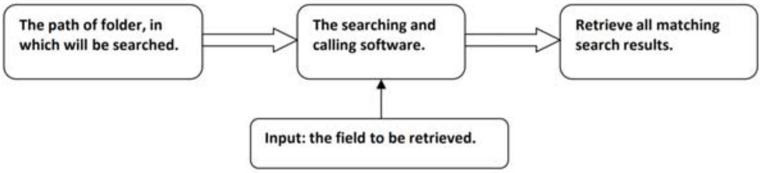

The path of folder, in which will be searched.
The searching and calling software.
Retrieve all matching search results.
Input: the field to be retrieved.

Figure 1. The process of system.

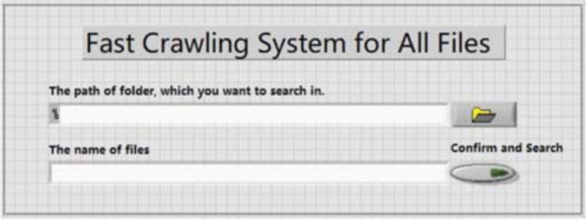

Fast Crawling System for All Files
The path of folder, which you want to search in.
The name of files
Confirm and Search

Figure 2. System operation interface.

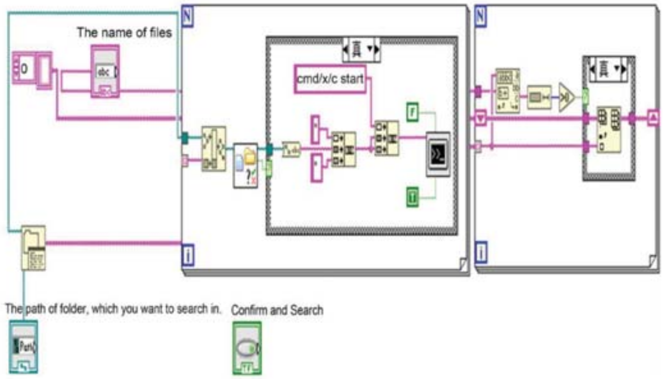
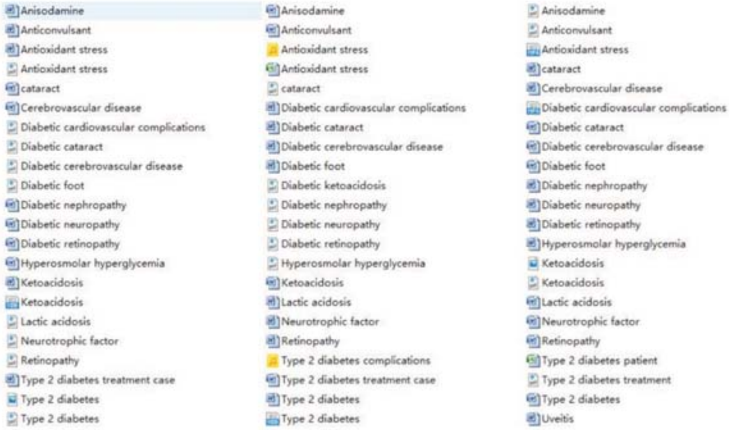


Figure 3. System program interface.

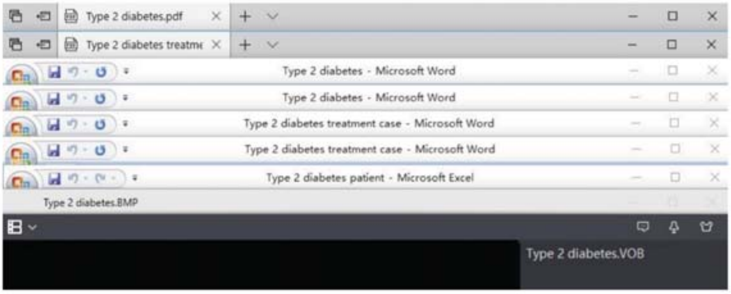


Anisodamine
Anticonvulsant
Antioxidant stress
Antioxidant stress
cataract
Cerebrovascular disease
Diabetic cardiovascular complications
Diabetic cataract
Diabetic cerebrovascular disease
Diabetic foot
Diabetic nephropathy
Diabetic neuropathy
Diabetic retinopathy
Hyperosmolar hyperglycemia
Ketoacidosis
Ketoacidosis
Lactic acidosis
Neurotrophic factor
Retinopathy
Type 2 diabetes treatment case
Type 2 diabetes
Type 2 diabetes

Anisodamine
Anticonvulsant
Antioxidant stress
Antioxidant stress
cataract
Diabetic cardiovascular complications
Diabetic cataract
Diabetic cerebrovascular disease
Diabetic foot
Diabetic ketoacidosis
Diabetic nephropathy
Diabetic neuropathy
Diabetic retinopathy
Hyperosmolar hyperglycemia
Ketoacidosis
Lactic acidosis
Neurotrophic factor
Retinopathy
Type 2 diabetes complications
Type 2 diabetes treatment case
Type 2 diabetes
Type 2 diabetes

Anisodamine
Anticonvulsant
Antioxidant stress
cataract
Cerebrovascular disease
Diabetic cardiovascular complications
Diabetic cataract
Diabetic cerebrovascular disease
Diabetic foot
Diabetic nephropathy
Diabetic neuropathy
Diabetic retinopathy
Hyperosmolar hyperglycemia
Ketoacidosis
Ketoacidosis
Lactic acidosis
Neurotrophic factor
Retinopathy
Type 2 diabetes patient
Type 2 diabetes treatment
Type 2 diabetes
Uveitis

Figure 4. Cluttered folder to be retrieved.



Type 2 diabetes.pdf
Type 2 diabetes treatme
Type 2 diabetes - Microsoft Word
Type 2 diabetes - Microsoft Word
Type 2 diabetes treatment case - Microsoft Word
Type 2 diabetes treatment case - Microsoft Word
Type 2 diabetes patient - Microsoft Excel
Type 2 diabetes.BMP
Type 2 diabetes.VOB

Figure 5. System operation result.

Thereby, the management efficiency of the file data is greatly improved, and the consumption of the operator's energy in the data sorting and data management is effectively reduced.

ACKNOWLEDGEMENTS

At that period, we nearly spent all our time in hospital. We did a lot of work in this field. At last, we achieve our goal and get progress. Thanks to all ones of our team, we finish this work successfully.

REFERENCES

1. Dan, M.I. (2018) Application of Office Automation in the Management of Documents and Archives. China Health Standard Management, 13, 6-8.
2. Wang, J., Chen, J.-L., Li, Z.-J. and Su, Y. (2018) Discussion on Management of Hospital Commercial Contract Based on Office Automation. Hospital Management Forum, 8, 73-74.
3. Fu, T. and Quan, Y. (2018) Paperless Hospital Management. Modern Hospital Management, 4, 82-84.
4. Tan, C. (2018) The Construction Research on OA Office System by Changsha Construction Investment & Development. Hunan University, Changsha.
5. Sun, Y.F. (2018) Design and Implementation of Office Automation System Based on SOA. Tianjin University, Tianjin.
6. Srivastava, H.S., Sivasankar, T. and Patel, P. (2018) An Insight into the Volume Component Generated from Risat-1 Hybrid Polarimetric Sar Data for Crop Biophysical Parameters Retrieval. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-5, 5, 209-214.
7. Ye, N., Walker, J.P., Rüdiger, C., Ryu, D. and Gurney, R.J. (2018) Remote Sensing; Investigators from Monash University Have Reported New Data on Remote Sensing (Surface Rock Effects on Soil Moisture Retrieval from L-Band Passive Microwave Observations). Journal of Technology & Science, 10, 33-43.

CHAPTER 8

"THE EFFECTIVENESS OF A MULTIMEDIA MESSAGING SERVICE REMINDER SYSTEM IN THE MANAGEMENT OF KNEE OSTEOARTHRITIS: A PILOT STUDY"

Gali Dar[1,2], Yaron Marx[3], Emma Ioffe[4], Einat Kodesh[1]

[1] Physical Therapy Department, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel
[2] Ribstein Center for Research and Sports Medicine, Wingate Institute, Netanya, Israel
[3] Maccabi Health Care Services, Afula and Tiberias, Israel
[4] Maccabi Health Care Services, Grand Canyon, Haifa, Israel

ABSTRACT

Background: Patient compliance to home exercise programs is significantly linked to improved treatment outcomes. Finding ways of encouraging patient conformity to these programs is imperative. For patients with osteoarthritis (OA) of the knee, a condition that causes pain, disability and

Citation: Dar, G., Marx, Y., Ioffe, E. and Kodesh, E. (2014). "The Effectiveness of a Multimedia Messaging Service Reminder System in the Management of Knee Osteoarthritis: A Pilot Study". International Journal of Clinical Medicine, 5, 483-489. doi: 10.4236/ijcm.2014.59067.

Copyright: © 2014 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

lessens the quality of life, exercise is essential for effective control of the condition. Aim: to investigate the effectiveness of multimedia messaging services (MMS) in improving patient adherence and functional outcome to home based exercise programs for patients suffering from knee OA. Methods: Fourteen patients diagnosed with knee OA and were referred to an exercise group therapy (for a total of six sessions) participated in this pilot study. The patients were randomly assigned to either the research or control group. The research group received MMS messages additional to the exercise sessions (video of exercises up to 10 seconds in length) to their mobile phone. Outcome measurement included the Western Ontario and McMaster Universities Arthritis Index (WOMAC) questionnaire, combined Focus On Therapeutic Outcomes (FOTO) questionnaire, Visual Analog Scale for Pain (VAS), Fear-Avoidance Beliefs Questionnaire (FABQ) and a general questionnaire. Results: The research group had baseline scores representing slightly higher disability, pain and fear avoidance than the control group as observed by the lower FOTO score and the higher FABQ, VAS and WOMAC scores. Analyzing the difference between initial and final scores revealed that the research group had a slightly higher perceived functional improvement. Conclusions: This study addressed the feasibility of short video messaging via mobile phones in increasing compliance to home exercise programs prescribed to patients suffering from knee OA. This pilot study provides an indication for the potential of success and a larger sample study should be conducted.

Keywords: Osteoarthritis, Multimedia, Physical therapy, Rehabilitation, Exercises

INTRODUCTION

Patient compliance to home exercise programs is significantly linked to improved treatment outcomes. Finding ways of encouraging patient conformity to these programs is imperative. Indeed, in a recent Israeli demographic study, it was found that compliance with self-exercise programs was one of the strongest predictors of functional outcome [1]. Exercise is essential for effective control of osteoarthritis (OA) of the knee, a condition that causes pain, disability and poor quality of life [2]. It had been shown that physical therapy and exercise are essential for improving function, decreasing pain and reducing the need for surgical intervention among OA patients [3] [4]. However, OA sufferers were more likely to maintain a supervised rather than home exercise program [5].

Self-exercise programs clash with patient compliance as they are intrinsically time consuming and often require patience and perseverance [1]. Forgetfulness and lack of motivation are common reasons for non-adherence [6]. Short message service (SMS) texting is a constructive approach of directly addressing the problem of forgetfulness and perhaps a way of reinforcing motivation [6]. When the patient is convinced that it is worth the time and effort, good compliance with exercise programs can be achieved [1]. Health care professionals have limited contact time with their patients, therefore, finding effective and efficient means of teaching and reinforcing self-responsibility for rehabilitation should be a goal. The last decade has seen a technological revolution in communication and the universality of mobile phones. Currently, with phones that allow messaging of both text and video, an opportunity has opened up for health professionals to maintain remote contact with patients between visits. Mobile or cellular phones have become a ubiquitous appliance. In developed countries such as the United States, mobile phone use is similar across socioeconomic backgrounds [7] [8].

In recent years, SMS text messaging has been proven to be effective in all niches of health related fields. Studies have confirmed the usefulness of SMS text messaging in improving recall of rehabilitation goals [9], monitoring glucose levels in diabetic patients [10] and weight loss programs [11]. Armstrong et al. (2009) [12] demonstrated that daily text messaging was an effective way of improving adherence to daily sunscreen use. Mobile phones, internet- based motivation and action support systems have also been shown to significantly increase and maintain the level of physical activity in healthy adults [13]. Worth noting is that older populations were just as likely to favor SMS text message reminders [14]. Telemedicine is an attractive means of complimenting conventional patient-practitioner contact as it is readily accessible (and rapidly becoming more so) to most people in all sectors of the community. It is relatively inexpensive, discreet and if set up on an automated system, requires minimal time of the health practitioner [7]. The ever increasing number of applications on mobile phones (including GPS, music, internet etc.) indicates the possibilities for more advanced telemedicine which will continue to expand and provide innovative means of meeting current challenges. Davalos et al. (2009) wrote a comprehensive review of the aspects related to the economics of telemedicine. Of note, they stressed that there is a lack of economic studies to date and those addressing economic issues usually consider at the cost of running the telemedicine program without fully accounting for its benefits [15].

Telemedicine may also aid physiotherapists by providing a way of reminding patients to execute their exercises (via SMS text messaging) and possibly increase patient motivation and interest by using video messaging to demonstrate the exercises. The use of multimedia messaging services (MMS) in encouraging compliance of home exercise programs is a simple and cost-effective method with potential to increase patient satisfaction and improve treatment outcomes. However, the effectiveness of SMS/MMS reminder messages to carry out physiotherapist prescribed home exercises on patient compliance and functional improvement has not been reported in the literature.

This research aims to investigate the usefulness of MMS video messaging in improving patient adherence and functional outcome in accordance with home based exercise programs for patients suffering from OA of the knee.

MATERIALS AND METHODS

Participants

Patients with OA of the knee were recruited from a physiotherapy group exercise class given at an out-patient clinic of Maccabi Health Services in Haifa, Israel. Patients identified as suitable for the study were randomly assigned to either the research group (who during the course of their participation in the exercise group therapy received MMS messages) or the control group (who did not receive MMS messages). Inclusion criteria included an orthopedic diagnosis of knee OA and at least a moderate competency in using mobile phone and opening MMS messages. The Helsinki Committee of Maccabi Health Services approved this study. Eligible patients read and signed an informed consent form prior to enrollment.

Procedures

The knee exercise group met once a week under the instruction of a physiotherapist with referred patients able to join at any stage and participate for a total of six sessions. Participants in the research group received MMS short video messages (up to 10 seconds in length) to their mobile phones every second day (excluding Saturday) with a visual reminder as to one of the home exercises prescribed at the group session (a verbal explanation, given by a physiotherapist, accompanied the visual video demonstration). A library of 12 MMS messages was created and messages were chosen arbitrarily from this library.

Outcome Measurement

Prior to the first group session, all participants (research and control groups) filled out the following questionnaires: Western Ontario and McMaster Universities Arthritis Index (WOMAC), combined Focus on Therapeutic Outcomes (FOTO), Visual Analog Scale for Pain (VAS), Fear-Avoidance Beliefs Questionnaire (FABQ) and a general questionnaire. At the final session, the patients filled out the same series of questionnaires. The aim of the questionnaires was to measure pain and function as well as home compliance of the exercise program and effectiveness of MMS messages in motivation.

Detailed descriptions of the questionnaires:

- *FOTO (Focus on Therapeutic Outcomes):* FOTO is a web-based short assessment of the functional level of the impaired body part, returning a score between 0 - 100. The score represents a percentage of the functional status of the impaired body part relative to a healthy patient, i.e. a score of 30 would indicate that the patient had only 30% of the ability compared to a healthy knee. A higher score signified high function or better outcome.
- *FABQ (Fear-Avoidance Beliefs Questionnaire):* The FABQ questionnaire was developed to assess avoidance behavior resulting from fear of pain [16]. The questionnaire consists of two subscales: FABQPA—questions relating to physical activity and FABQW—questions relating to work. A lower score signified lower fear avoidance.
- *VAS (Visual Analog Scale for Pain):* This scale was used to assess the severity of current or recent pain with a higher score corresponding to greater pain intensity. Patients were asked to respond on a scale of 0 to 10, the amount of pain felt, with a higher number corresponding to greater pain.
- *WOMAC (Western Ontario and McMaster Universities Arthritis Index):* WOMAC is a standard questionnaire used to evaluate patients with OA of the knee and hip [17], assessing pain, joint stiffness, physical function, and the social and emotional function

of a person with OA in determining the overall level of disability. The questionnaire consists of 24 questions for which the patient can choose responses on a scale of 0 to 4 with 0 corresponding to "not at all" and 4 corresponding to "very much". A lower score signified high function or better outcome.

- *Initial General Questionnaire*: This short questionnaire was designed for this study and consists of four questions: 1) level of physical activity in everyday life, 2) competency in handling a mobile phone, 3) level of expectation from the group exercise sessions and 4) perceived disability. The participants were asked to answer on a scale of 0 to 4 with 0 corresponding to "not at all" and 4 corresponding to "very much".
- *Final General Questionnaire*: This questionnaire was created for this study and was completed at the end of the final group exercise session. Both the research and control groups were presented with five questions aimed at assessing perceived improvement in function and usefulness of the group sessions and home exercises as well as compliance with the home exercise program. The research group was presented with an additional five questions aimed at assessing the perceived usefulness of the MMS messages in providing a reminder to perform the exercises, a reminder to correctly carry out the exercises as well as compliance to opening the MMS messages and exercising after opening the messages.

Analysis of Data

The statistical analysis was conducted using SPSS v.16.0. and Excel software. T test and Pearson correlation coefficients test were utilized to compare research and control groups. Significant difference for all measurements was set as $p < 0.05$.

RESULTS

Participants

Fourteen patients were recruited to the study; 9 were assigned to the control group and 5 to the research group. The research and control groups were composed of a similar mix of sexes (67% and 57% female, respectively) although the research group was on average, younger - 56 years old compared to 69 in the control group. The average baseline scores for the FOTO, FABQ, VAS and WOMAC questionnaires are shown in Figure 1. In general, it can be stated that the control group had baseline scores representing slightly lower disability, pain and fear avoidance than the research group, as observed by the higher FOTO score and the lower FABQ, VAS and WOMAC scores.

Final, Post Intervention

Data Patient Compliance

The research group had a similar perception of the usefulness of the home exercises as did the control group. The control group had a lower perceived usefulness of reminder messages. For research group specific questions, there was an enthusiastic response when asked about the usefulness of the MMS messages as reminders for performing and correctly carrying out the exercises. The patients were asked if it was beneficial to include this service in future group sessions. The average response in all cases was that the participants valued this as "very" useful. The reported opening and immediate carrying out of exercises after opening the MMS messages corresponded with the reported exercise compliance (2/3 times per week) (Table 1).

Functional Outcomes

The average final, post intervention scores for the FOTO, FABQ, VAS and WOMAC questionnaires are shown in Figure 2. No significant difference between groups was found as to the functional outcome indicators ($p > 0.05$).

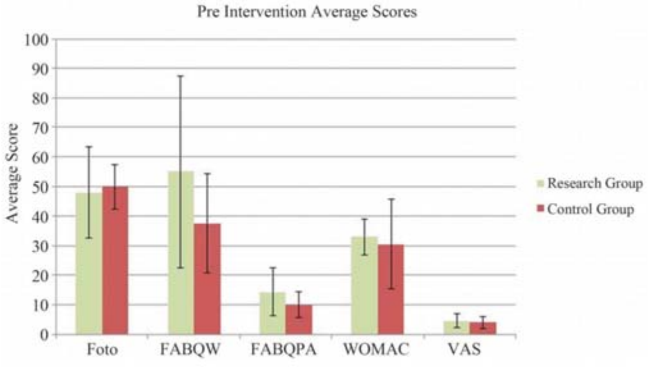


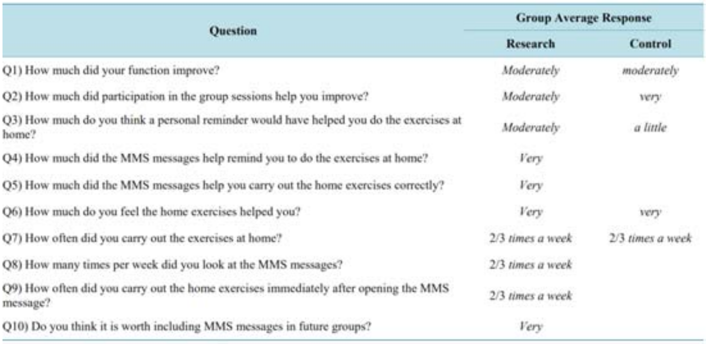
Figure 1. Average Baseline Questionnaire Scores.

However, a more informative measure of outcome differences between the two groups was achieved by analyzing the difference between initial and final scores. The group average of this pre and post intervention score difference for each patient is shown in Figure 3. Figure 3 is displayed as an absolute value so that comparison of any outcome measure between the two groups means that a greater value corresponded to a greater functional improvement. Comparing the research and control group's average response to the final general questionnaire scores showed that the research group exhibited a slightly higher perceived functional improvement.

DISCUSSION

This pilot study provides some indication that the use of MMS reminder messages could be effective in increasing compliance and functional outcome for patients with OA of the knee who were prescribed home exercises as part of their physiotherapy treatment. A comparison of pre and post intervention scores showed that for all out come measures, there was more improvement in the research group than the controls (although not significantly). The outcome measures for the research group were slightly poorer than the control group on average indicating that the research group was more functionally impaired at the onset of the study. The research group showed greater improvement than the control group in terms of the average absolute improvement in functional outcome measures between the onset and completion of the group exercise sessions. Whether this is a result of the MMS messaging or perhaps the fact that they started with a lower functional status making improvement more attainable or noticeable, will require further investigation.

Table 1. Response to Post Intervention General Questionnaire (empty lines questions are those which were not presented to the control group).



Question	Group Average Response	
	Research	Control
Q1) How much did your function improve?	*Moderately*	*moderately*
Q2) How much did participation in the group sessions help you improve?	*Moderately*	*very*
Q3) How much do you think a personal reminder would have helped you do the exercises at home?	*Moderately*	*a little*
Q4) How much did the MMS messages help remind you to do the exercises at home?	*Very*	
Q5) How much did the MMS messages help you carry out the home exercises correctly?	*Very*	
Q6) How much do you feel the home exercises helped you?	*Very*	*very*
Q7) How often did you carry out the exercises at home?	2/3 *times a week*	2/3 *times a week*
Q8) How many times per week did you look at the MMS messages?	2/3 *times a week*	
Q9) How often did you carry out the home exercises immediately after opening the MMS message?	2/3 *times a week*	
Q10) Do you think it is worth including MMS messages in future groups?	*Very*	

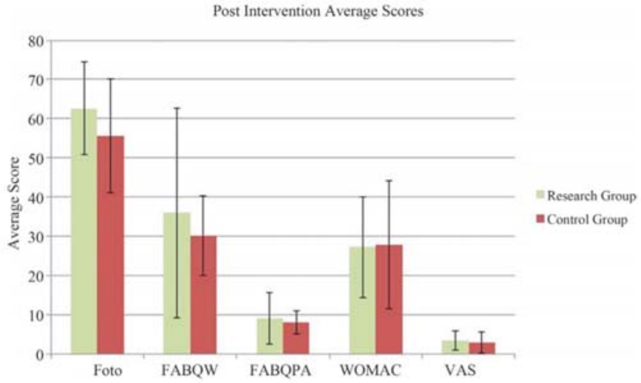


Figure 2. Average Post Intervention Questionnaire Scores.

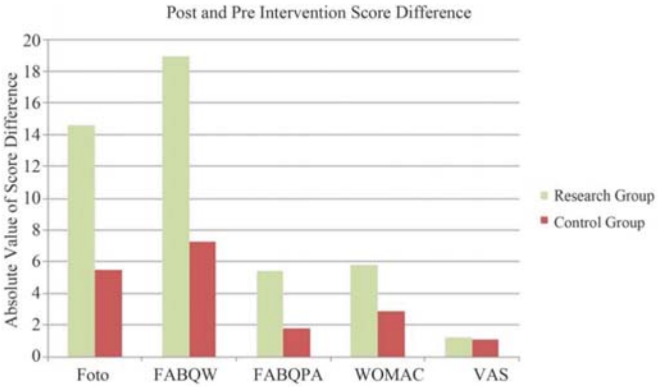


Figure 3. Absolute difference between post and pre intervention scores. Note: the actual difference in all cases was indicative of improvement.

The trend for greater functional improvement for participants receiving MMS reminders together with the success of SMS and MMS reminders had been demonstrated in other medical fields. Future studies may also benefit by the use of objective measures of treatment outcome such as physiotherapist assessment. This study used only subjective measures to assess pre and post intervention, i.e. questionnaires. In order to minimize the subjectivity of this method, several specific questionnaires were used including some which are internationally recognized. Other multimedia applications could possibly be advantageous in effectively maximizing patient treatment outcome, i.e. e-mail reminders may be used together with or in place of MMS messages. The internet could also be utilized as a forum for providing instruction videos reminding the patients how to correctly carry out the exercises and possibly a forum for online discussions regarding topics such as difficulties in performing the exercises or tracking progress. In this modern era, the use of computers has plays an important role in daily life, even within the older generation population; its potential within the realm of telemedicine should be developed.

CONCLUSION

This study addressed the feasibility of short video messaging via mobile phones in increasing compliance to home exercise programs prescribed to patients suffering from OA of the knee. This population can benefit from exercise and at the same time increase motivation and compliance to treatment.

STUDY LIMITATIONS

The major shortcoming of this study is the small number of participants which makes it difficult to gain statistically significant results. Therefore, this study should be regarded as a pilot study providing only an indication for the potential of success on a larger scale.

ACKNOWLEDGEMENTS

The authors thank Iluz limor and Amal kiwan (Maccabi Health Care Services, Grand Canyon, Haifa, Israel) for their contribution and support of this work.

REFERENCES

1. Deutscher, D., Horn, S., Dickstein, R., Hart, D., Smout, R., Gutvirtz, M. and Ariel, I. (2009) Associations Between Treatment Processes, Patient Characteristics, and Outcomes in Outpatient Physical Therapy Practice. Archives of Physical Medicine and Rehabilitation, 90, 1349-1363. http://dx.doi.org/10.1016/j.apmr.2009.02.005
2. Fransen, M., Crosbie, J. and Edmonds, J. (2001) Physical Therapy Is Effective for Patients with Osteoarthritis of the Knee: A Randomized Controlled Clinical Trial. Journal of Rheumatology, 28, 156.
3. Deyle, G., Henderson, N. and Matekel, R. (2000) Effectiveness of Manual Physical Therapy and Exercise in Osteoarthritis of the Knee. A Randomized Controlled Trial. Annals of Internal Medicine, 132, 173. http://dx.doi.org/10.7326/0003-4819-132-3-200002010-00002
4. Bennell, K., Hinman, R., Metcalf B., Buchbinder, R., McConnell, J., McColl, G., Green, S. and Crossley, K. (2005) Efficacy of Physiotherapy Management of Knee Joint Osteoarthritis: A Randomised, Double Blind, Placebo Controlled Trial. Annals of the Rheumatic Diseases, 64, 906. http://dx.doi.org/10.1136/ard.2004.026526
5. Chamberlain, M.A., Care, G. and Harfield, B. (1982) Physiotherapy in Osteoarthrosis of the Knees. A Controlled Trial of Hospital versus Home Exercises. Disability & Rehabilitation, 4, 101-106.
6. Chan, D., Lonsdale, C., Ho, P., Yung, P. and Chan, K. (2009) Patient Motivation and Adherence to Postsurgery Rehabilitation Exercise Recommendations: The Influence of Physiotherapists' Autonomy-Supportive Behaviors. Archives of Physical Medicine and Rehabilitation, 90, 1977-1982. http://dx.doi.org/10.1016/j.apmr.2009.05.024
7. Miloh, T. and Annunziato, R (2010) Adhering to Your Non-Adherent Patients: The Challenge of Non-Compliance. Acta Paediatrica, 99, 335-337. http://dx.doi.org/10.1111/j.1651-2227.2010.01702.x
8. Krishna, S., Austin Boren, S. and Balas, A. (2009) Healthcare via Cell Phones: A Systematic Review. Telemedicine and E-Health, 15, 231-240.
9. Culley, C. and Evans, J. (2010) SMS Text Messaging as a Means of Increasing Recall of Therapy Goals in Brain Injury Rehabilitation: A Single-Blind Within-Subjects Trial. Neuropsychological Rehabilitation, 20, 103-119. http://dx.doi.org/10.1080/09602010902906926

10. Kim, H.S. and Jeong, H.S. (2007) A Nurse Short Message Service by Cellular Phone in Type-2 Diabetic Patients for Six Months. Journal of Clinical Nursing, 16, 1082-1087. http://dx.doi.org/10.1111/j.1365-2702.2007.01698.x
11. Patrick, K., Raab, F., Adams, M.A., Dillon, L., Zabinski, M., Rock, C.L., Griswold, W.G. and Norman, G.J. (2009) A Text Message-Based Intervention for Weight Loss: Randomized Controlled Trial. Journal of Medical Internet Research, 11, e1. http://dx.doi.org/10.2196/jmir.1100
12. Armstrong, A., Watson, A., Makredes, M., Frangos, J., Kimball, A. and Kvedar, J. (2009) Text-Message Reminders to Improve Sunscreen Use: A Randomized, Controlled Trial Using Electronic Monitoring. Archives of Dermatology, 145, 1230-1236. http://dx.doi.org/10.1001/archdermatol.2009.269
13. Hurling, R., Catt, M., De Bonj, M., Fairley, B., Hurst, T., Murray, P., Richardson, A. and Sodhi, J. (2007) Using Internet and Mobile Phone Technology to Deliver an Automated Physical Activity Program: Randomized Controlled Trial. Journal of Medical Internet Research, 9, e7. http://dx.doi.org/10.2196/jmir.9.2.e7
14. Sahm, L., MacCurtain, A., Hayden, J., Roche, C. and Richards, H. (2009) Electronic Reminders to Improve Medication Adherence—Are They Acceptable to the Patient? Pharmacy World and Science, 31, 627-629. http://dx.doi.org/10.1007/s11096-009-9327-7
15. Davalos, M., French, M., Burdick, A. and Simmons, S (2009) Economic Evaluation of Telemedicine: Review of the Literature and Research Guidelines for Benefit-Cost Analysis. Telemedicine and E-Health, 15, 933-949.
16. Waddell, G., Newton, M., Henderson, I., Somerville, D. and Main, C.J. (1993) A Fear-Avoidance Beliefs Questionnaire (FABQ) and the Role of Fear-Avoidance Beliefs in Chronic Low Back Pain and Disability. Pain, 52, 157-168. http://dx.doi.org/10.1016/0304-3959(93)90127-B
17. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J. and Stitt, L.W. (1988) Validation study of WOMAC: A Health Status Instrument for Measuring Clinically Important Patient Relevant Outcomes to Antirheumatic Drug Therapy in Patients with Osteoarthritis of the Hip or Knee. The Journal of Rheumatology, 15, 1833-18340.

CHAPTER 9

VIRTUAL REALITIES IN THE TREATMENT OF MENTAL DISORDERS: A REVIEW OF THE CURRENT STATE OF RESEARCH

Christiane Eichenberg and Carolin Wolters

University of Cologne, Germany

INTRODUCTION

In the past decade, *virtual reality* **(VR)** technologies have been discussed as promising supplements in psychotherapy**.** Virtual realities enable users to interact in real time with computer-generated environments in three dimensions [1]. The fact that VR applications simulate real experiences and trigger anxiety, including physiological symptoms such as sweating or nausea, emphasizes their potential to replace conventional exposure therapy.

If users are to experience virtual environments as real, two conditions are required: immersion and presence. *Immersion* **describes a state of**

Citation: Christiane Eichenberg and Carolin Wolters (September 12th 2012). Virtual Realities in the Treatment of Mental Disorders: A Review of the Current State of Research, Virtual Reality in Psychological, Medical and Pedagogical Applications, Christiane Eichenberg, IntechOpen, DOI: 10.5772/50094.

Copyright: © 2012 by authors and IntechOpen. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

consciousness in which the user's awareness of the physical self declines due to an increasing involvement in the virtual environment. A sensation of immersion can be achieved by creating realistic visual, auditory or tactile stimulation. Additionally, the usage of specific output devices (e.g. data-goggles and monitors) and input devices (e.g. data gloves, voice recognition and eye tracking software) may facilitate the user's perception of immersion. The feeling of being physically immersed can result in a sense of *presence*, that includes a perception of the environment as being real, shutting out real-life stimuli and performing involuntary, objectively meaningless body movements such as ducking to avoid an object displayed in VR. Moreover, persons seem to experience a strong sense of control in VR. A study [2] showed that persons who were told to have control over the movements of an elevator but actually did not, rated their perceived control as high as those who in fact had control over the elevator.

Another technology that has been developed in the past years is referred to as *Augmented Reality* **(AR)**. AR describes the superimposition of virtual elements into the real world. Persons therefore see a visualization of the real world and virtual elements at the same time [3].Advantages of AR in comparison to VR may include an enhanced feeling of presence and reality, since the environment is in fact real. Additionally, AR might be less expensive, because the real world environment can be used as a scheme. Thus, the setting does not need to be entirely developed.

Research on the usage of VR and AR technologies in psychotherapy has mainly focused on behavioral therapy and was proven to be effective particularly in the treatment of specific phobias [4]. According to well-established behavior therapy theories, clients have to be exposed to fear inducing situations in order to treat phobias, because avoidance of fearful stimuli might stabilize the assumption that they are dangerous. Corrective experiences would thus be prevented. Two kinds of exposure can be implemented in therapy. While in-vivo exposure involves the immediate exposure to a fear-enhancing situation or object in reality, in-sensu exposure describes the mere imagination of the exposure to fearful stimuli. In terms of a graduated exposure, stimuli that trigger low levels of anxiety are usually presented first, increasing up to the client's most extreme fear, which is called "flooding" (in-vivo exposure) or "implosion" (in-sensu exposure).

BENEFITS AND COSTS OF APPLYING VR IN PSYCHOTHERAPY

As already mentioned, exposure therapy supported by VR technologies exceeds imaginative exposure by adding a sense of presence. Moreover, including VR applications in psychotherapy offers a series of advantages. These include the possibility of adjusting virtual environments to each client's specific needs and controlling what is presented to the client. In addition, VR enables the therapist to expose the client to conditions that might be unsafe or only accessible at high cost in the outside world, and to improve confidentiality by avoiding spectators [5]. Furthermore, therapists seem to consider VR exposure to be less aversive than in-vivo therapy [6]. Presumably, the same applies to patients. For instance, García-Palacios et al. [7] showed that only 3% of 150 participants suffering from specific phobia refused VR exposure, while 27% refused in vivo therapy.

Nevertheless, the usage of VR entails considerable costs. First of all, despite recent findings, some groups might be reluctant to the use of VR technologies and might therefore be excluded from treatment. Furthermore, the handling of VR applications requires a certain amount of training for therapists. Besides, therapists are tied to the position of VR equipment, since it is usually too unhandy to transport [1]. Additionally, equipment acquisition is rather expensive, even though costs have sunk dramatically in the past ten years [5]. Finally, clients might experience dizziness and nausea while undergoing a VR application, a syndrome referred to as simulation sickness [4]. But even though the cited costs have to be taken into account, a recent study [1] indicated that therapists perceive the benefits of VR supported psychotherapy to be outweighing potential costs.

Self-evidently, those costs should only be accepted on condition that VR applications are able to effectively treat mental disorders. The present article aims to outline recent findings in order to examine the effectiveness of usage of VR technologies in psychotherapy.

CURRENT STATE OF RESEARCH

Previous studies have mainly focused on the use of VR applications in the treatment of anxiety disordersand particularly specific phobias, such as fear of heights, fear of flying, fear of animals or social phobia. However, research has recently started to focus on the usage of VR in the treatment of other disorders as well, including eating disorders and sexual dysfunctions. In the following, an overview of the current state of research will be given. After briefly describing the search strategy, two meta-analyses that are concerned with the application of VR in the treatment of anxiety disorders will be presented. Subsequently, exemplary studies evaluating the effectiveness of VR-assisted psychotherapy of different specific disorders are summarized.

Method

In order to identify eligible studies, a search on the databases PsychInfo, PsychArticles and Pubmed was conducted. The search words *Virtual/ Augmented Reality, Exposure Therapy* **and** *effectiveness/ efficacy/ metaanalysis* **were entered alone and in combination with** *mental disorder* **and derivatives of the different terms for disorders, particularly** *acrophobia/ fear of heights/ aviophobia/ fear of flying/ arachnophobia/ fear of spiders/ social phobia/ fear of public speaking/ panic disorder/ posttraumatic stress disorder.* **To ensure the** currentness of the findings presented here, we focused on studies that were published within the past ten years, even though studies conducted before were not excluded if they contributed significantly to the current state of research.

Meta-analyses

Two current meta-analyses have been reported concerning the effectiveness of VR in the treatment of anxiety disorders. Parsons and Rizzo [8] analyzed *N*= 21 studies that used pre-post measurements but not necessarily a controlled study design. The authors found an average effect size of *d*=.95 (*SD*=.02) for the reduction of symptoms in VR-assisted therapy. The treatment of fear of flying (*d*= 1.5; *SD*=.05), and panic disorder with agoraphobia (*d*= 1.79; *SD*=.02) using VR applications accounted for the largest effect sizes concerning symptom reduction. They were followed by treatment of social phobia (*d*=.96; *SD*=.10), acrophobia (*d*=.93; *SD*=.06) and arachnophobia (*d*=.92; *SD*=.12), while the treatment of posttraumatic stress disorder (PTSD) by means of VR obtained the smallest effect size of *d*=.87 (*SD*=.01). In addition, a series of determining factors were assumed. These include the degree of immersion and presence, duration of disease and socio-demographic variables. However, due to a lack of data within the examined study, the authors were unable to make a valid statement about potential moderators.

Powers and Emmelkamp [9] examined *N*= 13 controlled studies, reverting to a more rigid design that excluded studies involving case reports, multiple components of treatment conditions, and an unequal amount of treatment sessions in the conditions compared. In general, effect sizes of VR exposure therapy were found to be large to very large, ranging from *d*=.85 to *d*= 1.67. A smaller effect size of *d*=.35 favored treatment with the aid of VR to in-vivo exposure and therefore demonstrates the superiority of VR in comparison to in-vivo treatment. Admittedly, studies considering therapy of specific phobias predominated.

Nevertheless, overall results prove that VR applications are highly effective in the treatment of anxiety disorders. However, difficulties common to the realization of meta-analyses, for instance a publication bias that favors publication of studies implying significant results, have to be taken into account. Moreover, small sample sizes as well as missing data about the point of time of follow-up ratings and therefore questionable lastingness of treatment effects, limit the meaningfulness of findings. Future research should include varied levels of immersion and ensure controlled study designs.

Exemplary studies of Various Syndromes

The meta-analyses presented here mainly focused on the effectiveness of VR as a supplement of behavior therapy for patients with anxiety disorders, some of the most frequently diagnosed psychological disorders. Nearly one out of five adults in the USA suffers from an anxiety disorder, whereat women are more often affected than men [10]. Therefore, the continuing development and evaluation of effective treatment methods seems crucial.

Anxiety disorders present the first syndrome category for which the use of modern media such as the Internet or VR technology as a setting for interventions was scientifically evaluated. They are usually assigned to the field of *behavior therapy*. Since anxiety disorders are frequently treated with the aid of exposure, they are suitable for VR settings. In contrast, *psychodynamic therapy* **concentrates more on relationship aspects. However, there are** conceptual considerations about how to integrate VR in psychodynamic therapy [11], and a few studies have already been conducted to examine the use of VR within the psychodynamic approach (e.g. [12]).

Anxiety disorders are classified differently within the two major diagnostic classification systems. While the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [13] sorts them within a separate chapter, the International Classification of Diseases (ICD-10) [14] includes them in the chapter „Neurotic, Stress and Somatoform Disorders". The latter distinguishes between the subgroups of phobic disorders (agoraphobia, social anxiety disorder, specific phobias) and other anxiety disorders (panic disorder, generalized anxiety disorders). In both classification systems, posttraumatic stress disorder is discussed along with anxiety disorders. In the following, the effectiveness of VR-assisted treatment of various syndromes is presented.

Fear of heights

Acrophobia, classified as a specific phobia of the naturalistic type, describes an extreme fear of heights. It involves the avoidance of various height-related situations, such as stairs, terraces, high buildings, bridges, or elevators. The fear of heights is widely spread: In a survey of more than 8000 adults, 20% stated that they had already experienced an exaggerated fear of heights in the past, which did not meet the criteria for acrophobia [15]. In the aforesaid study, the prevalence of acrophobia amounted to 5.3% and therefore closely followed the prevalence of fear of animals. While women usually tend to develop specific phobias substantially more often than men [12], merely 55 to 70% of acrophobic persons are female.

The first successful application of VR in the treatment of acrophobia was presented in a case study of an acrophobic student who was successfully treated using graded VR exposure [16]. A more extensive study including a sample of 20 students furnished further evidence for the effectiveness of VR-assisted treatment [17]. However, due to study limitations such as the absence of a control group, the further conclusions can only be drawn under reserve.

The first clinical trial of the effectiveness of VR in the treating acrophobia was conducted by Emmelkamp and collaborators [18]. In a within group design, ten patients were treated with two sessions of VR, followed by two sessions of exposure in-vivo. Acrophobic symptoms were measured before treatment, after VR treatment and after in-vivo exposure. Results showed that after being treated by the means of VR, exposure to real situations did not lead to any significant improvement on the Acrophobia Questionnaire (AQ) or the Attitudes Towards Heights Questionnaire (ATHQ). Unexpectedly, the research design had created a ceiling effect, insofar as the VR treatment effects left little space for improvement during exposure in-vivo.

In a randomized controlled trial (RCT) conducted by the same research group, effectiveness of exposure by the means of VR and in vivo were compared [19]. The places used in the exposure in vivo were reproduced in a virtual environment. Exposure was affected in a real or virtual shopping mall in Amsterdam, a fire escape, and a roof garden. *N*= 33 acrophobic persons underwent three weekly sessions of one hour each. Anxiety levels were reported on the Subjective Units of Disturbance-Scale (SUDS). Results demonstrated that both kinds of treatment were equally effective and improvements were maintained at a six months follow-up.

Krijn et al. [20] examined the effectiveness of different VR systems. *N*= 37 acrophobic subjects were treated either with three VR sessions administered by a head-mounted display (HMD) or by a computer animated virtual environment (CAVE) or were assigned to the waitlist control group. Results showed no differences in effectiveness between the different VR systems. The higher degree of presence that was experienced in the CAVE condition did not affect outcome measures. In a following study [21] the same research group analyzed the role of cognitive self-statements in VR exposure therapy. In a crossover design, *N*= 26 acrophobic persons were randomly assigned to two sessions of VR treatment followed by two sessions of VR treatment plus self-statements or vice versa. Results indicated that VR-assisted treatment reduced symptoms of fear of heights as well as behavioral avoidance and improved attitudes towards heights. However, cognitive self-statements did not additionally enhance effectiveness of VR.

Another study series concentrating on treatment of acrophobia with the aid of VR was conducted by Coelho and collaborators [22, 23]. Initially, the authors compared effects of treatment in a VR (*N*= 10) and a real environment (*N*= 5). Both groups showed equally large improvements on the Behavioral Avoidance Test (BAT), the ATHQ, and the AQ, even though treatment time was substantially lower in the VR condition. A following study with eight persons suffering from fear of heights revealed that movement during VR exposure enhances anxiety. One of the virtual settings used in these studies can be seen in Figure 1.

However, VR-assisted treatment of acrophobia is not only effective and time efficient, but additionally represents a series of advantages. Anxiety inducing situations such as being on bridges or high buildings can be experienced without any great logistic efforts. Therefore, difficulties of accessing the actual place and potential disturbances by pedestrians can be avoided.

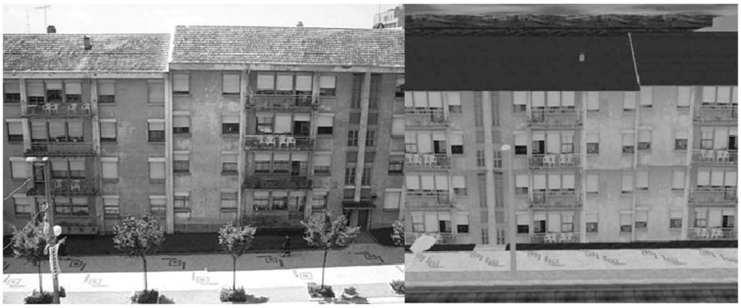


Figure 1. View from the real world (left) and the virtual reality system (right). Adapted from “Contrasting the Effectiveness and Efficiency of Virtual and Real Environments in the Treatment of Acrophobia” by C.M. Coelho, C.F. Silva, J.A. Santos, J. Tichon and G. Wallis, 2008, PsychNology, 6(2), p. 206. Copyright 2008 by PsychNology Journal. Adapted with permission.

Example case

Choi and collaborators [24] described the case of a 61-year old patient, who had been suffering from acrophobia for the past 40 years. He was not able to go up higher than the third floor of any apartment and therefore lived on the third floor on his 18-story building. In order to treat his acrophobia, the authors planned eight sessions of VR therapy that were supposed to take place three times a week and took about 30 minutes each.

The virtual environment was comprised of a steel tower which involved a lift within a steel frame structure that was open to all four sides. To enhance the sense of reality, sounds of wind and a moving lift were included, and the patient was isolated in a dark room in order to increase immersion. Prior to VR treatment, the patient received four sessions of relaxation training, including abdominal breathing and progressive muscle relaxation training, to be able to cope with body sensations during the VR sessions. Pretreatment assessment was completed and the patient accomplished a demo program to get used to VR.

In the first session, the patient stayed on the floor of the virtual lift to get accustomed to the environment. He was free to decide whether to go up on a higher floor or stay where he was. In this session, the patient went up to the fourth floor, experiencing dizziness and sweating and reporting 70 to 90 subjective units of disturbance (SUD). SUD was evaluated every two to five minutes. Whenever the patient stated to experience intense fear, he was instructed to relax. In the second and third session, the patient went up to the eighth floor, but still experienced high levels of SUD, breathlessness, and the sensation of falling down. After these sessions, the patient was already able to walk up to the eighth floor of his building for the first time in ten years. According to the patient, the virtual lift scared him more than going up his building in the real world. In the fourth session, the patient went up to 18th floor of the virtual tower, and then to the 25th floor, the top of the tower, in the fifth session. Even while looking down, the patient did not experience any particular symptoms and reported SUD scores below 30. After the sixth session, the patient claimed that he did not need VR anymore. The authors therefore changed treatment plans and assigned the patient to go up to the observatory of a mountain by cable car. Going up to and looking down from the observatory, the patient showed neither symptoms of anxiety nor avoidance. Subsequently, he suggested going up the highest building of Seoul. Looking outside from the elevator of this building, the patient expressed only little fear. Six months after the treatment, the patient stated that he did not have any fear of heights.

Fear of flying

Fear of flying, or *aviophobia*, is characterized by an intense fear of flying that often results in flight avoidance or experiencing substantial distress while flying. Acrophobia affects 10-20% of the general population and 20% of airline passengers consume alcohol or sedatives to deal with their fear of flying [25]. Most persons suffering from acrophobia fear a plane crash, while some fear being closed in and therefore often meet the DSM-IV criteria for claustrophobia. Further fears concern experiencing a panic attack and not being able to escape the situation or get medical attention, complying with the concept of panic disorder with agoraphobia, or a general fear of heights. Therefore, comorbidity with other anxiety disorders occurs very frequently.

The use of VR applications in the treatment of aviophobia could be advantageous to an exposure in-vivo because financial and logistical expenses are essentially lower. Furthermore, the privacy and confidentiality of a VR exposure in contrast to a regular flight should be emphasized.

The first RCT investigating the effectiveness of VR treatment of aviophobia was presented by Rothbaum et al. [26]. *N*= 49 participants were randomly assigned to VR exposure therapy, exposure in-vivo, or a waitlist control group. Both treatment groups received four sessions of anxiety management and four further sessions consisting of exposure to an airplane, either in reality or VR. The latter involved acoustic and visual simulations with the aid of a HMD, as well as vibration simulation. Exposure in vivo included preparation training at an airport as well as visualization of takeoff, flight and landing inside of an airplane. Both treatment groups showed significant symptom reduction on several standardized scales, while no improvements were observed for the control group. Effects remained stable after six and twelve months follow-up. However, flight situations differed between conditions, because an actual flight was not part of the exposure in-vivo. Moreover, both treatments were combined with anxiety management training with the result that treatment effects were not completely distinguishable. The findings were replicated in another sample of *N*= 75 aviophobic persons [27].

Another study compared VR exposure therapy with and without physiological feedback measures to self-visualization in *N*= 30 persons suffering from fear of flying [28]. Results showed significant improvements in flying behavior, physiological measures of anxiety, and standardized self-report measures of anxiety in the VR condition in contrast to imaginative exposure. Furthermore, the combination of VR treatment and biofeedback was more effective than VR treatment alone. The authors reported that after eight weeks of therapy, 20% of the patients in the imaginative condition, 80% of those in the VR condition and 100% of patients who received both VR treatment and biofeedback were able to fly again. In a follow-up study three years later, treatment effects were maintained [29].

Mühlberger et al. [30, 31] proved the effectiveness of VR-assisted treatment of aviophobia in a series of studies. *N*= 30 participants were randomly assigned to a VR treatment condition or a relaxation training group. While both treatment conditions resulted in significant symptom improvement, several outcome measures, including physiological fear responses, indicated larger effects of VR exposure therapy than self-visualization. In a following study, the research group demonstrated that one session of VR exposure therapy in combination with cognitive behavior therapy (CBT) achieved better results than CBT only and a control group (*N*= 45) [31]. Results remained stable at six months follow-up. Limitations of the study include the time spent with the therapist, which was much longer for the combined treatment than for CBT only. In addition, group assignment was not randomized. Nevertheless, the elucidated study demonstrates that VR-supported exposure can show persistent effects even after one single session. Furthermore, Mühlberger et al. [32] revealed that the completion of graduation flights might be important for long-term treatment effectiveness, but that the presence of a therapist is not necessarily required.

Comparing five sessions of VR exposure therapy to an attention placebo group, Maltby et al. [33] obtained mixed results. While the VR treatment condition was superior to the placebo condition on self-report instruments, BAT scores did not reveal any significant differences. Moreover, VR exposure was more effective on only one self-report measure at six months follow-up.

Furthermore, Krijn and collaborators [34] compared four sessions of VR exposure with four sessions of CBT and with five weeks of bibliotherapy, that involved reading a psycho-educative book about aviophobia. Results indicated that both VR treatment and CBT were effective and did not differ in symptom reduction. However, after undergoing an additional CBT program, including an exposure in-vivo, CBT group was superior to VR treatment group.

Finally, the efficacy of VR and computer-aided psychotherapy in the treatment of aviophobia was examined by Tortella-Feliu et al. [35].*N*= 60 participants were randomly assigned to the following conditions: VR exposure, computer-aided exposure with a therapist's assistance, and self-administered computer-assisted exposure. Results demonstrated that all three conditions were equally effective in reducing flying phobia, even after one year. The findings indicate that therapist involvement might be reduced in VR and computer-aided treatment.

Fear of spiders

According to the ICD-10 [15],arachnophobia is categorized within the group of zoophobias and is characterized by a persistent fear of spiders, an immediate anxiety response to exposure to a spider, and avoidance of spiders. The category of "bugs, mice, snakes or bats", which includes spiders, accounts for about 40% of specific phobias [36]. Approximately 3.5 to 6.1% of the general population suffers from arachnophobia, whereof the majority is constituted by women. Even though most arachnophobic persons recognize that their fear is unreasonable, daily life can be restrained. For instance, persons suffering from fear of spiders might depend on the help of others when confronted with a spider, or be restricted in choosing an apartment.

VR applications seem to represent a potential treatment method for arachnophobia. Rinck et al. [37] examined spider fearful persons' attention and motor reactions to spiders on a VR. The authors demonstrated that spider fearfuls show increased state anxiety, spend more time looking at spiders, and exhibit behavioral avoidance of spiders.

A first single case report examining the effectiveness of treating arachnophobia with the aid of VR was conducted by Carlin et al. [38]. They used VR as well as mixed reality, which involved touching real objects that can be seen in VR, to treat a 37-year old female suffering severe fear of spiders. After twelve weekly sessions of one hour each, measures of anxiety, avoidance, and behavior towards real spiders improved significantly.

In 2002, a RCT was conducted that compared VR exposure therapy group and a waitlist group of altogether *N*= 23 participants [36]. The VR treatment group received four one-hour sessions on average. Effects were assessed by the Fear of Spiders Questionnaire (FSQ), a BAT, and severity ratings effected by clinicians. Results demonstrated that 83% of participants who received VR treatment showed clinically significant improvement compared with 0% in the waitlist control group.

In a following study with *N*= 36 participants, it was demonstrated that VR combined with touching an object that resembles a spider was more effective than only VR exposure [39]. *N*= 36 phobic students were randomly assigned to one of three conditions: No treatment, VR exposure without any tactile stimulation, and VR exposure including tactile stimulation. After three sessions of VR exposure, the treatment groups showed less avoidance and lower levels of anxiety than the control group; and VR including tactile simulation was superior to VR without any tactile clues.

Michaliszyn et al. [40]found similar results comparing the effectiveness of VR treatment and in-vivo exposure to a waitlist condition. A total of *N*= 43 persons suffering arachnophobia were randomly assigned to the three conditions. Treatment groups received eight therapy sessions of one and a half hour each. Outcome measures included the Fear of Spiders Questionnaire, the Spider Phobia Beliefs Questionnaire (SBQ), a BAT, and the Structured Interview for DSM-IV (SKID). Both treatment groups showed clinically significant improvements in comparison to the waitlist control group, whereat in-vivo exposure was superior to VR treatment on the SBQ-F.

Furthermore, a study demonstrated that modified 3D computer games instead of actual VR software can be effective in the treatment of arachnophobia [41]. Modification of computer games could therefore represent a less expensive alternative to VR equipment.

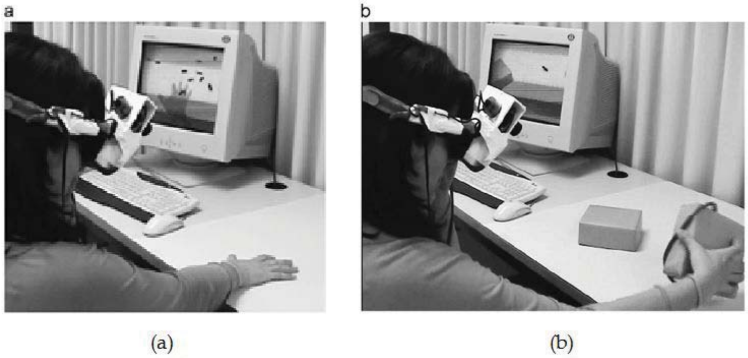


Figure 2. a) Participant putting her hand on the table and the cock- roaches crossing over it. (b) Participant searching for hidden cockroaches. Adapted from "A comparative study of the sense of presence and anxiety in an invisible marker versus a marker augmented reality system for the treatment of phobia towards small animals" by M.C. Juan and D. Joele, 2011, International Journal of Human-Computer Studies 69(6), p. 445. Copyright 2011 by Elsevier. Adapted with permission.

Research has also focused on the use of AR in treating phobia towards small animals. In doing so, virtual spiders or cockroaches are blended into the real world. In a first case study, a participant suffering from cockroach phobia showed significant decreases in fear and avoidance levels, being capable of approaching, interacting with, and killing real cockroaches following AR exposure and one month later [3]. In a following study evaluating the effectiveness of AR, nine participants with either spider or cockroach phobia were treated in a single session [42]. Firstly, progressively more virtual spiders or cockroaches were presented in the therapist's hand. Participants were asked to bring their hand closer to the animals. Subsequently, a box appeared which the participants had to pick up to see if there was an animal underneath. Finally, virtual animals had to be killed with an insecticide, flyswatter or dustpan and put into a box. After completion of the session, participants were asked to approach, interact and kill real spiders or cockroaches. All of the participants succeeded in doing so, showing considerable less avoidant behavior. A validation of the system used in these first studies demonstrated that all elements of the AR environment were highly fear inducing in *N*= 6 female participants with cockroach phobia. In addition, ratings of presence, reality and immersion obtained high scores [43]. In this AR system, visible markers were used to identify insecticide, flyswatter or dustpan approaching a virtual animal. To avoid this warning, a second version in which the markers were invisible was compared to the first one [44]. For an example of the AR setting, see Figure 2. Results indicated that the invisible marker-tracking system induced a similar or higher sense of presence and levels of anxiety and seems therefore superior to the visible marker-tracking system. In this context, it should me mentioned that we consider the killing of animals within the studies as ethically questionable.

Social phobia

Social phobia is defined as an unreasonable or excessive fear of social situations and the interaction with other people that automatically brings on feelings of self-consciousness, judgment, evaluation or inferiority [14]. Symptoms of social phobia include intense fear, blushing, sweating, a dry mouth, trembling, a racing heart and shortness of breath. There are two subtypes of social phobia: specific social phobia, that is limited to a small number of fear inducing situations, and generalized social phobia, that involves almost all social situations. Situations that may evoke fear include speaking in public, establishing contacts, protecting one's interests and being under scrutiny. Usually, persons suffering from social phobia are worried that their fear is being noticed by others. Social phobia is one of the most commonly observed mental disorders, showing a life-time prevalence of 13%, according to the USA National Comorbidity Survey [45].

Roy et al. [46] presented a clinical protocol to assess the effectiveness of VR treatment of social phobia, describing the study structure, assessment tools, and content of the therapy sessions. Four virtual environments were used to reproduce situations inducing high levels of anxiety in social phobics: Performance, intimacy, scrutiny, and assertiveness. In a preliminary study, the effectiveness of VR treatment was demonstrated in N= 10 persons suffering from social phobia in a between-subjects design. In a following study conducted by the same research group [47], the same virtual environments were used to examine the effectiveness of VR exposure in N= 36 social phobics. Participants were assigned to either VR treatment or cognitive-behavioral group therapy (CBGT). After twelve weeks of therapy, both treatment groups showed clinically and statistically significant improvement. In a more recent RCT, the effectiveness of VR treatment, a combination of CBT and VR exposure and a waitlist control condition were compared in N= 45 participants diagnosed with social phobia [48]. Results indicated a significant reduction of anxiety on all questionnaires for both treatment groups in contrast to the waitlist control group.

Furthermore, a few studies have focused specifically on the effectiveness of virtual environments in treating *public speaking anxiety*. Harris et al. [49] showed that four VR treatment sessions of 15 minutes each (see table 1) reduced self-reported anxiety as well as physiological reactions significantly in eight students suffering from public speaking anxiety in comparison to six students in a waitlist control group. In an open clinical trial, the effectiveness of four sessions of anxiety management and four subsequent therapy sessions was examined, using a virtual audience in *N*= 10 participants diagnosed with social phobia [50]. As a result, self-report measures indicated lower levels of public speaking anxiety at post treatment and three months follow-up. However, participation rates of giving a free speech to an actual audience did not differ before and after the treatment. A larger sample size of *N*= 88 persons with public speaking anxiety was used in a RCT that compared the effectiveness of CBT, VR and CBT combined, and a waitlist control condition [51]. Results demonstrated significant improvements of both treatment groups on self-rated anxiety during a behavioral task and four out of five anxiety measures in contrast to the control group. At one year follow-up, results remained stable.

Table 1. Procedure of VR treatment for public speaking anxiety [49]

Initial interview	Besides self-report instruments measuring social anxiety, a voice test sample was recorded while the participants answered a question and read a paragraph. The heart rate was measured during the speaking test and a brief relaxation exercise.
Session 1	Participants stood at a podium with a microphone, looking around a virtual empty auditorium to get accustomed to the environment. Subsequently, participants were asked to talk about their public speaking anxiety.
Session 2	Participants were asked to say the American Pledge of Allegiance. The auditorium was gradually filled with people, and applause was used to encourage participants. The pledge was repeated, with the virtual audience applauding at the end of the recitation.
Session 3	Participants were asked to deliver a 2-min speech with a small light on the clipboard. The room was gradually filled with audience, people were speaking to each other, laughing, asking the speaker to speak louder, and applauding at the end of the speech. Afterwards, the speech was repeated.
Session 4	Participants were asked to give the same or another speech. Manipulations of the scenario were made as in session 3.

Other studies have concentrated on specific aspects of treating social phobia with the aid of VR. For instance, Ter Heijden and Brinkmann [52] evaluated speech detection and recognition techniques in comparison to a human control condition in a VR surrounding. Interactions were observed in two phobic and 24 healthy persons. Results indicated that automatic speech techniques often did not show any significant differences compared to manual speech. Therapist workload of entering speech content might therefore be minimized in VR treatment.

Another study brought the aspect of presence within VR exposure into focus [53]. The relationship of three components of presence (spatial presence, involvement, and realness), fear ratings during VR, and treatment effectiveness were evaluated in *N*= 41 participants suffering from social phobia. The authors found an association between total presence as well as realness subscale scores and fear-ratings during treatment, while only scores on the involvement subscale were able to significantly predict treatment outcome.

R environments may also facilitate research on specific aspects of social phobia. For example, Cornwell et al. [54] used a VR setting to examine physiological reaction of persons diagnosed with social anxiety disorder in social-evaluative threat situations. Participants were asked to deliver a short speech in front of a virtual audience. In this way, no actual audience has to be recruited in order to realize study designs of that kind.

Panic disorder

Around 5% of the US Americans suffer from panic disorder in their lifespan [55]. Panic disorder is diagnosed if a panic attack, including symptoms such as sweating, palpitations, trembling, nausea, derealization and depersonalization, results in consistent concern about having additional attacks, worries about its consequences or behavioral changes. Persons suffering from panic disorder frequently develop agoraphobic avoidance behaviors. Agoraphobia refers to anxiety about being in places or situations from which escape might be difficult or in which help may not be available in case of having an unexpected panic attack, such as being in a crowd, on a bridge, train or the like. Therefore, agoraphobia often has to be included in the treatment of panic disorder as well.

Vincelli et al. [56] presented a treatment protocol called Experiental Cognitive Therapy (ECT), which integrates VR in order to treat panic disorder and agoraphobia. Its effectiveness was demonstrated in *N*= 12 patients, who were randomly assigned to an ETC group and therefore undergoing VR exposure, a CBT group or a waitlist control group. Results indicated that eight sessions of ECT and twelve sessions of CBT equally reduced the number of panic attacks, the level of depression and state and trait anxiety.

In a following study examining the effectiveness of VR in the treatment of panic disorder, *N*= 40 participants received either four sessions of cognitive therapy including VR exposure, or twelve sessions of a panic control program [57]. Results indicated that both treatments were equally effective. However, findings did not remain stable at six months follow-up, where participants of the panic disorder program showed higher overall functioning. Botella et al. [58] used a more rigid study design in order to compare *N*= 37 persons receiving nine sessions of either CBT with VR exposure or CBT with in-vivo exposure and a waitlist control group. Both treatment groups obtained equal symptom reductions in comparison to the waiting list, and results were maintained at twelve months follow-up. A following study using a between-subjects design compared eleven sessions of CBT including exposure in-vivo with CBT and VR exposure in *N*= 28 participants diagnosed with panic disorder [59]. All participants additionally received antidepressant medication, and a BAT was applied to assess treatment effects. Results revealed that both treatments were equally effective, and results remained stable after three months. Findings were replicated in a RCT by the same research group in *N*= 27 participants with panic disorder and agoraphobia [60].

A later study evaluated effectiveness of interoceptive exposure in a virtual environment, simulating physical sensations through audible stimulation such as rapid heartbeat and panting, and visual stimulations such as blurry or tunnel vision [61]. Results indicated that both IE using VR and traditional IE significantly reduced symptoms of panic disorder, and that results were maintained or even improved at three months follow-up. Finally, Meyerbröker et al. [62] showed that varied levels of presence by using either a CAVE or a HMD did not influence effects of VR treatment of panic disorder.

Obsessive-compulsive disorder

Obsessive-compulsive disorder **(OCD)** is a debilitating mental disorder that is characterized by either obsessions, compulsions, or both. According to the DSM-IV, obsessions are defined as recurrent and persistent thoughts, impulses, or images that may cause anxiety, including the obsession of contamination, need for symmetry or aggression [13]. Compulsions refer to repetitive behaviors, such as hand washing, ordering, and checking, or mental acts such as praying, counting, or repeating words silently, that are performed to respond to an obsession or to rules in order to reduce distress. Lifetime prevalence rates are estimated about 2% worldwide.

While the benefits of computer-based assessment and treatment of OCD has already been demonstrated [63], only preliminary data concerning the use of VR in the treatment of OCD is available. A South Korean research group presented first results of VR exposure therapy of OCD [63]. N= 33 participants with OCD and n= 30 healthy controls navigated through a virtual environment, consisting of a training, distraction, and main task phase. Anxiety rates as well as decreased ratio of anxiety during the main task were significantly higher in participants with OCD than healthy controls. VR may therefore function as anxiety-provoking and a potential treatment tool for OCD. The same virtual environment was used to examine its potential efficacy in assessing OCD in n= 30 patients with OCD and n= 27 matched healthy controls [64]. Results indicated that OCD patients had significantly greater difficulties with compulsive checking than controls, and that task performance was positively correlated with self-reported symptoms as well as interviewer-rated measures of OCD. Another study by the same research group demonstrated that anxiety levels of *N*= 24 healthy participants decreased as a result of performing virtual arrangement tasks three times with three-day intervals [65]. However, the amount of anxiety reduction depended on the type of task, and only the Symmetry, Ordering and Arrangement Questionnaire (SOAQ) showed significant correlation with anxiety. Nevertheless, VR seems to be a potential device for the assessment and treatment of persons with symptoms of arranging compulsion.

Posttraumatic stress disorder

Posttraumatic stress disorder **(PTSD)** is a serious condition that persons experiencing a traumatic event may suffer from. According to the ICD-10, the traumatic event needs to be exceptionally threatening or catastrophic and would distress most people. Such disasters can be either manmade, as it is the case in war, torture, or sexual abuse, or they can be natural disasters, such as earthquakes, accidents, or life-threatening diseases. Criteria for PTSD include intrusions such as flashbacks and repeating dreams, avoidance of situations similar to the traumatic event, loss of memory about certain aspects of the event, and symptoms of hyperarousal. Experiencing psychological distress right up to PTSD is common among military members who are constantly confronted with threatening situations. Approximately 18% of warfighters returning from Iraq and 11% returning from Afghanistan were screened positive for PTSD.

Some authors suggest the application of new treatment approaches such as VR exposure, reasoning that conventional therapy approaches may be rejected by war veterans due to stigmatization and that in-vivo exposure is not possible. In fact, situations that caused the traumatization are difficult to frequent, but according to traumatherapy, this is neither necessary nor indicated [66]. On the contrary, exposure to virtual settings that are reconstructing traumatizing situations is ethically questionable, which is demonstrating by the following scenarios.

A case report describes the first application of VR for a Vietnam veteran suffering from PTSD [67]. As a result of VR treatment, he significantly improved on all PTSD measures and those gains remained stable at six months follow-up. A following open clinical trial demonstrated the effectiveness of VR in ten male Vietnam veterans diagnosed with PTSD [68]. They underwent eight to 16 sessions of VR exposure in two virtual environments: a virtual helicopter flying over Vietnam, and a clearing surrounded by jungle. Participants showed significant PTSD symptom reductions on the Clinician Administered PTSD Scale (CAPS) at six months follow-up, declaring symptom reductions ranging from 15 to 67% in an interview. Self-reported intrusion symptoms as measured by the Impact of Event Scale were significantly lower at three months follow-up in comparison to baseline, but not at six months follow-up. Another open trial of VR in the treatment of *N*= 21 Vietnam veterans was conducted by Ready and collaborators [69], simulating virtual environments in response to participants' memories. Participants showed significant symptom reductions at three and six-months follow-up, even though two participants experienced an increase in symptoms during VR exposure. A RCT using a small sample size was presented by the same research group [70]. Eleven Vietnam veterans were assigned to either ten sessions of VR exposure or present-centered therapy, utilizing a problem-solving approach. Results indicated no significant differences between treatment groups at posttreatment and six months follow-up.

Furthermore, several studies examined the use of a virtual environment to treat veterans returning from "Operation Iraqi Freedom" who suffered from PTSD. The "Virtual Iraq/ Afghanistan" environment was adapted from the Microsoft® X-box game "Full Spectrum Warrior". Scenarios include a Middle Eastern city and a Humvee driving down a desert highway. Auditory, visual, olfactory and vibrotactile stimulation such as gunfire, weather conditions, the smell of burnt rubber and the sensation of a moving car can be adjusted. Some examples of Virtual Iraq/Afghanistan are shown in Figure 3.

Several case reports were conducted [e.g. 71, 72]. The first clinical trial assessing the effectiveness of exposure using "Virtual Iraq" indicated clinically and statistically significant symptom reduction in *N*= 20 participants [73]. In addition, McLay and collaborators [74] presented a RCT comparing the efficacy of VR exposure therapy and usual CBT in a sample of *N*= 19 Iraq veterans diagnosed with PTSD. Within the VR condition, seven out of ten participants improved at least 30% on the CAPS, while only one out of nine within the CBT condition showed similar improvement. The effectiveness of "Virtual Iraq" was also supported by Reger et al. [75] in 24 Iraq and Afghanistan veterans. Moreover, a pilot study focusing on the combination of VR exposure and cognitive enhancing medication has shown promising results [76].

Another VR environment was created to treat Portuguese survivors of the 1961-1974 wars in Africa. Subsequently to a case study [77], Gamito and colleagues [78] examined the effectiveness of a VR war environment to imaginal exposure and a waiting list condition. Participants in the VR condition showed significant reduction of depressive and anxiety symptoms.

An increased incidence of PTSD was also detected among the survivors of the attacks of September 11, 2001. Consequently, Difede and Hoffmandeveloped a virtual environment simulating jets crashing into the World Trade Center, people jumping to their deaths from the buildings, and towers collapsing. A study revealed that participants in a VR condition (*n*= 9) showed significantly greater improvement on CAPS scores than the waitlist control group (*n*= 8) [79]. Findings were replicated in a following study [80].

In a further study, a VR surrounding was developed in order to treat victims of a terrorist bus bombing in Israel. The potential effectiveness of “BusWorld” was demonstrated in a study examining 30 asymptomatic participants, who showed significantly higher mean subjective units of discomfort scores (SUDS) with increasingly distressful scenarios. Treatment of a 29-year-old victim of a bus bombing using VR resulted in significant reduction of PTSD symptoms as measured by the CAPS [81].

Another field of application in the treatment of PTSD by the means of VR exposure is made up by motor vehicle accident survivors. Saraiva et al. [82] presented a case study describing positive outcomes of VR exposure of a 42-year-old female in the aftermath of a serious vehicle accident. Findings were confirmed by Beck et al. [83], who demonstrated significant reductions of re-experiencing, avoidance, and emotional numbing in six persons reporting subsyndromal PTSD after completing ten sessions of VR treatment.

A new approach of treating PTSD was introduced by Fidopiastis et al. [84]: As aforementioned, AR, referred to as Mixed Realities (MR) by the authors, are supposed to blend virtual content into the real world, which means that computer-generated objects can be superimposed on the real-world environment. In a pilot study, first promising effects of MR in the assessment of PTSD by capturing the patient’s interaction with the simulated environment were demonstrated. Riva et al. [85] further advanced the approach of MR by presenting the paradigm of Interreality, which is supposed to bridge the virtual and real world by using activity sensors, personal digital assistants or mobile phones.

However, even though treatment of persons suffering PTSD is crucial, study designs using VR seem questionable with regard to ethical concerns. Exposing war veterans or victims of terror attacks to simulated war scenarios is contra-indicated according to current research on trauma therapy. Certain phases of traumatherapy such as stabilization and development of a therapeutic relationship have to precede the processing of the traumatic experience [86]. In the studies presented here, none of these phases were considered so that VR exposure bore the risk of retraumatization. Besides, if the virtual setting does not in detail project the traumatic event, renewed traumatisation is risked. To date, long-term effects of exposing persons with PTSD to virtual environments are mostly unknown, because efficacy studies rarely collect follow-up data. Therefore, even though the use of VR technology seems feasible in the treatment of PTSD, ethical concerns and aspects with regard to therapy indication always need to be considered.

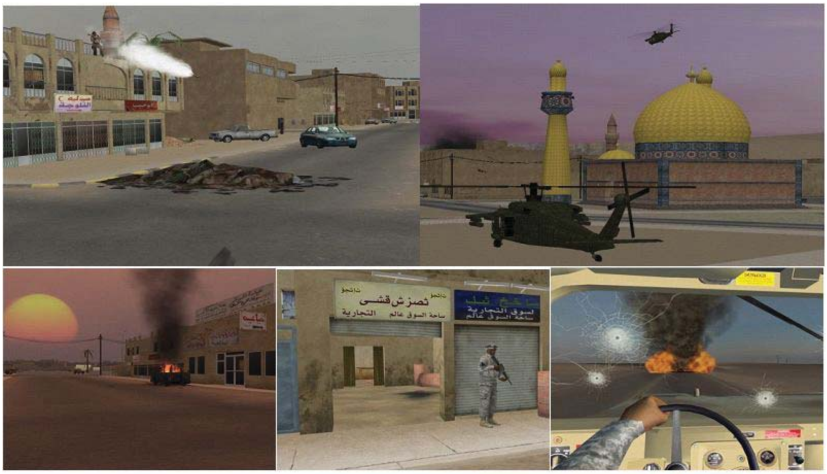


Figure 3. Virtual Iraq/ Afghanistan scenarios. Courtesy of Virtually Better Inc. and University of Southern California, Institute for Creative Technologies.

Other applications

Virtual environments have also been applied in the assessment, treatment and research of other mental disorders such as eating disorders, sexual dysfunctions, schizophrenia, attention deficit disorder, and addictions. In the following, a cursory overview of VR treatment approaches in those clinical pictures will be provided.

The first use of VR in treating eating disorders was accomplished by an Italian research group [87]. VR programs focused on the improvement of body image, body satisfaction and physical acceptance in obese patients and reduction of perfectionisms, body dissatisfaction and negative attitudes towards the body in anorectic patients. A few controlled studies demonstrated the effectiveness of VR in the treatment of eating disorders [88-90].

In contrast, research on the effectiveness of VR in the treatment of sexual dysfunction is still in an experimental stage. In one study, VR was integrated into psychodynamic psychotherapy for the treatment of erectile dysfunction and premature ejaculation in N= 160 men [12]. VR seemed to help to work through events and associations that were creating the sexual problems. Positive outcomes remained stable after one year follow-up [91].

Furthermore, VR has been used to assess attention impairments in order to diagnose Attention Deficit Disorder (ADD). A VR classroom scenario was created, in which children had to perform attention task while being distracted by classroom noises, activities occurring outside the window, or persons passing by [92]. In a clinical trial conducted by the same research group, it was demonstrated that the system could reliably distinguish between children with ADD and healthy controls.

Another virtual environment was developed to treat people with schizophrenia [93]. VR scenarios can be individually tailored to simulate patients' hallucinations, such as voices or walls appearing to close in, to teach patients to ignore hallucinations in real life. But even though VR treatment might be an effective adjunct in the treatment of schizophrenics, indications for VR exposure have to be carefully pondered.

The treatment of addiction by means of VR seems to be a promising area as well. Similar to exposure therapy in specific phobias, repeatedly showing cues of alcohol or tobacco should lead to extinction of craving. Virtual environments presenting virtual cigarettes [94] or bottles and glasses of alcohol [95] were able to significantly decrease craving.

IMPLICATIONS FOR RESEARCH AND THERAPY

Hereafter, the findings presented here will be discussed with regard to their implications for research and therapy.

Research

The studies examining the efficacy of VR treatment in psychotherapy that were conducted up to that point are various with respect to their designs, treatment methods, and results. A multitude of case reports and pilot studies with questionable generalizability were published to demonstrate that VR can be an effective tool in the treatment of mental disorders. To provide a clearer overview of the studies proving the effectiveness of VR in anxiety disorders, the study designs and results of all controlled trials were listed in table 2. All controlled trials that examined the VR or AR treatment of at least one group of participants suffering from an anxiety disorder and that used a standardized outcome measure of anxiety were included. As the table shows, most RTCs have been effected in the field of aviophobia. Particularly with reference to specific phobias, considerable systematic research has been conducted in the past years. However, while the effectiveness of VR and AR exposure in treating specific phobias seems to be proven, the application of VR in more complex disorders like panic disorder, obsessive-compulsive disorder, and PTSD needs to be further evaluated. Treatment protocols in this field of research are still in an experimental phase and lack controlled studies to prove their effectiveness. In addition, the majority of studies examining the effects of VR-based treatment combines different treatment approaches and therefore makes it difficult to analyze VR outcomes separately. Future research should also work out which groups of patients benefit most from VR and how environments can be adapted to patients' needs.Additionally, comparable outcome measures such as behavioral avoidance tests should be included in future studies. Finally, sample sizes are often too small to generalize study findings and longer-term catemneses are frequently missing.

Table 2. Overview of VR treatment outcome studies

Study	**Clinical Sample**	*N*	**Design**	**No. sessions**	**Results**
Fear of heights					
Coelho et al. (2008)	Acrophobia	15	Between-subjects	3	VR and in vivo exposure were equally effective, despite shorter treatment times of VR
Emmelkamp et al. (2001)	Acrophobia	10	Within-subjects	2	Exposure in vivo did not lead to any significant improvements after VR exposure
Emmelkamp et al. (2002)	Acrophobia	33	RCT	3	VR and in vivo exposure were equally effective; results stable after 6 months
Krijn et al. (2004)	Acrophobia	37	RCT	3	VR administered by HMD and CAVE were equally effective; results stable after 6 months
Krijn et al. (2007)	Acrophobia	26	RCT	4	Self-statements did not additionally enhance effectiveness of VR treatment
Fear of flying					
Krijn et al. (2007)	Aviophobia	59	RCT	4	VR treatment and CBT were equally effective and superior to bibliotherapy

Malty et al. (2002)	Aviophobia	45	RCT	5	VR treatment was superior to attention placebo group on self-report measures, but not avoidance test; results not stable after 6 months
Mühlberger et al. (2001)	Aviophobia	30	RCT	1	VR treatment and relaxation training were equally effective
Mühlberger et al. (2003)	Aviophobia	45	RCT	1	VR treatment in combination with CBT was more effective than CBT alone
Mühlberger et al. (2006)	Aviophobia	30	RCT	1	Presence of a therapist did not influence effectiveness of VR treatment
Rothbaum et al. (2000)	Aviophobia	49	RCT	8	VR and in vivo exposure were equally effective in comparison to a waitlist control group; results stable after 6 and 12 months
Rothbaum et al. (2006)	Aviophobia	75	RCT	8	VR and in vivo exposure were equally effective in comparison to a waitlist control group; results stable after 6 to 12 months
Tortella-Feliu et al. (2011)	Aviophobia	60	RCT	6 max.	VR exposure, computer-aided exposure with a therapist's assistance, and self-administered computer-assisted exposure were equally effective; results stable after 12 months
Wiederhold & Wiederhold (2003)	Aviophobia	30	RCT	8	VR exposure in combination with biofeedback was more effective than VR exposure alone
Fear of spiders/ cockroaches					

Bochard et al. (2006)	Arachnophobia	11	Within-subjects	5	Modified 3D computer games were effective in the treatment of arachnophobia
Garcia-Palacios et al. (2002)	Arachnophobia		Between-subjects	4 on average	83% of the VR exposure group showed clinically significant improvement, in comparison to 0% of the waitlist control group
Hoffman et al. (2003)	Arachnophobia	36	RCT	3	VR treatment including tactile stimulation was more effective than VR without tactile stimulation; both treatment groups were superior to waitlist control group
Juan et al. (2005)	Arachnophobia, cockroach phobia	9	Open trial	1	AR treatment significantly reduced participants' fear and avoidance
Michaliszyn (2010)	Arachnophobia	43	RCT	8	VR and in vivo exposure groups showed clinically significant improvement in comparison to waitlist control group
Social Phobia					
Anderson et al. (2003)	Social Phobia	10	Within-subjects	8	The combination of VR and anxiety management resulted in reduction of public speaking anxiety on self-report; stable at 3 months follow-up
Harris et al. (2002)	Fear of public speaking	14	Between-subjects	4	VR treatment reduced self-reported anxiety and physiological reactions significantly in comparison to waitlist control group
Klinger et al. (2005)	Social Phobia	36	RCT	12	VR treatment and CBT showed equally significant improvements in anxiety and avoidance behavior

Price et al. (2011)	Social Phobia	41	RCT	8	Involvement score predicted therapy outcome
Robillard et al. (2010)	Social Phobia	45	RCT	16	VR treatment and combination of CBT and VR were both effective in comparison to waitlist control group
Roy et al. (2000)	Social Phobia	10	Between-subjects	12	VR treatment and CBT equally showed significant improvements in anxiety and avoidance behavior
Wallach et al. (2009)	Fear of public speaking	88	RCT	12	CBT as well as VR and CBT combined resulted in significant improvements of self-rated anxiety and 4 out of 5 anxiety measures in contrast to waitlist control group; results stable at 12 months follow-up
OCD					
Kim et al. (2008)	OCD	63	Matched between-subjects	1	Participants with OCD experienced significantly higher anxiety, but also showed a higher decreased ratio of anxiety than healthy controls
Panic disorder					
Botella et al. (2007)	Panic disorder	37	RCT	9	CBT including VR exposure and CBT including exposure in vivo resulted in equal symptom reductions in comparison to waitlist control group; results stable at 12 months follow-up

Choi et al. (2005)	Panic disorder	40	RCT	12	CBT including VR exposure and a panic disorder program were equally effective; results did not stable at 6 months follow-up
Penate et al. (2008)	Panic disorder	27	Matched between subjects	11	CBT including VR exposure and CBT including exposure in vivo were equally effective in addition to antidepressive medication
Pérez-Ara et al. (2010)	Panic disorder		Between- subjects	8 max.	Interoceptive exposure using VR and traditional interoceptive therapy equally reduced symptoms; results stable at 3 months follow-up
Pitti et al. (2008)	Panic disorder	28	Matched between-subjects	11	CBT including VR exposure and CBT including exposure in vivo were equally effective in addition to antidepressive medication
Vincelli et al. (2003)	Panic disorder	12	RCT	9	VR treatment and CBT equally reduced the number of panic attacks, the level of depression and state and trait anxiety
PTSD					
Beck et al. (2007)	subsyndromal PTSD	6	Within-subjects	10	Motor vehicle accident survivors showed significant reductions of re-experiencing, avoidance, and emotional numbing after VR treatment

Difede et al. (2006)	PTSD	17	Between-subjects	14	Survivors of 9/11 undergoing VR exposure showed significantly greater improvement on CAPS scores than waitlist control group
Difede et al. (2007)	PTSD	21	Quasi-experimental	14 max.	Survivors of 9/11 undergoing VR exposure showed significantly greater improvement on CAPS scores than waitlist control group
Gamito et al. (2010)	PTSD	10	Between-subjects	12	Portuguese war veterans in the VR condition showed significant reduction of depressive and anxiety symptoms in comparison to waitlist control group
Ready et al. (2006)	PTSD	14	Open trial	20 max.	Vietnam veterans showed significant symptom reductions at 3 and 6 months follow-up; 2 participants experienced an increase in symptoms during VR exposure
Ready et al. (2010)	PTSD	11	RCT	10	No significant differences between VR and present-centered therapy at posttreatment and 6 months follow-up in Vietnam veterans
Rizzo et al. (2009)	PTSD	20	Open trial	10	Participants of “Virtual Iraq” showed clinically and statistically significant symptom reductions
Reger et al. (2011)	PTSD	24	Open trial	3-12	Significant symptom reduction in Iraq or Afghanistan active duty soldiers

Rothbaum et al. (2001)	PTSD	10	Open trial	8-16	Vietnam veterans showed significant symptom reductions on the CAPS at 6 months follow-up; self-reported intrusion symptoms were significantly lower at 3 but not at 6 months follow-up

Alongside the realization of further outcome studies, future research should focus on underlying cognitive and physiological processes of VR exposure. Moreover, the role of the therapist-patient-relationship should be further investigated. Although some studies indicate that the assistance of a therapist might be reduced (e.g. [35]), the consequences of a changing role of the therapist still need to be explored. For instance, the exposure of war veterans to frightening war scenarios might impair trust towards the therapist and therefore influence treatment outcome.

Therapy

A significant number of studies has furnished evidence for the effectiveness of using VR in psychotherapy. However, if therapists decide to include VR into treatment sessions, they should act in accordance with certain guidelines in order to abet positive outcomes and minimize negative treatment effects. To date, just a few treatment manuals have been published. For example, Rothbaum et al. [95] presented an abbreviated treatment manual for exposure therapy of acrophobia. VR was used to replace conventional exposure as a component of behavioral therapy. According to the manual, treatment sessions should include symptom assessment, breathing retraining, cognitive restructuring, hyperventilation exposure and VR exposure. The authors recommend arranging VR settings as follows:

- Sitting on plane, engines off
- Sitting on plane, engines on
- Taxiing
- Takeoff
- Smooth flight
- Landing
- Thunderstorm and turbulent flight

Another treatment manual was developed by Spira et al. [96]. The authors describe in detail twelve steps to treat combat-related PTDS with the aid of meditation, biofeedback, and VR.Furthermore, Bouchard et al. [98] outlined a treatment manual for VR exposure therapy of specific phobias, can be used with different VR software. In approximately eight sessions, patients are supposed to overcome their fears and stop avoidance behaviors by participating in cognitive restructuring and graduated VR exposure. In addition, guidelines to enhance the sense of presence and minimize potential negative side effects of immersion are provided.However, even though first publications are promising, more evidence-based treatment manuals focusing on specific syndromes are required in order to advance VR usage in psychotherapy.

CONCLUSION

The current state of research presented in this article furnishes considerable evidence for the effectiveness of virtual and augmented environments in the treatment of several mental disorders. However, VR treatment is not yet part of ordinary mental health care. Possible explanations for that could be:

- *Costs:* **Acquisition of VR equipment is (still) expensive, and training is needed to apply VR tools.**
- *Reservations against technology:* **A myriad of therapists have reservations regarding modern technologies** and therefore do not consider using them. German studies demonstrated a relatively high readiness to make use of therapy that integrates text messages or e-mails, but not VR [99, 100].
- *Limited indications:* **VR exposure might be contraindicated in patients with PTSD or co-morbid mental disorders. Despite the potential benefits of using modern technologies** in psychotherapy, indications with regard to the patient and specific disorder always need to be balanced. In some cases, VR treatment might not be as efficient as conventional therapy.

On the other hand, in the case of obvious indication of VR treatment, therapists should be open with respect to embedding VR technologies into therapy. Those who apply VR in therapy should be aware that VR tools always have to be part of a broader therapy plan and only complement, but cannot replace the skills of well-trained clinicians. Advantages of VR treatment include:

- Cost reduction: With further technological advancements as well as increased amounts of research on the effectiveness and applicability of VR and AR, technical and financial costs of those tools will probably be reduced [1].
- Lower logistic efforts: Using VR usually reduces logistic and therefore financial costs in the long term, because no real places have to be accessed in order to expose patients to fear inducing situations.
- Controllability of settings: Virtual scenescan be easily controlled and adjusted to each patient`s needs.
- Therapy motivation: Use of VR might increase therapy motivation, especially in younger patients.For these reasons, integration of virtual environments into day-to-day clinical practice will hopefully be extended in the future. According to a representative survey of the German population [101], 15.7% (n= 375) of the respondents would "maybe" make use of VR treatment in case of suffering from a phobia, and 7.4% (n= 177) estimated the use of VR as "rather" or "very probable". The results indicated that persons under the age of 35 who were already familiar with modern technologies were most likely to consider VR therapy. It can therefore be concluded that a certain group of the population would take advantage of an expanded offer of VR in psychotherapy.

REFERENCES

1. Segal R, Bhatia M, Drapeau M (2011) Therapists' Perception of Benefits and Costs of Using Virtual Reality Treatments. Cyberpsychology, Behavior, and Social Networking 14(1- 2): 29–34.
2. Hobbs CN, Kreiner DN, Honeycutt MW, Hinds RM, Brockman CJ (2010) The Illusion of Control in a Virtual Reality Setting. North American Journal of Psychology, 12(3): 551-564.
3. Botella CM, Juan MC, Banos RM, Alcaniz M, Guillén V, Rey B (2005) Mixing realities? An application of augmented reality for the treatment of cockroach phobia. Cyberpsychology & Behavior, 8(2): 162-171.
4. Eichenberg C (2011) Application of „Virtual Realities" in Psychotherapy: Possibilities, Limitations and Effectiveness. In J.-J. Kim (ed.), Virtual reality (pp 481-496). Rijeka: InTech.
5. Glantz K, Rizzo A, Graap K (2003) Virtual Reality for Psychotherapy: Current Reality and Future Possibilites. Psychotherapy: Theory, Research, Practice, Training, 40(1/2): 55-67.
6. Garcia-Palacios A, Hoffman HG, See SK, Tsay A, Botella C (2001) Redefining therapeutic success with virtual reality exposure therapy. CyberPsychology & Behavior 4: 341–8.
7. García-Palacios A, Botella C, Hoffman H, Fabregat S. (2007) Comparing Acceptance and Refusal Rates of Virtual Reality Exposure vs. In Vivo Exposure by Patients with Specific Phobias. CyberPsychology & Behavior 10 (5): S. 722–724.
8. Parsons TD, Rizzo A (2008) Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A meta-analysis. Journal of Behavior Therapy and Experimental Psychiatry, 39: 250-261
9. Powers MB, Emmelkamp PMG (2008) Virtual reality exposure therapy for anxiety disorders: A metaanalysis. Journal of anxiety disorders 22 (3): 561-569
10. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE (2005) Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Archives of General Psychiatry, 62: 617-627
11. Eichenberg, C. (2007). Der Einsatz von „Virtuelle Realitäten" in der Psychotherapie: Ein Überblick zum Stand der Forschung. Psychotherapeut, 52, 5, 362-367.

12. Optale G, Marin S, Pastore M, Nasta A, Pianon C. (2003) Male Sexual Dysfunctions and Multimedia Immersion Therapy (Follow-Up). CyberPsychology & Behavior 6(3): 289-294.
13. American Psychiatric Association. (2000) Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington, DC: Author.
14. World Health Organization. (2008). ICD-10: International statistical classification of diseases and related health problems (10th Rev. ed.). New York, NY: Author.
15. Curtis, GC, Magee, WJ, Eaton, WW, Wittchen, H-U & Kessler, RC (1998) Specific fears and phobias. Epidemiology and classification. The British Journal of Psychiatry 173: 212- 217.
16. Rothbaum BO, Hodges LF, Kooper R, Opdyke D, Williford J, North M (1995) Virtual reality graded exposure in the treatment of acrophobie: a case report. Behavior Therapie 26: 547-554.
17. Rothbaum BO (1995) Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. The American Journal of Psychiatry, 152 (4): 626- 628.
18. Emmelkamp P, Bruynzel M, Drost L, van der Mast C (2001) Virtual realitiy treatment in acrophobia: a comparison with exposure in vivo. Cyberpsychologie and Behavior 3: 335- 341.
19. Emmelkamp P, Krijn M, Hulsbosch L, de Vries S, Schuemie MJ, van der Mast C (2002) Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. Behavior Research and Therapie 4: 509-516.
20. Krijn M, Emmelkamp PM, Biemond R, de Wilde de Ligny C, Schuemie MJ, van der Mast CA (2004) Treatment of acrophobia in virtual reality: the role of immersion and presence. Behav Res Ther 42(2): 229-39.
21. Krijn M, Emmelkamp PMG, Olafsson RP, Schuemie MJ, van der Mast CA (2007) Do self-statements enhance the effectiveness of virtual reality exposure therapy? A comparative evaluation in acrophobia. CyberPsychol Behav 10: 362–370.
22. Coelho CM, Santos JA, Silva C, Wallis G, Tichon J, & Hine, TJ (2008) The role of selfmotion in Acrophobia Treatment. CyberPsychology & Behavior, 11(6): 723-725.
23. Coelho, CM, Silva, CF, Santos JA, Tichon J, & Wallis G (2008) Contrasting the Effectiveness and Efficiency of Virtual and Real

Environments in the Treatment of Acrophobia. PsychNology, 6(2): 203-216.

24. Choi YH, Jang DP, Ku JH, Shin MB, Kim SI (2001) Short-term treatment of acrophobia with virtual reality therapy (VRT): A case report. Cyberpsychology & Behavior, 4(3): 349- 354.
25. Wiederhold BK, Gevirtz RN, Spira JL (2001) Virtual Reality Exposure Therapy vs. Imagery Desensitization Therapy in the Treatment of Flying Phobia. In G. Riva, C. Galimberti (Eds), Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age (Kap. 14). Amsterdam: IOS Press.
26. Rothbaum, BO, Hodges L., Smith S, Lee JH (2000) A controlled study of virtual reality exposure therapy for fear of flying. J Consult Clin Psychol, 68: 1020–1026.
27. Rothbaum BO, Zimand E, Hodges L, Lang D, Wilson J (2006) Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. Behav Ther 37: 80–90.
28. Wiederhold B, Gevirtz R, Spira J (2001) Virtual reality exposure therapy vs. imagery desensitization therapy in the treatment of flying phobia, in G. Riva, C. Galimberti (eds). Towards CyberPsychology: Mind, Cognition, and Society in the Internet Age. Amsterdam: IOS Press. pp 254-272.
29. Wiederhold BK, Wiederhold MD (2003) Three year follow-up for virtual reality exposure for fear of flying. Cyberpsychol Behav 6: 441-445.
30. Mühlberger A, Herman MJ, Wiedemann G, Ellgring H, Pauli P (2001) Repeated exposure off light phobics to flights in virtual reality. Behav Res Ther 39: 1033-1050.
31. Mühlberger A, Widemann G, Pauli P (2003) Efficacy of a one-session virtual reality exposure treatment for fear of flying. Psychother Res 13: 323-336.
32. Mühlberger A, Weik A., Pauli P, Wiedemann G (2006) One-session virtual reality exposure treatment for fear of flying: 1-Year follow-up and graduation flight accompaniment effects. Psychotherapy Research 16(1): 26-40.
33. Maltby N, Kirsch I, Mayers M, Allen GJ (2002) Virtual reality exposure therapy fort he treatment of fear of flying: a controlled investigation. J Consult Clin Psychol 70: 1112-1118.

34. Krijn M, Emmelkamp PMG, Olafsson RP, et al. (2007) Fear of flying treatment methods: virtual reality exposure vs. cognitive behavioral therapy. Aviat Space Environ Med 78: 121–128.

35. Tortella-Feliu M, Botella C, Labres J, et al. (2011) Virtual Reality Versus ComputerAided Exposure Treatments for Fear of Flying. Behavior Modification 35(1): 3-30.

36. Garcia-Palacios A, Hoffman H, Carlin A, Furness TA III, Botella C (2002) Virtual reality in the treatment of spider phobia: A controlled study. Behaviour Research and Therapy, 40, 983-993.

37. Rinck M, Kwakkenbos L, Dotsch R, Wigboldus DHJ, Becker ES (2010) Attentional and behavioural responses of spider fearfuls to virtual spiders. Cognition & Emotion 24(7): 1199–1206.

38. Carlin AS, Hoffmann HG, Weghorst S (1998) Virtual reality and tactile augmentation in the treatment of spider phobia: a case study. Behaviour Research and Therapie 35: 153-158.

39. Hoffmann HG, García-Palacios A, Carlin A, Furness III TA (2003). Interfaces that heal: Coupling real and virtual objects to treat spider phobia. International Journal of humancomputer interaction, 16(2): 283-300.

40. Michaliszyn D, Marchand A, Bouchard S, Martel M-O, Poirier-Bisson, J (2010) A randomized controlled clinical trial of in virtuo and in vivo exposure for spider phobia. Cyberpsychology, Behavior and Social Networking 13(6): 689-695.

41. Bouchard S, Côté S, St-Jaques J, Robillard G, Renaud P (2006) Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. Technology and Health Care 14: 19-17.

42. Juan MC, Alcaniz M, Monserrat C, Botella C, Banos RM, Guerrero B (2005) Using augmented reality to treat phobias. IEEE Computer Graphics and Application 05: 31-37.

43. Bretón-López J, Quero S, Botella C, García-Palacios A, Banos RM, Alcaniz M (2010) An augmented reality system validation for the treatment of cockroach phobia. Cyberpsychology, Behavior and Social Networking 13: 705-710.

44. Juan MC, Joele D (2011) A comparative study of the sense of presence and anxiety in an invisible marker versus a marker augmented reality system for the treatment of phobia towards small animals. International

Journal of Human-Computer Studies 69(6): 440–453.

45. Kessler, RC, McGonagle KA, Zhao S, et al. (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Archives of General Psychiatry 51: 8–19.

46. Roy S, Klinger E, Légeron P, Lauer F, Chemin I, Nugues P (2003) Definition of a VRBased Protocol to Treat Social Phobia. CyberPsychology & Behavior 6: 411-420.

47. Klinger E, Bouchard S, Légeron P, Roy S, Lauer F, Chemin I, Nugues P (2005) Virtual Reality Therapy Versus Cognitive Behavior Therapy for Social Phobia: A Preliminary Controlled Study. Cyberpsychology and Behavior 1: 76-88.

48. Robillard G, Bouchard S, Dumoulin S, Guitard T, Klinger E (2010) Using virtual humans to alleviate social anxiety: Preliminary report from a comparative outcome study. Annual Review of CyberTherapy and Telemedicine 8: 46-48.

49. Harris SR, Kemmerling RL, North MM (2002) Brief Virtual Reality for Public Speaking Anxiety. CyberPsychology & Behavior 5(6): 543-550

50. Anderson, Page L.; Zimand, Elana; Hodges, Larry F.; Rothbaum, Barbara O. (2005): Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. In: Depress. Anxiety 22 (3), S. 156–158.

51. Wallach HS, Safir MP, Bar-Zvi M (2009) Virtual Reality Cognitive Behavior Therapy for Public Speaking Anxiety: A Randomized Clinical Trial. Behavior Modification 33(3): 314– 338.

52. Ter Heijden N, Brinkman W-P (2011) Design and evaluation of a virtual reality exposure therapy system with free speech interaction. Journal of CyberTherapy and Rehabilitation 41(1): 41-56.

53. Price M, Mehta N, Tone EB, Anderson PL (2011) Does engagement with exposure yield better outcomes? Components of presence as a predictor of treatment response for virtual reality exposure therapy for social phobia. Journal of Anxiety Disorders 25(6): 763– 770.

54. Cornwell BR, Heller R, Biggs A, Pine DS, Grillon C (2011) Becoming the Center of Attention on Social Anxiety Disorder: Startle Reactivity to a Virtual Audience during Speech Anticipation. J Clin Psychiatry 72: 942-948.

55. Grant BF, Hasin DS, Stinson FS, et al. (2006) The epidemiology of DSM-IV panic disorder and agoraphobia in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 67: 363–74.

56. Vincelli F, Choi YH, Molinari E, et al. (2001) A VR-based multicomponent treatment for panic disorders with agoraphobia. Studies in Health Technology and Informatics 81: 544–550.

57. Choi YH, Vincelli F, Riva G, Wiederhold BK, Lee JH, Park KH (2005) Effects of group experiental cognitive therapy for the treatment of panic disorder with agoraphobia. CyberPsychol Behav 8: 387-393.

58. Botella C, García-Palacios A, Villa H, et al (2007) Virtual reality exposure in the treatment of panic disorder and agoraphobia: a controlled study. Clin Psychol Psychother 14: 164-175.

59. Penate W, Pitti CT, Bethencourt JM, de la Fuente J, Gracia R (2008) The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. Int J Clin Health Psychol 8: 5-22.

60. Pitti CT, Penate W, de la Fuente J, et al. (2008) Agoraphobia: combined treatment and virtual reality. Preliminary results. Actas Esp Psiquiatra 36: 94-101.

61. Pérez-Ara MA, Quero S, Botella C, Banos R, Andreu-Mateu S, García-Palacios A, Bréton-Lopez J (2010) Virtual reality interoceptive exposure for the treatment of panic disorder and agoraphobia. Annual Review of CyberTherapy and Telemedicine 8: 61-64.

62. Meyerbröker K, Morina N, Kerkhof G, Emmelkamp PMG (2011) Virtual Realitiy Exposure Treatment of Agoraphobia: a Comparison of Computer Automatic Virtual Environment and Head-Mounted Display. Annual Review of Cybertherapy and Telemedicine 9: 51-56.

63. Kim K; Kim C-H, Cha KR, Park J, Han K, Kim YK, et al. (2008) Anxiety Provocation and Measurement Using Virtual Reality in Patients with Obsessive-Compulsive Disorder. CyberPsychology & Behavior 11 (6), S. 637–641.

64. Kim, K, Kim SI, Cha KR, Park J, Rosenthal MZ, Kim J-J, et al. (2010) Development of a computer-based behavioral assessment of checking behavior in obsessive-compulsive disorder. Comprehensive Psychiatry 51(1): 86–93.

65. Kim, K, Roh D, Kim, SI, Kim C-H (2012) Provoked arrangement

symptoms in obsessive–compulsive disorder using a virtual environment: A preliminary report. Computers in Biology and Medicine 42(4): 422–427.

66. Redemann L (2003). Die psychodynamisch imaginative Traumatherapie (PITT). ZPPM 1(2): 1-8.
67. Rothbaum BO, Hodges L, Alarcon R, et al. (1999) Virtual reality exposure therapy for PTSD Vietnam Veterans: a case study. J. Trauma. Stress 12: 263–271.
68. Rothbaum BO, Hodges L, Ready D, et al. (2001) Virtual reality exposure therapy for Vietnam veterans with posttraumatic stress disorder. J Clin Psychiatry 62: 617-622
69. Ready, David J.; Pollack, Stacey; Rothbaum, Barbara Olasov; Alarcon, Renato D. (2006): Virtual Reality Exposure for Veterans with Posttraumatic Stress Disorder. In: Journal of Aggression, Maltreatment & Trauma 12 (1-2), S. 199–220.
70. Ready DJ, Gerardi RJ, Backschneider AG, Mascaro N, Rothbaum BO (2010). Comparing Virtual Reality Exposure Therapy to Present-Centered Therapy with U.S. Vietnam Veterans with PTSD. Cyberpsychology, Behavior and Social Networking 13(1): 49-54.
71. Gerardi M, Rothbaum BO, Ressler K, Heekin M, Rizzo A (2008) Virtual reality exposure therapy using a virtual Iraq: Case report. J. Traum. Stress 21(2): 209–213.
72. Tworus R, Szymanska S, Illnicki S (2010). A Soldier Suffering from PTSD Treated by Controlled Stress Exposition Using Virtual Reality and Behavioral Training. Cyberpsychology, Behavior, and Social Networking 13(1): 103-107.
73. Rizzo AA, Reger G, Difede J, et al. (2009) Development and clinical results from the virtual Iraq exposure therapy application for PTSD. IEEE Explore: Virtual Rehabilitation 2009.
74. McLay RN, Wood DP, Webb-Murphy JA, Spira JL, Wiederhold MD, Pyne JM, Wiederhold BK (2011) A Randomized, Controlled Trial of Virtual Reality-Graded Exposure Therapy for Post-Traumatic Stress Disorder in Active Duty Service Members with Combat-Related Post-Traumatic Stress Disorder. Cyberpsychology, Behavior, and Social Networking 14(4): 223–229.
75. Reger GM, Holloway KM, Candy C, Rothbaum BO, Difede JA, Rizzo AA, Gahm GA (2011) Effectiveness of virtual reality exposure therapy

for active duty soldiers in a military mental health clinic. J. Traum. Stress 24(1): 93–96.

76. Rothbaum BO, Rizzo A, Difede J (2010) Virtual reality exposure therapy for combatrelated posttraumatic stress disorder. In: Annals of the New York Academy of Sciences 1208(1): 126–132.

77. Gamito P, Oliveira J, Morais D, et al. (2007) War PTSD: a VR pretrial case study. Annual Review of Cybertherapy & Telemedicine 5: 191–198.

78. Gamito P, Oliveira J, Rosa P, Morais D, Duarte N, Oliveira S, Saraiva T (2010) PTSD Elderly War Veterans: A Clinical Controlled Pilot Study. Cyberpsychology, Behavior, and Social Networking 13(1): 43-48.

79. Difede J, Cukor J, Patt I, Giosan C, Hoffman H. (2006) The Application of Virtual Reality to the Treatment of PTSD Following the WTC Attack. Ann. N.Y. Acad. Sci.1071: 500-501.

80. Difede J, Cukor K, Jayasinghe N, Patt I, Jedel S, Spielman L, Giosan C, Hoffman HG (2007) Virtual Reality Exposure Therapy for the Treatment of Posttraumatic Stress Disorder Following September 11, 2001. J Clin Psychiatry 68(11): 1639-1647.

81. Freedman SA, Hoffman HG, García-Palacios A, Weiss PL, Avitzour S, Josman N (2010) Prolonged Exposure and Virtual Reality Enhanced Imaginal Exposure for PTSD following a Terrorist Bulldozer Attack: A Case Study. Cyberpsychology, Behavior, and Social Networking 13(1): 95-101.

82. Saraiva T, Gamito P, Oliveira J, Morais D, Pombal M, Gamito L, Anastácio M (2007) The use of VR exposure in the treatment of motor vehicle PTSD: A case-report. Annual Review of CyberTherapy and Telemedicine 5: 199-205.

83. Beck JG, Palyo SA, Winer EH, Schwagler BE, Ang EJ (2007) Virtual Reality Exposure Therapy for PTSD Symptoms After a Road Accident: An Uncontrolled Case Series. Behavior Therapy 38 (1): 39–48.

84. Fidopiastis C, Hughes CE, Smith E (2009) Mixed reality for PTSD/TBI assessment. Annual Review of CyberTherapy and Telemedicine 7: 216-220.

85. Riva G, Raspelli S, Algeri D, Pallavicini F, Gorini A, Wiederhold BK, Gaggioli A (2010) Interreality in Practice: Bridging Virtual and Real Worlds in the Treatment of Posttraumatic Stress Disorders. Cyberpsychology, Behavior, and Social Networking 13(1): 55-65.

86. Fischer G (2000) Mehrdimensionale Psychodynamische Traumatherapie MPTT. Manual zur Behandlung psychotraumatischer Störungen. Heidelberg: Asanger.

87. Riva, G. (1998) Virtual reality in psychological assessment: The Body Image Virtual Reality Scale. Cyber Psychology and Behavior, 1: 37–44.

88. Ferrer-García M, Gutiérrez-Maldonado J (2012) The use of virtual reality in the study, assessment, and treatment of body image in eating disorders and nonclinical samples. A review of the literature. Body Image 9(1): 1-11.

89. Perpiná C, Botella C, Banos RM, Marco H, Alcaniz M, Quero S (1999) Body image and virtual reality in eating disorders: Is exposure to virtual reality more effective than the classical body image treatment? CyberPsychology & Behavior, 2: 149–159.

90. Riva G, Bacchetta M, Baruffi M, Molinari E (2001) Virtual Reality–Based Multidimensional Therapy for the Treatment of Body Image Disturbances in Obesity: A Controlled Study. Cyberpsychology & Behavior 4: 511-526.

91. Riva G, Bachetta M, Baruffi M, Molinari E (2002) Virtual-reality-based multidimensional therapy for the treatment of body image disturbances in binge eating disorders: A preliminary controlled study. IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society, 6: 224–234.

92. Rizzo A, Buckwalter J, Bowerly Tvan der Zaag C, Humphrey L, Neumann U, Chua C, Kyriakakis C, van Rooyen A, & Sisemore D. (2000) The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits. CyberPsychology and Behavior 3: 483–501.

93. Nowak R. (2002, June 26) VR hallucinations used to treat schizophrenia. New Scientist. Retrieved from http://www.newscientist.com/news/news.jsp?id_ns99992459.

94. Girard B, Turcotte V, Bouchard S, Girard B (2009) Crushing Virtual Cigarettes Reduces Tobacco Addiction and Treatment Discontinuation. CyberPsychology & Behavior, 12(5): 477-483.

95. Lee J-H, Kwon H, Choi J, Yang B-H (2007) Cue-exposure therapy to decrease alcohol craving in virtual environment. CyberPsychology & Behavior 10(5): 617-623.

96. Rothbaum BO, Hodges L, Smith S (1999) Virtual Reality Exposure Therapy Abbreviated Treatment Manual: Fear of Flying Application. Cognitive and Behavioral Practice 6: 234- 244.

97. Spira JL, Wiederhold BK, Pyne J, Wiederhold MD (2006). Virtual Reality Treatment Manual. In Virtuo Physiologically-Faciliated Graded Exposure Therapy in the Treatment of Recently Developed Combat-related PTSD. San Diego: The Virtual Reality Medical Center.

98. Bouchard S, Robillard G, Larouche S, Loranger C (2012). Description of a Treatment Manual for In Virtuo Exposure with Specific Phobia. In: C. Eichenberg (ed.) (2012). Virtual Realities. Rijeka: InTech.

99. Eichenberg, C. & Kienzle, K. (2011). Psychotherapeuten und Internet: Einstellung zu und Nutzung von therapeutischen Online-Angeboten im Behandlungsalltag. Psychotherapeut.

100. Eichenberg, C. & Molitor, K. (2011). Stationäre Psychotherapie und Medien: Ergebnisse einer Befragungsstudie an Therapeuten und Patienten. Psychotherapeut, 2, 162-70. DOI: 10.1007/s00278-011-0833-4.

101. Eichenberg C, Brähler E: Das Internet als Ratgeber bei psychischen Problemen: Eine bevölkerungsrepräsentative Befragung in Deutschland. Psychotherapeut. DOI 10.1007/s00278-012-0893-0.

SECTION 3:

MULTIMEDIA TRANSMISSION IN WIRELESS NETWORKS

CHAPTER 10

CLUSTERING IN WIRELESS MULTIMEDIA SENSOR NETWORKS

Pushpender Kumar and Narottam Chand

Department of Computer Science and Engineering, National Institute of Technology, Hamirpur, India

ABSTRACT

Wireless Multimedia Sensor Networks (WMSNs) are comprised of small embedded audio/video motes capable of extracting the surrounding environmental information, locally processing it and then wirelessly transmitting it to sink/base station. Multimedia data such as image, audio and video is larger in volume than scalar data such as temperature, pressure and humidity. Thus to transmit multimedia information, more energy is required which reduces the lifetime of the network. Limitation of battery

Citation: P. Kumar and N. Chand, "Clustering in Wireless Multimedia Sensor Networks," Journal of Sensor Technology, Vol. 3 No. 4, 2013, pp. 126-132. doi: 10.4236/jst.2013.34019

Copyright: © 2013 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

energy is a crucial problem in WMSN that needs to be addressed to prolong the lifetime of the network. In this paper we present a clustering approach based on Spectral Graph Partitioning (SGP) for WMSN that increases the lifetime of the network. The efficient strategies for cluster head selection and rotation are also proposed as part of clustering approach. Simulation results show that our strategy is better than existing strategies.

Keywords: Wireless Multimedia Sensor Network; Clustering; Spectral Graph Partitioning; Eigenvector; Eigenvalue

INTRODUCTION

Recent developments in wireless communication and embedded technology have made the wireless sensor network (WSN) possible. Wireless sensor networks are constituted of large number of low-cost, low-power and less communication bandwidth tiny sensor nodes. The sensors, which are randomly deployed in an environment, are required to collect data from their surroundings, process the data and finally send it to the sink through multi hops [1]. Traditional WSNs collects the scalar data such as temperature, pressure, etc. and transmit it to the sink. WSN has potential to design many new applications for handling emergency, military and disaster relief operations that require real time information for efficient coordination and planning [2].

Wireless multimedia sensor network (WMSN) uses cheap CMOS (Complementary Metal Oxide Semiconductor) camera and microphone sensors which can acquire multimedia information. WMSN consists of camera sensors as well as scalar sensors. Camera sensors can retrieve much richer information in the form of images or videos and hence provide more detailed and interesting data about the environment [3,4]. The multimedia content has the potential to enhance the level of information collected, compared with scalar data. Multimedia content produces immense amount of data to transmit over WMSN, which is limited in terms of power supply, communication bandwidth, memory, etc. In a large-scale network, if all the nodes have to communicate their data to their respective destination, it will deplete their energy quickly due to the long-distance, large volume of data and multi-hopnature of the communication. This will also lead to network contention. The clustering is a standard approach for achieving efficient and scalable control in these networks [5].

Clustering results in a number of benefits. It facilitates distribution of control over the network. It saves energy and reduces network contention by enabling locality of communication. Nodes communicate their data over shorter distances to their respective cluster head (CH). The cluster head aggregates these data into a smaller set of meaningful information. Not all nodes, but only the cluster heads need to communicate with their neighbouring cluster heads and sink/base station. Figure 1 shows the clustering of nodes in a general WSN. In this paper, we have utilized spectral graph partitioning (SGP) technique based upon eigenvalues proposed by Fiedler to form clustering in WMSN [6]. SGP method has been used in many applications such as image segmentation, social networks, etc. [7].

The spectral graph partitioning (SGP) algorithm is based on second highest eigenvalues of particular graph.

The second highest eigenvalue of the Laplacian matrix corresponding to different eigenvectors, is used to partition the graph into two parts. Within a cluster, a node with highest eigenvalue is selected as cluster head. In case of WMSN, large volume of sensed data is generated, therefore, such clustering can be utilized to reduce the volume and number of data transmissions through data aggregation. Simulation experiments have been performed to evaluate the performance of proposed method and compare it with the existing technique.

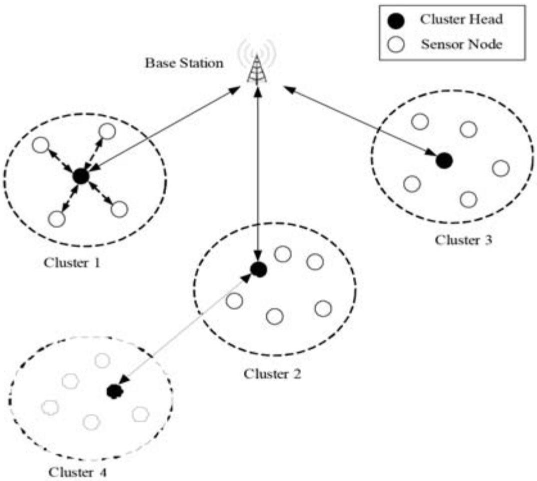


Figure 1. Clustering of SNs in WSN.

The second highest eigenvalue of the Laplacian matrix corresponding to different eigenvectors, is used to partition the graph into two parts. Within a cluster, a node with highest eigenvalue is selected as cluster head. In case of WMSN, large volume of sensed data is generated, therefore, such clustering can be utilized to reduce the volume and number of data transmissions through data aggregation. Simulation experiments have been performed to evaluate the performance of proposed method and compare it with the existing technique. The rest of the paper is organized as follows. Section 2 reviews the related work. General SGP strategy for clustering has been presented in Section 3. Section 4 describes the use of SGP for WMSN. In Section 5 we present the results of performance evaluation of the method and Section 6 concludes the paper.

RELATED WORK

The nodes are often grouped together into disjoint and mostly non-overlapping groups are called clusters. Clusters are used to minimize communication latency and improve energy efficiency. Leader of every cluster is often called the cluster head (CH) and generally has to perform more functions as compared to normal sensor node.

M. Qin et al. suggested voting-based clustering algorithm (VCA) [8] that enhances the criteria for cluster selection and combines load balancing consideration together with topology and energy information. VCA addresses inefficient cluster formation using a voting scheme, which enables the nodes to exchange information about their local network view. This method assumes synchronization among the nodes. Similar to WCA [9], the time required for the nodes to gather information about all other nodes depends on the network size and is not constant. B. Elbhiri et al. [7] suggested spectral classification based on near optimal clustering in wireless sensor networks (SCNOPC-WSN) algorithm. This algorithm deals with the clustering problem in WSN. Energy aware adaptive clustering protocol is used for the bi-partitioning spectral classification and it guarantees robust clustering. SCNOPC-WSN also deals with the optimization of the energy dissipated in the network.

Banerjee and Khuller [10] suggested hierarchical clustering algorithm based on geometric properties of the wireless network. A number of cluster properties such as cluster size and the degree of overlap, which are useful for the management and scalability of the hierarchy, are also considered while grouping the nodes. In the proposed scheme, any node in the WSN can initiate the cluster formation process. Initiator with least node ID will take precedence, if multiple nodes started cluster formation process at the same time.

Bandyopadhyay and Coyle [11] proposed EEHC which is a distributed, randomized clustering algorithm for WSNs with the objective of maximizing the network lifetime. CHs collect the sensor reading in their individual clusters and send an aggregated report to the base station. Their technique is based on two stages—initial and extended. In EEUC [12], the hot-spot problem in multihop networks is solved using cluster with unequal size. CHs that are closed to the base station tend to die faster, because they relay much more traffic than remote nodes. Setting smaller cluster sizes to the close CHs preserves their energy. Additional improvement for multihop networks is presented in [13], using a separation between the data gathering and aggregation task and the forwarding task. Spectral graph partitioning algorithm partitions the graph using the eigenvectors of the matrix obtained from the graph. SGP obtains data representation in the low-dimensional space that can be easily clustered. Eigenvalues and eigenvectors provide a penetration into the connectivity of the graph.

SGP FOR CLUSTER FORMATION

Spectral graph partitioning technique is based on eigenvalues and eigenvector of the adjacency matrix of graph to partition the graph. The methods are called spectral, because they make use of the spectrum of the adjacency matrix of the data to cluster the points [14]. Spectral methods are widely applied for graph partitioning. Spectral graph partitioning is a powerful technique and also is being used in image segmentation and social network analysis. SGP divides the graph into two disjoint groups, based on eigenvectors corresponding to the second smallest eigenvalue of the Laplacian matrix [15].

Let G (V, E) is an undirected graph where V represents the set of vertices (sensor nodes) and E represents the set of edges connecting these vertices. Each vertex is identified by an index $i \in \{1, 2, \cdots, N\}$. The edge between node i and node j is represented by e_{ij}. The graph can be represented as an adjacency matrix. The adjacency matrix A of graph G having N nodesis the N × N matrix where the non-diagonal entry a_{ij} is the number of edges from node i to node j, and the diagonal entry a_{ii} is the number of loops at node i. The adjacency matrix is symmetric for undirectedgraphs [16].

The adjacency matrix A is defined as

$$A=\left[a_{ij}\right]=\begin{cases}1 & \text{edge weight between node } i \text{ and node } j\\ 0 & \text{otherwise}\end{cases}$$

We also define degree matrix D for graph G. The degree matrix of a graph gives the number of edges between node i to another node [17]. The degree matrix is a diagonal matrix which contains information about the degree of each node. It is helpful to construct the Laplacian matrix of a graph. The degree matrix D for G is a N × N square matrix and is defined as

$$D=\left[\deg_{ij}\right]=\begin{cases}\text{total weight of edges incident to node } i\\ 0\end{cases}$$

The Laplacian matrix is formed from adjacency matrix and the degree matrix. The Laplacian matrix of the graph G having N vertices is N × N square matrix and is represented as

L=D-A

The normalized form of Laplacian matrix can be written as

$$\Upsilon(i,j)=\begin{cases}1 & \text{if } i=j \text{ and } \deg_j \neq 0\\ -\dfrac{1}{\sqrt{\deg_{\mathrm{i}}\deg_{\mathrm{j}}}} & \text{if } i \text{ and } j \text{ are adjacent}\\ 0 & \text{otherwise}\end{cases}$$

The eigenvalues of matrix Υ are denoted by, λ_i i= Λ N such that $\lambda_1 \leq \lambda_2 \leq \ldots\ldots \leq \lambda_n$ Laplacian matrix has the property $\Upsilon \cdot X=\lambda \cdot X$ where X is the eigenvector of the matrix and λ is the eigenvalue of the matrix. Laplacian matrix plays important role in spectral graph theory. λ_1 represents the number of subgraphs in the network. The second smallest eigenvalue λ_2 is referred to the algebraic connectivity and its corresponding eigenvector is usually referred to as the Fiedler Vector [14,18].

We choose the eigenvector values corresponding to the second highest eigenvalue λ_2. Second highest eigenvalue (λ_2) divides the graph into two subgraphs. G is divided into two subgraphs G^- and G^+, where G^+ and are the set of vertices related to the new subgraphs. contains nodes corresponding to positive eigenvalues and G^- contains nodes corresponding to negative eigenvalues. The set of vertices is defined by

$$N=N^+\bigcup N^-$$

and

$$N^{+} \bigcap N^{-} = \phi$$

where $N = |G|$, $N^{+} = |G^{+}|$ and $N^{-} = |G^{-}|$.

CLUSTER FORMATION IN WMSN

SGP technique can be used for dividing the network into clusters. SGP has many advantages as compared to other clustering algorithms. SGP partitions the graph on the basis of eigenvalues and eigenvector adjacency matrix. If the graph is partitioned into more than two subgraphs, apply SGP technique recursively. These properties make SGP technique a better option for multimedia data clustering where large volume of data is transmitted between nodes and CH.

In our proposed method, clustering of WMSN has been done on the basis of Spectral Graph Partitioning technique [19]. Each node sends short message to sink which contains the location information of the node. On the basis of this information, the sink constructs the adjacency matrix and degree matrix and then constructs the Laplacian matrix. The eigenvector corresponding to second smallest eigenvalue (called Fiedler Vector) is used to partition the WMSN. The location of each node may be found by GPS or any other localization method [18,20].

Steps for Clustering

1) Construct a graph G of the given sensor network.
2) Construct the normalized Laplacian matrix as

$$\Upsilon(i,j) = \begin{cases} 1 & \text{if } i = j \text{ and } \deg_j \neq 0 \\ -\dfrac{1}{\sqrt{\deg_i \deg_j}} & \text{if } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Where $\deg_i$ is the degree of node i

3) From the Laplacian matrix Υ of the graph compute the eigenvalues and eigenvector of Laplacian matrix.
4) Select the second smallest eigenvalue λ_2 of Laplacian matrix Υ.

5) Choose eigenvector value corresponding to the eigenvalue λ_2.

Divide the graph G into two subgraphs G^+ and G^- where G^+ contains nodes corresponding to positive eigenvalues and G^- , contains nodes corresponding to negative eigenvalues.

After the first iteration of above method, the whole network is divided into two clusters based on the eigen values of the nodes. Table 1 shows the two partitions/ clusters for a given graph shown in Figure 2. After first iteration cluster 1 contains all the nodes with positive eigenvector values and another cluster 2 contains nodes having negative eigenvector values. Cluster 1 has five nodes with positive value of eigenvector and the nodes are A, B, C, D and F. Cluster 2 has five nodes that have negative eigenvector values and the nodes are E, G, H, I and J.

Only two clusters are formed in first iteration. The larger size clusters can be further divided into two different clusters by applying the algorithm recursively. This process continues until maximum intra-node distance within a cluster is less than $R/2\sqrt{2}$ where R is the transmission range of the sensor node. When intranode distance is remaining $R/2\sqrt{2}$ two nodes in neighbouring clusters can communicate in one hop. After applying the algorithm recursively, the given network is divided into four clusters as shown in Figure 3. Table 2 shows the formed cluster after first iteration.

It has been observed that both the clusters have higher intra-node distance than $R/2\sqrt{2}$, so apply the algorithm to both the clusters. After applying the algorithm cluster 1 is portioned into two different clusters. This algorithm is also applied to cluster 2 of Figure 3. Thus the whole network is divided into four clusters as shown in Figure 4. Table 3 shows different nodes present in the each cluster

Cluster Head Election

The clustering algorithm divides the whole network into clusters. The next step is election of cluster head for each cluster. As per the property of SGP, the least eigenvector value of node signifies that the node is well connected to the other nodes within the cluster as well as it is connected to cluster [21].

For initial cluster head election, we chose the least eigenvector value among the nodes within cluster, Table 4 represents the eigenvector values of the cluster and Table 5 shows the elected cluster heads in different clusters on the basis of eigenvector values. Therefore, we compare the eigenvector values of the cluster and choose the least eigenvector node as a cluster head.

$$\text{Cluster Head} = \text{Least}\left|\text{Eigenvector}\right|$$

Figure 5 shows the flow chart for cluster head selection and rotation. Figure 6 shows the elected cluster heads for different clusters. Cluster head rotation must take place when residual energy (E_{res}) of the cluster head node falls below the threshold value (E_{th}). The present cluster head declares the election process by sending a message that contains its E_{res} to all the cluster members. The cluster members whose residual energy is greater than E_{res} responds to this message by sending the residual energy to the cluster head.

The new cluster head is elected based upon CH Candidacy Factor (CF) defined as

$$CF_i = \frac{E^i_{res}}{D_i}$$

Where E^i_{res} is the residual energy of node i, Di is the distance between node i and current cluster head. If (x_{ch}, y_{ch}) and (x_i, y_i) are the location coordinates of current cluster head and node i, respectably, then

$$D_i = \sqrt{(x_{ch} - x_i) + (y_{ch} - y_i)}$$

A node with highest value of CF is elected as next cluster head.

PERFORMANCE EVALUATION

In this section the performance of SGP has been evaluated through simulation. The simulation has been performed in Matlab2013a. The performance of SGP protocol is compared with HEED [22]. The performance metrics include percentage of nodes acting as cluster heads and ratio of single node cluster and network lifetime. Ratio of single node cluster indicates that the ratio of number of clusters having single nodes to the total number of clusters. If the ratio of single node clusters in the network is high, it may lead to early energy dissipation. Figure 7 illustrates the percentage of cluster with more than one node. High single node cluster (the cluster head) may lead to reduce the network lifetime. Single node cluster arise when a node is forced to represent itself. The figure shows that HEED produces a higher percentage of non-single node clusters than the SGP. Percentage of nodes acting as cluster head describes the fraction of total sensor nodes in percentage, acting as cluster head in the sensing field. In WMSN, smaller percentage of nodes should act as CH, creating a cluster which has enough number of sensor nodes (cluster members). High values indicate that numbers of clusters are present in the network with small size of clusters. Figure 8 shows the percentage of cluster heads selected by both HEED and SGP (for different number of nodes) are identical.

CONCLUSION

This paper has proposed an approach to deal with the clustering problem in a given wireless multimedia sensor network. We have explained the details of the proposed clustering algorithm. It divides the network into two clusters and partitions the network till we get the clustering, where nodes in neighbouring clusters can communicate using single hop. A cluster head elections technique based on eigenvector has also been proposed. Simulation results show that our proposed algorithms perform better than those HEED algorithms.

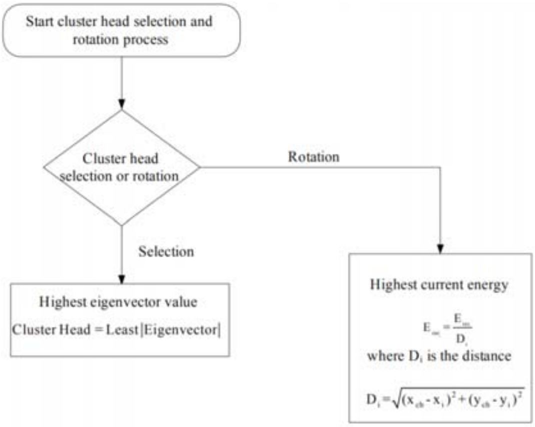

Start cluster head selection and rotation process
Cluster head selection or rotation
Rotation
Selection
Highest eigenvector value
Cluster Head = Least|Eigenvector|
Highest current energy
$E_{m} = \frac{E_{m}}{D_i}$
where D_i is the distance
$D_i = \sqrt{(x_{ch} - x_i)^2 + (y_{ch} - y_i)^2}$

Figure 5. Flow chart for selection and election of cluster head.

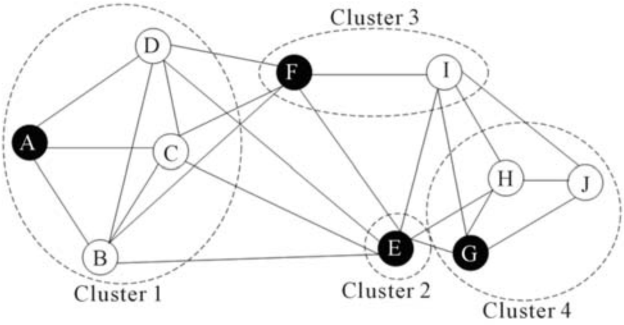


Figure 6. Election cluster heads of given graph.

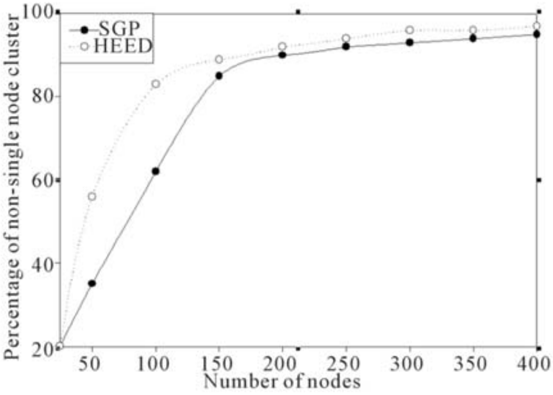


Figure 7. Percentage of non-single node cluster.

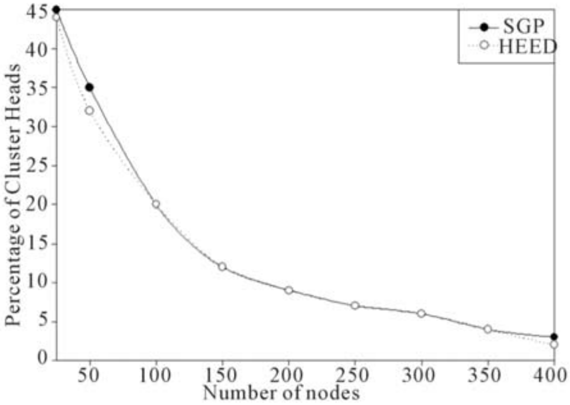


Figure 8. Percentage of selected cluster heads.

REFERENCES

1. F. Akyldiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey," Computer Networks, Vol. 38, No. 4, 2002, pp. 393-422.
2. S. Taruna and M. R. Tiwari, "An Event Driven Energy Efficient Data Reporting System for Wireless Sensor Networks," Vol. 2, No. 2, 2013, pp. 70-75.
3. I. F. Akyildiz, T. Melodia and K. R. Chowdhury, "A Survey on Wireless Multimedia Sensor Networks," Computer Networks, Vol. 51, No. 4, 2007, pp. 921-960.
4. P. Kumar and N. Chand, "Clustering in Wireless Multimedia Sensor Networks Using Spectral Graph Partitioning," International Journal of Communication, Network and System Sciences, Vol. 6, No. 3, 2013, pp. 128-133.
5. M. Demirbas, A. Arora, V. Mittal and V. Kulathumani, "A Fault-Local Self-Stabilizing Clustering Service for Wireless Ad Hoc Networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 17, No. 9, 2006, pp. 912- 922.
6. B. Auffarth, "Spectral Graph Clustering," Course Report. http://wwwlehre.inf.uos.de/~bauffart/spectral.pdf
7. B. Elbhiri, S. El Fkihi, R. Saadane and D. Aboutajdine, "Clustering in Wireless Sensor Network Based on Near Optimal Bi-partitions," 6th EURO-NF Conference on Next Generation Internet (NGI), 2010, pp. 1 -6.
8. M. Qin and R. Zimmermann, "VCA: An Energy Efficient Voting Based Clustering Algorithm for Sensor Networks," Journal of Universal Computer Science, Vol. 13, No. 1, 2007, pp. 87-109.
9. M. Chatterjee, S. K. Das and D. Turgut, "WCA: A Weighted Clustering Algorithm for Mobile Ad hoc Networks," Journal of Cluster Computing (Special Issue on Mobile Ad hoc Networks), Vol. 5, No. 2, 2002, pp. 193-204.
10. S. Banerjee and S. Khuller, "A Clustering Scheme for Hierarchical Control in Multi-Hop Wireless Networks," IEEE INFOCOM, 2001, pp. 1028-1037.
11. S. Bandyopadhyay and E. Coyle, "An Energy Efficient Hierarchical Clustering Algorithmfor Wireless Sensor Networks," 22nd Annual Joint Conference of the IEEE Computer and Communications Societies

(INFOCOM), Vol. 3, 2003, pp. 1713-1723.

12. C. Li, M. Ye, G. Chen and J. Wu, “An Energy Efficient Unequal Clustering Mechanism for Wireless Sensor Networks,” 2nd IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS), 2005, pp. 125-132.
13. Y. He, Y. Zhang, Y. Ji and S. X. Shen, “A New Energy Efficient Approach by Separating Data Collection and Data Report in Wireless Sensor Networks,” International Conference on Communications and Mobile Computing, 2006, pp. 180-192.
14. A. Bertrand and M. Moonen, “Distributed Computation of the Fiedler Vector with Application to Topology Inference in Ad Hoc Networks,” Internal Report KU Leuven ESAT-SCD, 2012.
15. Laplacian Matrix Wikipedia. http://en.wikipedia.org/wiki/Laplacian_matrix
16. Adjacency Matrix Wikipedia. http://en.wikipedia.org/wiki/Adjacency_matrix
17. A. Savvides, C. Han and M. B. Srivastva, “Dynamic FineGrained Localization in Ad Hoc Networks of Sensors,” 7th International Conference on Mobile Computing and Networking (MOBICOM), 2001, pp. 166-179.
18. D. Lu, N. Nan and X.-X. Huang, “Clustering Based Spectrum Allocation Scheme in Mobile Ad Hoc Networks,” Bulletin of Advanced Technology Research (BATR), Vol. 5, No. 12, 2011, pp. 37-41.
19. N. Bulusu, J. Heidemann and D. Estrin, “GPS-Less Low Cost Outdoor Localization for Very Small Devices,” IEEE Personal Communication Magazine, Vol. 7, No. 5, 2000, pp. 28-34.
20. A. Savvides, C. Han and M. B. Srivastva, “Dynamic FineGrained Localization in Ad Hoc Networks of Sensors,” 7th International Conference on Mobile Computing and Networking (MOBICOM), 2001, pp. 166-179.
21. O. Younis and S. Fahmy, “Distributed Clustering in AdHoc Sensor Networks: A Hybrid, Energy Efficient Approach,” IEEE Transactions on Mobile Computing, Vol. 3, No. 4, 2004, pp. 366-379.

CHAPTER 11

A DYNAMIC LINK ADAPTATION FOR MULTIMEDIA QUALITY-BASED COMMUNICATIONS IN IEEE_802.11 WIRELESS NETWORKS

Ahmed Riadh Rebai and Mariam Fliss

Wireless Research Group, Texas A&M University at Qatar, Qatar

INTRODUCTION

Assuming that the IEEE 802.11 Wireless Local Area Networks (WLANs) are based on a radio/infrared link, they are more sensitive to the channel variations and connection ruptures. Therefore the support for multimedia applications over such WLANs becomes non-convenient due to the compliance failure in term of link rate and transmission delay performance. Voice and broadband video mobile transmissions (which normally have strict bounded transmission delay or minimum link rate requirement) entail

Citation: Ahmed Riadh Rebai and Mariam Fliss (March 7th 2012). A Dynamic Link Adaptation for Multimedia Quality-Based Communications in IEEE_802.11 Wireless Networks, Multimedia - A Multidisciplinary Approach to Complex Issues, Ioannis Karydis, IntechOpen, DOI: 10.5772/36405.

Copyright: © 2012 by authors and IntechOpen. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

the design of various solutions covering different research aspects like service differentiation enhancement (Rebai et al., 2009), handoff scheme sharpening (Rebai, 2009a, 2009b, 2010) and physical rate adjustment. The core of this chapter focuses on the last facet concerning the link adaptation and the Quality of Service (QoS) requirements essential for successful multimedia communications over Wi-Fi networks. In fact, the efficiency of rate control diagrams is linked to the fast response for channel variation. The 802.11 physical layers provide multiple transmission rates (different modulation and coding schemes). The original 802.11 standard operates at 1 and 2 Mbps (IEEE Std. 802.11, 1999). Three high-speed versions were added to the original version. The 802.11b supports four physical rates up to 11 Mbps (IEEE Std. 802.11b, 1999). The 802.11a provides eight physical rates up to 54 Mbps (IEEE Std. 802.11a, 1999). The last 802.11g version, maintains 12 physical rates up to 54 Mbps at the 2.4 GHz band (IEEE Std. 802.11g, 2003). As a result, Mobile Stations (MSs) are able to select the appropriate link rate depending on the required QoS and instantaneous channel conditions to enhance the overall system performance. Hence, the implemented link adaptation algorithm symbolizes a vital fraction to achieve highest transmission capability in WLANs. “When to decrease and when to increase the transmission rate?” are the two fundamental matters that we will be faced to when designing a new physical-rate control mechanism. Many research works focus on tuning channel estimation schemes to better detect when the channel condition was improved enough to accommodate a higher rate, and then adapt their transmission rate accordingly (Habetha & de No, 2000; Qiao et al., 2002). However, those techniques usually entail modifications on the current 802.11 standard. In (del Prado Pavon & Choi, 2003), authors presented a motivating rate adaptation algorithm based on channel estimation without any standard adjustment. However, this scheme supposes that all the transmission failures are due to channel errors and not due to multi-user collisions.

Another way to perform link control is based on local Acknowledgment (Ack) information for the transmitter station (Qiao & Choi, 2005). Consequently, two new techniques (Chevillat et al., 2003; Kamerman & Monteban, 1997) where accepted by the standard due to their efficiency and implementation simplicity. In fact, the source node tries to increase its transmission rate after successive successful Ack responses, and therefore they do not involve any change for the 802.11 standard. Moreover and as it was demonstrated by (Sangman et al., 2011) a fine and excellent physical link adjustment will carry out a qualityaware and robust routing for mobile multihop ad hoc networks. A good study (Galtier, 2011) was recently addressed regarding the adaptative rate issues in the WLAN Environment and highlighted the high correlation between the Congestion Window (CW) of the system, and the rate at which packets are emitted. The given analytical approach opens the floor and shows that the different mechanisms that have been implemented in the MAC systems of WLAN cards have strong correlations with other transmission parameters and therefore have to be redesigned with at least a global understanding of channel access problems (backoff and collisions) and rate adaptation questions.

In this chapter we propose a new dynamic time-based link adaptation mechanism, called MAARF (Modified Adaptive Auto Rate Fallback). Beside the transmission frame results, the new model implements a Round Trip Time (RTT) technique to select adequately an instantaneous link rate. This proposed model is evaluated with most recent techniques adopted by the IEEE 802.11 standard: ARF (Auto Rate Fallback) and AARF (Adaptive ARF) schemes. Thus, we are able to achieve a high performance WLAN transmission. Consequently, we can extend this approach in various Wi-Fi modes to support multimedia applications like voice and video tasks.

The rest of the chapter is organized as follows. Section 2 offers a literature survey on related link-adjustment algorithms and the actual used ones. Section 3 is dedicated to the new proposed MAARF method and its implementation details. Simulation results will be given in Section 4 to illustrate the link quality improvement of multimedia transmissions over WiFi networks and to compare its performance with previous published results (Kamerman & Monteban, 1997; Lacage et al.,2004). We show how the proposed model outperforms previous approaches (ARF and AARF) because of its new time-based decision capability in addition to Ack count feature.

REVIEW OF THE CURRENT RATE-CONTROL APPROACHES

First we recall that the standard IEEE802.11 (IEEE Std. 802.11a, 1999; IEEE Std. 802.11b, 1999; IEEE Std. 802.11g, 2003) includes various versions a/b/g and allows the use of multiple physical rates (from 1Mbps to 54Mbps for the 802.11g). Therefore, several studies have been made to develop mechanisms which lead to adapt transmission attempts with the best physical available rate depending on the estimated channel condition to avoid transmission failures with Wi-Fi connections. The most important issues that should be taken into account and are responsible for the design of a reliable rate adaptation mechanism are:

- The channel condition variation due to a packet transmission error which results to multiple retransmissions or even a transmission disconnection.
- The channel sensitivity against interferences (Angrisani et al., 2011) due to disturbing incidences, additive random noises, electromagnetic noises, the Doppler effect, an accidental barrier or natural phenomena.
- The packet emissions latency which affects the autonomy of mobile stations in case of transmission error (since the communication duration is extended and the energy consumption becomes more important).
- The MSs Mobility leads to a distances change, hence, to an appropriate mobility management protocol over Wi-Fi connections.

Depending on the instantaneous channel quality, a rate adjustment will be always needed to achieve better communication performance with respect for multimedia QoS requirements.

Since WLAN systems that use the IEEE802.11g (IEEE Std. 802.11g, 2003) physical layer offer multiple data rates ranging from 1 to 54 Mb/s, the link adaptation can be seen as a process of switching or a dynamic choosing mechanism between different physical data rates corresponding to the instantaneous channel state. In other words, it aims to select the ‘ideal’ physical rate matching the actual channel condition. The best throughput can be larger or smaller than the current used one. The adequate rate will be chosen according to the instantaneous medium conditions. There are two criteria to properly evaluate this adaptation/adjustment: the first is the channel quality estimation; secondly is the adequate rate selection.

The estimation practice involves a measurement of the instantaneous channel states variation within a specific time to be able to predict the matching quality. This creates a large choice of indicator parameters on the medium condition that may include the observed Signal to Noise Ratio (SNR), the Bit Error Rate (BER), and the Received Signal Strength Indicator (RSSI). Those various physical parameters express instantaneous measurements operated by the 802.11 PHY card after completion of the last transmission.

Regarding the rate selection formula, it entails a first-class exploitation of channel condition indicators to better predict the medium state and then fit/adjust the suitable physical rate for the next communication. Consequently, this process will reduce packets' retransmissions and the loss rate. Bad channel-quality estimation would result in performance degradation. Thus, inaccurate assessments resulting from a bad choice of medium state indicators give rise to inappropriate judgments on the instantaneous conditions and cause deterioration on the observed performance. Therefore, this estimation is essential to better support multimedia services and maximize performance and the radio channel utilization.

Accordingly, during packets transmission, the corresponding MS may increase or decrease the value of its physical rate based on two different approaches:

- With the help of accurate channel estimation, the MS will know precisely when the medium conditions are improved to accommodate a higher data rate, and then adapt its transmission rate accordingly. However, those techniques (Habetha & de No, 2000; Qiao et al., 2002) require efforts to implement incompatible changes on the 802.11 standard. Another research work (del Prado Pavon & Choi, 2003) have presented a very interesting data rate adapting plan based on RSSI measurements and the number of transmitted frames for an efficient channel assessment without any modification on the standard. On the other hand, this plan operates under the assumption that all transmission failures are due to channel errors. Thus, it will not work efficiently in a multi-user environment where multiple transmissions may fail due to collisions and not only to the channel quality.
- The alternative way for the link adaptation is to carry out decisions based exclusively on the information returned by the receiver. In 802.11 WLANs, an acknowledgment (ACK) is sent by the receiver after the successful data recovery. Only after receiving an ACK frame the transmitter announces a successful transmission attempt. On the other hand, if an ACK is either incorrect or not received, the sender presumes a data transmission failure and reduces its actual data rate down to the next available physical rate-slot. In addition, the transmitter can increase its transmission rate after assuming a channel condition enhancement by receiving a specific number of consecutive positive ACKs. These approaches (Qiao & Choi, 2005; Chevillat et al., 2003) do

not require changes on the actual Fi-Wi standard and are easy to deploy with existing IEEE 802.11 network cards.

Various additional techniques have been proposed in the literature to sharpen the accuracy of the rate adaptation process and improve the performance of IEEE 802.11 WLANs. The authors in (Pang et al., 2005) underlined the importance of MAC-layer loss differentiation to more efficiently utilize the physical link. In fact, since IEEE 802.11 WLANs do not take into account the loss of frames due to collisions, they have proposed an automatic rate fallback algorithm that can differentiate between the two types of losses (link errors and collisions over the wireless link). Moreover it has been shown in (Krishnan & Zakhor, 2010) that an estimate of the collision probability can be useful to improve the link adaptation in 802.11 networks, and then to increase significantly the overall throughput by up to a factor of five. In (Xin et al., 2010) the authors presented a practical traffic-aware active link rate adaptation scheme via power control without degrading the serving rate of existing links. Their basic idea consists to firstly run an ACK based information exchange to estimate the upper power bound of the link under adaptation. Then by continuously monitoring the queue length in the MAC layer, it would be easy to know whether the traffic demand can be met or not. If not, the emitting power will be increased with respect to the estimated power upper-bound and will switch to a higher modulation scheme. A similar strategy was presented in (Junwhan & Jaedoo, 2006) that provides two decisions to estimate the link condition and to manage both the transmission rate and power.

Several research works (Haratcherev et al., 2005; Shun-Te et al., 2007; Chiapin & Tsungnan, 2008) have implemented a cross-layer link adaptation (CLLA) scheme based on different factors as: the number of successful transmissions, the number of transmission failures, and the channel information from the physical layer to determine actual conditions and therefore to adjust suitably transmission parameters for subsequent medium accesses. As well in (Chen et al., 2010) a proper-designed cross application-MAC layer broadcast mechanism has been addressed, in which reliability is provided by the application layer when broadcasting error corrections and next link rate adaptations (resulting from the MAC layer).

Another approach (Jianhua et al., 2006) has been developed where both packet collisions and packet corruptions are analytically modeled with the proposed algorithm. The models can provide insights into the dynamics of the link adaptation algorithms and configuration of algorithms parameters. On the other hand, in (An-Chih et al., 2009) the authors presented a joint adaptation of link rate and contention window by firstly considering if a proper backoff window has been reached. Specifically, if the medium congestion level can be reduced by imposing a larger backoff window on transmissions, then there may be no need to decrease the link rate, given that the Signal to Interference-plus-Noise Ratio (SINR) can be sustained. In the rest of this section we decide to provide details and discuss only the two main currently-implemented techniques.

Auto Rate Fallback (ARF)

Auto Rate Fallback (Kamerman & Monteban, 1997) was the first rate-control algorithm published and quickly adopted/integrated with the Wi-Fi standard. It was designed to optimize the physical rate adjustment of the second WLANs generation (specifically for 802.11a/g versions which allow multi-hop physical rates). The ARF technique is based simply on the number of ACKs received by the transmitting MS to determine the next rate for the next frame transmission. This method does not rely on hidden out-layer information, such as a physical channel quality measurement like (i.e. the SNR value). Thus, it was easy to implement and fully compatible with the 802.11 standard. In fact, after a fixed number of successful transfers equal to 10 or the expiration of a timer T initially launched the ARF increments the actual physical transmission rate from R_i to a higher rate R_{i+1} among those allocated by the standard. In other words, the ARF mechanism decides to increase the data rate when it determines that channel conditions have been improved. Unlike other algorithms reported in the literature, the ARF detection is not based on Physical layer measurements upon a frame delivery. Basically it simply considers the medium status improvement by counting the number of consecutive successful transmissions made or the timer (T) timeout. This timer is defined as the maximum waiting delay which will be launched by the MS each time it switches between the given data rates. Once this timer ends without any rate swap the MS will be testing a higher available rate. This practical implementation is considered as second alternative to adapt the best transmission rate since it covers the case that medium conditions are excellent and favorable to adopt a higher rate and the counter of consecutive successful transmissions will never reach the desired value (10) due to other failures. This case is very common in such wireless networks where a transmission failure is not only due to an inadequate rate.

In addition, the next transmission must be completed successfully immediately after a rate increase otherwise the rate will be reduced instantaneously (back to the old smaller value), and the timer T will be reset. In fact, the mechanism has estimated that the new adopted rate is not adequate for next network transmissions. Also after any two consecutive failures the algorithm automatically reduces its actual rate until it reaches again a number of 10 consecutive ACKs or the expiration of timer T. In this way, ARF detects the deterioration of the channel quality based on two consecutive failed transmissions, and chooses to back out to the previous rate. Figure 1 summarizes the operation of the ARF and shows the corresponding flow diagram.

While ARF increases the transmission rate at a fixed frequency (each 10 consecutive ACKs) to achieve a higher system throughput, this model has two main drawbacks:

- Firstly, this process can be costly since a transmission failure (produced by an unsuitable rate increasing decision made by the ARF mechanism) reduces the overall throughput. Specifically, for a steady channel status (stable characteristics) ARF will try periodically to switch to a higher rate by default which leads to unnecessary frame transmission failure and reduces the algorithm efficiency.
- Secondly, ARF is unable to stabilize the rate variations. In fact if the channel conditions deteriorate suddenly, the ARF mechanism will be unable to respond fast to these changes and to suit the current state. It will carry out numerous transmission failures so that it reaches the desired rate value. Therefore, this algorithm cannot cope with rapid medium status changes.

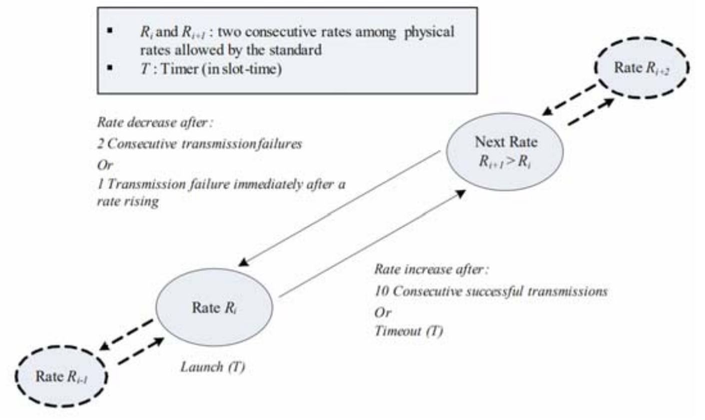

▪ R_i and R_{i+1} : two consecutive rates among physical rates allowed by the standard
▪ T : Timer (in slot-time)
Rate R_{i+2}
Rate decrease after:
2 Consecutive transmissionfailures
Or
1 Transmission failure immediately after a rate rising
Next Rate
$R_{i+1} > R_i$
Rate increase after:
10 Consecutive successful transmissions
Or
Timeout (T)
Rate R_i
Launch (T)
Rate R_{i-1}

Figure 1. The ARF flow diagram

Adaptive Auto Rate Fallback (AARF)

To overcome the given shortcomings, a new approach called Adaptive Auto Rate Fallback (AARF), was proposed (Lacage et al., 2004). It is based on the communication history and aims to reduce unnecessary rate variations caused by a misinterpretation of the channel state. Thus, this method controls the time-making process by using the Binary Exponential Backoff (BEB) technique (the same used by the CSMA/CD and CSMA/CA access mechanisms).

Therefore, when a packet transmission fails just after a rate increase, a lower rate is chosen for next transmission attempts. In addition, the number of consecutive successful transmissions n required for the next rate-switching decision will be multiplied by two (with a limit of $n_{max} = 50$). Similar to the old version in a rate decrease caused by two consecutive frames transmission errors, this value is reset to $n_{min} = 10$. The flow diagram in Figure 2 briefly explains the operation of AARF.

Consequently, this new version dynamically controls the number of positive ACKs needed for the rate control. Thus, AARF overcomes the old ARF version in case of a long steady channel conditions by eliminating needless and continuous rate-rising attempts. However it keeps the same disadvantage of the old implementation in case of rapid changes produced on the channel state.

Figure 3 illustrates the behavior of both ARF and AARF approaches for a time period equal to 0.4s needed for 230 data frames. Various physical rates were adopted in this experiment (1, 2, 5.5 and 11Mbps for 802.11b). During this experimentation, we set channel conditions supporting the use of the physical rate R_3 (5.5Mbps) for data transmission. We note that the period between two successive attempts is increased using the AARF technique while the ARF mechanism is trying regularly to increment the current rate to a higher value each ten successive successful transmissions. For example, within the time interval [0.2s, 0.25s] AARF doesn't create any unnecessary rate-switching effort, while ARF carries out three attempts. Likewise, the AARF algorithm considerably has reduced the number of produced errors due to bad decisions (3/4 of errors were removed compared to those generated by the ARF mechanism).

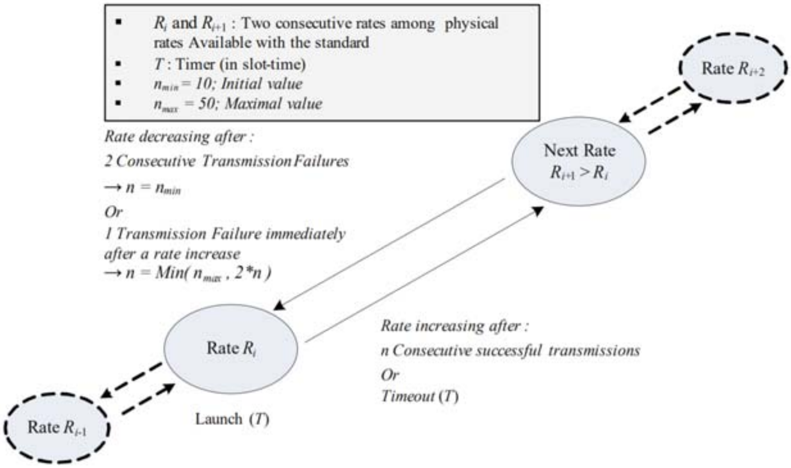

R_i and R_{i+1} : Two consecutive rates among physical rates Available with the standard
T : Timer (in slot-time)
n_{min} = 10; Initial value
n_{max} = 50; Maximal value
Rate decreasing after :
2 Consecutive Transmission Failures
→ $n = n_{min}$
Or
1 Transmission Failure immediately after a rate increase
→ $n = Min(n_{max}, 2*n)$
Rate R_{i+2}
Next Rate $R_{i+1} > R_i$
Rate R_i
Rate increasing after :
n Consecutive successful transmissions
Or
Timeout (T)
Rate R_{i-1}
Launch (T)

Figure 2. The AARF flow diagr

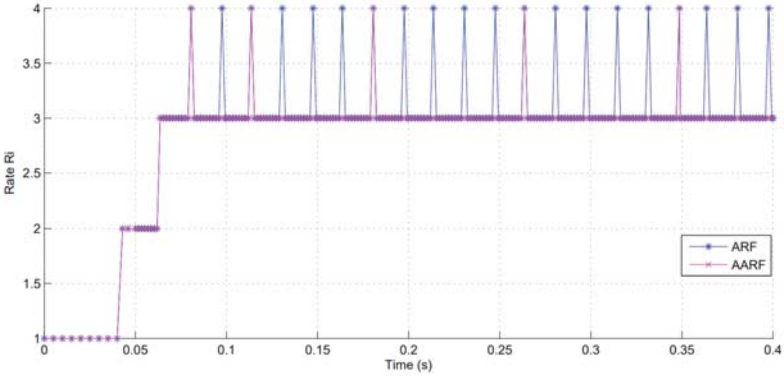


Figure 3. ARF and AARF performance evaluati

Discussion

We have shown in the above study that the actual rate selection algorithms ARF and AARF do not conduct to an accurate decision when the channel is relatively noisy. Despite the given transmission enhancements both models still need improvement and refinement since they cannot react instantly to sudden changes of the channel state. In addition, an interval of time is needed to reach the maximum throughput in case of ‘ideal’ medium condition. Thus, these mechanisms do not represent optimal solutions for the physical link adaptation in noisy and ’ideal’ environments.

Indeed, at a slow channel quality variation AARF is more suitable than ARF as it proceeds to the elimination of unnecessary rate increases. And thereafter, it decreases greatly the number of lost packets while relying on already rate exchanges made previously. However, this improvement is still insufficient since the decision criterion depends only on the nature of acknowledgments (ACKs), whereas this parameter no longer provides sufficient information about the instantaneous channel state. As result AARF need a high latency to reach the maximum throughput. In other words, a negative ACK (or lack of transmission success) is interpreted only by medium quality deterioration. However, this phenomenon may be caused by other networks anomalies (destination not reachable, collision occurred with another data frame, bad CRC, etc.).

It is also observed that when the competing number of stations is high, packet collisions can largely affected the performance of ARF and make ARF operate with the lowest data rate, even when no packet corruption occurs. This is in contrast to the existing assumption that packet collision will not affect the correct operation of ARF and can be ignored in the evaluation of ARF. Therefore, ARF and AARF can only passively react to the signal quality experienced at the receiver. In some occasions, we need to actively improve the signal quality in order to make the transmission rate to meet the traffic demand, even when the link length is a little large. This enhancement will optimize the overall performance and typically will demonstrate a practical effectiveness for multimedia transmissions over Wi-Fi WLANs.

Accordingly, in the next section we propose a new rate adaptation technique to improve the decision based on instantaneous channel conditions while respecting and still complying with the 802.11 standard. In addition, the new approach will be compared with those currently deployed. Simulation results will be also presented to demonstrate the enhancement of the proposed technique compared to those currently presented. Also parameters optimization of the new mechanism will be carried out to be then considered during next scenario simulations.

PROPOSED ADAPTIVE RATE CONTROL TECHNIQUE

The main idea of the proposed method is to introduce a new channel status assessment parameter which cooperates with the number of ACKs to provide an efficient and accurate prediction of instantaneous channel conditions and subsequently to improve the actual rate adjustment mechanism. A logical way to cope with the slow accommodation characteristics of statistics-based feedback methods is to look for methods that use faster feedback, i.e., feedback that quickly provides up-to-date information about the channel status. Such a feedback — the RTT — has been theoretically discussed in (Rebai et al., 2008), but so far, to our knowledge, it has not been used in a practical implementation. We use this RTT measurement in the proposed 802.11 radio to enhance multimedia performance, and also to provide feedback information about the channel conditions that the MAC layer requires. In this section, we first define the new parameter which will be required for the system design. Next, we will describe its implementation and the principle of operation.

Round Trip Time (RTT)

Reliable transport protocols such as Transport Control Protocol (TCP) (Tourrilhes, 2001) were initially designed to operate in 'traditional' and wired networks where packet losses are mainly due to congestion. However, wireless networks introduce additional sorts of errors caused by uncontrolled variations of the medium.

Face to the congestion problems, TCP responds to each loss by invoking congestion control algorithms such as Slow Start, Congestion Avoidance and Fast Retransmission. These techniques have been introduced in different versions of the TCP protocol (TCP Reno, TCP Tahoe, etc.).. These proactive algorithms consist to control the Congestion Window (CW) size based on observed errors. Another TCP-Vegas version has been proposed by (Kumar & Holtzman, 1998; Mocanu, 2004) and rapidly has been adopted by the TCP protocol since it includes an innovative solution designed for preventive systems. In fact, it performs a CW size adjustment based on a fine connection status estimation achieved by a simple measurement of the TCP segment transmission delay. This delay is called Round Trip Time (RTT) and represents (as illustrated in Figure 4) the time period between the instant of issuing a TCP segment by the source noted te and the reception time of the corresponding ACK noted $t_{r.}$

If the measured RTTs will have larger values, the TCP protocol infers network congestion, and reacts accordingly by reducing the congestion window size (symbolized by the number of sent frames and their size). If the values of observed RTTs become smaller, the protocol concludes an improvement on the medium conditions and that the network is not overloaded anymore. Therefore, it proceeds dynamically to increment the CW size, and thus a good operating performance will be achieved based on the new Vegas-version technique.

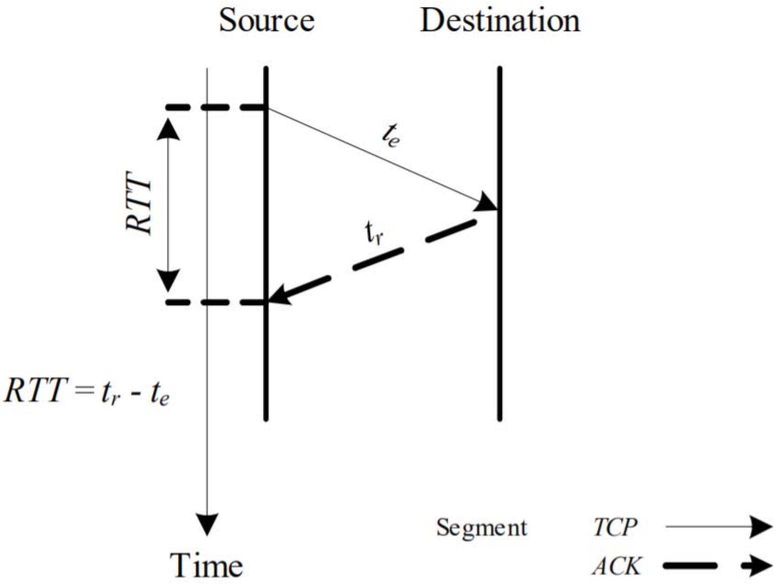


Figure 4. The RTT delay computation

The RTT Parameter Integration

An interesting information and immediate channel observation will be deducted after each data frame transmission by means of RTT measurement and calculation. This feature represents the innovative part of the new control algorithm to adjust the data rate based on the channel capacity. A first integration attempt has been presented in (Rebai et al., 2008) and a rate adaptation design has been proposed. In this chapter, we implement an enhanced mechanism called Modified Adaptive Auto Rate Fallback (MAARF) which aims to predict the medium conditions and minimize the unnecessary loss of data. It chooses the appropriate rate value needed for the next transmission according to the measured RTT value. In fact, it performs a match between the observed value of RTT and the physical rate selection.

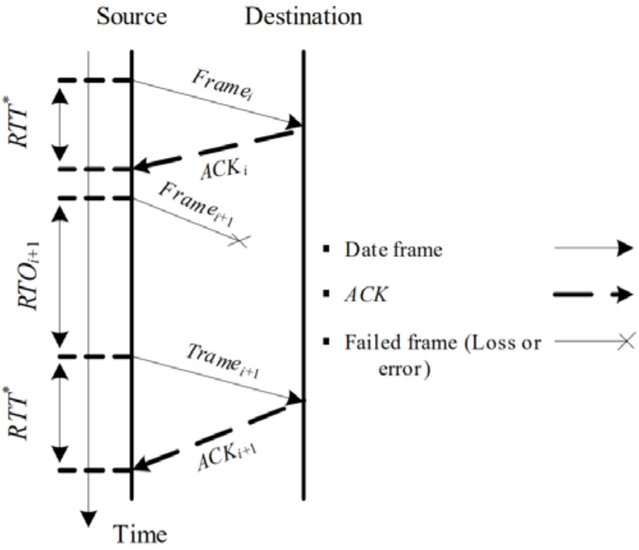


Figure 5. The date frames transmissions

Furthermore we define two types of RTT. The first variety is the observed value directly measured from the channel following the frame sending and called instantaneous RTT denoted by RTT*. The second, denoted by RTT_i, is a theoretical value computed based on the sending rate R and the data frame size. During a successful transmission of a frame i resulting the receipt of the associated ACK_i, a value of RTT* is calculated. We introduce an associated recovery timer, called "Retransmission Time Out" and noted RTO_i, which will detect receipt/loss of a data frame. Based on this parameter, the transmitter detects the loss of a frame i in case of no receipt of the corresponding ACK_i till the expiration of the RTO_i timer. In this case, the last issued frame i will be retransmitted (see Figure 5).

Similarly, we introduce an interval defined by [RTT_i^+ , RTT_i^-] adjacent to the theoretical value RTT_i corresponding to the rate Ri. If the observed value of RTT* belongs this interval (i.e. close enough to the theoretical expected value RTT_i) the channel conditions will be considered insignificant and do not require change of the current rate Ri for next transmissions. In other words, if the value of RTT* belongs the interval [RTT_i^+ , RTT_i^-] the link quality is assumed stable and therefore it is suitable that the MS will transmit using the same current rate Ri since the measured RTT* is considered close to the expected RTT_i value. Outside this window, channel changes are presumed:

- Improved, if the RTT* value is less than RTT_i^+ since the corresponding ACK frame was received earlier than expected and therefore channel conditions had got better.
- Degraded, if the RTT* value is greater than RTT_i^- since the received ACK frame was delayed and then we assume that the risk of data loss increases.

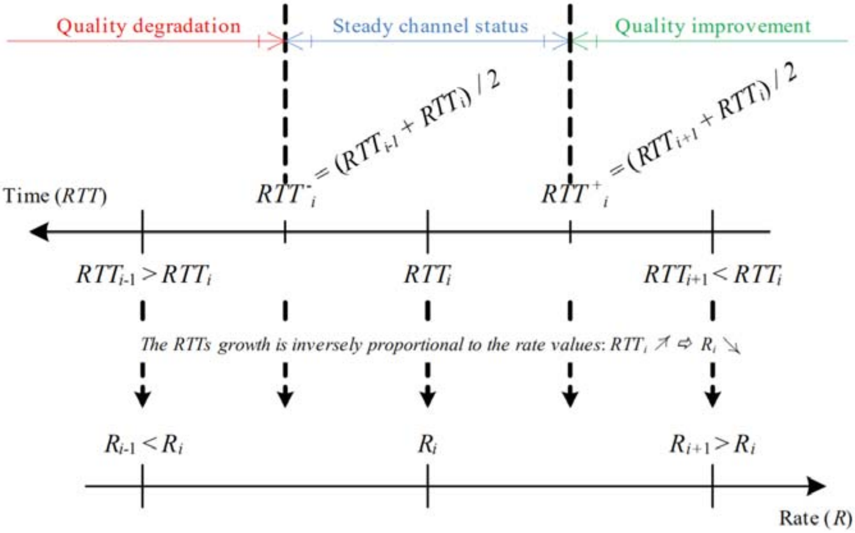


Figure 6. The algorithm parameters set

In both cases, the MS must then change its emission rate and adapt it according to these instantaneous channel state interpretations. We point that $RTT_i^+ < RTT_i^-$ based on the statement $RTT_{i+1} < RTT_i < RTT_{i-1}$ since $R_{i+1} > R_i > R_{i-1}$. The calculation of the parameters values $+ RTT_i$ and $- RTT_i$ is associated to the RTT_i value as defined in the Equation 1 and 2.

$$RTT_i^- = (RTT_{i-1} + RTT_i)/2 \tag{1}$$

$$RTT_i^+ = (RTT_{i+1} + RTT_i)/2 \tag{2}$$

As stated in Figure 6 the RTT_i^- value will be the middle of the interval [RTT_{i-1}, RTT_i], and similarly, RTT_i^+ will have the centre value of the interval [RTT_{i+1} , RTT_i].

Modified Adaptive Auto Rate Fallback (MAARF): Principle of operation

Subsequent to each successful frame transmissions, we compare the variation between instantaneous RTT* and theoretical RTT_i values. More specifically, we test if the value of RTT* has exceeded $RTT_i^{\ +}$ and $RTT_i^{\ -}$ bounds or no. However, the according rate adjustment decision will be taken after several observations of RTT* samples:

- As the number of consecutive successful transmissions did not reach a value of n (required as in AARF algorithm for the next rate-switching decision – it is initially initialized to n_{min}) we perform the following tests:
- If the observed RTT* value is less than $RTT_i^{\ +}$ ($RTT^* < RTT_i^{\ +}$) during h successive transmissions and the maximum speed (54 Mbps) is not yet acquired, then the instant RTT* is considered smaller than the RTT_i value and rather close to the RTT_{i+1} one. Subsequently, the MAARF technique switches to a higher bit rate R_{i+1} starting the next attempt (since an improvement of the channel characteristics was interpreted).
- If the value of RTT* is greater than $RTT_i^{\ -}$ ($RTT^* > RTT_i^-$) for the last g transmission attempts and the lower rate (6 Mbps) is not yet reached, this implies that the instantaneous RTT* value is larger than the expected RTT_i and relatively close to RTT_{i-1}. Thus, MAARF detect an early deterioration of the link quality and therefore we reduce the current rate Ri-1 for future communications.
- If the value of RTT* remains between the two theoretical bounds [$RTT_i^{\ +}$, $RTT_i^{\ -}$] (i.e. $RTT_i^{\ +} < RTT^* < RTT_i^{\ -}$) then the rate will be kept and stay invariant R_i (MAARF assumes a steady state for subsequent network transmissions).
- Similar to the AARF algorithm, when the number of consecutive successful transmissions reaches the desired value (which can be at any given time 10, 20, 40 or 50), we switch to a higher throughput without consideration of the observed RTT* values.

Analogically, when a transmission fails (no acknowledgment received within the RTO_i value) MAARF modifies values of the decision-making parameters (n, h and g) as follows:

If a transmission error is occurred just after a rate increase, it will be then decremented. In addition, as shown in Equation 3 the number of successful transmissions n that should be attained for the next rise will be doubled with a limit value equal to n_{max}.

$$n = Min(2 * n ; n_{\max}) \quad (3)$$

- If two consecutive errors are detected the MAARF mechanism reduces the current rate, while resetting the value of successful transmissions to the minimum one ($n = n_{min}$) for the next rising attempt.
- The same backoff control technique used for the parameter n adaptation is employed as well for the parameters h and g adjustment. In fact, these two variables will be dynamically adapted and will vary between the upper and lower limits to maintain a rigorous decision to increment/decrement the data rate.
- When a transmission error occurs just after a rate increase decision caused by an interpretation of the RTT* value, the current rate will be reduced and the h value (as shown in Equation 4) will be multiplied by two as the upper limit h_{max} is not reached.

$$h = Min(2 * h ; h_{\max}) \quad (4)$$

In other words, during successful transmissions the condition ($RTT^* < RTT_i^+$) should verified using the new value of h to be able to increment the rate.

Likewise, if a transmission error was detected immediately after a rate decrease decision based on comparisons between RTT* and RTT_i^- values. Then the rate will be raised to its old value and the responsible g parameter (see Equation 5) will be doubled if its value does not reach g_{max}.

$$g = Min(2 * g ; g_{\max}) \quad (5)$$

- This means that the condition $RTT^* > RTT_i^-$ should be established using the new value of g during subsequent transmissions to be able to decrease again the rate.
- Both of the above parameters will also be reset identically to the parameter n after two consecutive transmission failures as follows:

$$h = h_{\min} ; g = g_{\min}$$

The MAARF setting

In this section we detail the new parameters designed for the MAARF algorithm. The IEEE 802.11 standard defines within its 802.11a and 802.11g versions, different physical rate values which can reach 54Mbps. Thus, we setup:

- R_i: the current data rate that varies from the following shown values {6, 9, 18, 12, 24, 36, 48, 54}Mbps.
- RTT* : is the observed value when sending a frame (measured from the transmission channel after receiving the corresponding ACK).
- RTT_i: the theoretical time computed between the frame sending time to the ACK receipt instant. It reflects the channel occupation and does not include the waiting time to access the medium by the transmitter. It is given by Equation 6 as follows:

$$RTT_i = t_{em.Frame} + t_{propag} + t_{treat.Receiver} + SIFS + t_{em.ACK} + t_{propag} + t_{treat.Emitter} \tag{6}$$

- with, t_{em} is the emission time (Data Frame or ACK), t_{propag} is the propagation time over the transmission medium and t_{treat} is the treatment time of each received frame

In practice, this value will be represented only by the data frame transmission delay as shown in Equation 7. This approximation is made because of the negligibility of the other delays compared to the chosen value.

$$RTT_i \approx t_{em.Frame} = \frac{Frame.size}{R_i} \tag{7}$$

- RTO_i (Retransmission Time Out): is a recovery controlling timer after a frame loss. Its value is assigned based on RTT_i (see Equation 8).

$$RTO_i = 2 * RTT_i \tag{8}$$

- RTT_i^+ and RTT_i^- : the two decisional parameters (the RTT_i borders) which their values are chosen for each used rate Ri as defined in Equations 1 and 2.
- h: is the rate-increase responsible variable and it belongs to the interval $[h_{min}, h_{max}] = [4, 16]$.

- g: is the rate-decrement responsible variable and it belongs to the interval $[g_{min}, g_{max}] = [2, 8]$.
- n: is the already used parameter by the AARF technique. It represents the number of successive successful transmissions and belongs to the interval $[n_{min}, n_{max}] = [10, 50]$.

Finally, we illustrate the detailed MAARF functioning in Figure 7.

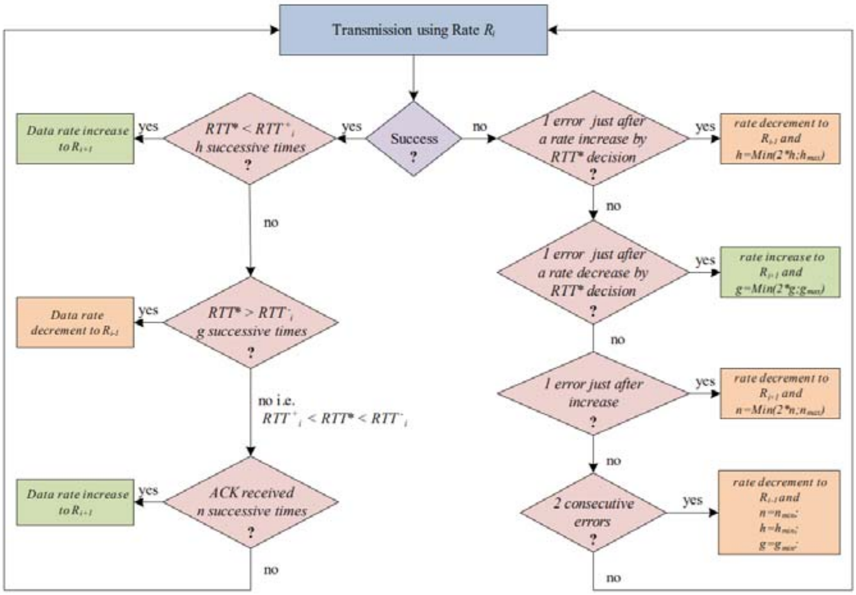

Transmission using Rate R_i
Data rate increase to R_{i+1}
yes
$RTT^* < RTT^+_i$ h successive times ?
yes
Success ?
no
1 error just after a rate increase by RTT* decision ?
yes
rate decrement to R_{i-1} and h=Min(2*h;h_{max})
no
no
1 error just after a rate decrease by RTT* decision ?
yes
rate increase to R_{i+1} and g=Min(2*g;g_{max})
Data rate decrement to R_{i-1}
yes
$RTT^* > RTT^-_i$ g successive times ?
no
1 error just after increase ?
yes
rate decrement to R_{i-1} and n=Min(2*n;n_{max})
no i.e. $RTT^+_i < RTT^* < RTT^-_i$
no
Data rate increase to R_{i+1}
yes
ACK received n successive times ?
2 consecutive errors ?
yes
rate decrement to R_{i-1} and $n=n_{min}$; $h=h_{min}$; $g=g_{min}$
no
no

Figure 7. The Transition diagram of the new MAARF algorithm

RESULTS AND PERFORMANCE EVALUATION

The algorithms were implemented using the C language on a Unix based operating-system environment (gcc/terminal MAC) to be then easily integrated into the network simulator.

We conducted various tests using the following configuration:

- The number of sent frames is 100 frames (approximately 0.5 seconds).
- The size of each data frame is equal to the 802.11g minimum frame size (=1200 bytes).
- An initial data rate of Ri is 6Mbps (up to 54Mbps).

- Failure of an ACK return reflects a transmission failure: packet loss, RTO expired or error detected by the CRC.
- A returned ACK by the receiver indicates a successful transmission only if it is received before the RTO expiration.
- The current value of RTT (RTT*) is read/measured after each ACK reception.

Several scenarios have been considered to evaluate the performance of the proposed algorithm compared to the other versions (ARF and AARF).

Optimization of Algorithm Parameters

This first experiment is designed to study and optimize the decision-making parameters of the new algorithm: h (counting the number of successive times in which $RTT^* < RTT_i^+$) and g (reflecting the number of consecutive times that $RTT^* > RTT_i^-$). We discuss the values of h_{min} and g_{min}. In Figure 8, we show the implementation results of different MAARF algorithm configurations for various parameters values. These results express the chosen physical rate for each transmitted frame in the network.

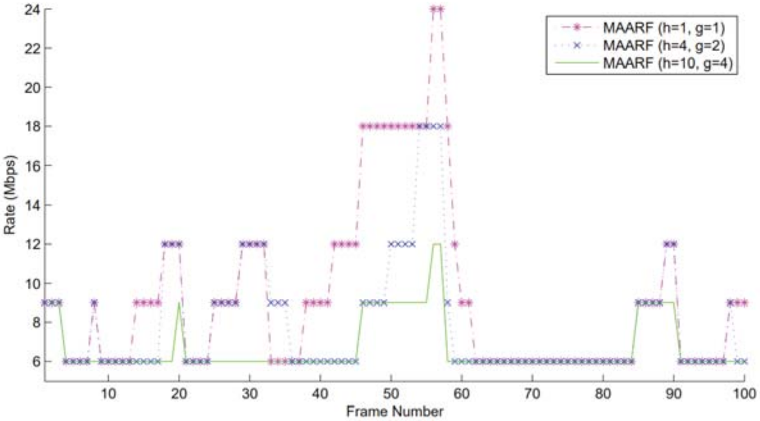


Figure 8. Rate adaptation for different MAARF configurations

We note that by choosing low values of g and h ($h_{min} = g_{min} = 1$), MAARF makes quick decisions to increment and decrement the physical rate. In fact, it becomes sensitive for channel variations and adapts sinusoidal regime. On the other hand, by choosing large initial values of the g and h parameters (h_{min}=10 and g_{min}=4) the algorithm does not respond effectively to significant quality deviations and reacts as AARF. Therefore, we point out that the best initial and rigorous values of g and h with which MAARF gives the best results are respectively 4 and 2.

Test regimes

Unbalanced channel state

We compare now the new scheme against the AARF technique (currently used by the 802.11 WLANs) during unstable channel conditions (random improvement/degradation of the medium state). In Figure 9, we present the corresponding results graph and we clearly notice an efficient reaction of the MAARF technique against channel changes. In fact, the new algorithm detects faster the medium availability by adjusting its physical rate value starting from the 4th frame, while AARF reacts only from the 10th frame. We also note the MAARF ability to dynamically respond against medium interferences dissimilar to the AARF mechanism. In addition, we conclude a significant improvement that has been reached (about 26%) regarding the mean value of recorded rates. In fact, we measure 6.1Mbps and 7.6Mbps respectively for the AARF and MAARF techniques.

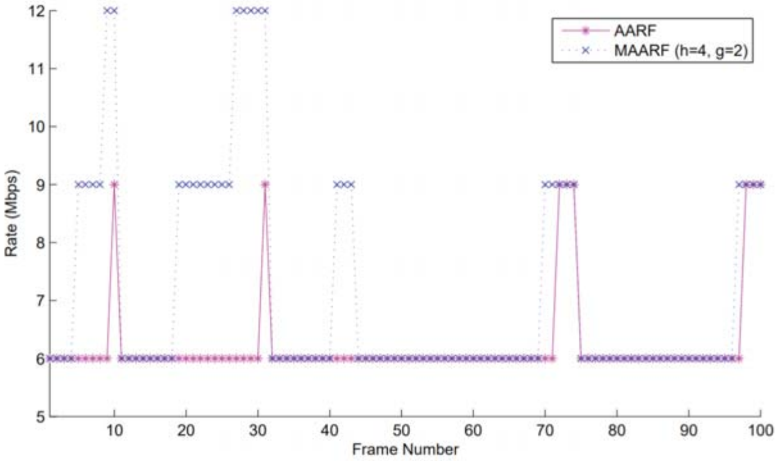


Figure 9. Rate adaptation in transitory regime

Steady channel state

We assume in this case that only positive acknowledgments will be returned to the transmitter following the frame sending (packet transmissions without losses). Thus RTT* values recorded from the medium will be close to those of the theoretical corresponding RTT_i values.

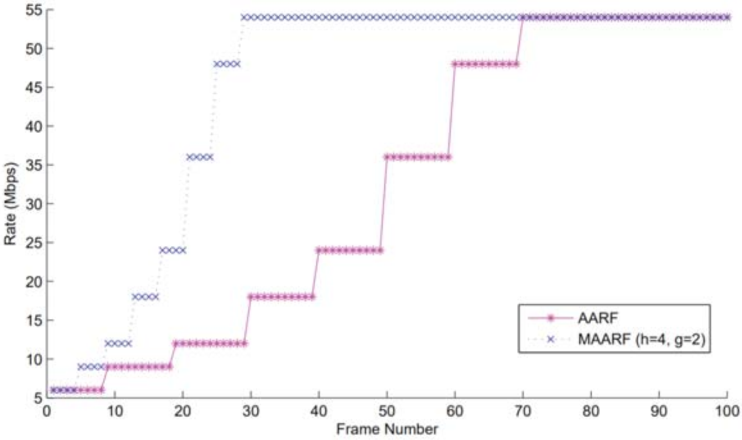


Figure 10. Rate adaptation within steady regime

This first scenario shows a huge variation in terms of selected physical data rate and the overall mean value between the two techniques. In fact, a clear throughput enhancement is obtained (41%) since we traced as a mean rate value during the simulation time, 32.04Mbps for the AARF and 45.48Mbps for the new algorithm. According to the results in Figure 10, the maximum data rate (54Mbps) is reached earlier by the new algorithm as it detects the channel condition improvement (from frame No. 28) and thus takes advantage of the large possible rate values. However the AARF technique reaches the maximum rate later on (only from the frame No. 70). This was caused by the fact that AARF is required to wait at least 10 positive acknowledgments at each rate hop.

Mobile Environment Situation

A fourth simulation on the rate adaptation is conducted within a variable channel regime. In the case of 802.11 WLAN the medium quality variations are very fast and totally random. This is reflected by intervals where the channel conditions improve rapidly, separated by those where the medium state deteriorates suddenly. We note from Figure 11 that the new technique adapts the same rate as obtained by AARF; however MAARF is more agile and predictive of the medium communication conditions for the data rate rise decision. When transmission errors take place, both methods pass at a lower rate almost at the same time.

The average rate value obtained for both AARF and MAARF mechanisms is equal, respectively, to 6.72Mbps and 7.89Mbps. As a result, an improvement of 17% was reached due to the responsive capability and the fast adaptability of the new link control mechanism.

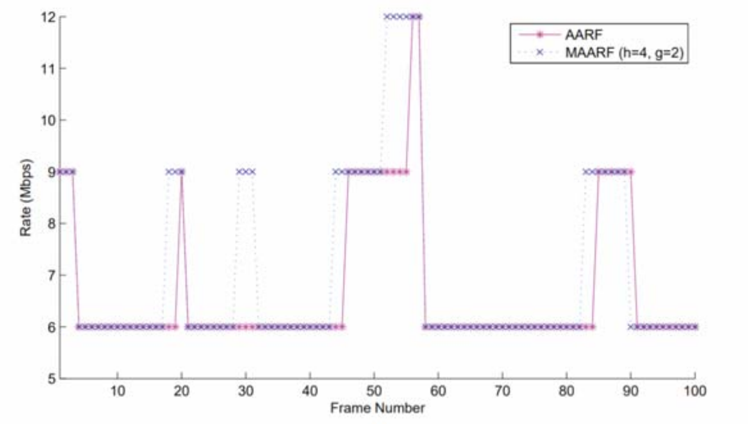


Figure 11. Rate adaptation for instantaneous and unpredictable channel conditions

Network simulations under NS-2 platform

The results obtained during the new MAARF algorithm implementation have shown that it is possible:

- To estimate the channel conditions through the observed RTT* values.
- To detect/avoid packet losses before they happen.
- To take the necessary decisions faster than current mechanisms.

We have revealed in this study that it is no longer necessary to wait 10 or more consecutive ACKs to adjust the theoretical rate as it was deployed by classical algorithms. We compare in Figures 12 and 13 the results obtained by applying the new MAARF mechanism and the current AARF for the same channel conditions. These results are reflecting, respectively, the computed rate mean value and the number of frames errors depending on the transmitted packets number.

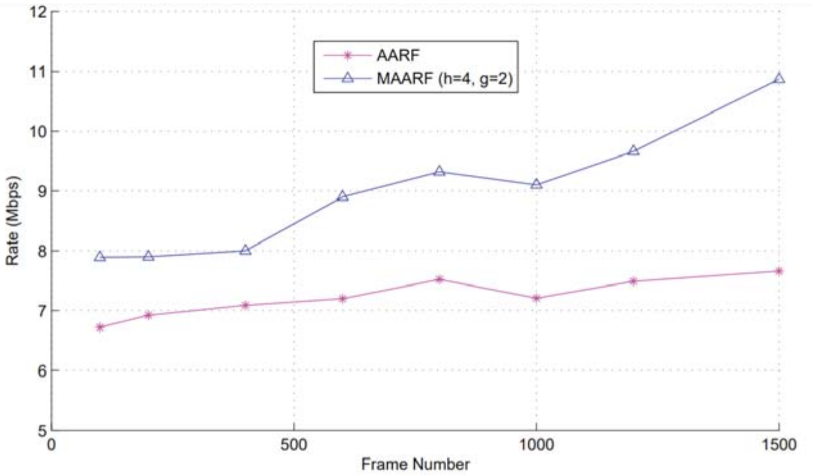
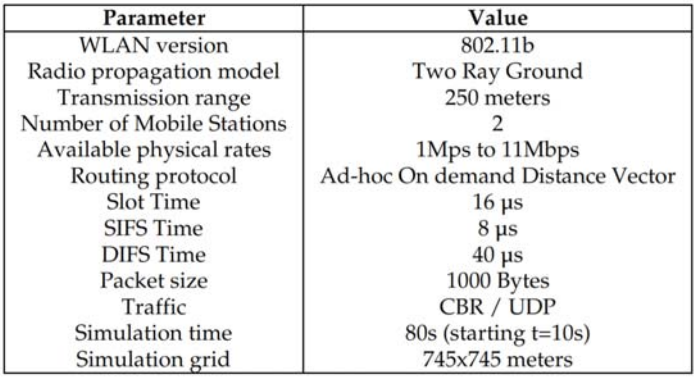


Figure 12. Observed throughput for both AARF and MAARF mechanisms

Table 1. NS-2 simulation parameters

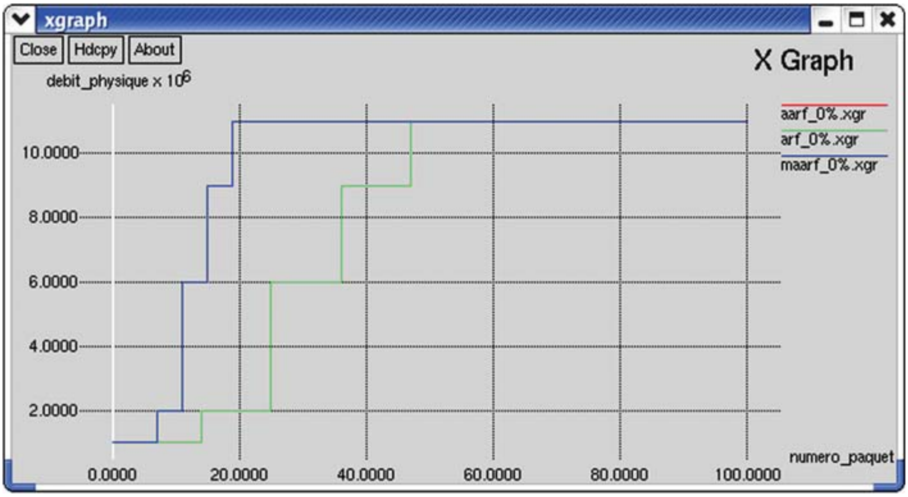


Parameter	Value
WLAN version	802.11b
Radio propagation model	Two Ray Ground
Transmission range	250 meters
Number of Mobile Stations	2
Available physical rates	1Mps to 11Mbps
Routing protocol	Ad-hoc On demand Distance Vector
Slot Time	16 µs
SIFS Time	8 µs
DIFS Time	40 µs
Packet size	1000 Bytes
Traffic	CBR / UDP
Simulation time	80s (starting t=10s)
Simulation grid	745x745 meters

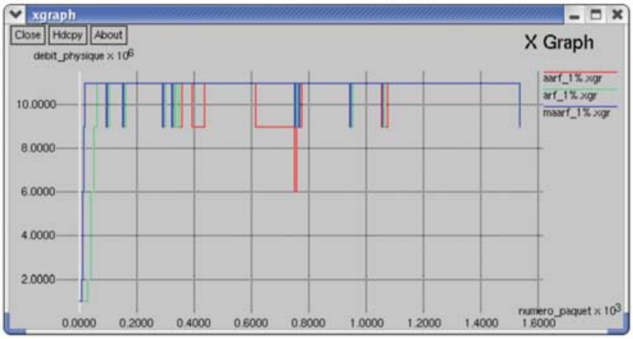
We show based on the conducted experimentations, the improvements are distinguished and very clear in terms of the overall throughput and packet errors. We also confirm the MAARF algorithm performance by simulating the new technique on the Network Simulator NS-2 platform (Network Simulator-II, 1998). Figure 14 outlines various simulation arrangements performed on NS-2. We easily confirm the initial results by varying the bit error rate. The simulation parameters are summarized in Table 1.

First we note that the chosen traffic for the carried simulations was Constant Bit Rate (CBR) over the Transport-layer User Datagram Protocol (UDP). In fact, the CBR service category is used for connections that transport traffic at a constant bit rate, where there is an inherent reliance on time synchronization between the traffic source and destination. CBR is tailored for any type of data for which the end-systems require predictable response time and a static amount of bandwidth continuously available for the life-time of the connection. These applications include services such as video conferencing, telephony (voice services) or any type of on-demand service, such as interactive voice and audio.

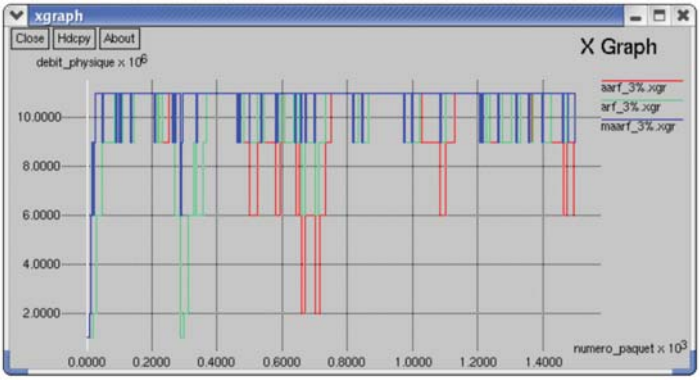
The obtained results verify the initial theoretical observations and validate the efficiency and adaptability of the new mechanism for both slow and rapid fluctuations of the transmission channel quality. In absence of errors (as shown in Figure 14.a.), MAARF reacts quickly and rises the higher allowed rate before the current techniques. While varying the Bit Error Rate (BER) value during simulation scenarios (Figures 14.b., 14.c. and 14.d.) the physical rate adjustment corresponding to MAARF is more suitable and faster than other tested algorithms. This outcome is clearly confirmed in Figures 14.e. and 14.f. by measuring the achieved throughput for two different BER values. Accordingly, we note an enhanced response from MAARF against the channel quality variations compared to the other two techniques.



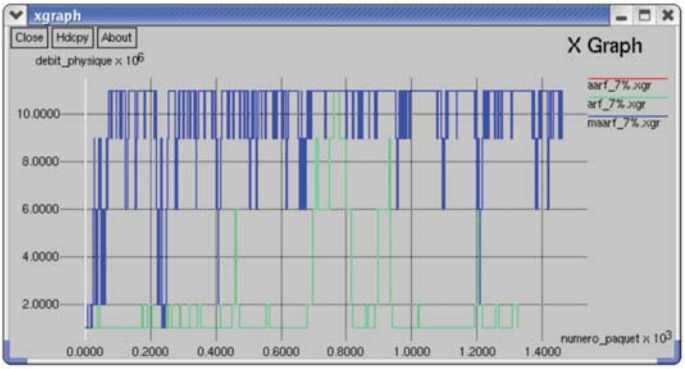
a. Physical rate tuning for ideal channel



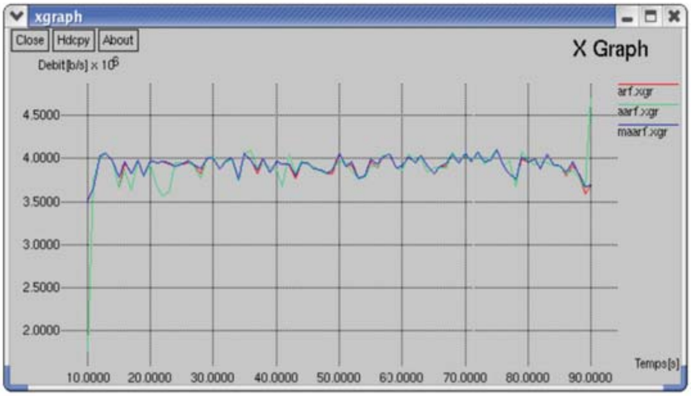
b. Physical rate tuning with BER=1%



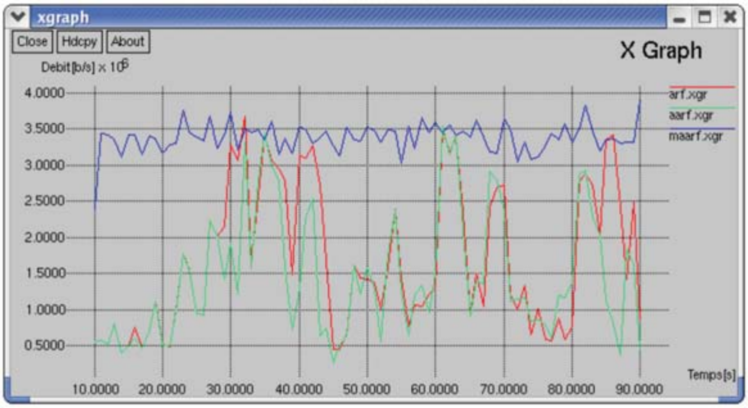
c. Physical rate tuning with BER=3%



d. Physical rate tuning with BER=7%



e. Measured throughput with BER=1%



f. Measured throughput with BER=5%

Figure 14. Experimental NS-2 simulation results by varying the BER value

CONCLUSIONS

General Remarks

The IEEE 802.11 standard defines several MAC-level parameters (settings) that are available for tuning at the side of the Wireless Network Interface Card (NIC). The most important parameter available for adjustment is the transmit rate. Each rate corresponds to a different modulation scheme with its own trade-off between data throughput and distance between the

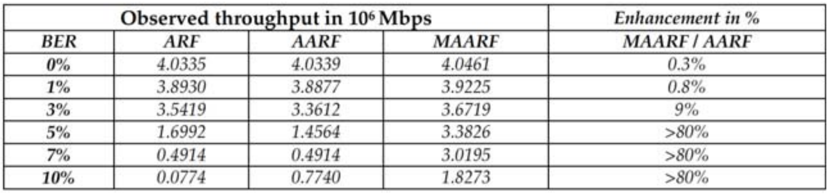
MSs. In order to decide which rate is optimal at each specific moment, a control algorithm needs information about the current link conditions. Since it is difficult to get this information directly, most of the MAC rate-control algorithms use statistics-based feedbacks, for example, ACK count. We have shown, through a deep study of the currently used rate control mechanisms, that the main disadvantage of this indirect feedback is that it is inherently slow, causing communication failures when the link conditions degrade rapidly (e.g., when the user moves fast). The short-term dropouts are normally handled by frame retransmissions. This is acceptable for download applications whose key requirement is a flagrant data throughput. However it leads to a significant increase in (average) packet delay and in the jitter due to the variations in the number of retransmissions. Streaming applications are very sensitive to long packet delays and high jitter, and less sensitive to the overall throughput (of course when this throughput is larger than the minimum value required by the application). Consequently, streaming applications achieve poor performances under a standard rate control (like ARF and AARF techniques).

Hence, we have proposed a new algorithm noted Modified Auto Rate Fallback (MAARF). This technique implements a new decisional variable called Round Trip Time (RTT) which complies and cooperates with the basic parameter (number of returned ACKs). This new parameter is designed to make a good estimate of the instantaneous channel quality (observation of the channel state after each transmitted frame), and choose the adequate rate accordingly.

Based on the simulation results we have shown a remarkable improvement in data throughput and physical rate control. In fact, the proposed MAARF mechanism provides higher values (about 17% to 58%) in comparison with those resulting from conventional algorithms. Table 2 presents an observed throughput summary of the MAARF scheme compared to existing algorithms (ARF and AARF) and gives an overview on the rate control enhancement for different BERs. The overall throughput observed within MAARF is much higher than other mechanisms when the BER reaches high values (12 times when the error rate is 10%).

In conclusion simulation experiments were performed on the new dynamic time-based link adaptation mechanism and the corresponding results have shown the quality improvement on the transmission link. The results also demonstrated that the proposed mechanism outperforms the basic solution in terms of providing support to both acknowledgment based and time-based rate control decisions. Therefore MAARF meets the desired objectives by being able to reduce errors resulting from bad rate adjustment and then satisfy the transmission of multimedia applications in terms of required QoS.

Table 2. Enhancement ratio of the MAARF technique



Observed throughput in 10^6 Mbps				*Enhancement in %*
BER	*ARF*	*AARF*	*MAARF*	*MAARF / AARF*
0%	*4.0335*	*4.0339*	*4.0461*	*0.3%*
1%	*3.8930*	*3.8877*	*3.9225*	*0.8%*
3%	*3.5419*	*3.3612*	*3.6719*	*9%*
5%	*1.6992*	*1.4564*	*3.3826*	*>80%*
7%	*0.4914*	*0.4914*	*3.0195*	*>80%*
10%	*0.0774*	*0.7740*	*1.8273*	*>80%*

The MAARF Working Out Performance for Real-Time Video Streaming

In the context of the Voice-over-IP and real-time video streaming (Video conferencing) which represent the most end-user demanded multimedia streaming applications to be supported over wireless connections, the proposed technique is supposed to enhance the video quality transmission by keeping the same compression degree, and so, avoiding the cross layer design implemented by (Haratcherev et al., 2005) which involves an interaction between the application and MAC layers by assuming that the video encoder can also adapt to the link quality by changing the compression degree for example, and thus modifying the data rate. This so-called Coupled scheme is based on a cross-layer signaling by letting the rate control loops - of the MAC-layer and the Video coder - be mutually associated. In such way, the video coder will poll frequently for the current and predicted rate. For performance evaluation purpose, the same real experiment as performed in (Haratcherev et al., 2005) was carried out by streaming a video file between two laptops, both running Linux. The 802.11a cards used are based on the Atheros AR5000 chipset, and the card driver implements the discussed rate control algorithms. One of the laptops had a fixed position and the other one is following a predetermined track. The track consists of three parts: reaching from the room to a specific start position in the hallway, and waiting until certain time elapses (10s). Then keep moving with the laptop three times up and down the hallway (60s). Finally we back again into the room, where the fixed laptop lies (20s). We have evaluated four cases: ARF, AARF, MAARF and a Coupled version of the hybrid rate control (responsive MACadaptation using radio-SNR and packet loss statistics) as described in

(Haratcherev et al., 2004). Each experiment took 90 seconds and we have compared the quality of the received videos using the Peek-Signal-to-Noise-Ratio (PSNR), a commonly used metric in video compression.

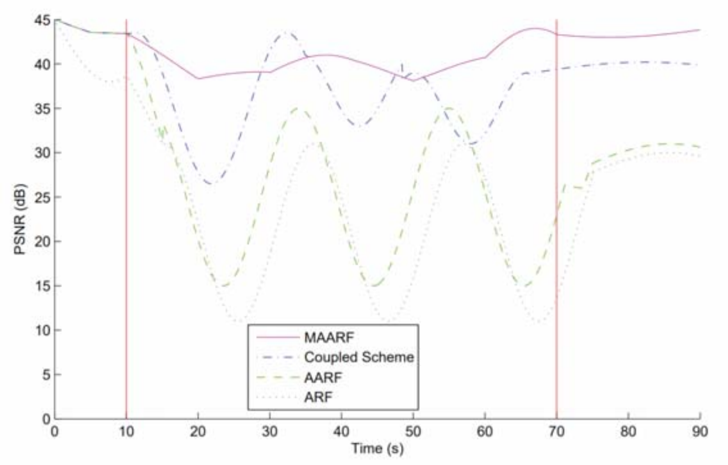
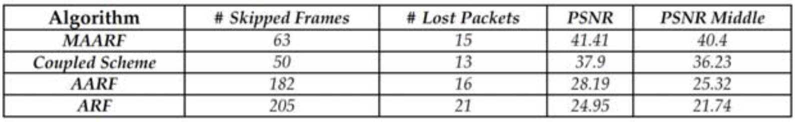


Figure 15. PSNR measurements during the 90s real-time video streaming experiments

Table 3. Summary of algorithms results



Algorithm	# Skipped Frames	# Lost Packets	PSNR	PSNR Middle
MAARF	63	15	41.41	40.4
Coupled Scheme	50	13	37.9	36.23
AARF	182	16	28.19	25.32
ARF	205	21	24.95	21.74

In Figure 15 the quality (PSNR) is shown for the whole experiment (90s). In the left part (0−10s) the channel conditions are excellent since a high quality was fulfilled with all cases. The middle part is best described as having continuous variation. The right part has good channel conditions again. As we can notice, the MAARF case has a higher PSNR in almost all cases. Table 3 summarizes the number of skipped frames (by the encoder), the number of lost packets, the average PSNR and the average PSNR in the unstable period (between 10 and 70s). Although the lost packets counts are not the same in the four cases, it does not justify the difference in PSNR values. In the ARF and AARF cases, the total number of skipped frames is much higher, as we expected. Looking at the average PSNR in the 10−70s period, we conclude that MAARF outperforms the other techniques based on an enhanced channel assessment. As a result, the proposed video streaming experiments better illustrate the improvement made using MAARF by reducing the number of lost packets and skipped frames. Both effects resulted in a high visual quality opposed to the cases of ACK-based rate control algorithms. Meanwhile the Coupled cross-layer signaling scheme had also very encouraging results in terms of packets loss and skipped frames which let us consider a warning concept between the video encoder and the MAARF decision in the MAC layer. Therefore, a second version of the MAARF technique including a cross-layer signaling solution will be investigated in the next research step. This extended adaptation would be able to further avoid packet losses and especially to prevent skipped frames of the emerging voice/video streaming services. Of course, it would have an adverse effect on the visual quality of the decoded video however we will need to further carry out real streaming video experiments over a wireless 802.11 link between MSs to conclude if the cross-layer signaling between the MAC-layer and the video coder will lead to a visual quality increase in term of measured PSNR.

REFERENCES

1. Angrisani, l.; Napolitano, A. & Sona, A. (2011). VoIP over WLAN: What about the Presence of Radio Interference?, VoIP Technologies, Shigeru Kashihara (Ed.), ISBN: 978-953-307-549-5, InTech, pp. 197-218, Feb. 2011
2. An-Chih, L.; Ting-Yu, L. & Ching-Yi, T. (2009). ARC: Joint Adaptation of Link Rate and Contention Window for IEEE 802.11 Multi-rate Wireless Networks, The 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks SECON '09, Rome, Italy, July 2009.
3. Chen, X.; Wan, Y. & Lu, J. (2010). Cross-layer link rate adaptation for high performance multimedia broadcast over WLANs, in Proc. IEEE GLOBECOM Workshops GCWorkshops'10, pp. 965-969, Miami, FL, USA, Dec. 2010
4. Chevillat, P.; Jelitto, J.; Barreto, A. N. & Truong, H. (2003). A Dynamic Link Adaptation Algorithm for IEEE 802.11a Wireless LANs, in Proc. IEEE International Conference on Communication ICC'03, Anchorage, AK, May 2003
5. Chiapin W. & Tsungnan L. (2008). A Cross-Layer Link Adaptation Algorithm for IEEE 802.11 WLAN with Multiple Nodes, in Proc. IEEE Asia-Pacific Services Computing Conference APSCC '08, pp. 1161-1167, Yilan, Taiwan, Dec. 2008
6. del Prado Pavon, J. & Choi, S. (2003). Link Adaptation Strategy for IEEE 802.11 WLAN via Received Signal Strength Measurement, in Proc. IEEE International Conference on Communication ICC'03, Anchorage, AK, May 2003
7. Galtier, J. (2011). Adaptative Rate Issues in the WLAN Environment, Advances in Vehicular Networking Technologies, Miguel Almeida (Ed.), ISBN: 978-953-307-241-8, InTech, pp. 187-201, April 2011
8. Habetha, J. & de No, D. C. (2000). New Adaptive Modulation and Power Control Algorithms for HIPERLAN/2 Multihop Ad Hoc Networks, in Proc. European Wireless (EW'2000), Dresden, Germany, Sept. 2000
9. Haratcherev, I.; Langendoen, K.; Lagendijk, I. & Sips, H. (2004). Hybrid Rate Control for IEEE 802.11. In ACM International Workshop on Mobility Management and Wireless Access Protocols MobiWac'04, pp. 10–18, Philadelphia, USA, Oct. 2004

10. Haratcherev, I.; Taal, J.; Langendoen, K.; Lagendijk, I. & Sips, H. (2005). Fast 802.11 link adaptation for real-time video streaming by cross-layer signaling, In Proc. IEEE International Symposium on Circuits and Systems ISCAS'05, pp. 3523-3526, Vol. 4, Kobe, May 2005

11. IEEE Standard 802.11 (1999), Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11, 1999

12. IEEE Standard 802.11a (1999), Supplement to Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz Band, IEEE Std. 802.11a-1999, Sept. 1999

13. IEEE Standard 802.11b (1999), Supplement to Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-speed Physical Layer Extension in the 2.4 GHz Band, IEEE Std. 802.11b-1999, 1999

14. IEEE Standard 802.11g (2003), Supplement to Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Further Higher Data Rate Extension in the 2.4 GHz Band, IEEE Std. 802.11g-2003, June 2003

15. Jianhua, H.; Kaleshi, D.; Munro, A. & McGeehan, J. (2006). Modeling Link Adaptation Algorithm for IEEE 802.11 Wireless LAN Networks, The 3rd International Symposium on Wireless Communication Systems ISWCS '06, pp. 500-504, Sep. 2006

16. Junwhan, K. & Jaedoo, H. (2006). Link Adaptation Strategy on Transmission Rate and Power control in IEEE 802.11 WLANs, in Proc. IEEE Vehicular Technology Conference VTC'06, Montreal, QC, Canada, Sept. 2006

17. Kamerman, A. & Monteban, L. (1997). WaveLAN-II: A High-Performance Wireless LAN for the Unlicensed Band, Bell Labs Technical Journal, pp. 118–133, Summer 1997

18. Krishnan, M. N. & Zakhor, A. (2010). Throughput Improvement in 802.11 WLANs Using Collision Probability Estimates in Link Adaptation, in Proc. IEEE Wireless Communications and Networking Conference WCNC'10, Sydney, Australia, April 2010

19. Kumar, A. & Holtzman, J. (1998). Performance analysis of versions of TCP in WLAN, Indian Academy of Sciences Proceedings in

Engineering Sciences, Sadhana, Feb. 1998

20. Lacage, M.; Manshaei, M. H. & Turletti, T. (2004). IEEE802.11 Rate Adaptation: A Practical Approach. INRIA Research Report, number 5208, May 2004
21. Mocanu, S. (2004). Performance Evaluation of TCP Reno and Vegas. Technical report CR256, Laboratoire d'Automatique de Grenoble, Département Télécom, July 2004 Network Simulator II (1998), ns-2, available from http://www.isi.edu/nsnam/ns/
22. Pang, Q.; Leung, V.C.M. & Liew, S.C. (2005). A rate adaptation algorithm for IEEE 802.11
23. WLANs based on MAC-layer loss differentiation, in Proc. International Conference on Broadband Networks BroadNets'05, pp. 659-667, Vol. 1, Boston, MA, USA, Oct. 2005
24. Qiao, D.; Choi, S. & Shin, K. G. (2002). Goodput Analysis and Link Adaptation for IEEE 802.11a Wireless LANs, IEEE Transactions on Mobile Computing, vol. 1, pp. 278–292, Oct. 2002
25. Qiao, D. & Choi, S. (2005). Fast-Responsive Link Adaptation for IEEE 802.11 WLANs, in Proc. IEEE International Conference on Communication ICC'05, Sept. 2005
26. Rebai, A. R.; Alnuweiri, H.; Hanafi, S. (2009). A novel prevent-scan Handoff technique for IEEE 802.11 WLANs, in Proc. International Conference on Ultra Modern Telecommunications & Workshops ICUMT '09, St. Petersburg, Russia, Oct. 2009
27. Rebai, A. R.; Fliss, M.; Jarboui, S. & Hanafi, S. (2008). A new Link Adaptation Scheme for IEEE 802.11 WLANs, in Proc. IEEE New Technologies, Mobility and Security NTMS'08, Nov. 2008
28. Rebai, A. R.; Haddar, B.; Hanafi, S. (2009). Prevent-scan: A novel MAC layer scheme for the IEEE 802.11 handoff, In Proc. International Conference on Multimedia Computing and Systems ICMCS '09, pp. 541-546, Ouarzazate, Morocco, April 2009
29. Rebai, A. R.; Hanafi, S.; Alnuweiri, H. (2009). A new inter-node priority access enhancement scheme for IEEE_802.11 WLANs, in Proc. International Conference on Intelligent Transport Systems Telecommunications ITST'09, pp. 520-525, Lille, France, Oct. 2009
30. Rebai, A. R.; Rebai, M. F.; Alnuweiri, H.; Hanafi, S. (2010). An enhanced heuristic technique for AP selection in 802.11 handoff procedure, In

Proc. IEEE International Conference on Telecommunications ICT'10, pp. 576-580, Doha, Qatar, April 2010

31. Sangman, M.; Moonsoo, K. & Ilyong, C. (2011). Link Quality Aware Robust Routing for Mobile Multihop Ad Hoc Networks, Mobile Ad-Hoc Networks: Protocol Design, Xin Wang (Ed.), ISBN: 978-953-307-402-3, InTech, pp. 201-216, Jan. 2011

32. Shun-Te, W.; Wu, J.-L.C.; Chung-Ching, D. & Chun-Yen, H. (2007). An adaptation scheme for quick response to fluctuations in IEEE 802.11 link quality, in Proc. IEEE Region 10 Conference TENCON'07, Taipei, Oct. 2007

33. Tourrilhes, J. (2001). Fragment Adaptive Reduction: Coping with Various Interferers in Radio Unlicensed Bands", in Proc. IEEE International Conference on Communication ICC'01, Helsinki, Finland, July 2001

34. Xin, A.; Shengming, J. & Lin T. (2010). Traffic-aware active link rate adaptation via power control for multi-hop multi-rate 802.11 networks, in Proc. IEEE International Conference on Communication Technology ICCT'10, pp. 1255-1259, Nanjing, Nov. 2010

CHAPTER 12

A MULTIMEDIA AND VOIP-ORIENTED CELL SEARCH TECHNIQUE FOR THE IEEE 802.11 WLANS

Ahmed Riadh Rebai[1], Mariam Fliss[1] and Saïd Hanafi[2]

[1]Texas A&M University at Qatar – Doha, Qatar
[2]University of Valenciennes et du Hainaut-Cambrésis, France

INTRODUCTION

With the development and widespread of diverse wireless network technologies such as wireless local area networks (WLANs), the number of mobile internet users keeps on growing. This rapid increase in mobile internet users records a phenomenal growth in the deployment of the IEEE 802.11 WLANs (IEEE Std 802.11, 1999) in various environments like universities (Corner et al., 2010; Hills & Johnson, 1996), companies,

Citation: Ahmed Riadh Rebai, Mariam Fliss and Saïd Hanafi (March 7th 2012). A Multimedia and VoIP-Oriented Cell Search Technique for the IEEE 802.11 WLANS, Multimedia - A Multidisciplinary Approach to Complex Issues, Ioannis Karydis, IntechOpen, DOI: 10.5772/36801.

Copyright: © 2012 by authors and IntechOpen. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

shopping centers (Bahl et al., 2001) and hotels. This category of networks then will be the underlying basis of ubiquitous wireless networks by decreasing infrastructure costs and providing stable Internet connectivity at anytime and anywhere (Kashihara et al., 2011; Kunarak & Suleesathira, 2011). Hence many believe that are expected to be part of the integrated fourth generation (4G) network. At the same time, voice over IP (VoIP) is expected to become a core application in the ubiquitous wireless networks, i.e., the next generation cell-phone. Recently, many users have easily used VoIP communication such as Skype (Skype, 2003) in wireless networks. However, in the mobility context WLANs become not appropriate to the strict delay constraints placed by many multimedia applications, and Mobile Stations (MSs) cannot seamlessly traverse the 802.11 Access Points (APs) during VoIP communication due to various factors such as the inherent instability of wireless medium and a limited communication area. MSs are required to find and then to associate with another AP with acceptable signal quality whenever they go beyond the coverage area of the currently associated AP. The overall process of changing association from one AP to another is called handoff or handover process and the latency involved in the process is termed as handoff latency.

Thus, even if an MS can avoid communication termination at handoff, the following problems must also be resolved to maintain VoIP communication quality during movement. First, the timing to initiate handover is also a critical issue. In fact, late handover initiation severely affects VoIP communication quality because the wireless link quality suddenly degrades. Second, how to recognize which AP will be the best choice among available APs is an issue of concern. Thus, to meet the lofty goal of integrating the next generation networks and to maintain VoIP communication quality during movement, the above requirements must be satisfied.

In fact, in 802.11 networks, the handoff process is partitioned into three phases: probing (scanning), re-authentication and re-association. According to (Mishra et al., 2003; Bianchi et al., 1996) the handoff procedure in IEEE 802.11 normally takes hundreds of milliseconds, and almost 90% of the handoff delay is due to the search of new APs, the so-called probe delay. This rather high handoff latency results in play-out gaps and poor quality of service for time-bounded multimedia applications. Other than the latency concern, the MS association with a specific AP is based only on the Received Signal Strength Indicator (RSSI) measurement of all available APs. This naïve procedure needs to be tuned since it leads to undesirable results (many MSs are connected to a few overloaded APs). The handoff process should take into account other context-based parameters, i.e. the load of APs.

In this chapter, we propose an optimized VoIP-oriented version of the Prevent Scan Handoff Procedure (PSHP) scheme (Rebai et al., 2009a, 2010, 2011) that will decrease both handoff latency and occurrence by performing a seamless prevent scan process and an effective nextAP selection. Basically, the IEEE 802.11 PSHP technique reduces the probe phase and adapts the process latency to support most of multimedia applications. In fact, it decreases the delay incurred during the discovery phase significantly by inserting a new Pre-Scan phase before a poor link quality will be reached. The available in range APs are kept in a dynamic list which will be periodically updated. As a complementary proposition, the authors in (Rebai et al., 2009b) integrated an effective AP selection based on Neighbor Graph (NG) manipulation and a new heuristic function that employs multiple-criteria to derive optimized search. Furthermore so far, to our knowledge, no adaptive techniques with crosslayering approach have been addressed, on transmitting real-time applications during a handover process. However various cross-layer adaptive rate control methods coupled with the MAC link adaptation have been presented for voice/video applications. Analyzing the opportunity from the literature, this research study focuses on IEEE802.11 handoff optimization using codec adaptation mechanism based on both parameters: codec type and packet size. Through real experiments and performance evaluation, we show effectiveness of the optimized PSHP draft which accomplishes a VoIP transmission over an 802.11 link without interruptions when altering between available APs. The rest of the chapter is organized as follows. Section 2 presents overview of handoff procedure performed in IEEE 802.11 WLANs and discusses related works. We present in section 3 operation details and several simulation results of PSHP Medium Access Control (MAC)-Layer handoff method (Rebai et al., 2011). The new VoIP-oriented PSHP technique and its experimental analysis are shown in section 4 followed by concluding remarks in section 5.

BACKGROUNDS AND RELATED WORKS

The Handoff Process in IEEE 802.11

Typically the 802.11 WLAN was originally designed without the consideration of mobility support, MAC layer handoff mechanism enables

MSs to continue their communication between multiple nearby APs. However, in regard to the mobility in the WLAN, there exists a problem to support the VoIP applications. When a MS performs handoff to the other AP, it should suffer from significant handoff latency causing the service degradation of the VoIP service where the typical VoIP application requires maximum 20~50ms packet arrival time. First we define the following terms: the coverage area of an AP is termed by Basic service set (BSS). Extended service set (ESS) is an interconnection of BSSs and wired LANs via distributed system (DS) as shown in Figure 1.

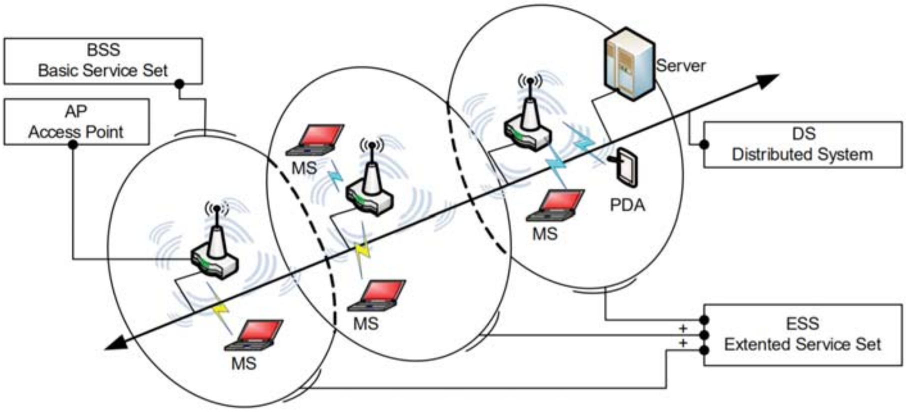


Figure 1. The IEEE 802.11 infrastructure mode

The inter-cell commutation can be divided into three different phases: detection, probing (scanning) and effective handoff (including authentication and re-association). In order to make a handoff, the MS must first decide when to handoff. A handoff process in IEEE 802.11 is commonly initiated when the Received Signal Strength (RSS) from current AP drops below a pre-specified threshold, termed as handoff threshold in the literature (Mishra et al., 2003; Raghavan et al., 2005). Using only current AP's RSS to initiate handoff might force the MS to hold on to the AP with low signal strength while there are better APs in its vicinity. As shown in Figure 2, when a handoff is triggered, an AP discovery phase begins and a MAC layer function called scan is executed. A management frame called De-authentication packet is sent, either by the mobile station before changing the actual channel of communication which allows the access point to update its MS-affiliation table, either by the AP which requests the MS to leave the cell. Since there is no specific control channel for executing the scan, the MS has to search for new APs from channel to channel by temporarily interrupting its association with the old AP. The scan on each channel can be performed by either passively listening to beacon signals or actively exchanging probe messages with new APs. After a new AP is found and its RSS exceeds Delta-RSS over the old AP, the MS will change its association to the new AP and a re-authentication phase begins. During the passive scan mode the MS listens to the wireless medium for beacon frames which provide the MS with timing and advertising information. Current APs have a default beacon interval of 100ms (Velayos & Karlsson, 2004). Therefore, the passive scan mode incurs significant delay. After scanning all available channels, the MS performs a Probe phase (used in active mode) only for the selected AP. As mentioned the polled AP is elected only based on RSSI parameter. The 802.11k group (IEEE Std. 802.11k, 2003) works on improving the choice of the next AP taking into account the network. In the active scan mode, the MS sends a Probe Request packet on each probed channel and waits MinChannelTime for a Probe Response packet from each reachable AP. If one packet at least is received, the MS extends the sensing interval to MaxChannelTime in order to obtain more responses and the channel is declared active. Thus, the waiting time on each channel is irregular since it is controlled by two timers (not as passive scan). The selected AP exchanges IEEE 802.11 authentication messages with the MS.

During this phase one of the two authentication methods can be achieved: Open System Authentication or Shared Key Authentication. Detailed authentication packets exchange has been addressed in (Rebai et al., 2011). After that the MS is authenticated by the AP, it sends Re-association Request message to the new AP. At this phase, the old and new APs exchange messages defined in Inter Access Point Protocol IAPP (IEEE Std. 802.11F, 2003). Furthermore, once the MS is authenticated, the association process is triggered. The Cell's information is exchanged: the ESSID and supported transmission rates.

During these various steps, the MS will be not able to exchange data with its AP. Based on values defined by the IEEE 802.11 standard, it is observed that the re-authentication which comprises an authentication and an association spends no more than 20ms on average, but a scanning delay may take between 350 and 500 ms and increase considerably the overall handoff latency (Mishra et al., 2003; Bianchi et al., 1996). An additional process is involved when the MS needs to change its IP connectivity (Johnson et al., 2004). In such a scenario, the MS needs to find a new access router. Also, the address binding information has to be updated at the home agent and corresponding agent (Cornall et al., 2002).

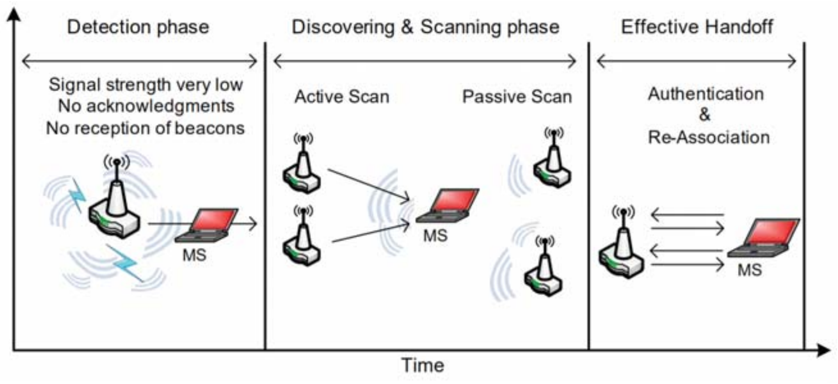


Figure 2. The 802.11 handover phases

Literature Survey

The scan phase of handover is the most costly in terms of time and traffic since it includes the probe and the channel switching sub-phases. Explicitly, the switching time is negligible and varies between 40 and 150µs as it was identified in (IEEE Std. 802.11i, 2004). Within the passive scan the interval between beacons is 100ms of IEEE 802.11b with 11 channels and 802.11a with 32 channels, the average latency will be respectively 1100ms and 3200ms. On the other hand, the time incurred with an active scan can be determined by the MinChannelTime and MaxChannelTime values. The MinChannelTime value should be large enough to not miss the proberesponse frames and obeys the formula given in Equation 1.

$$MinChannelTime \geq DIFS + (CW \times SlotTime) \qquad (1)$$

where, DIFS is the minimum waiting time necessary for a frame to access to the channel. The backoff interval is represented by the contention window (CW) multiplied by SlotTime. Regarding the authentication and authentication phases latency are proportional to the number of messages exchanged between the AP and the MS and limited to the medium access time which depends on the traffic in the cell (such management frames have no special priority). In (Velayos & Karlsson, 2004) these delays are estimated to less than 4ms in absence of a heavy traffic in the new selected cell. Thus, numerous schemes have been proposed to reduce the handoff scan delay in the 802.11 WLANs.

An interesting handoff scheme, called SyncScan, is proposed (Ramani & Savage, 2005) to reduce the probe delay by allowing a MS to monitor continuously the nearby APs and to record RSSs from available channel. Essentially, this technique replaces the existing large temporal additional costs during the scan phase. The absence delay of the MS with its current channel is minimized by synchronizing short listening periods of other channels (see Figure 3). In fact, the MS synchronize its next channel probing with the transmission of other APs beacons on each channel. By switching regularly and orderly on each channel, the MS reduces its disconnection delay with its actual AP. However, the SyncScan process suffers from regular additional interruptions during the MS absence when exploring other channels. These errors are very costly in terms of packets loss and skipped frames for timebounded applications. Moreover, this extra charge will affect all MSs even those that will never proceed to a handoff.

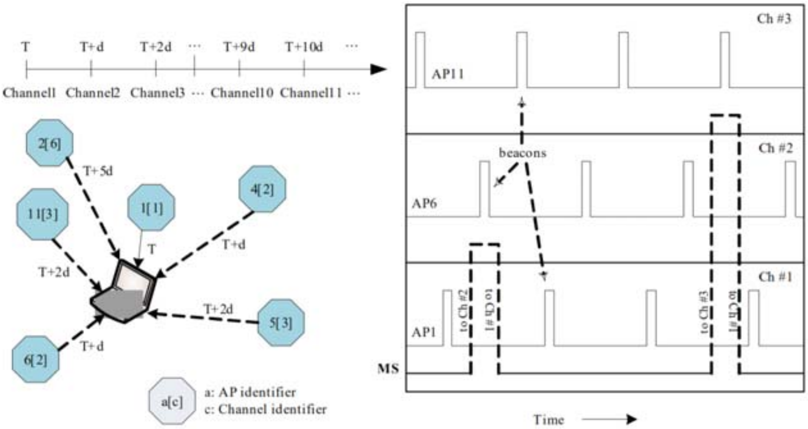


Figure 3. The SyncScan mechanism

In (Waharte et al., 2004) authors propose an innovative solution to optimize the AP's exploration during the scan phase based on the use of sensors operating on the 802.11 network. As shown in Figure 4, these sensors are arranged in cells and spaced 50 to 150 meters. These sensors have a role to listen to the network using beacons sent periodically by in-range APs. When the MS should change its actual cell, it performs a pre-scan operation which involves the sending of a request query to the sensors. Only sensors that have received this request (in range of the MS) react by sending the list of APs that they have identified. We figure out that this solution is effective in terms of the next-AP choice and the consequent results have improved significantly the standard handoff scheme. However, it is very expensive and has an extra cost by causing an additional load of unnecessary network traffic due to the sensor use. Moreover, this method is a non-compliant solution with the actual, 802.11 networks and requires radical changes to adapt it.

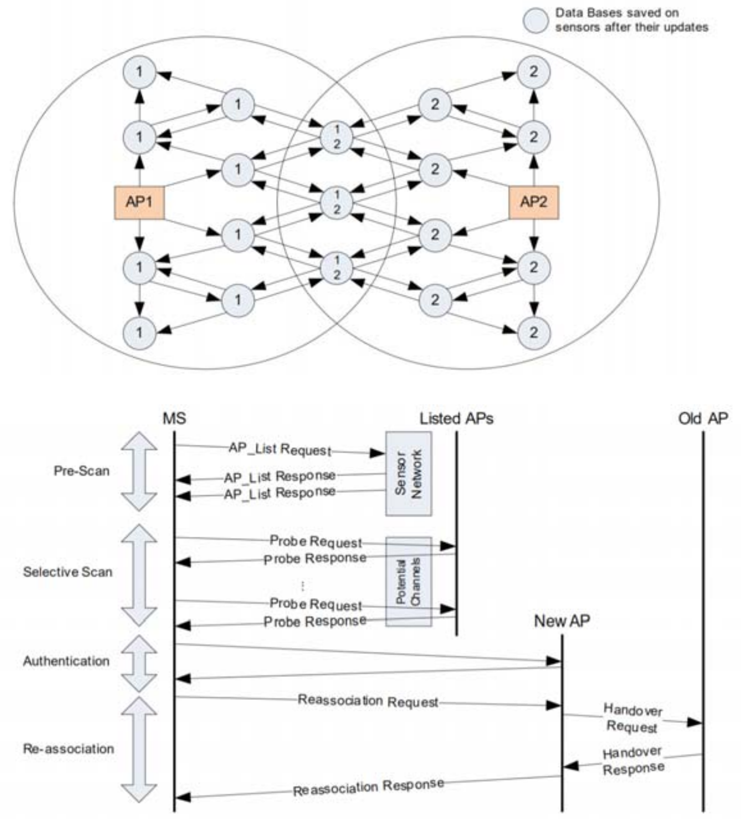
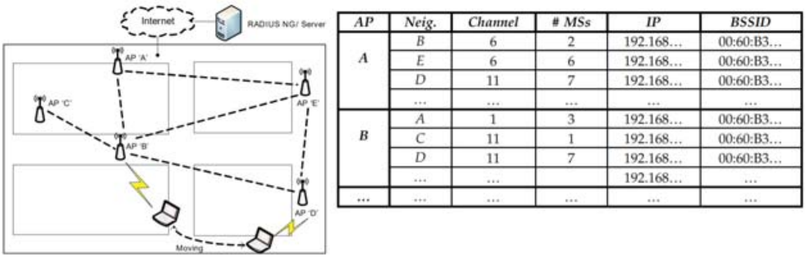


Figure 4. Sensors-based scan handoff technique

In (Huang et al., 2006), the authors proposed a selective scan technique in the IEEE 802.11 WLAN contexts that support the IAPP protocol (IEEE Std. 802.11F, 2003) to decrease the handover latency. This mechanism, as shown in Figure 5, reduces the scan time of a new AP by combining an enhanced Neighbor Graph 'NG' (Kim et al., 2004) scheme and an enhanced IAPP scheme. If a MS knows exactly its adjacent APs - provided by the NG RADIUS server (Radius, RFC 2865 & 2866) - it can use selective scanning by unicast to avoid scanning all channels. They enhanced the NG approach by putting the MS to power-saving mode (PSM) to pre-scan neighboring APs. This solution reduces, in a remarkable way, the total latency of the handoff mechanism. On the other hand, it requires that the MS must have knowledge on the network architecture and its adjacent APs to be able to employ selective scan. In addition, we should take into account the number of packets added by the IAPP that may affect the current traffic. Moreover, we note that all data packets have been sent to the old AP and then routed to the new selected AP before the link-layer is updated, which corresponds to a double transmission of the same data frames in the network. Thus, it greatly increases both the collision and the loss rates in 802.11 wireless networks.



AP	Neig.	Channel	# MSs	IP	BSSID
A	B	6	2	192.168...	00:60:B3...
	E	6	6	192.168...	00:60:B3...
	D	11	7	192.168...	00:60:B3...
	...	...	...	...	...
B	A	1	3	192.168...	00:60:B3...
	C	11	1	192.168...	00:60:B3...
	D	11	7	192.168...	00:60:B3...
	...	...		192.168...	...
...	...	...	...	...	...

Figure 5. The NG-based handover architecture

In (Chintala et al., 2007) the authors proposed two changes to the basic algorithm of IEEE 802.11 which reduce significantly the handover average latency using inter-AP communications during the scan phase. In the first proposed scheme (as shown in Figure 6), called Fast Handoff by Avoiding Probe wait (FHAP), additional costs incurred during this phase are reduced by forcing the potential APs to send their probe response packets to the old AP using the IAPP protocol and not to the MS which was sent the probe request. Therefore, the MS will avoid the long probe wait and then the packets loss is considerably reduced without any additional cost in the network. Consequently, during the probe phase the MS switches between channels and sends the probe request. Then it switches back to its actual AP to receive the probe responses.

However, based on this approach we note that the handover threshold should be adjusted so that the MS can communicate with its AP after the probe phase. Also, while the probe response packets are received via the current AP and not on their respective channels, the MS will not be able to measure the instantaneous values of RSSI and then evaluate the sensed channels quality. In (Chintala et al., 2007; Roshan & Leary, 2009) the authors improved their first solution by proposing a new mechanism called Adaptive Preemptive Fast Handoff (APFH). The APFH method requires that the MS predetermines a new AP before the handover begins. Then, the handover threshold is reached, the MS avoids the discovery phase and triggers immediately the re-authentication phase. This process will reduce the total handover latency to the re-association/authentication delay. The APFH method splits the coverage area of the AP depending on the signal strength in three areas: safe zone, gray zone and handover zone. As its name indicates, the safe zone is the part of the coverage area where the MS is not under a handover threat. The gray area is defined as an area where the handover probability is high. Therefore the MS begins collecting information on a new best AP. This second mechanism removes the entire handover latency and respects the strict VoIP transmission constraints.

Many research works were done on the network-layer regarding the challenge to support the mobility in IP networks – i.e. IPv6 (Johnson et al., 2004; Cornall et al., 2002). A detailed review of the most relevant methods and a deep time study of the handoff process have been presented in (Rebai et al., 2009a, 2009b; Ramani & Savage, 2005).

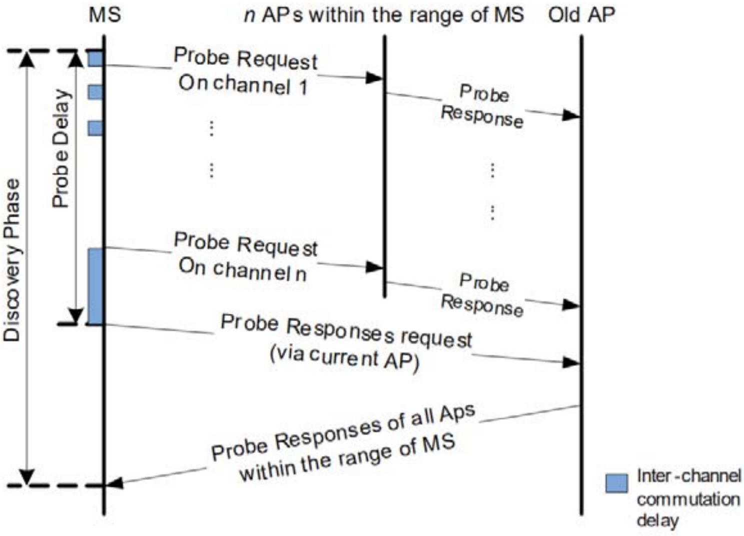


Figure 6. The discovery phase of the FHAP method

Prevent-Scan Handoff Procedure (PSHP)

In (Aboba, 2003) it has showed that the typical handoff latency in IEEE 802.11b with IAPP network may take a probe delay of 40 to 300ms with a constant IAPP delay of 40ms. To allow the IAPP protocol to reduce this delay, we firstly have imposed (Rebai et al., 2009b) that the MS must authenticate with the first AP of the ESS. The IAPP based preauthentication (Orinoco T.B., 2000) is achieved even before MS enters into the discovery state, thus it does not contribute to the handoff latency. Then we have proposed (Rebai et al., 2011) to define a new threshold other than the existing handoff threshold. It is called Preventive RSSI and termed by ($RSSI_{prev}$) and defined in the given Equation 2.

$$RSSI_{prev} = RSSI_{\min} + (RSSI_{\max} - RSSI_{\min})/2 \qquad (2)$$

where, $RSSI_{max}$ indicates the best link quality measured between a MS and its AP. As its name implies, the $RSSI_{prev}$ is a value of the link quality above which the MS is not under the threat of imminent handoff. Starting from this threshold the proposed algorithm detects the mobility of a MS and triggers the next-AP search which can offer a better link quality.

A continuous pre-scan process is generated in which the MS should switch channels and wait for beacons from potential APs. Since the switching and waiting delays are greater than the maximum retransmission time of 802.11 frames (4ms), time-bounded packets may be dropped by the MS (not able to acknowledge them). To overcome this drawback, the algorithm let the MS announces a Power Saving Mode (PSM) before switching channels (Baek & Choi, 2008). This causes the AP to buffer packets until the MS returns to its channel and resets the PSM mode. These buffers will not be overfilled during the PSM mode (very short in duration). In addition the pre-scan is programmed so that it does not disturb the existing traffic flow between the MS and its AP. Figure 7 presents the new state machine for a MS showing the various amendments that we have added to the basic algorithm.

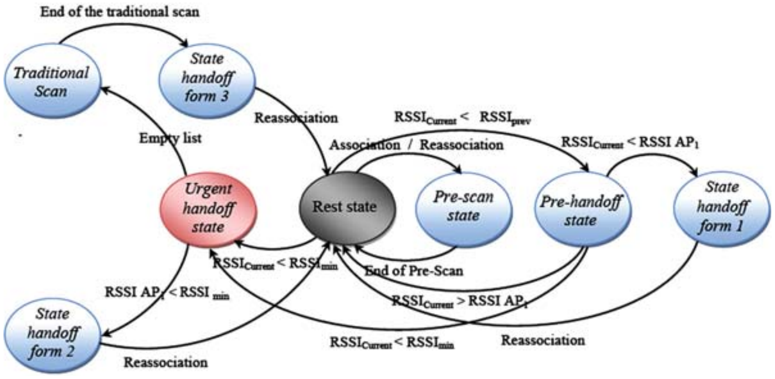


Figure 7. State machine of the PSHP procedure performed by the mobile station

The major advantage of the proposed scheme resides in the periodicity of checking for a new AP offering a better quality of link for forthcoming transmissions. The pre-scan phase is launched each α defined by the following Equation 3:

$$\alpha = \left[\left(T_{switch} + MaxChannelTime\right) \times N\right] \times 1.5 \quad (3)$$

where, T_{switch} is the switching delay from one channel to another, and N is the number of available channels.

In the PSHP scheme we enumerate three forms of handoff that can be happened depending on network conditions. Initially, the MS is in standby state. If the RSSI value associated to the current AP degrades and reaches the RSSIprev, then the MS switches to the pre-handoff state to check its dynamic list. It will try to find out a new AP with a corresponding RSSI value higher than the actual one. If such value exists, the MS switches to a 'handoff form1' state and performs a re-association procedure with the chosen AP. Otherwise, the MS returns to its standby state. If the measured RSSI value with the current AP is deteriorating suddenly and reaches the minimum bound (handover threshold), then the MS passes directly from the standby state to the 'urgent handover' state. In such state the MS must decide whether to perform the second or the third form of handover depending only on the instantaneous data of the dynamic list. If the first-listed AP has a value of RSSI greater than the handover threshold, then the MS switches to the 'handover form2' state. If such case does not exist, the MS switches to the 'handover form3' state in which it carries out a classical 802.11 handoff with a traditional scan procedure. We note that during the first two handoff occurrences (form1 and form2) the overall latency is equal only to the re-association delay, while using the 'handoff form3' is rare in practice (after carrying out pre-scan cycles).

Multiple-criteria AP selection technique

A second PSHP add-on mechanism for the AP selection was proposed initially in (Rebai et al., 2010). The proposed techniques aims to choose the most adapted AP from available Aps for the next handover occurrence. The basic procedure is considering only the RSSI value as an indicator of the AP quality. This naïve procedure, leads to the undesirable result that many MSs are connected to a few APs, while other APs are underutilized or completely idle. In (Chou & Shin, 2005) the authors argued that the login data with the APs can reflect the actual situation of handovers given discrete

WLAN deployment. As an example (see Figure 8), two WLANs may be very close to each other but separated by a highway or a river. Conversely, if the user is moving fast (e.g. in a train), handover may need to take place among WLANs that are far apart, i.e. among non-neighbor APs. Thus, the context user history allows us to better predict the probability of the user's next movement.

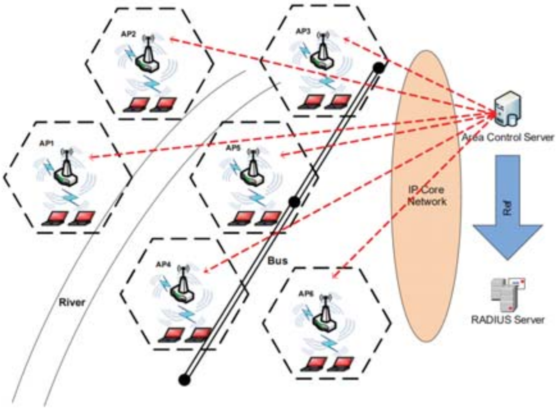


Figure 8. Example of an IEEE 802.11 infrastructure-mode WLAN

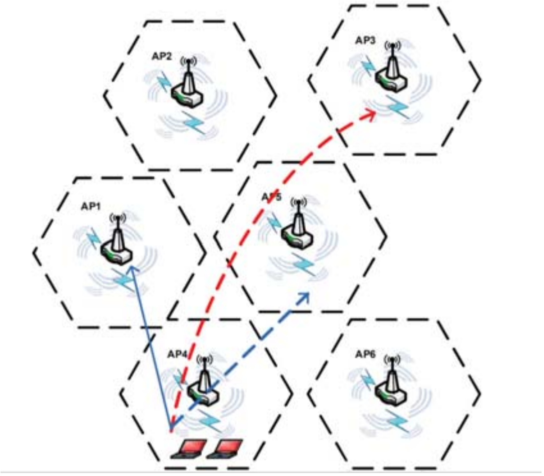


Figure 9. Handoff scenario for a MS moving towards AP3

A new network-configuration method - that differs from the RSSI constrained process (Shin et al., 2004) – has been proposed by introducing three new network parameters to optimize the next-AP selection during a WLAN handoff procedure (Rebai et al., 2010). The first parameter reproduces the number of MSs associated with an AP to exploit the overload factor of APs. The second parameter counts the handoff occurrence between the actual AP and each potential AP, and so represents significant history-based information. This counter is incremented by one each time a handoff occurs between the corresponding two APs. It includes the location and other context-based information useful for the next AP-selection. The third parameter reproduces the number of neighbors APs of the next handover chosen AP. In other words, is reflects the number of 2-hop neighbors of the current AP through a potential AP. This “look-ahead” parameter is added to improve the choice of the next AP to maintain long-term connections. A full numerical optimization approach for the next-AP selection was discussed in (Rebai et al., 2010) as well as its application results.

To elucidate the concept further, Figure 9 shows a sample handoff situation for a mobile moving from AP4 towards AP3. The standard 802.11 approach would handoff the mobile station (MS) to AP1 based on RSSI strength only. Then after few attempts it may switch the mobile to AP5. The proposed approach would result in a handoff directly to AP5 based on better handoff history criteria between AP4 and AP5 (high occurrence because of the bus trail), and more 2-hop neighbors through AP5. As the number of MSs was increased, with a bias towards moving to AP3, this add-on heuristic technique demonstrates clear advantage over the simple RSSI only approach and reduces considerably the handovers over the WLAN.

PSHP Simulation Results

In this section, the performance of the proposed scheme PSHP is evaluated and compared to the basic handoff scheme (currently used by most network interface cards) and other significant works founded in (Velayos & Karlsson, 2004; Ramani & Savage, 2005; Chintala et al., 2007). The handoff latencies of all schemes for different traffic loads are presented. This is followed by discussion on the total amount of time spent on handoff for all schemes. The effect of the proposed schemes on real time traffic is explored and weighed against the basic handoff scheme. We used C++ to simulate the new 802.11 handoff versus other described techniques. The IEEE 802.11b (IEEE Std. 802.11b, 1999) networks are considered for testing the schemes. The total number of the probable channels is assumed to be 11 channels (number of all the legitimate channels used in USA for 802.11b). We employed a total of 100 APs and 500 MSs to carry out the simulations. The other parameters are outlined in Table 1.

Table 1. Simulation Parameters

Parameter	Value
Speed of MS	0.1 – 15 m/s
Mobility Model	Random Way Point
MinChannelTime / MaxChannelTime	7/11 ms
Switch Delay	5 ms
Handoff Threshold	-51 dB
Pre-Scan Threshold	-45 dB

In general, all the suggested solutions to optimize the handoff process aim to reduce the total latency below 50ms (International Telecom Union [ITU], 1988) mainly for multimedia applications. The proposed PSHP solution aims to be conforming to this restriction by reducing the total handoff delay incurred in 802.11 WLANs. We choose a free propagation model for the mobile stations. Thus, in performed simulations the received signal strength indicator value is based on the distance between a MS and its AP (RSSI-based positioning) as shown in (Kitasuka et al., 2005; Rebai et al., 2011). The adopted mobility model is based on the model of random mobility "Random Way Point Mobility Model" presented in (Boudec, 2005). The same moving model has been also adopted in other algorithms (Ramani & Savage, 2005; Huang et al., 2006; Kim et al., 2004; Chintala et al., 2007).

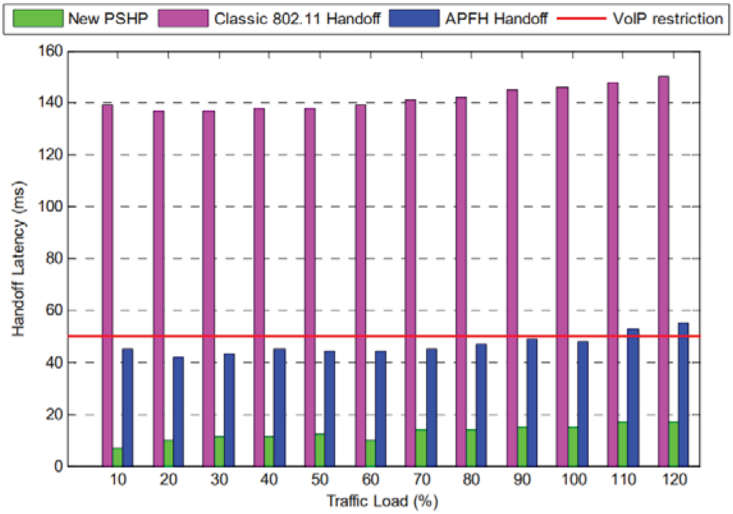
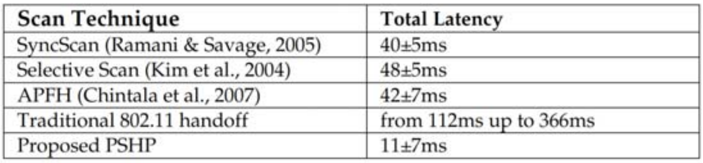


Figure 10. Handoff Latency versus Traffic Loads

Figure 10 shows the average handoff latency against different traffic loads for the three tested schemes. The APFH scheme achieves 67.62% delay improvement while the new PSHP method attains 95.21% improvement versus the basic 802.11 handoff scheme. The handoff latency of the classical approach is consistent with the simulation results in (Velayos & Karlsson, 2004) with similar parameters. Also, we point out by observing Figure 8 that the average handoff latencies for the PSHP and APFH schemes are both under 50ms which is well within VoIP constraints. However, PSHP performs the best and the minimal handoff delay compared to the APFH (Chintala et al., 2007). This remarkable improvement is reached since the new procedure performs a cyclic pre-scan phase before carrying out a handoff and most of handoffs are accomplished early by detecting the premature quality deterioration. As in (Ramani & Savage, 2005; Chintala et al., 2007) the traffic load is computed by dividing the number of active MSs (the MSs having data to transmit) over the maximum number of MSs transmitting on one AP's cell. The maximum number of active MSs is equal to 32 in IEEE 802.11 WLANs.

Based on the given results in (Mishra et al., 2003; Ramani & Savage, 2005; Kim et al., 2004) of related handoff techniques, we draw the following Table 2 resuming the total handoff delays for corresponding proposed mechanisms. We figure out a significant reduction achieved by the new PSHP algorithm compared to other solutions, and more specifically with the basic handover mechanism. We also note that the SyncScan solution satisfies the time-bounded applications. However, the selective scanning method occasionally exceeds the required QoS limits. This result is due to the inefficacity of the NG graph technique to manage all network topology changes due to the continuous MS mobility. With PSHP the total latency is reduced to the re-authentication phase (≈11ms). This delay may reach 18ms because in some simulated cases a handoff occurrence is triggered during a pre-scan cycle.

Table 2. Average latencies of different handoff procedures



Scan Technique	Total Latency
SyncScan (Ramani & Savage, 2005)	40±5ms
Selective Scan (Kim et al., 2004)	48±5ms
APFH (Chintala et al., 2007)	42±7ms
Traditional 802.11 handoff	from 112ms up to 366ms
Proposed PSHP	11±7ms

Figures 11 and 12 evaluate the performance of the APFH technique – the best known solution in literature – and the new PSHP scheme against

VoIP traffic. The packet interarrival time for VoIP applications is normally equal to 20ms (Chen et al., 2004), while it is also recommended that the inter-frame delay to be less than 50 ms (ITU, 1988; Chen et al., 2004). This restriction is depicted as a horizontal red line at 50ms in Figures 11 and 12. A node with VoIP inter-arrival time is taken and the corresponding delays are shown. The vertical green dotted lines represent a handoff occurrence. The traffic load for the given simulations was fixed to 50% and the number of packets sent to 600 (≈2.5s). We note that handoff occurrences are not simultaneous for the two simulated patterns. The MSs adopting the new PSHP algorithm detect the quality deterioration with their corresponding AP earlier than the APFH process. We note that both techniques respect the time constraint of real-time applications on recorded inter-frame delays without exceeding the required interval (50ms). However, this constraint is better managed by the new approach and the inter-packet periods are more regular and smaller. As discussed before, the PSHP handoff latency is only reduced to re-authentication delay if all handovers occur under form1 and 2. If PSHP form3 is performed, the latency is equal to the delay incurred in legacy 802.11 scanning all channels in addition to re-authentication delay.

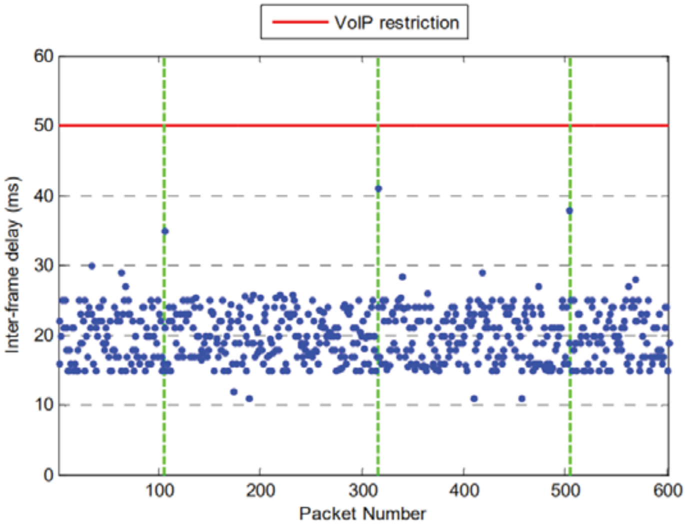


Figure 11. Inter-frame delay in APFH (Chintala et al., 2007)

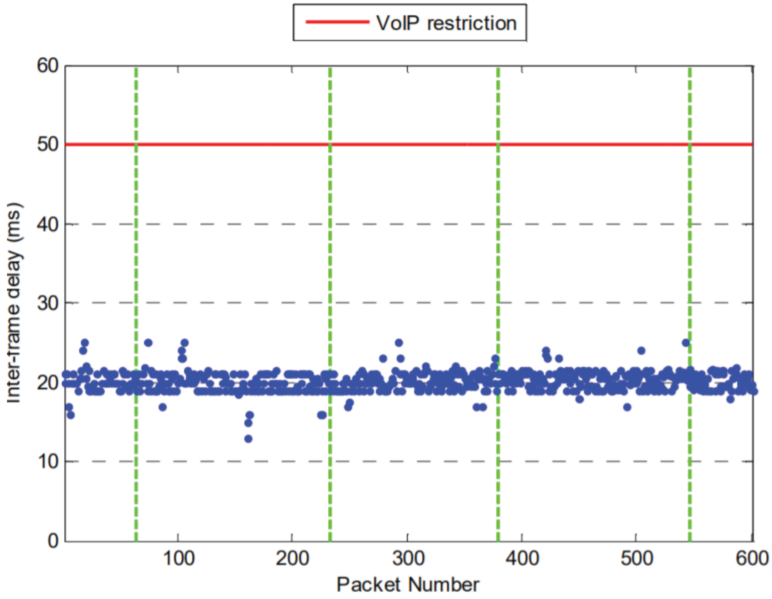


Figure 12. Inter-frame delay in PSHP

To emphasize the last assertion we present in Figures 13 and 14, respectively, a count of handoff occurrences for both APFH and PSHP schemes according to the traffic load and the detailed number of the various handoff forms related to the new PSHP technique. We set the simulation time to 10s for each considered traffic load.

By comparing values obtained by the two algorithms in Figure 13, we easily point out that the APFH technique (Chintala et al., 2007) performs less handovers in the network than the proposed PSHP scheme. This result can be explained by the adoption of the new form of preventive Handover (termed form1). Using this new form, a MS will not wait for a minimum quality recorded equal to the handoff threshold to trigger a handover. This new technique detects early the link quality deterioration with its current AP and performs an AP-switching to potentially improve link conditions. Therefore, the periodic pre-scan adopted by the new technique offers new opportunities to enhance the link quality between a MS and its AP and a significant reduction of the total handover delay. Indeed, with the pre-scan cycle the MS can discover other APs that have a better value of RSSI than provided by the current AP and provide the means to make more intelligent choices before and during a handover. The new algorithm PSHP has a better choice for the next AP by collecting periodic RSSI measurement. Thus, the decision is earlier and more beneficial when a handover is performed (rather than relying on a single sample as in usual schemes). Consequently, the extra number of PSHP occurrences versus APFH procedure happenings is compensated by an early choice of next AP with a better offered quality.

In Figure 14, the vertical red lines represent the executed number of form1 handoffs. Blue lines represent the number of handoffs taken under the second and third form, i.e. urgent handoffs. Recall that handoff under the first form is started when the RSSI value degrades below the $RSSI_{prev}$ and above the handoff threshold. Handoffs of the second and third form start only if RSSI value is degraded below the handoff threshold. In Figure 14 we figure out for most traffic loads, urgent handoffs occur less frequently than handoffs of form1. We also state that the proposed algorithm presents true opportunity to improve link quality since most of handoff occurrences are executed before that the RSSI value degrades below the handoff threshold. Accordingly, we conclude that almost half of accomplished handoffs are done under the new first form, which explains the delay reduction of PSHP since the first form decreases the related latency considerably and improves the link quality between the MS and its current AP.

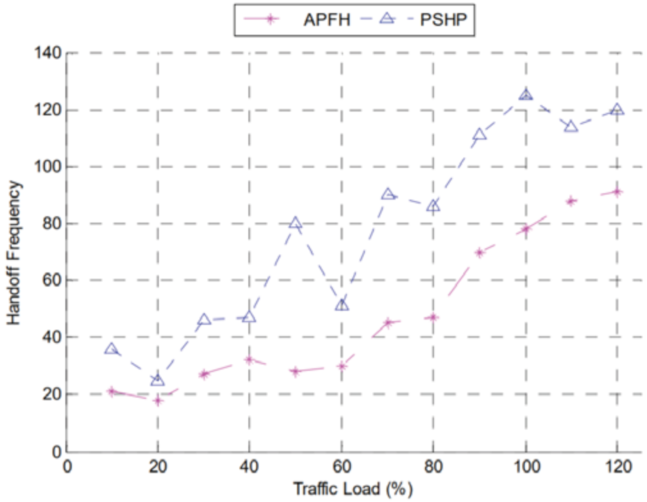


Figure 13. Handoff Frequency

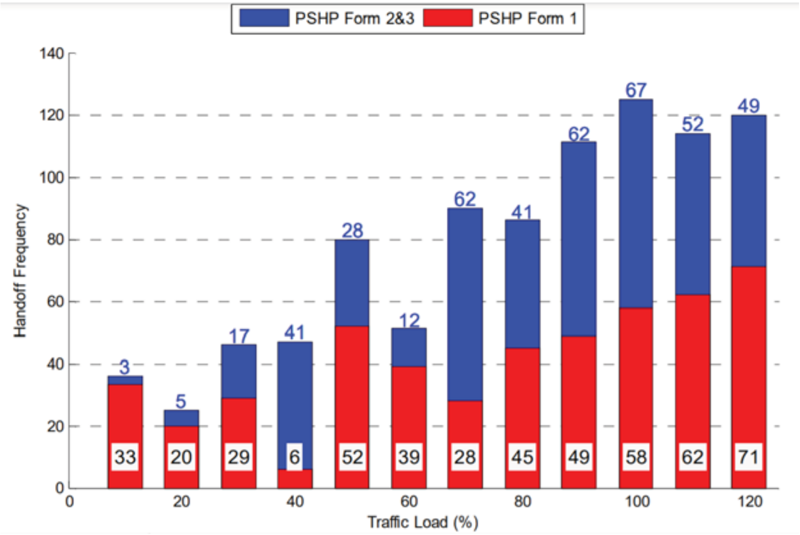
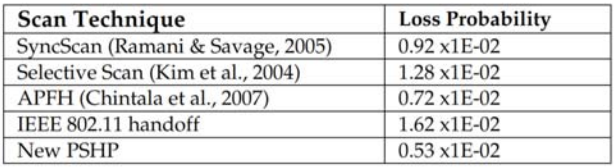


Figure 14. Occurrence of Handoff forms in PSHP

Table 3 shows the average probability of data packets being dropped and caused mainly by handoff procedure for the three schemes (APFH, PSHP, and the classic 802.11 approach). We also add the obtained result in (Ramani & Savage, 2005; Kim et al., 2004) for SyncScan and SelectiveScan, respectively. For comparison purposes, the traffic load for all nodes is divided into real-time and non real-time traffic with a ratio of 7.5/2.5. Other than errors caused by handoff occurrences, the real-time data packets are dropped also if the interframe delay exceeds 50ms. The simulation time for each traffic type is 10s (equivalent to about 2500 frames). Clearly, PSHP outperforms the other three schemes and the basic 802.11 as long as the traffic load is limited. The loss probability value of the new PSHP technique is divided by two compared to these obtained by SyncScan and SelectiveScan methods and by three of that accomplished by the standard 802.11 scheme.

Table 3. VoIP packet's loss



Scan Technique	Loss Probability
SyncScan (Ramani & Savage, 2005)	0.92 x1E-02
Selective Scan (Kim et al., 2004)	1.28 x1E-02
APFH (Chintala et al., 2007)	0.72 x1E-02
IEEE 802.11 handoff	1.62 x1E-02
New PSHP	0.53 x1E-02

In conclusion, periodic scanning also provides the means to make more intelligent choices when to initiate handoff. The new PSHP can discover the presence of APs with stronger RSSIs even before the associated AP's signal has degraded below the threshold. In addition, the pre-scan phase does not affect the existing wireless traffic since the corresponding MS will carry out a pre-scan cycle after declaring the PSM mode to buffer related packets.

Evaluation of the new add-on AP-selection heuristic

As mentioned above we add new context-based parameters for the next AP choice when a handover is triggered in the network by a MS. The result technique is not dependent on the used handoff method. Thus, we integrate the new developed heuristic function with both the classic and the proposed switching algorithm. Specifically, in the standard 802.11 method the next AP selection will be performed after the scan phase on the found APs by choosing one based on the new objective function. Regarding the PSHP procedure this choice will be performed after each pre-scan cycle only on APs belonging the associated dynamic list. This function is also performed for both handoff form2 and form3. The only algorithm modification in PSHP form1 handoff process is that the objective function is performed only on listed APs that have an RRSI value greater than the actual RSSI measured between the MS and its actual AP. By adopting this condition we always maintain the main purpose of the PSHP which is an earlier selection of a new AP that offers a better link quality. Therefore, the modified PSHP will not choose automatically the first best AP in the list. However, it will select from existing AP that maximizes the objective function and also offers a better channel link quality. We set the same simulation parameters as given in Table 1. However, we add geographic constrains by influencing some MS-AP link qualities depending on AP initial positions and by introducing initial specific values for the CNX parameter to illustrate the already performed MS-journeys in the network and a random primary associations between MSs and the given set of APs. The simulated mobility model regarding the MS moves is no longer “Random Way Point”. To be closer to realistic networks and to better assess our mechanism we switch to the “Random Direction” Mobility Model which forces mobile stations to travel to the edge of the simulation area before changing direction and speed. We choose this model because of its inclusion simplicity and instead of the “City Section” Mobility Model – which represents streets within a city. By including these constrain, we evaluated of the proposed heuristic combined with handoff schemes. In Figure 15 we resume the handoff occurrences for both classic and modified handoff schemes for the standard 802.11 and the PSHP techniques according to the traffic load. We set the simulation time to 10s for each considered traffic load. We point out a perceived reduction for handoff occurrences for both schemes when using the proposed heuristic procedure during the next AP selection. The produced results with the PSHP procedure are clearly enhanced in term of handoff count by integrating the new add-on heuristic technique. This result reflects the pay effect of the new objective function that accomplishes a better AP choice for the next inter-cell commutation, and consequently, improves the total number of handoff happening by reducing worse AP selections that was based only on RSSI-measurement decisions.

The detailed number of the various handoff forms related to the extended PSHP technique is shown in Figure 16. As well as in Figure 14, the vertical red and blue lines represent, respectively, the executed number of form1 handoffs and the count of handoffs taken under the second and third form (called also urgent handoffs). We figure out that handoffs form1 – performed when the RSSI value degrades below the RSSIprev threshold – are more triggered using the modified PSHP. We note that the proposed algorithm detects earlier the MS path and direction based on supplementary context-based information, and as a result, chooses quicker the best AP that improves the link quality and offers a continuous channel connection. Accordingly, 72% of accomplished handoffs are done under the first form of PSHP that decrease considerably the total latency and improves the link quality. As discussed before, data packets are dropped mainly by the handoff procedure and the violation of VoIP restrictions. Table 4 summarizes the data loss average probability for both classic 802.11 and PSHP approaches. As settled before the simulation time is 10s. The traffic load for MSs is equally combining real-time and non real-time traffic. The given results are the average of simulated values by varying the traffic load (from lower to higher loads).

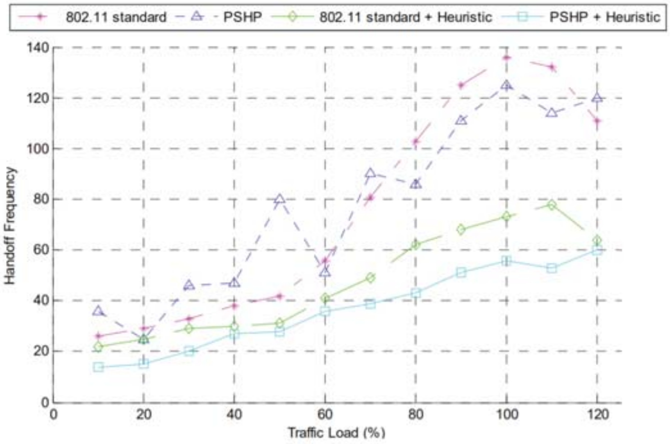


Figure 15. WLAN Handoff's frequency

Table 4. Packet's loss with heuristic selection

Scan Technique	Loss Probability
Standard 802.11 handoff	1.62 x1E-02
PSHP	0.53 x1E-02
IEEE802.11+heuristic selection	0.78 x1E-02
PSHP + heuristic selection	0.32 x1E-02

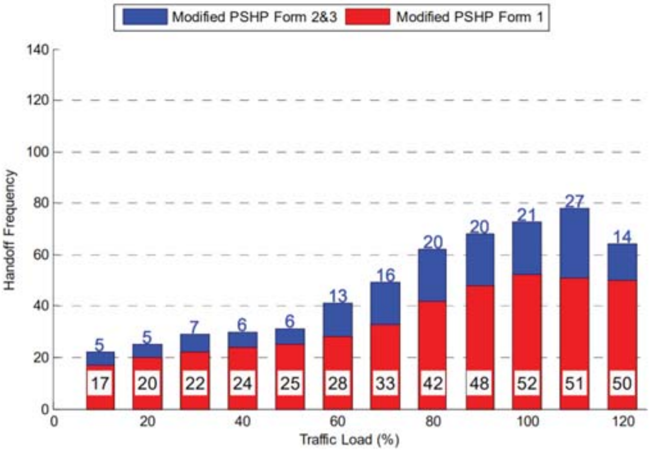


Figure 16. Handoff Occurrence in PSHP

We note that the modified PSHP version is outperforming the standard scheme. The reduced number of handoffs and also the high percentage of form1 handoffs lead to minimize the packet loss caused by handoff procedures. Thus, we can conclude that the loss probability value obtained by the new PSHP integrating the heuristic technique includes mainly dropped packets associated to a higher traffic load and not linked to the lack of respect of QoS constrains. In this section we have established that the proposed merit function used to evaluate network performance based on user preferences was adopted to find the best possible nextAP for the MS and to determine the optimal target AP based on a heuristic prediction process.

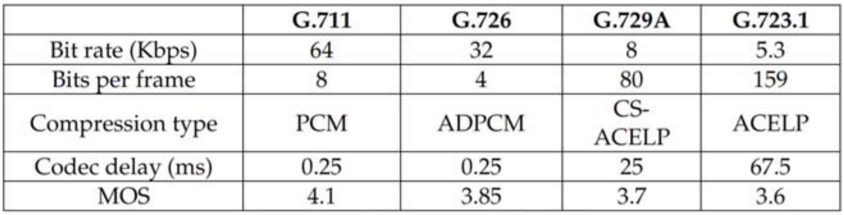
A NEW CROSS-LAYER SIGNALING APPROACH FOR VOIP-ORIENTED PSHP MECHANISM

Although the layered architecture of the network model is designed for wired networks and it served that purpose well, it is still not efficient enough for wireless networks (Srivastava & Motani, 2005). Consequently, the wired network layering structure is not sufficient enough to support the objective of transmitting real-time applications over WLAN. There have been several methods and algorithms designed to improve the performance of wireless network for real-time transmission. However, some studies and surveys showed that cross-layer approach has a great impact on improving transmission performance of real-time traffic over wireless networks, and thus over WLAN. The concept of cross layer approach is that it allows the network layers to communicate and exchange information with each other for better performance (Ernesto et al., 2008). On the other hand, since no specific codec can work well in all network conditions (Karapantazis & Pavlidou, 2009), developing codec adaptive techniques have been proposed. Although developing this mechanism is still in its early stage (Myakotnykh & Thompson, 2009), different adapting codec rate schemes were proposed particularly for real-time applications in wired, wireless, or WLAN networks. Codec rate adaptation technique is defined as a technique that adjusts codec parameters or changes it to another codec with lower rate when the network gets congested. Codec parameters that can be considered in this technique are: packet size and compression rate (Myakotnykh & Thompson, 2009).

Besides, adaptive rate approaches have been implemented using different constant bit-rate codec or variable bit-rate codec, such as AMR (Servetti & De Martin, 2003) and Speex (Sabrina & Valin, 2008). Moreover, it was shown that adaptive approaches perform better than constant bit rate (Servetti & De Martin, 2003; Sabrina & Valin, 2008). Table 5 (Karapantazis & Pavlidou, 2009), below illustrates parameters of different codecs. It was also concluded that for WLAN adapting the packet rate of a codec is sufficient with remaining the same codec type. Thus, changing packet size is an important parameter and would produce results of better quality (Myakotnykh & Thompson, 2009); hence it is considered in the approach.

Hence, our objective is to develop a cross-layering approach between MAC and Application layers. An agent will be designed and positioned as a mean of implementing the approach, which mainly aims to reduce VoIP delay and packet loss over WLAN, and therefore achieving better quality of VoIP. This section will focus on addressing a cross layer signaling technique in WLAN during handoff event by using the codec adaptive technique.

Table 5. Voice codec Parameters



	G.711	G.726	G.729A	G.723.1
Bit rate (Kbps)	64	32	8	5.3
Bits per frame	8	4	80	159
Compression type	PCM	ADPCM	CS-ACELP	ACELP
Codec delay (ms)	0.25	0.25	25	67.5
MOS	4.1	3.85	3.7	3.6

Cross Layer Adaptive Agent (CLAA)

In order to improve the QoS of VoIP during a handoff procedure, a cross-layer agent is proposed to allow the communication between MAC and Application layers. This new Cross Layer Adaptive Agent (CLAA) monitors the MAC handoff state changes and then adapts the suitable packet size and the codec type in the Application layer (Figure 17).

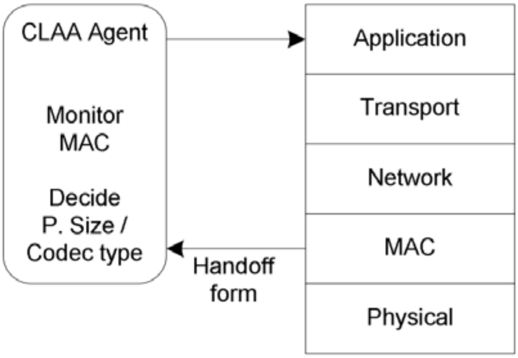


Figure 17. Cross Layer Adaptive Agent (CLAA) model

The main function of the agent is to detect if there is a change in the PSHP state machine traced in the MAC layer. If such change occurs, it will inform the Application layer to act accordingly and better compensate either by the packet size at the codec algorithm in a dynamic manner in order to minimize channel congestion in WLAN or by codec type change. The other key point is that the agent tries to resolve congestion that the MS suffers locally firstly. If the congested MS condition is getting worse, then a second decision-phase will assist the agent to reduce the congestion. We point that the CLAA agent is implemented only on local mode (sender side) since a global mode involves a collaboration between the sender and the receiver nodes. Such operations entail the receiver to send back to the sender information regarding the voice quality through its current AP which leads for extra network congestions during a handoff procedure. Figure 18 shows a descriptive flow chart of the local CLAA function.

Therefore, the agent chooses to resolve the handoff process issue locally at the sender side. This phase is in an open loop monitoring MAC layer and observes if any changes on the PSHP state/handover form occurred. If so, then the agent decides a new packet size/codec type and informs the Application layer to adjust. In fact, when MS detects a potential handoff (PSHP pre-handoff state) with one new cell (selected from the listed APs) offering a better quality, it initiates the handoff form1 state and simultaneously the CLAA agent decides to reduce the packet size to reduce congestion and packet loss during this short handoff period. However if a transition from pre-handoff state to an urgent handover state is triggered by the PSHP mechanism, the CLAA will change moderately the codec type. By doing so the resultant bit rate is lowered to prevent from large delay transmissions and enhance the overall received voice quality. Of course the codec type adaptation has a minor effect to decrease the received VoIP quality; however it will eliminate the RTP stream interruption on the receiver.

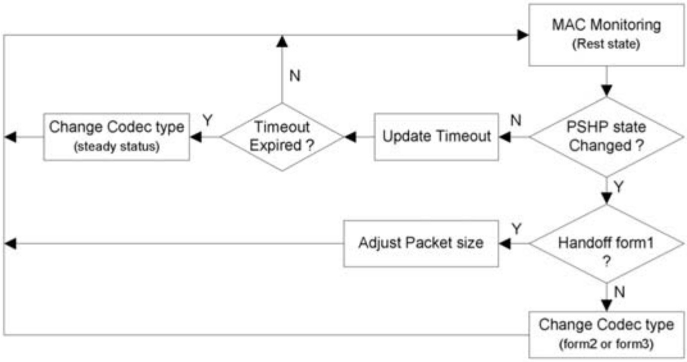


Figure 18. CLAA flowchart and principle of operation

On the other hand, following to an inter-cell occurrence involving a codec alteration a threshold will be set, thus no consecutive changes will happen at the Application layer. If the agent detects a long steady status (Rest state in the PSHP state machine) within a predefined timeout, then codec type would be changed to a higher one with larger bit rate. The codec techniques are sorted based on their bit rate (as shown in Table 5) for the agent to select accordingly.

Performance Evaluation and Discussions

The testbed used in the performance tests of the proposed VoIP-oriented PSHP method is shown in Figure 19. The mobile client's 802.11b driver has been slightly modified to provide transport of the necessary cross-layer related information during handoffs. The Chariot Console [NetIQ] was used to measure the handoff performance under real-time and multimedia traffic. In all tests, three laptops (MS1 MS2, and MS3) were running VoIP sessions towards three different PC hosts (RTP Stream, Chariot G.711u script).

The measurements were taken for movements between different APs (a likelihood path is shown in Figure 19 using a dotted blue line). Both of MS1 and MS2 are implementing PSHP + the Heuristic function while MS3 uses only a standard PSHP driver. The MS1 includes also the designed VoIP-oriented CLAA Agent. All MSs were initially located in AP1, and then moved around the university lab rooms and the hallway following a predefined path and using a constant 1~2m/s speed over potential seven

APs. The experiment consisted mainly to walk through the hallway (from a start point to a far end point), then to lab room#1 and finally back to the start point. In Figure 19 the red lines represent the physical separation between university rooms and the green dotted line corresponds approximately to the AP's coverage area. Several handoffs were performed over the network during multiple three-minute experimentations for each MS.

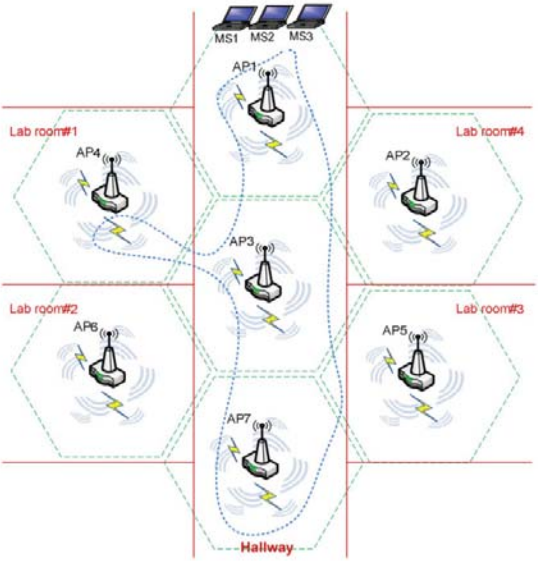


Figure 19. Testbed for the MS's movement and network topology

The clients MS1 and MS2 roam between APs identically: from AP1 to AP3 at around 17th second in both tests, from AP3 to AP7 at around the 31st sec., from AP7 to AP4 at around 96th sec., and back to AP1 at around the 124th sec. Since MS3 was using a standard PSHP driver (without an enhanced heuristic AP selection), it performed dissimilar handoffs approximately as follows: from AP1 to AP2 at 21st sec., then from AP2 to AP5 at 35th sec., from AP5 to AP3 at 46th sec., from AP3 to AP7 at 58th sec., from AP7 to AP6 at 69th sec., back to AP7 at 72nd sec., then from AP7 to AP4 at 104th sec., from AP4 to AP3 at 126th sec., and finally to AP1 at 155th sec. It is important to note that the initial transmission rate and codec type values are set to 11Mbps and G.711 respectively for all simulated MSs.

What can be observed from Figures 20 and 21 are the very small one-way delay and the very small packet loss performance achieved by MS1 compared to the other mobile nodes. Indeed, the CLAA integration has reduced significantly the total packet delay (caused by handoffs) by 64.5% from the PSHP+Heuristic mechanism and by 87.9% from the standard PSHP method which affects considerably the overall quality of received voice streams. Analogically, the number of the lost packets decreases to just 247 (from 538 and 2067 resulted through PSHP+Heuristic and PSHP implementations, respectively) during the three-minute RTP streaming. The high number of lost frames accomplished by standard PSHP arises from multiple handoff occurrences during the continuous MS movement between available Aps.

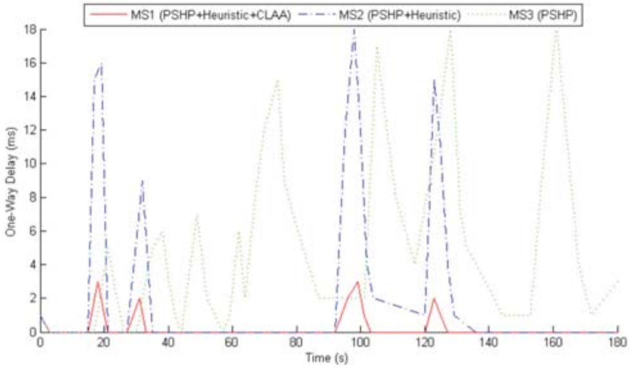


Figure 20. Measured one-way delay from MS drivers

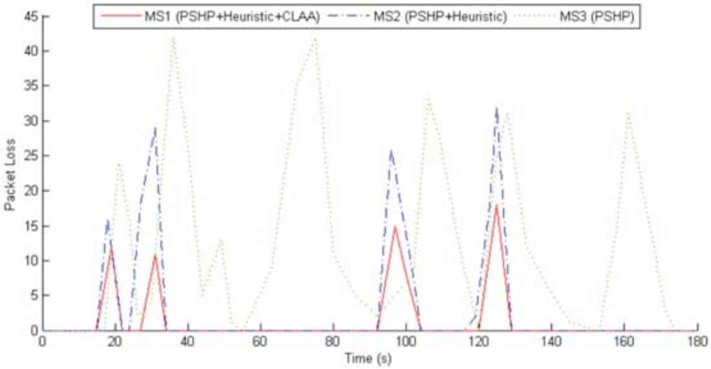


Figure 21. Packet Loss given by the three different MS handoff

From the result given by Figure 22 the overall throughput attained by MS1 suffered a degradation of only ≈ 6.6% while values of 11.8% and 14.3% were accomplished by MS2 and MS3 respectively. To determine the quality of VoIP under packet loss, the most common metric is the Mean Opinion Score (MOS) (IUT, 1996), which evaluates the effect of bursty loss on VoIP perceived quality (the Overall Voice Quality). In a MOS test, the listeners rate audio clips by a score from 5 to 1, with 5 meaning Excellent, 4 Good, 3 Fair, 2 Poor, and 1 Bad. In fact, voice and video communications quality usually dictates whether the experience is a good or bad one. Therefore, besides the qualitative description we hear, like ‘quite good’ or ‘very bad’, there is a numerical method of expressing voice and video quality given by the MOS which provides a statistical indication of the perceived quality of the media received after being transmitted and eventually compressed using VoIP codecs.

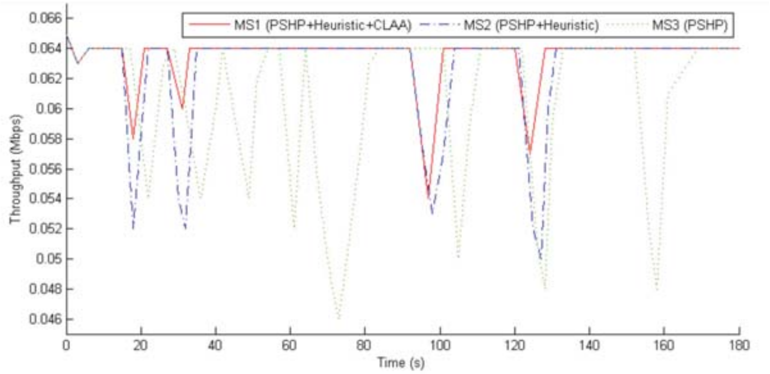


Figure 22. Throughput measurements versus VoIP streaming elapsed time

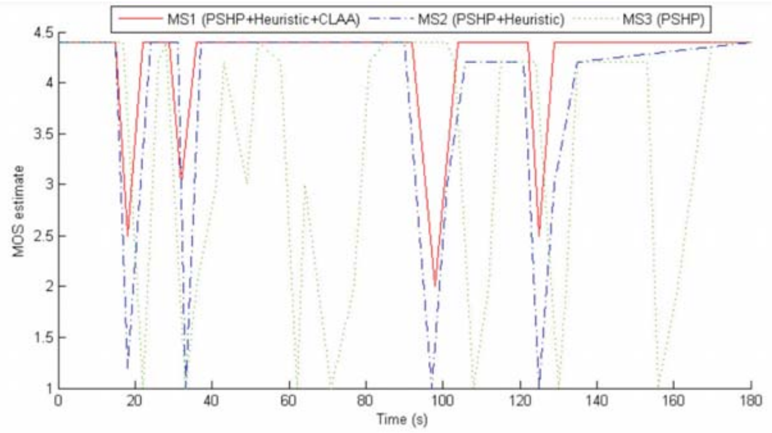


Figure 23. MOS estimate of received voice quality

The MOS estimate of conducted test experiments shows in Figure 23 that the call was not interrupted with all MSs; It only suffered substantial quality degradation with a low peak at MOS=1, and quickly restored its initial quality (MOS=4).

Based on the above VoIP session experiments, we notice that MS1 (PSHP+Heuristic+CLAA) roams between different APs and all related test results are enhanced. In fact, the MOS mean values obtained by MS1, MS2 and MS3 handoff implementations are 3.86, 3.35 and 3.04 respectively; hence a considerable enhancement of 15.2% from PSHP+Heuristic and 28.6% from the standard PSHP was achieved by the new VoIP-oriented technique. This is due to the fact that the integrated CLAA agent cooperates between the MAC and Application layers and then contributes in shortening the total handoff latency during movements by adjusting the packet size and the codec type when needed. Thus, it preserves efficiently the VoIP session and maintains a satisfactory aggregate throughput. As also verified by the MOS estimate, the minimum measured MOS value during the MS1 test is equal to 2 (which match 'Poor' quality). This value was reached only one time (around the 98th second). However using the other two handoff versions a minimum MOS value of 1 (symbolizes Bad quality) was measured several times over the real streaming experiments.

CONCLUSIONS

VoIP over WLAN applications are rapidly growing due to the features they offer over the traditional public switched phones and their support symbolize at present an emerging challenge for 802.11-based WLANs. However, the integration of these two technologies still facing quality challenges to meet the quality obtained from the traditional telephony system. Besides, mobile stations in WLAN suffer the continuous inter-cell handoff issue, which affects the quality of the perceived voice. In order to keep a VoIP communication several commitments should be satisfied: eliminating communication termination, initiating appropriate handover based on reliable handover triggers and selecting the next AP with good link quality.

We firstly highlighted some of the technical challenges and related literature on the ongoing research force, especially focusing on approaches for enabling multimedia transit, as well as convenient and effective handover over IEEE802.11 mobile environments. Then we have revisited the PSHP handoff technique. As demonstrated, the continuous scanning PSHP technique offers significant advantages over other schemes by minimizing the time during which an MS remains out of contact with its AP and allowing handoffs to be made earlier and with more confidence. The result is a staggering 95% reduction of handoff latency compared to the typical procedure. As a second contribution we took into account additional network-based parameters to drive a better next-AP choice. This new add-on profit function is used to insert new factors reflecting resource availability, location, and other context-based information. Thus, the overall network performance is improved by electing from available APs, the one that increases the benefit of the next handoff occurrence.

In particular, this chapter presented another PSHP version satisfying between user requirement and network conditions and avoiding unnecessary handoffs as well. The policy is to minimize handoff delay for real time service and to reach an acceptable level for nonreal time services. A pre-scanning phase is periodically activated to consider whether the handoff should be triggered. During the AP selection procedure, the heuristic function is adopted to find candidate APs satisfying preference of a user and minimizing the overall delay. PSHP is using four handoff metrics, RSS, (AP-extensibility) number of neighbors, (load) number of users per AP, and historical traffic (old occurred handoffs) as inputs to determine the optimal target network.

Furthermore, one of the challenges in the next generation of wireless communications is the integration of existing and future wireless technologies and supporting transparent and seamless vertical handoffs (between different networks standards and technologies) without degrading the QoS between these heterogeneous networks. Hence, this research work proposed a Cross-Layer Adaptive Approach (CLAA) in order to enhance the QoS of VoIP over WLAN with help of an agent. The Cross layering concept has been shown to have a great impact on the performance of the wireless networks. Adapting code parameters in the Application layer according to the network condition has also shown better performance of real-time applications. Thus, the new scheme would be easily extended to cover internetworks handoff decision toward universal 4G ubiquitous access.

REFERENCES

1. Aboba, B. (2003). Fast handoff issues. IEEE-03-155rO-I, IEEE 802.11 Working Group, Mars 2003
2. Baek, S. & Choi, B.D. (2008). Performance analysis of power save mode in IEEE 802.11 infrastructure WLAN. International Conference on Telecommunications ICT 2008, St. Petersburg, Russia, 2008
3. Bahl, P.; Balachandran, A. & Venkatachary, S. (2001). Secure Wireless Internet Access in Public Places. IEEE International Conference on Communications, vol.10, pp. 3271-3275, June 2001
4. Bianchi, G.; Fratta, L. & Oliveri, M. (1996). Performance Evaluation and Enhancement of the CSMA/CA MAC Protocol for 802.11 Wireless LANS. The 7th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), vol. 2, pp. 392 – 396, 1996
5. Boudec, J.Y. (2005). On the Stationary Distribution of Speed and Location of Random Waypoint. The IEEE Transactions on Mobile Computing, vol. 4, pp. 404-405, July 2005
6. Chen, Y.; Smavatkul, N. & Emeott, S. (2004). Power management for VoIP over IEEE 802.11
7. WLAN. The IEEE WCNC 2004, vol.5, pp.1648–1653, March 2004
8. Chintala, V. M. & Zeng, Q.A. (2007). Novel MAC Layer Handoff Schemes for IEEE 802.11
9. Wireless LANs. The IEEE Wireless Communications and Networking Conference, WCNC, Mars 2007
10. Chou, C.T. & Shin, K.G. (2005). An Enhanced Inter-Access Point Protocol for Uniform Intra and Intersubnet Handoffs. IEEE Transactions on Mobile Computer, vol. 4, no. 4, July 2005
11. Cornall, T.; Pentland, B. & Khee, P. (2002). Improved Handover Performance in Wireless Mobile IPv6. The 8th International Conference on Communication Systems (ICCS), vol. 2, pp. 857-861, Nov. 2002
12. Corner, D.; Lin, J. & Russo, V. (2010). An Architecture for a Campus-Scale Wireless Mobile Internet. Technical Report CSD-TR 95-058, Purdue University, Computer Science Department, 2010
13. Ernesto, E.; Nicolas, V. W.; Christophe, C. & Khalil, D. (2008). Introducing a cross-layer interpreter for multimedia streams, Computer Networks, Vol. 52 (6), pp. 1125-1141, April 2008

14. Huang, P.J.; Tseng, Y.C. & Tsai, K.C. (2006). A Fast Handoff Mechanism for IEEE 802.11 and IAPP Networks. The 63rd IEEE Vehicular Technology Conference, VTC Spring, 2006

15. Hills, A. & Johnson, D. (1996). A Wireless Data Network Infrastructure at Carnegie Mellon University. IEEE Personal Communications, vol. 3, pp. 56–63, Feb. 1996

16. IEEE Standard 802.11 (1999), Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11, 1999

17. IEEE Standard 802.11b (1999), Supplement to Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-speed Physical Layer Extension in the 2.4 GHz Band, IEEE Std. 802.11b-1999, 1999

18. IEEE Standard 802.11F (2003). IEEE Trial-use Recommended Practice for Multi-Vendor Access Point Interoperability via An Inter-access Point Protocol across Distribution Systems supporting IEEE 802.11 Operation. IEEE Std 802.11F, July 2003

19. IEEE Standard 802.11i (2004), Part 11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Security Enhancements. Supplement to IEEE 802.11 Standard, June 2004

20. IEEE Standard 802.11k (2003). Radio Ressource Management. IEEE Standard 802.11- 2003 International Telecommunication Union (1988). General Characteristics of International Telephone Connections and International Telephone Circuits. ITU-TG.114, 1988

21. International Telecommunication Union (1996). Subjective performance assessment of telephoneband and wideband digital codecs, Recommendation P.830, Telecommunication Standardization Sector of ITU, Geneva, Feb. 1996

22. Johnson, D.; Perkins, C. & Arkko, J. (2004). Mobility Support in IPv6. RFC 3775 (Proposed Standard): ftp.rfc-editor.org in-notes rfc3775.txt, June 2004

23. Karapantazis, S. & Pavlidou, F. (2009). VoIP: A comprehensive survey on a promising technology, Computer Networks, Vol. 12, pp. 2050-2090, 2009

24. Kashihara, S.; Niswar, M.; Taenaka, Y.; Tsukamoto, K.; Yamaguchi, S. & Oie, Y. (2011). Endto-End Handover Management for VoIP

Communications in Ubiquitous Wireless Networks, VoIP Technologies, Shigeru Kashihara (Ed.), ISBN: 978-953-307-549-5, InTech, pp. 295-320, Feb. 2011

25. Kim, H.; Park, S.; Park, C.; Kim, J. & Ko, S. (2004). Selective Channel Scanning for Fast Handoff in Wireless LAN using Neighboor Graph. ITC-CSCC, July 2004
26. Kitasuka, T.; Hisazumi, K.; Nakanishi, T. & Fukuda, A. (2005). Positioning Technique of Wireless LAN Terminals Using RSSI between Terminals. The 2005 International
27. Conference on Pervasive Systems and Computing (PSC-05), pp. 47-53, Las Vegas, Nevada, USA, June 2005
28. Kunarak, S. & Suleesathira, R. (2011). Predictive RSS with Fuzzy Logic based Vertical Handoff Decision Scheme for Seamless Ubiquitous Access, Mobile Ad-Hoc Networks: Protocol Design, Xin Wang (Ed.), ISBN: 978-953-307-402-3, InTech, pp. 261-280, Jan. 2011
29. Mishra, A.; Shin, M. & Arbaugh, W. (2003). An empirical analysis of the IEEE 802.11 MAC layer handoff process. ACM SIGCOMM Computer Communication Review, vol. 33: 93-102, April 2003
30. Mishra, A.; Shin, M.; Petroni, N.; Clancy, T. & Arbaugh, W. (2004). Proactive Key Distribution Using Neighbor Graphs. IEEE Wireless Communications Magazine, vol. 11, pp. 26-36, Feb. 2004
31. Myakotnykh, E. S. & Thompson, R. A. (2009). Adaptive Speech Quality Management in Voice-over-IP Communications, Fifth Advanced International Conference on Telecommunications, Venice, Italy, May 2009.
32. NetIQ Chariot Console, available at: http://www.netiq.com/products/chr/default.asp.
33. Orinoco (2000). Inter Access Point Protocol (IAPP). Technical Bulletin TB 034/A, Feb. 2000
34. Radius, RFC 2865 et 2866, http www.ietf.org rfc rfc2865.txt, http www.ietf.org rfc rfc2866.txt
35. Raghavan, M.; Mukherjee, A.; Liu, H.; Zeng, Q-A. & Agarwal, D. P. (2005). Improvement in QoS for Multimedia Traffic in Wireless LANs during Handoff. The 2005
36. International Conference on Wireless Networks ICWN'05, Las Vegas, USA, pp. 251-257, June 2005

37. Ramani, I. & Savage, S. (2005). SyncScan: Practical Fast Handoff for 802.11 Infrastructure Networks. The IEEE Infocom'05, vol. 1, pp. 675-684, March 2005.
38. Rebai, A. R.; Alnuweiri, H.; Hanafi, S. (2009). A novel prevent-scan Handoff technique for IEEE 802.11 WLANs, in Proc. International Conference on Ultra Modern
39. Telecommunications & Workshops ICUMT '09, St. Petersburg, Russia, Oct. 2009
40. Rebai, A. R.; Haddar, B.; Hanafi, S. (2009). Prevent-scan: A novel MAC layer scheme for the IEEE 802.11 handoff, In Proc. International Conference on Multimedia Computing and Systems ICMCS '09, pp. 541-546, Ouarzazate, Morocco, April 2009
41. Rebai, A. R. & Hanafi, S. (2011). An Adaptive Multimedia-Oriented Handoff Scheme for IEEE 802.11 WLANs, International Journal of Wireless & Mobile Networks IJWMN, pp. 151-170, Vol. 3, No. 1, Feb. 2011
42. Rebai, A. R.; Rebai, M. F.; Alnuweiri, H.; Hanafi, S. (2010). An enhanced heuristic technique for AP selection in 802.11 handoff procedure, In Proc. IEEE International Conference on Telecommunications ICT'10, pp. 576-580, Doha, Qatar, April 2010
43. Roshan, P. & Leary, J. (2009). 802.11 Wireless LAN Fundamentals. CISCO Press., ISBN No.1587050773, 2009
44. Sabrina, F. & Valin, J.-M. (2008). Adaptive rate control for aggregated VoIP traffic, in Proc. Of GLOBECOM'08, pp. 1405–1410, 2008
45. Servetti, A. & De Martin, J. C. (2003). Adaptive Interactive Speech Transmission Over 802.11 Wireless LANs". In Proc. Int. Workshop on DSP in Mobile and Vehicular Systems, Nagoya, Japan, April 2003
46. Shin, M.; Mishra, A. & Arbaugh, W. A. (2004). Improving the Latency of 802.11 Hand-offs
47. using Neighbor Graphs. The ACM MobiSys Conference, Boston, MA, USA, pp. 70-83, June 2004
48. Skype Limited. (2003), available at: http://www.skype.com
49. Srivastava, V.; Motani, M. (2005). Cross-Layer Design: A Survey and the Road Ahead, IEEE Communications Magazine, Vol. 43, No. 12., pp.112-119, 2005

50. Velayos, H. & Karlsson, G. (2004). Techniques to Reduce IEEE 802.11b MAC Layer Handover Time. IEEE ICC 2004, vol. 7, pp. 3844-3848, June 2004

51. Waharte, S.; Ritzenthaler, K. & Boutaba, R. (2004). Selective active scanning for fast handoff in wlan using sensor networks. IEEE International Conference on Mobile and Wireless Communication Networks (MWCN'04), Paris, France, October 2004

CHAPTER 13

UBIQUITOUS CONTROL FRAMEWORK FOR DELIVERING PERCEPTUAL SATISFACTION OF MULTIMEDIA TRAFFIC

K. L. Eddie Law[1] and Jacek Ilow[2]

[1]Kirin Cloud Solutions, Ltd Hong Kong
[2]Dalhousie University Canada

INTRODUCTION

With the latest computing technology, people can store information including audio, video, and data on the Internet. And the advanced networking protocol designs can easily enable interactive communications among users, applications, and services through wireless tablets and mobile devices (Satyanarayanan, 2001). However, people like to connect to the Internet while moving, the connectivity may vary during the course of an active session of a user. Although many portable device can provide high-speed

Citation: K. L. Eddie Law and Jacek Ilow (February 17th 2012). Ubiquitous Control Framework for Delivering Perceptual Satisfaction of Multimedia Traffic, Advanced Topics in Multimedia Research, Sagarmay Deb, IntechOpen, DOI: 10.5772/37802.

Copyright: © 2012 by authors and Intech-Open. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

connectivity, a user may move into area with weak signal or thin bandwidth, which possibly hinder the reception of satisfactory multimedia traffic. In today's business, it is important to make sure the satisfaction of content subscribers. Thus, it is adamant to develop a ubiquitous control framework to offer perceptual satisfaction of multimedia subscribers.

The ubiquitous computing platform should be designed for quality control of multimedia and data context through the Internet. The framework should manage the network and computing resources, such that the delivered information should at least meet the expected minimal perceptual quality of multimedia traffic stream of an end-user. In this chapter, a few basic design parameters for justifying the performance and design of the control framework will be elaborated. Quantitatively, different Quality of Service (QoS) parameters, e.g., packet loss rate, have been widely used for session transmission control. But for visual evaluation, the terminology known as Quality of Experience (QoE) has recently been widely used. QoE is a measure of performance expectations of the end-user; it may augment QoS by providing the quantitative link to user perception. Indeed, the only way to know how customers see your business is to look at it through their eyes.

Nowadays, due to the widespread use of mobile devices, the rapidly increasing demand on network resources impacts the underlying investment on high-performance hardware devices, which then affects the cost of a network architecture. Then, on the other hand, higher visual quality can then impact the number of subscribers and subsequently the top line income of a networking firm. As a result, a good visual quality control system with effective utilization of network resources for a networking firm is desirable. In the following, we shall elaborate the QoS and QoE design issues. Then a multimedia control framework will be proposed to offer satisfactory perceptual to ubiquitous multimedia subscribers. Its improvements will be thoroughly discussed.

VISUAL METRICS AND CONTROL FRAMEWORKS

Measuring Metrics

From the signal processing algorithmic design, multimedia sessions may consist of voices, images, videos, and data. Different signal types use different encoding/decoding algorithms for storage and transmissions. For example, MPEG (Motion Picture Experts Group) is a family of standards used for coding audio-visual information, e.g., movies, video, music, in a digital compressed format. The JPEG (Joint Photographic Experts Group), GIF (Graphic Image File), BMP (bitmap) are examples of image encoding data formats. Among them, bitmap image takes more memory spaces with sharper imaging quality because it has 256 quantization levels for each of the three base colors. An JPEG image is coded with a lossy Discrete Cosine Transform (DCT). It uses less memory space with a lower visual quality.

Peak signal-to-noise ratio (PSNR) is an easy-to-use error measurement metric, and is widely used for providing quantitative evaluation of receiving multimedia quality. Indeed, the PSNR ratio is more or less a subjective measurement technique, and it may fail to reflect what appear in images. As shown in Fig. 1, two images with identical PSNRs. But the one shown in Fig. 1(b) appears to give inferior visual quality (Winker & Mohandas, 2008). This is due to the local accumulation of errors on some nearby pixels. With the nonlinear functionality of retina in human vision system, the perceived quality can be drastically misleading. With the errors spread evenly across all pixels in the image, the one shown in Fig. 1(a) may be considered with better encoded quality

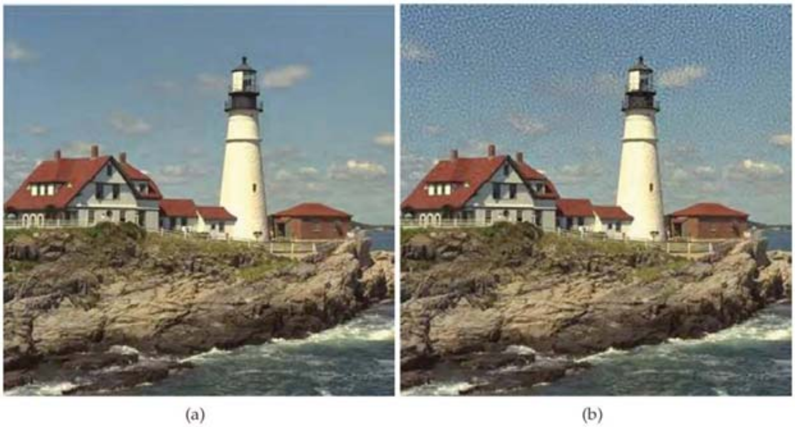
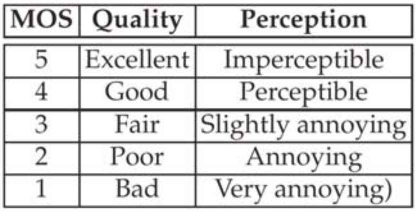


Figure 1. Human vision system and images with identical PSNRs (Winker & Mohandas, 2008).

As a result, PSNR may not be able to reflect the visual perceptual quality of multimedia content. That is, some perceptually poor and appealing images may have identical PSNRs (Grega et al., 2008). At this current moment, there is no conclusive measure that should be commonly accepted as the right measure for quantifying QoE. Hence, the Mean Opinion Score (MOS) is then recommended by the International Telecommunication Union (ITU)

(ITU-T Recommendation P.800, 1996). It is a subjective measure, and a number of users should rate the quality using a five-point scale from 1 to 5, inclusively, as listed in Table 1. The MOS is the arithmetic mean of all individual scores for judging the quality of a delivering video.

Table 1. Mean Opinion Score (MOS).



MOS	Quality	Perception
5	Excellent	Imperceptible
4	Good	Perceptible
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying)

Then from the perspective of network architectural design, the quality of a transmission session through the Internet is usually characterized by the term Quality of Service (QoS) using parameters, such as packet loss rate, transmission bandwidth, queue length, jitter, and delay etc. Each of these parameters can be used for performance analysis. Currently, there are a few standardized QoS associated network designs, for example, the Differentiated Service (Blake et al., 1998) system model through the recommendation of the Internet Engineering Task Force (IETF). In general, network operators have to monitor and manage network resources properly in order to deal with network congestion problems and packet loss issues so as to meet different QoS requirements. Network-introduced errors may be the sources of decoded signal errors. For example, as observed from the two rugby team pictures shown in Fig. 2, both of them suffer identical overall loss conditions in networks. The errors are spread across all pixels in Fig. 2(a) which gives a more appealing appearance. However, the errors are localized in Fig. 2(b), which may irritate the acceptance of a subscriber.

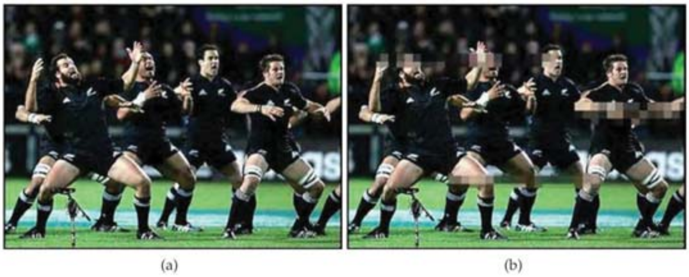


Figure 2. Network condition: 1% packet loss rate, 10 ms delay, 50 µs jitter, 500 kbps bandwidth (Winker & Mohandas, 2008).

Through these observations, QoE and QoS may be related but not in a linear way. What can be the proper way to assert the quality of online multimedia service? Objective evaluation methods are simpler because user inputs are usually not required. For example, data can be retrieved from network-level measurements, e.g., packet loss rate, or media-level measurements, e.g., PSNR, as the input for quality assessment. However, as discussed, they may not be able to judge delivered multimedia quality.

The subjective method using MOS for QoE can be a feasible solution. However, each subscriber may give a completely different MOS result. Furthermore, the same person under emotional stress can give a completely different score. As a result, we can consult some nonlinear functional methods for judging delivering visual quality. And some parameters are not considered in these methods. For example, the loss of volume is not considered by the Perceptual Evaluation of Speech Quality (PESQ), which is recommended by the ITU for determining the quality of a speech signal, in order to make the model tractable. Also, the latency between viewers and the video is not considered in Video Quality Measurement (VQM) model (Rajagopalan, 2010). Higher-order variations, i.e., the burstiness, of end-to-end delay and loss are not considered in many models for assessing VoIP quality.

Similar to PESQ, there are a few subjective methods that can assist in judging visual quality. One of the them is called Visual Differences Predictor (VDP). It is used to characterize the retina response curve. Although the computation complexity is relatively high, it can be used for quantifying image quality based on a reference image.

Examples of quality control frameworks

There are some basic architectural designs (Agarwal et al., 2008; Huang et al., 2008; Lum & Lau, 2002) for serving multimedia traffic adaptively according to varying network condition. The common goal is to improve the visual quality of multimedia traffic at recipients. In (Lum & Lau, 2002), proxy server or intermediate network server is used to relay and re-adapt information content for changed networking condition. As of today, there are a large amount of proxy video servers deployed across the Internet today. However, most of these proxies are for relay purpose only.

As reported in (Agarwal et al., 2008), a controlled testbed for experimenting video traffic delivery using peer-to-peer (P2P) streaming has been used. The results have indicated that tested P2P streaming systems carry significant overhead (up to 35% over the video stream size) with an average start-up delay of about 11 sec. Besides, an additional video buffering time of 30 sec is needed to combat packet arrival jitters for video playback. Despite these drawbacks, the P2P systems are robust regarding peer churn, and generate larger captured P2P bandwidth than using an over-provisioned server. Furthermore, they have found that quantitative measure such as PSNR, which is used among many P2P video streaming research reports, can not properly reflect the QoE.

Another investigation on P2P streaming can be found in (Huang et al., 2008). The paper offers a generic design framework, and identifies different building blocks in a system, for example, the file segmentation strategy, replication strategy, content discovery and management, piece/chunk selection policy, transmission strategy and authentication. The goal is to achieve a scalable system with efficient replication strategies for offering user-level satisfaction. A new fluency index has then been introduced as a performance measure for evaluating the health of the systems and the user satisfaction. Typically, the index measures the fraction of time a user spent watching a movie with respect to the total time spent on both the waiting and watching times. This design closely relies on the underlying network performance, instead of attempting to interpret and serve different types of content information. That is, the accuracy of the fluency index regarding perceptual quality has not been examined in the paper.

Then in (Law & Leung, 2003), a set of application programming interfaces (APIs) has been designed for programmable nodes in networks. This implies nodes on the Internet can function together in the form of loosely-coupled computing devices. This indicates that adaptation of traffic can be made inside the Internet. But the operation details for programs to execute must be carefully controlled by network and content service providers. In general, overlay networks provide better security control to network providers and code distribution flexibility to application providers. However, the response times may be slightly longer than those with programmable node concepts. For example, BitTorrent is one simple broadcasting mechanism for code distribution across end-users' computers, which operate as virtual network servers. To advance the design, structured overlays, such as using Distributed Hash Tables (DHTs) (Stoica et al., 2001), can be used.

In (Chen et al., 2009), a proposed framework with QoE consideration known as OneClick is proposed. Its operations is trivial. The client informs the server system directly regarding receiving perceptual quality. Upon viewing the multimedia content, when a user finds the receiving quality of content annoying, he or she can click certain button repeatedly to indicate his or her dissatisfaction. Therefore, a session with a larger number of clicks indicates a poorer receiving perceptual quality. The OneClick design can be considered as a reciprocal of Mean Opinion Score (MOS) (ITU-T Recommendation P.800, 1996). The OneClick may offer real-time response to the server system, although it has not been examined (Chen et al., 2009).

MULTIMEDIA AGENT FRAMEWORK

A network infrastructure is shown in Fig. 3. Initially the traffic communicating between a mobile client C and service provider S is traveling over the Path 1 as shown. Upon moving, the client C may have arrived at another location, and the communicating path has been switched to Path 2. The associated networking parameters might have completely changed, for example, the bottleneck link bandwidth and propagation delay, etc. As a result, the amount of data flow and traveling latency could be completely different, which can then impact the perceptual quality of receiving multimedia traffic.

Our proposed quality control framework is called Multimedia Agent Framework (MAF). The goal is to adapt traffic content to changing network constraints dynamically. The foundation of the framework is based on agent technology. A few basic components are defined for the system to operate properly. In Fig. 3, a functional connection is consisted of at least a content provider S, a mobile client C, and two agents, which sit at the edges on the Internet. Depending on the direction of the traffic flow, the agents are generically named the Ingress Agent (IA) and Egress Agent (EA). For the depicted traffic flow from a server to a client, the agents connecting to the source and destination are called ingress agent and egress agent respectively. Since wireless connections are not always stable, a mobile client may encounter different perceptual experiences while traveling. For example, the access bandwidth between C and egress agent EA2 is 54 Mbps for 802.11g, while the one between C and EA1 can be 11 Mbps for 802.11b. And in this paper, we assume that the bottleneck always occurs at the wireless link between egress agent and client.

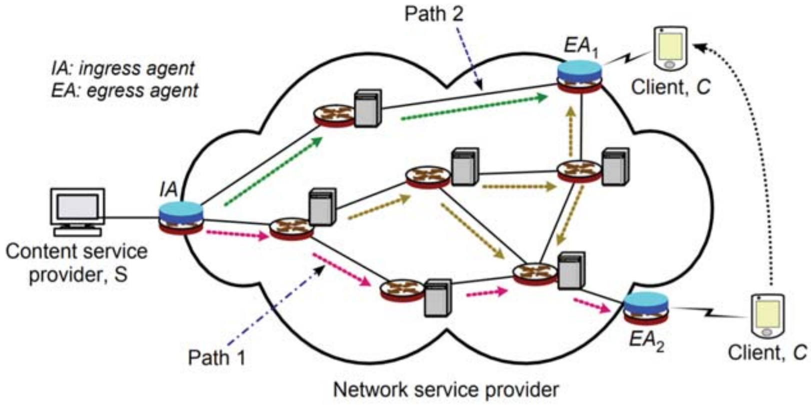


Figure 3. An operating model of the multimedia agent framework.

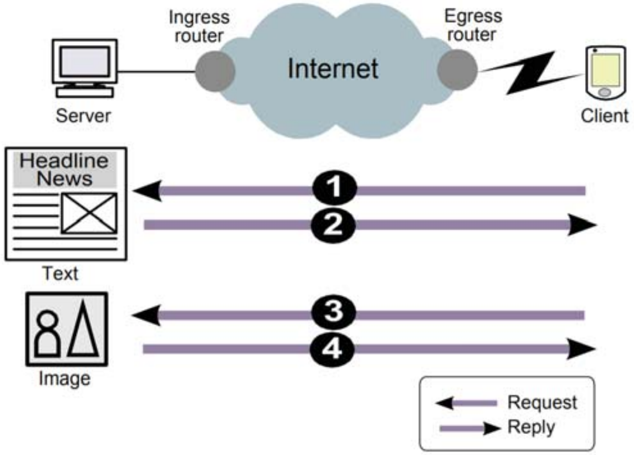
To sustain a satisfactory user experience of a mobile client under changing networking conditions, traffic context can be adjusted accordingly (Banavar & Bernstein, 2002; Noble, 2000; Lum & Lau, 2002) through the proposed quality control framework. Apart from the ingress and egress agents, the framework allows content alternation within the Internet upon permitted by both the content service providers and subscribers. But at the current stage, we focus on the basic operating model which functions between the two agents.

Communication model

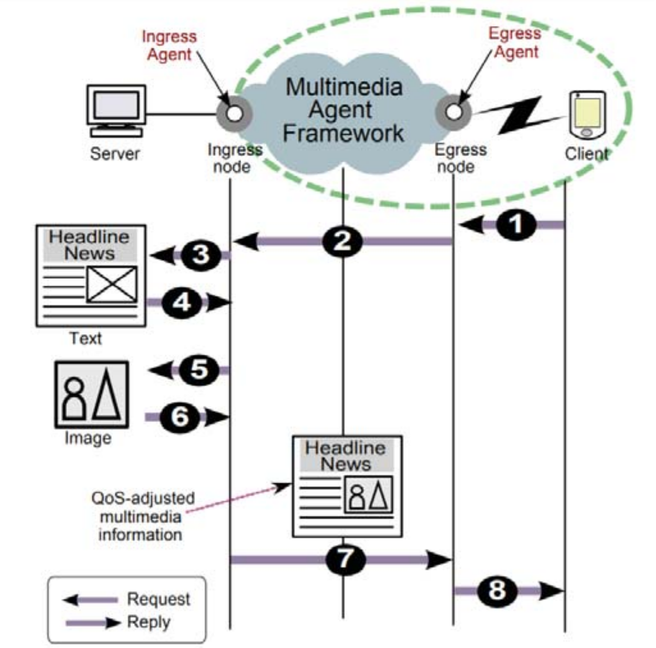
A traditional client-server transaction model is shown in Fig. 4(a). Suppose that a client tries to retrieve a web page, which is consisted of one HTML text document, and one picture file. Two separate HTTP GET requests from the client can be used to obtain the two files from the server. But in multimedia agent framework, a high-level conceptual design feature is enabled through the two agents, as shown in Fig. 4(b). Upon receiving the HTTP GET request from client, the egress agent also delivers information regarding the associated resource constraint of the wireless link to the ingress agent. Then the ingress agent can represent the client to collect both files and examine if the latest available resources are sufficient to receive both files in perfect condition. If not, the ingress agent can adjust the file context to meet the latest available QoS constraints for the client.

To have the framework worked as expected, structural control message flows between the egress and ingress agents have been designed. And these control messages are called "capsules" messages. In order to create a lightweight and effective design, multiple operating phases are introduced, which include: initialization, QoS negotiation, and provision phases. When a client moves from one access point to another access point, packets may be lost temporarily due to incorrect routing. This error can be reduced if the path change indication can be saved within the Internet. At the moment, this layer 3 operation is not investigated in this paper. Our current focus is on the changes of resource constraints, whereas multimedia information may not be able to provide the expected quality at the recipient.

For a newly moving-in mobile user, an egress agent may not be aware of any existing active connections. It starts to take notice, if this client sends a new request, retransmits a request, or the agent receives an incoming redirected messages from a server through the mobile IP protocol (Perkins, 2002; Johnson et al., 2004; Law & Lau, 2010). In either of these cases, the egress agent should start an initialization phase. That is, the egress agent attempts to establish and register a relationship with the corresponding ingress agent. The next stage is to begin a negotiation phase. At the current stage, the information being passed between the agents is for QoS monitoring. This enables the ingress agent to make decision on service class selection with an estimated performance for a specific traffic type. In future, information regarding service subscription should be integrated. This indicates that the right of use of a service class must also be verified if it is covered in its paid services. Then afterward, based on measured QoS parameters, the ingress agent makes decision if any modifications should be applied to the traffic, which should be delivered with satisfaction expectation of the subscriber. The role of QoS monitoring plays an important role in determining the method of adaptations which should be carried out. With the ubiquity of wireless devices, browsers have been the common tools for accessing different information on the Internet. It becomes naturally and important that the multimedia agent framework should start extending the protocol into REpresentational State Transfer (RESTful) model in future.



(a) Traditional client-server model.



(b) Conceptual model of multimedia agent framework.

Figure 4. Network transaction models for multimedia traffic.

Delivering Content Adaptation

The multimedia agent framework can tailor the content to meet the QoE expectations of subscribers. At the moment, content adaptations are carried out at the edges of the Internet, i.e., the ingress and egress agents. The adaptation should depend on the business contract between content and network providers, which may be outside the scope of this chapter. Collectively, we call these nodes the adaptation nodes. For carrying out meaningful operations in these nodes, then a number of information must be learnt and communicated among the ingress and egress agents. As shown in Fig. 5, the types of information should include: 1) user's QoE expectations, 2) properties of content material, 3) latest network status, 4) available network access interfaces of devices.

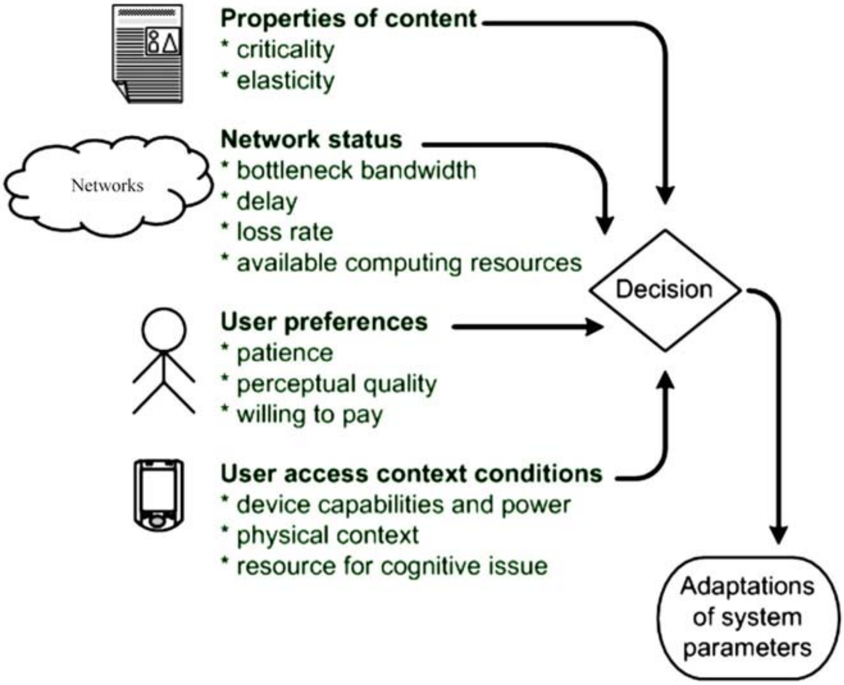


Figure 5. Decision parameters.

User Preferences And Expectation

In order for the infrastructure to work seamlessly and meet the QoEs of subscribers, their preferences and expectations should be set at the initial phases and passed back to the content providers. The associated agents can obtain this information for adaptation purpose, if needed. There are certainly other methods available for this type of information retrieval, e.g., the Service Level Agreement (SLA) through policy-based management. The capability description can follow the defined syntax structure, such as the Composite Capability/Preference Profile (CC/PP) from World-wide Web Consortium (W3C), or Media Feature Sets from IETF.

Some parameters can assist the network to adapt to the QoE expectation of the multimedia traffic:

- Does a subscriber want the content to be retrieved as quickly as possible? – In video streaming, significant variations in delivery time result in jerky video and choppy audio. The resulting video has a lower QoE value than a smoothly playback video.
- Does the subscriber want the critical content to be retrieved unfailingly? – Ranking or priority of content may be desirable.

For multimedia information, more thorough query should be carried out. Questions may have to be asked regarding, for example, the acceptable choices for picture sizes and compression granularities, etc. These collected data can then be combined into a set of meta-information.

Content: Meta-Information

The purpose of having meta-information in the setting is to assist all components in the framework to parse and retrieve desired parameters as quickly as possible. For data files, meta-information may contain file size, version, title, language, and authors. For multimedia content, additional meta-information may include minimal required and desired transmission bit rates, display size, compression ratio, and encoding methods. These extra data can assist the adaptation agents to carry out appropriate operations, if needed. Extensible Markup Language (XML) can be a possible choice for embedding the meta-information regarding the requirements of services and QoEs of subscribers.

Network status

In general, it is common to characterize a network path between two end-systems using available channel bandwidth, end-to-end delay (or round-trip delay), and packet loss rate. With adaptation capability, computation power can be added to indicate if an adaptation node can handle a large number of connections simultaneously. The agents in network may offer computation services for information being delivered from a sender to a receiver. These parameters are the traditional network layer QoS parameters.

Delays and Available Resources

For the framework to operate smoothly, we should establish methods to measure available resources in the networks. For example, in a client-server model, packets from server are relayed from one node to other until they reach the egress node that the client is connecting to. In the following, we examine different delay components incurred along the network infrastructure. A packet has to spend times and delays while traveling through nodes and links along the path, respectively. The delay components are additive, and the aggregate delay across the networks is the summation of delay values in various nodes and connecting links.

Typically, four different types of delays are introduced in networks:

- Propagation delay, f_{Gi} , for a link i;
- Transmission delay, $f_T(s, B)$, for packet size s and available bandwidth B;
- Processing delay, $f_C(c_p, c_b ,s)$, where c_p for incoming packet processing, c_b for outgoing interface determination given a packet with size s;
- Queuing delay, f_{Qj}, for the queue at node j.

The total one-way delay across the networks is the summation of all these delay components of all nodes and links along a path, as shown in (1). With these parameters, other network characteristics such as bandwidth and computation power can be implicitly reflected in the transmission and processing delay components, respectively

$$T_D = \sum_{i=1}^{m} \left(f_{T_i}(s, B_i) + f_{G_i}\right) + \sum_{j=1}^{n} \left(f_{C_j}(c_{p_j}, c_{b_j}, s) + T_{Q_j}(\phi_j)\right) \tag{1}$$

for m links and n nodes.

Adaptation for real-time delivery

Subscribers always expect information being retrieved should arrive briefly after they click the service requests. But they have no idea if they have moved into regions with poor connectivities. Real-time communications are more desirable features for some mobile users, e.g., stock traders. Therefore, in this case, the content carried by the late arrivals of these packets may become unimportant. Hence, in the framework, a parameter W is known as "expected real-time constraint." A subscriber should set the W according to his or her limit of patience on waiting time in the user preference profile; if not, it can be assigned to certain default wait time in system.

Suppose there is a connection being established between a subscriber and a server. A proper path has already determined in the initial setup phase and meta-information has been exchanged. If the total round-trip delay between the client and server is T, such that the forwarding and returning delays are identical, then we have $T = 2T_D$ from Eqn. (1). The expected real-time constraint of a client should not be shorter than the round-trip delay, i.e., $W < T$; otherwise, the retrieved multimedia content can never meet the QoE of the subscriber.

In the case that $W > T$, the TD may vary due to other traffic in networks. It may have chances to violate the real-time constraint requirement. In this scenario, the agent in the framework acts and attempts to adapt the content in order to meet the W requirement. For example, a sudden change of connecting speed, for example, from wired to wireless access link, has happened. Then content adaptation should be carried out in the network core, for example, by reducing the amount of multimedia traffic with compression and reduced frame size, in order to meet the real-time constraint. For example, the packet size has been modified, and the final delivered packet size is changed from s to s_A bytes when it reaches the subscriber.

In general, many multimedia session is composed of more than one traffic stream. In the example shown in Fig. 6, there are three traffic types in one session, and the importance of each of them is ranked. The rank 1 traffic may contain critical data of size 6,600 bytes; the rank 2 traffic may contain compressible multimedia traffic; and the rank 3 traffic carries unimportant data traffic. The ideal curve indicates that the available bandwidth changes linearly. When there is sufficient bandwidth, all traffic in this session can get the network without any adaptations, i.e., when bandwidth is larger than or equal to 50,000 bytes. But the available bandwidth starts decreasing linearly, the rank 3 data traffic should be removed, and then the rank 2 multimedia traffic should be adapted. The size reduction of the rank 2 traffic due to adaptation is not continuous. This hence leads to the staircase structure as shown. When the available bandwidth is small and then only the rank 1 critical data traffic must always be kept for delivery. This happens in this graph when sA is below 6,600 bytes, and both the ideal and real curves should stay flat.

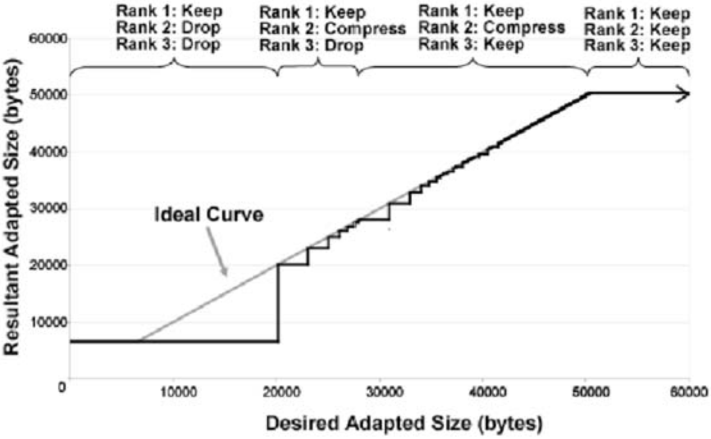


Figure 6. Desired adapted sizes.

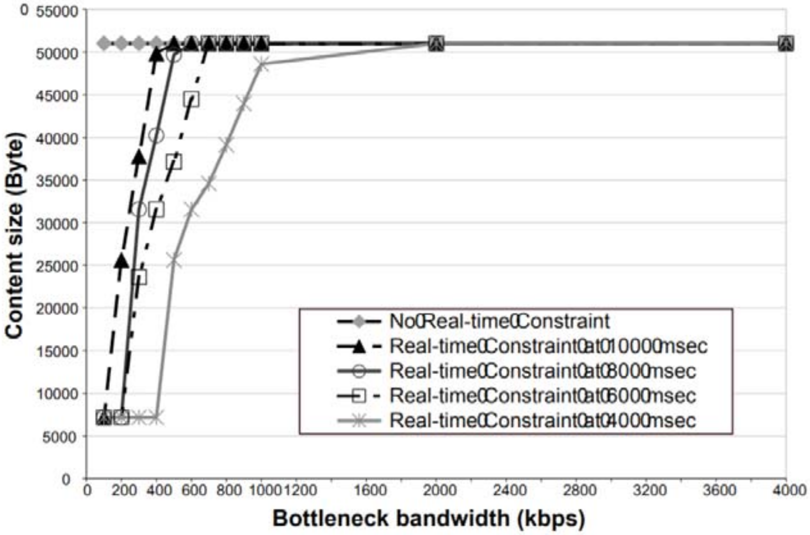
Testbed Evaluation

A testbed consists of ten nodes has been set up to validate the QoS control framework for multimedia traffic. Experiments have been carried out to confirm if the proposed QoS control framework can adapt traffic content to meet the expectation of subscriber. One set of experiments is to examine the delivery of web pages with real-time information component. Another set of experiments is to adapt video stream to meet the link bandwidth constraint in real-time. For both experiments, classification of traffic types have been preset for carrying out expected component adaptation accordingly.

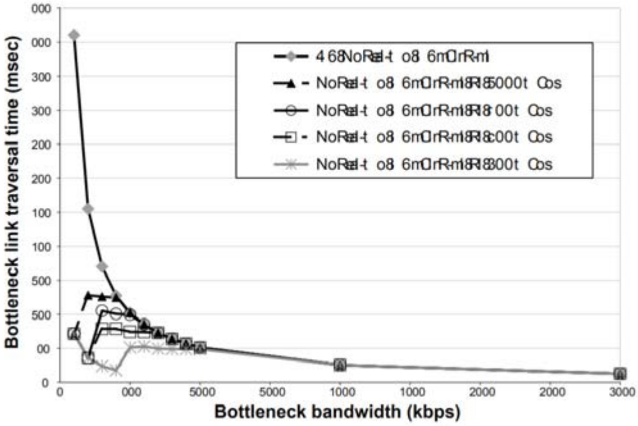
Real-time Delivery of Web Pages

In this set of experiments, a web page consists of multiple informative components is sent every 2.5 seconds. As listed in Table 2, these components are pre-classified into three ranking classes based on their relative importance. Rank 1 information is considered the most important, and it should be sent to the subscriber whenever possible. Then there is a picture in the page. It belongs to rank 2 class, and can be compressed with lower resolutions upon needed. The other rank 3 components in the HTML page are for creating an appealing look of the page only. They are not as important and can be dropped if resources are running low.

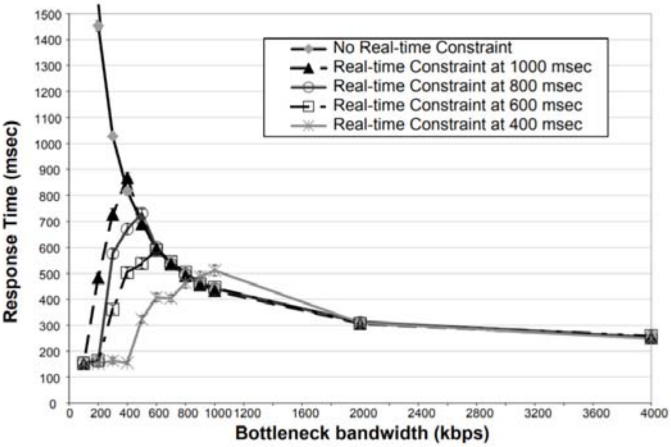
During the experiments, the ingress agent computes the desired adapted size, sADAPTED, after learning the limiting bottleneck bandwidth through receiving the result capsules from the associated egress agent. Measured data are averaged through the last 100 samples. Certainly, wireless access link bandwidth varies due to the mobility of the user, which is simulated through changing the Linux traffic control function. Different access bandwidths are used for testings, which include 115 kbps, 1 Mbps, and 2.4 Mbps. Furthermore, we have imposed different real-time constraints for the delivery of this web page information. The patience limit of a user is interpreted as the real-time delivery constraint. For the tests, the constraints with values of 1000, 800, 600, and 400 msec are used in testbed. This limit setting indicates that the user shall re-click or reload the page when the time is reached. This is the goal of the framework to deliver at least the rank 1 traffic to the subscriber within this time constraint.



(a) Adapting sizes of web pages.



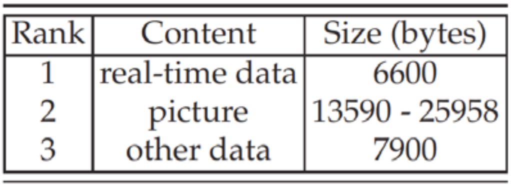
(b) Link traversal times.



(c) Response times.

Figure 7. Web page delivery: a) delivery size; b) traversal time; c) response time.

Table 2. Sizes of various components in a web page: original picture size is 25958 bytes, maximally compressed size is 13,590 bytes.



Rank	Content	Size (bytes)
1	real-time data	6600
2	picture	13590 - 25958
3	other data	7900

In Fig. 7(a), the measured sADAPTED is shown under different user patience limits and bottleneck bandwidth. Fig. 7(b) is obtained by dividing the measured sADAPTED with the bottleneck link bandwidth, thereby giving the bottleneck link traversal duration values. When selective delivery is triggered, the bottleneck link traversal time remains constant while the bandwidth is shrinking. This trend continues until only the rank 1 component is delivered. Thus, the bottleneck link traversal is the reciprocal of the bottleneck bandwidth multiplied by the size of the rank 1 component. Fig. 7(c) shows the resultant response time across the networks. If there is no limit on the patience threshold, the response time exponentially increases when the bottleneck bandwidth linearly decreases. In fact, when the bottleneck bandwidth is reduced to 100 kbps, the response time surges to 12, 016 ± 339 msec due to the exponential increase in the backlog of data waiting to be delivered to the client at the last-mile link.

The response time characteristics under various user patience limits are similar. In general, the response time increases exponentially when the bottleneck bandwidth decreases. This trend continues until they reach their respective adaptation thresholds. From then on, the response time decreases linearly with the reduction of bandwidth. When the bandwidth decreases even further, the response time reaches a point beyond which ranks 2 and 3 in-page components are dropped. The response time stays flat at about 150 milliseconds.

Real-Time Delivery of Images

The next set of experiments is to delivery images across networks. The bottleneck link is the access link of a subscriber. The real-time constraint W is passed from the client preference list to the server, and it can be discovered by the agents in networks. They can detect the existent of real-time constraint of a communication session, and the delivery of images based on the meta-information. If there are multiple good quality paths existing between the two agents, the ingress agent can then have multiple choices in carrying out appropriate operations inside the infrastructure. An exemplary operational detail of using a single path computation in the infrastructure is shown in Fig. 8. The goal of the infrastructure is to fit the image delivery to receiver within the duration of W sec.

The image scaling operation starts when the ingress agent receives the incoming Lena BMP images, as shown in Fig. 9(a). In the experiments, the ingress agent converts them into JPEG images, forwards them along a path, and the running times are measured against the time constraints. Regularly, the egress agent reports the measured bottleneck link bandwidth to the ingress agent. Through the measurements, the image compression ratio can be estimated through operations in the infrastructure. Different scaling parameter, λ, can be set between 1 and 16 inclusively, where 16 is the best quality and 1 is the worst. The next question is to determine if the generated images with selected compression ratio can meet the real-time and bandwidth constraints. Then it leads to design of an initial calibration process. The goal of the process is to find the lowest acceptable bound of the compression ratio for a subscriber.

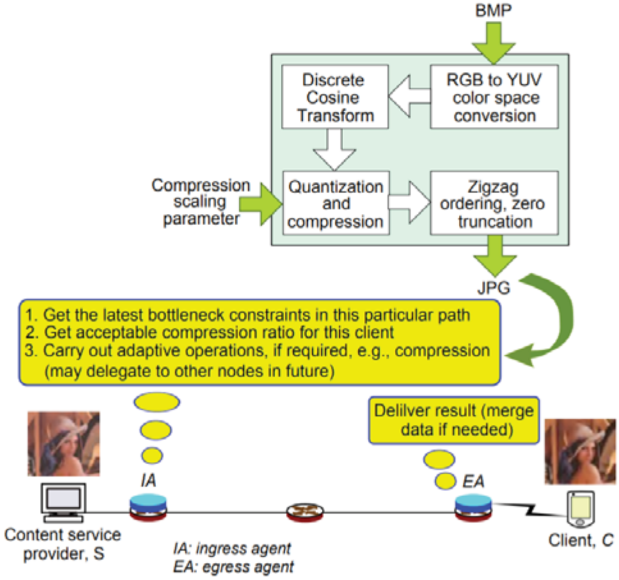
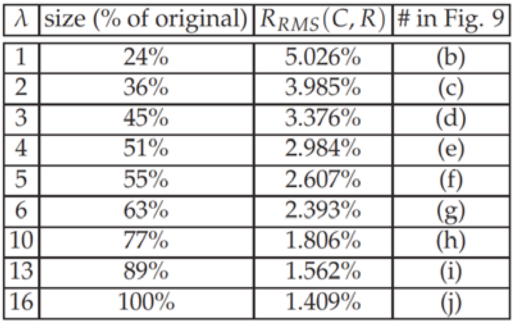

BMP
Discrete Cosine Transform
RGB to YUV color space conversion
Compression scaling parameter
Quantization and compression
Zigzag ordering, zero truncation
JPG
1. Get the latest bottleneck constraints in this particular path
2. Get acceptable compression ratio for this client
3. Carry out adaptive operations, if required, e.g., compression (may delegate to other nodes in future)
Delilver result (merge data if needed)
IA
EA
Content service provider, S
IA: ingress agent
EA: egress agent
Client, C

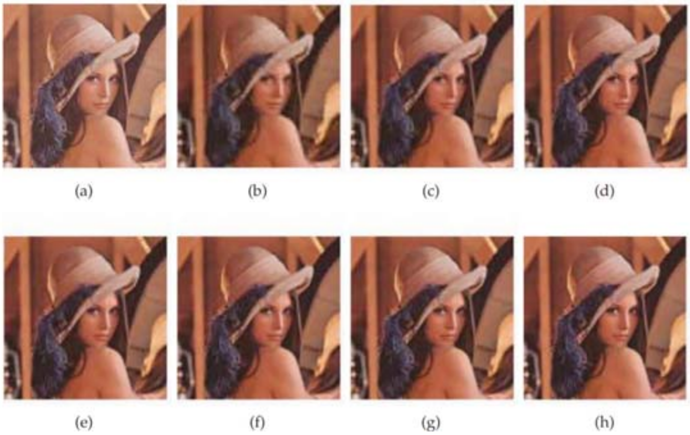
Figure 8. Real-time image scaling operations.

Table 3. Visual quality and $R_{RMS}(C, R)$.



λ	size (% of original)	$R_{RMS}(C, R)$	# in Fig. 9
1	24%	5.026%	(b)
2	36%	3.985%	(c)
3	45%	3.376%	(d)
4	51%	2.984%	(e)
5	55%	2.607%	(f)
6	63%	2.393%	(g)
10	77%	1.806%	(h)
13	89%	1.562%	(i)
16	100%	1.409%	(j)

With different compression ratios, sizes of resulting compressed images are listed in Table 3. For example, the visual qualities of some compressed images are shown from Fig. 9(b) to Fig. 9(j). Through the initial calibration process, a subscriber can actually select the worst quality image that he or she can accept, and inform the content provider. With this information, the infrastructure can then be aware of the minimum QoE expectation of the subscriber. And from the Table, an ingress agent can pick a lowered but acceptable bounding value of the compression scaling parameter to meet the latest networking conditions, if they have just turned worse. There are two parameters that should be considered before carrying out compression operations. The first one is to determine if the real-time constraint can be met. The second is to determine if the image quality can meet the quality expectation of a user. Both conditions must be met; otherwise, the images are dropped at ingress agent, because there are no good reasons to send an unacceptable low-quality images.



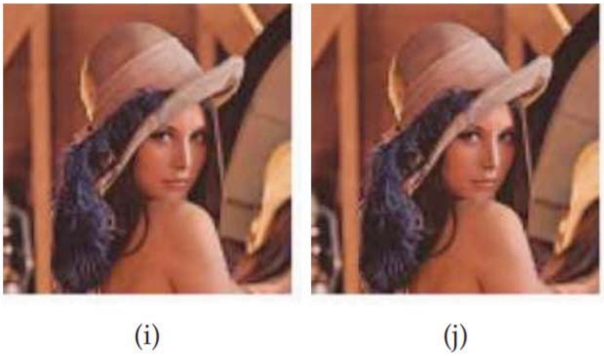


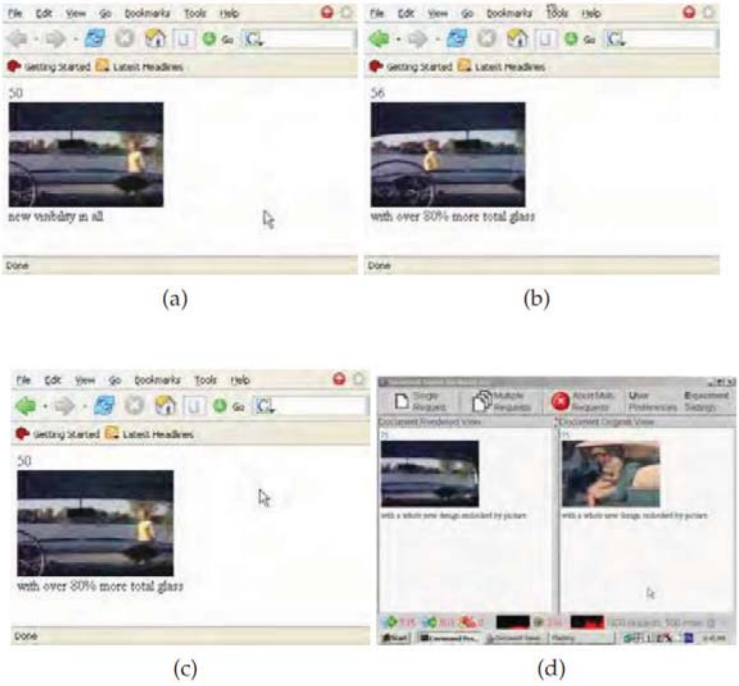
Figure 9. Visual quality of compressed images: (a) original in bitmap format, (b)-(j) compressed in JPEG formats

Real-Time Delivery of Streaming Content

Multimedia applications, such as IPTVs, and online conference meetings, belong to the real-time streaming traffic class with temporal relationship among sending information. Typically, a streaming movie session may consist of three traffic types: video, voice, and/or data such as the subtitle. Actions are spread across multiple consecutive video frames. Apart from the blocking effects in images, video motion may not be reconstructed properly if some motion vectors for decoding are lost. As a result, subscriber could have trouble interpreting the video content. When a video is transmitted across networks, some video frames can be dropped. Then, long overdue frame group not displaying at all causes viewer waiting for long delay, and the content may appear out-of-order. Hence, when there are more frames dropped, the reproducing video becomes jerky and gives poor perceptual interpretation.

It is important that our proposed framework can offer satisfactory experience of real-time video delivery. In the experiments, video frames are extracted from a theatrical screen advertisement promoting the 1955 Chevrolet models. The advertisement movie has long sequences of car movements. Inter-frames smoothness can be observed easily, and the perceived quality can be assessed easily.

For demonstration purpose, this movie clip does not carry voice component which is actually replaced by subtitle. Each frame in video is a 24-bit RGB bitmap image with a size of 63,414 bytes and a visual dimension of 176 × 120 pixels. The sizes of the accompanying subtitles range from 97 to 120 bytes. Compression ratio may differ frame by frame depending on the networking conditions. With the highest compression ratio, compressed frame images have sizes ranging from 8,806 bytes (13.89% of the originals) to 12,674 bytes (19.99%). Two frames are extracted every second for creating one new accompanying subtitle. Eighty-eight frames are extracted. At the server, the video clip is replayed continuously in the experiments. The relationship between compressed size and compression scaling parameter, λ, is encoded using meta-information. Thus, the network infrastructure can efficiently choose the closest adapted size, s_A, with an appropriate compression parameter λ. Then the network nodes execute compression operations according to λ. The real-time delivery scheme of the video experiment is based on a 2-level ranking levels: 1) subtitle is classified as rank 1, R_1 , 2) frame image is classified as rank 2, R_2; where rank 1 has higher priority than rank 2 traffic.



50

new visibility in all

(a)

56

with over 80% more total glass

(b)

50

with over 80% more total glass

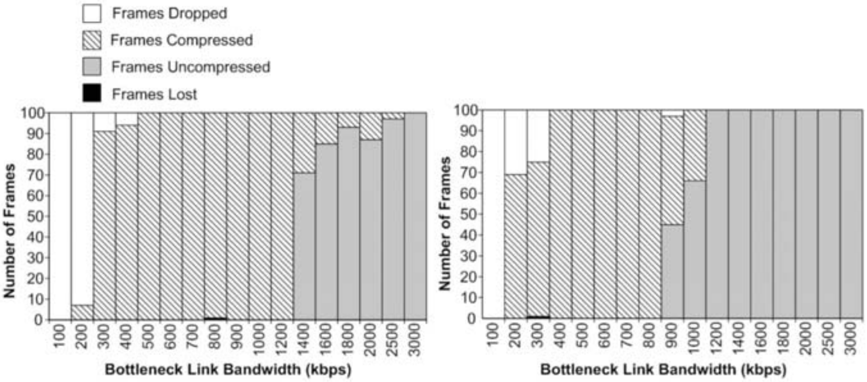
(c)

(d)

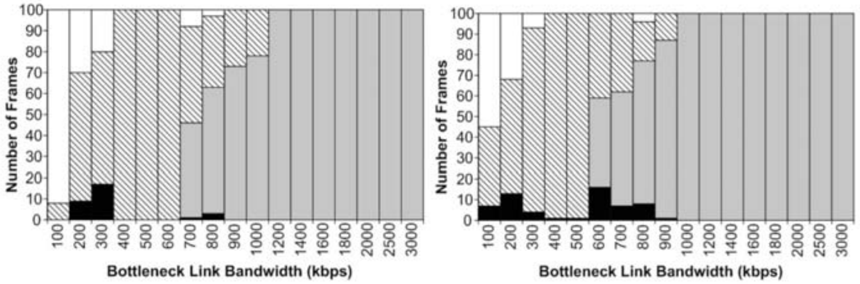
Figure 10. (a) Frame 50, (b) frame 56, (c) time at frame 56, picture at frame 50, (d) comparing viewer with traffic load monitoring.

For video communication, the ingress agent can have choices on delivering frames across the networks. Depending on the latest measured network conditions, the agent can send frames unchanged, compressed, or simply drop the frames. Certainly, frames sent may still get dropped inside the networks due to congestion, and these are called frame lost events. Two snapshots of the movie are shown in Fig. 10(a) and 10(b). When the network delay is too long with a real-time delay constraint being applied. Then frames are dropped at the ingress agent to reduce the traffic load in testbed. For example, as shown in Fig. 10(c), the movie is stuck at frame number 50, but the subtitle has arrived at frame 56. This implies that the proposed pervasive infrastructure can enforce certain minimal QoS quality. In this case, the receiver is still able to see the latest subtitle section, or he or she can listen to the voice part if the voice section exists. An MPEG-1 viewer at the subscriber has been created to compare a locally stored copy with the remotely retrieving video. The viewer, as shown in Fig. 10(d), also reports the packet loss and CPU consumption.

For performance measurements, the link speed of the simulated wireless access shrinks from 3 Mbps to 100 kbps. With different real-time constraints applied, different operations on video frames have been carried out. If bandwidth is abundant, frames are sent uncompressed, as in all constraint cases with link speed of 3 Mbps. But when the link rate reduces, and the real-time constraints is shorter, then more frames are compressed, or dropped. It is always important to observe that, for the cases reported in Fig. 11, the amount of frame losses inside the networks are small. This satisfies the goal of the proposed QoS control framework for multimedia traffic, that is, to deliver as much information as possible to the subscribers.



(a) Real-time constraint of 400 msec. (b) Real-time constraint of 600 msec.



(c) Real-time constraint of 800 msec. (d) Real-time constraint of 1000 msec.

Figure 11. Video stream modifications.

CONCLUSION

In this chapter, an agent-based quality control framework for delivering satisfactory multimedia traffic across the Internet has been designed. The framework is currently built on top of existing networking protocols. The major components in the platform consists of ingress and egress agents. QoS monitoring capsules are regularly sent between agents in order to enable the ingress agent to adapt the content information, while meeting the requirements and expectations of subscribers and end-users. At the current phase, communicating protocol designs between the two agents at the edges of networks have been tested. Furthermore, different traffic types have been thoroughly experimented. Through multimedia content classification, whether it is for real-time or non-real-time, important or unimportant, traffic can be sent to subscribers to meet their expected multimedia quality in our framework. And in future, more thorough investigation shall be carried out to enable routers inside the Internet to assist and relieve the loads of the ingress and egress agents.

REFERENCES

1. Satyanarayanan M. (2001). Pervasive computing: vision and challenges. IEEE Personal Communications, Vol. 8, No. 4, pp. 10-17.
2. Winkler S. & Mohandas P. (2008). The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics, IEEE Trans. Broadcasting, Vol. 54, No. 3, pp. 1–8
3. Chen K.-T.; Tu C.-C.; Xiao W.-C. (2009). OneClick: A Framework for Measuring Network Quality of Experience, IEEE Infocom'09, Brazil.
4. Grega M.; Janowski L.; Leszczuk M.; Romaniak P.; Papir Z. (2008) Quality of Experience Evaluation for Multimedia, Telecommunication Review, pp. 142-158, No. 4, 2008, Poland.
5. Methods for Subjective Determination of Transmission Quality, Inter. Telecommunication Union (ITU-T)
6. Rajagopalan R. (2010). Video Quality Measurements for Mobile Networks. Openwave. http://openwave.com/sites/default/files/docs/solutions/Video Quality MeasurementsWPfinal0710.pdf
7. Rahrer T.; Fiandra R.; Wright S. (2006). Triple-play Services Quality of Experience (QoE) Requirements, DSL Forum, Architecture & Transport Working Group, Technical Report TR-126
8. Chen B. & Cheng H.H. (2010). A Review of the Applications of Agent Technology in Traffic and Transportation Systems, IEEE Trans. Intelligent Transportation Systems, Vol. 11, No. 2
9. Lee J.-S. & Hsu P.-L. (2007). Implementation of a Remote Hierarchical Supervision System Using Petri Nets and Agent Technology, IEEE Trans. Systems, Man, Cybernetics – Part C, Vol. 37, No. 1
10. Blake S.; Black D.; Carlson M.; Davies E.; Wang Z.; Weiss W. (1998). An Architecture for Differentiated Services, RFC 2475. Internet Engineering Task Force, 1998.
11. Law K. L. E.; Leung R. (2003). Design and Implementation of Active Network Socket Programming, Microprocessors and Microsystems J., Vol. 27, Issues 5-6, pp. 277-284, June 2003.
12. Agarwal S.; Singh J. P.; Mavlankar A.; Bacchichet P.; Girod B. (2008). Performance of P2P Live Video Streaming Systems on a Controlled Test-bed, Tridentcom'08, Mar. 18-20, 2008, Innsbruck, Austria.
13. Huang Y.; Fu T. Z. J.; Chiu D.-M.; Lui J. C. S.; Huang C. (2008). Challenges, Design and Analysis of a Large-scale P2P-VoD System,

ACM Sigcomm’08, Aug. 17-22, 2008, Seattle, USA.

14. Jrad Z. E.-F.; Benmammar B.; Correa J.; Krief F.; Mbarek N. (2005). A User Assistant for QoS Negotiation in a Dynamic Environment Using Agent Technology, 2nd IFIP Inter. Conf. Wireless Optical Communications Networks (WOCN), pp. 270-274
15. Pratistha I. M.; Zaslavsky A.; Cuce S.; Dick M. (2005). Improving Operational Efficiency of Web Services with Mobile Agent Technology, IEEE/WIC/ACM Inter. Conf. Intelligent Agent Technology, pp. 725-731
16. Banavar G. & Bernstein A. (2002). Software infrastructure and design challenges for ubiquitous computing, Communications of the ACM, Vol. 45, No. 12, pp. 92-96
17. Noble B. D. (2000). System support for mobile, adaptive applications, IEEE Personal Communications, pp. 44-49
18. Lum W. Y. & Lau F. C. M. (2002). A context-aware decision engine for content adaptation,” IEEE Pervasive Computing, pp. 41-49
19. Byers J. W.; Considine J.; Mitzenmacher M.; Rost S. (2004). Informed Content Delivery Across Adaptive Overlay Networks, IEEE/ACM Trans Networking, Vol. 12, No. 5
20. Perkins C. (2002). IP Mobility Support for IPv4, IETF, RFC 3344
21. Johnson D.; Perkins C.; Arkko J. (2004). IP Mobility Support for IPv6, IETF, RFC 3775
22. Law K. L. E. & Lau C. (2010). Mobility Service Agent, 3rd IEEE Workshop on Wireless and Internet Services (WISe), in conjunction with 35th IEEE Conference on Local Computer Networks (LCN)
23. Daly S. (1993). The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity, Digital Images and Human Vision, Editor: A.B. Watson, pp.179-206, MIT Press, Cambridge MA
24. Stoica T.; Morris R.; Karger D.; Kaashoek F.; Balakrishnan H. (2001). Chord: A scalable Peer-To-Peer lookup service for internet applications, ACM SIGCOMM’01, Aug. 2001.

SECTION 4:

MULTIMEDIA APPLICATIONS IN EDUCATION

CHAPTER 14

RELIABILITY AND VALIDITY OF ONLINE INDIVIDUALIZED MULTIMEDIA INSTRUCTION INSTRUMENT FOR ENGINEERING COMMUNICATION SKILLS

Atef F. Mashagbh[1], Rosseni Din[2], M. Khalid M. Nasir[2], Lilia Halim[2], Rania Ahmad Al-Batainah[1]

[1]Al al-Bayt University, Mafraq, Jordan
[2] UniversitiKebangsaan Malaysia, Selangor, Malaysia

ABSTRACT

The utilization of Multimedia Instruction (MI) in teaching and learning is growing rapidly. The combination of various media assists educational reform, and is important to the improvement of education outputs. The MI use has been a challenge to educators especially in Jordan. This study aimed to re-calculate the reliability and validity of online individualized MI instrument in a new Online Individualized Multimedia Instruction

Citation: Mashagbh, A. , Din, R. , Nasir, M. , Halim, L. and Al-Batainah, R. (2019), Reliability and Validity of Online Individualized Multimedia Instruction Instrument for Engineering Communication Skills. Creative Education, 10, 3041-3048. doi: 10.4236/ce.2019.1012228.

Copyright: © 2019 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

(OIMI) framework for engineering communication skills. In this study, this model designates the multimedia instruction as one of the latent variables, to be measured by six observed variables, which are modality, contiguity, personalization, coherence, redundancy, and signaling. Data collected and tested from 166 engineering learners. Confirmatory factor analysis using AMOS was conducted to obtain three best-fit measurement models. The results showed evidence of a five-dimension measurement model for MI except for coherence. This result enlightens the model, which includes explanations of Mayer's Cognitive Theory of MI and multimedia instructional in Practice.

Keywords Multimedia Instruction, Learning Style, OIMI

INTRODUCTION

Engineering field is accountable for major industrial, technological, and economical progress in human history. The engineer who design and build things plays a major role in keeping the society running smoothly. Well qualified engineers are always needed in a growing society. The format of engineering communications can vary widely, from summaries of calculations, to short technical messages, to oral presentations, to drawings describing data or machinery. In Jordan, a number of engineering branches and specializations are taught by the public and private universities. Each university tries to offer unique specializations to attract students enrolment. Nevertheless, studying abroad continued even after the establishment of the engineering colleges in Jordan (Aqlan et al., 2010). These issues due to the reality where engineering education traditionally relies more on technical skills and less communication skills (Corrello 2012). Engineering instructors seem to conduct inappropriate teaching techniques in order to develop engineering student communication skills (Nasir et al., 2018; Baharudin et al., 2018). The study sought to examine the reliability and validity of an instrument use to measure Multimedia Instruction (MI) constructs. MI is used to deliver communication skills course for engineers. A research question of whether the measurement scale for MI is construct-valid was formed to guide the study.

PAST STUDY

Multimedia Instruction (MI)

Multimedia learning is the learning that "occurs when people build mental representations from words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video)" (Mayer, 2009). Mayer had investigated a number of instructional design principles, which had suggested ways of creating multimedia presentations intended to promote multimedia learning. Mayer's Cognitive Theory of MI acknowledged that for successful learning to occur, information must be synthesized into a coherent model of knowledge and integrated into long-term memory all within the working memory store.

Multimedia has remarkable impact on learning (Zainal et al., 2018; Gabarre et al. 2018; Ahmad et al. 2016; Azizul & Din, 2016). It has emerged in various form of recourses and equipment which can be use to aid the instructor and learner effort in ensuring effective learning environment (Ahmad et al., 2016). When meeting all these conditions, there would be an opportunity to increase efficiency and effectiveness of engineering communication skills specifically in Al al-Bayt University in Jordan.

The usage of MI in engineering education has changed the practice of structured engineering. MI in the circumstance of structured engineering enhances the engineering process. It can be used in the design and construction of a building including safety management in construction work, and computer-aided design and construction. The MI through video conferencing, shared-screen computing and remote multimedia links on construction projects could have a significant impact on inter-professional communication in engineering education (Di Gironimo et al., 2013). In short, this approach for learning with high-quality instructional materials can reduce lecture time and learners as well as instructors efforts (Khalid & Quick, 2016).

Learning Styles

According to Romanelli at el., (2009) "learning styles" are "characteristic cognitive, effective, and psychosocial behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment". Mismatches exist between common learning styles of engineering students and traditional teaching styles of engineering professors. In consequence, students become bored and inattentive in class, and do poorly on tests. According to Felder (2002), learning-style model classifies students according to the ways they receive and process information on a number of scales pertaining. A model intended to be particularly applicable to engineering education. Sensing and intuitive learners, visual and verbal learners, active and reflective learners sequential and global learners are describe in engineering student learning styles model (Felder & Silverman, 1988; 2002). Din (2010, 2017) also acknowledged that learning style can be divided into social and sensory learning style.

METHODOLOGY

This study uses a quantitative research approach via a survey distributed at a University in Jordan. It is about measuring engineering students' level of MI. It was constructed based on reviewed literature specifically grounded on Mayer's Cognitive Theory of Multimedia Learning (Mayer 2010; 2009).

Respondents

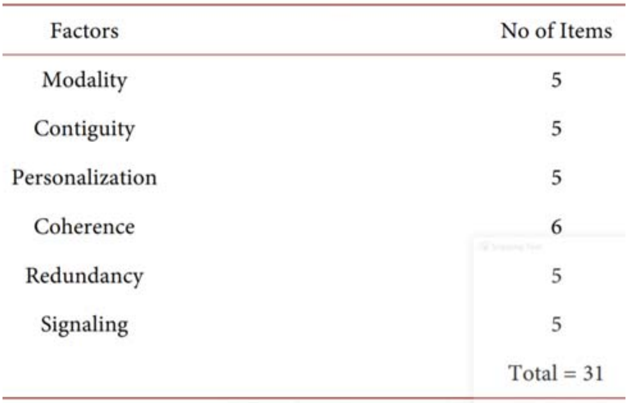
The sample was 166 engineering learners from Al-Bayt University in Jordan who enrolled in the first semester of an academic year. They were selected purposively from four major engineering courses: communication skills course; the provisions of the building; skills practice of the profession course; technical skills course. The sample size was still within the acceptable range (Hoyle, 1995).

Instrument

The MI item was developed specifically based on Mayer's Cognitive Theory of Multimedia Learning (Mayer 2010; 2009). The instrument consists of thirty-one items; five Items for each indicator (modality, contiguity, personalization, redundancy, signaling), while six items were for coherence measure. The measurements scale is a Likert-type scale, which has 1 to 5 scales; 1 equals "strongly disagree" and 5 equals "strongly agree." 1 represents the lowest and most negative impression on the scale, 3 represents an adequate impression, and 5 represents the highest and most positive impression. In addition, a response category for "Not Applicable" was added for each Likert item. Table 1 shows the contents of the MI measure after the content validation.

CFA was conducted on the hypothesized six-factor structure model using AMOS model-fitting program. The program adopted maximum likelihood estimation to generate estimates in the full-fledged measurement model. At the beginning, this model indicates the latent variable, MI, to be measured by six observed variables (modality, contiguity, personalization, coherence, redundancy, and signaling). The construct MI was indicated by six measured indicators and was identified. It had more degrees of freedom than the paths to be estimated.

Table 1. MI factors



Factors	No of Items
Modality	5
Contiguity	5
Personalization	5
Coherence	6
Redundancy	5
Signaling	5
	Total = 31

To assess the fit of the measurement model, the analysis relied on a number of descriptive fit indices, which included the 1) relative chi-square (χ2 /df), 2) comparative fit index (CFI), 3) Tucker-Lewis coefficient (TLI), and 4) root mean square error approximation (RMSEA). Hair et al. (2006) suggest the use of relative chi-square (chi-square/df) as a fit measure.

RESULTS

The estimated six-factor model for MI using the data drawn from 153usable and completed responses. A value of approximately .08 or less for RMSEA shows a reasonable error of estimation. The items from each scale were assumed to load only on the respective latent variable. The results indicate that the parameters ranging from .11 to .97. In the MI case, all of the coefficients were acceptable (>.7) except for the Coherence indicator which was .11 (Hair et al. 2006; Arbuckle 1997; James et al. 2006). The CFI (.922) exceed the threshold of .90 indicating a good fit (Hair et al. 2006, Arbuckle 1997, James et al. 2006), while the TLI (.870) fit indicators did not exceed the threshold of .90, indicating a poor fit (Hair et al. 2006; Arbuckle 1997; James et al. 2006). The root-mean square error of approximation (RMSEA = .160) was >.08, Chi-square (χ2) was 44.083 with degree of freedom (9) and p value = 0 (normally acceptable at p > .05) reflecting a possible fit problem (Hair et al. 2006; Arbuckle 1997; James et al. 2006).

Since the hypothesized model was found to be contaminated (p value = 0 and TLI (.870) is less than .9), the model was revised. The revised model was achieved after examining the modification indices in order to correlate the measurement errors of the signaling with contiguity, as well as correlate the measurement errors of the redundancy with personalization, and contiguity and modality factors. After a number of iteration to fit the model the results reflect a possible fit problem, yet no possible modifications could be made. As suggested by Aryee and Lee (2005) the researcher decided to delete a factor with the lowest loading where the "coherence" factor was dropped. To validate the likelihood of the revised five-factor model, another CFA was applied on the same sample. The magnitude of the factor loadings in the revised model was substantially significant with CFI = .997, TLI =. 986 and chi-square = 3.971. The parameters were free from offending estimates, ranging from .72 to. 99. The CFI (.998) and TLI (.983) fit indicators exceeded the threshold of .90, indicating a very good fit (Hair et al. 2006, Arbuckle 1997). The root-mean square error of approximation (RMSEA = .72) also indicated a good fit (Hair et al. 2006).

In this revised model, the chi-square (χ2) with a value of 1.784, with degree of freedom (1) successfully met the required threshold of .05) hence indicates that the test failed to reject the hypothesized model. The procedures established the model in Figure 1 as the validated confirmatory measurement model. The cronbach alphas for the five sub-constructs after CFA are range from .814 to .879 (Modality = .839, Contiguity = .814, Personalization = .879, Redundancy = .873 and Signaling = .813) while cronbach alpha for the whole section measures = .927. It is worth to mention cronbach alpha for coherence" factor which was dropped was .651. Overall result indicates that the test failed to reject the hypothesized model. Thus, the procedures established that the model in Figure 1 was a validated confirmatory measurement model.

DISCUSSION

The study was able to validate the MI model, which is measured by five observed variables: modality, contiguity, personalization, redundancy, and signaling. As proposed in literature, the study offered evidence that (five out of the six dimensions excluding the coherence indicator) the five-dimension measurement model did generate from data collected from Al al-Bayt university engineering learners. The result did not establish any basis, which can be used to claim that the MI model is incorrect. Thus, MI measurement model can be explained by five factors namely, redundancy, contiguity, personalization, modality, and signaling. This result was consistent with several literatures on MI (Gerjets at el. 2004; Ibrahim & Callaway, 2012a; 2012b; Sorden, 2012; Clark & Mayer, 2011).

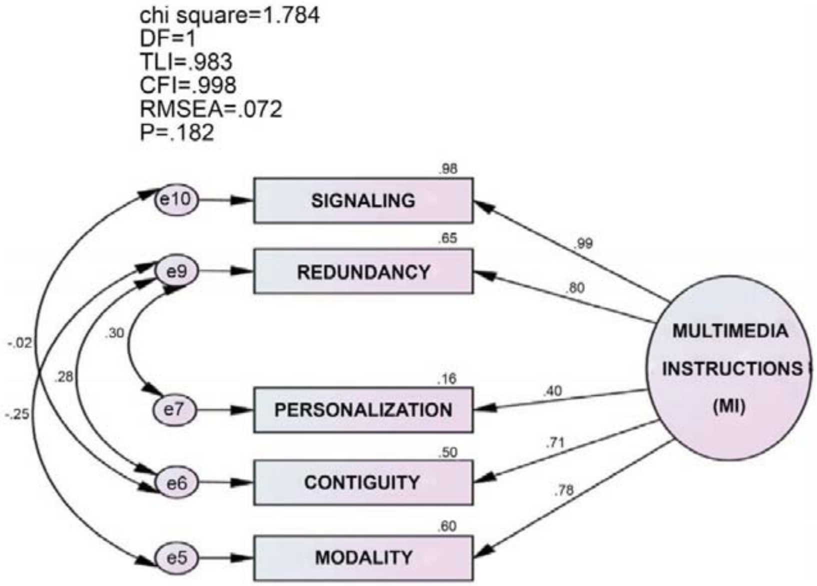


Figure 1. Revised CFA measurement model for MI.

This conception represents a major adjustment in the way engineering faculties have usually developed engineering communication skills. The results of the present study are relevant to give insights for theorists, learners, academic staff and knowledge management system designers and developers towards the goal of achieving effective learning and teaching environment for engineering communication skills. In addition, these findings would assist engineering learners with differentiated learning style preferences to learn and practice engineering communication skills knowledge by integrating MI theories into the learning environment via Blackboard Course Management System especially in Jordan higher learning education.

ACKNOWLEDGEMENTS

This publication and research were funded by the Faculty of Education and Universiti Kebangsaan Malaysia under the project code GG-2019-046 and PP-FPEND-2019, DCP-2017-202/3, GG2018-011 respectively. The authors would also like to express special thanks to all personalized education research team members, students, researchers, graduate and research assistants for their help and contribution in conducting this project.

REFERENCES

1. Ahmad, M., Badusah, J., Mansor, A. Z., Abdul Karim, A., Khalid, F., Daud, M. Y., Din, R., & Zulkefle, D. F. (2016). The Application of 21st Century ICT Literacy Model among Teacher Trainees. Turkish Online Journal of Educational Technology, 15, 151-161.
2. Aqlan, F., Al-Araidah, O., & Al-Hawari, T. (2010). Quality Assurance and Accreditation of Engineering Education in Jordan. European Journal of Engineering Education, 35, 311-323. https://doi.org/10.1080/03043797.2010.483608
3. Azizul, S. M. J., & Din, R. (2016). Teaching and Learning Geometry Using Geogebra Software via Mooc. Journal of Personalized Learning, 2, 39-50.
4. Baharudin, H., Nasir, M. K. M., Yusoff, N. M. R. N., & Surat, S. (2018). Assessing Students' Course Satisfaction with Online Arabic Language Hybrid Course. Advanced Science Letters, 24, 350-352. https://doi.org/10.1166/asl.2018.12005
5. Clark, R. C., & Mayer, R. M. (2011). E-Learning and the Science of Instruction (3rd ed.). San Francisco, CA: Pfeiffer. https://doi.org/10.1002/9781118255971
6. Di Gironimo, G., Mozzillo, R., & Tarallo, A. (2013). From Virtual Reality to Web-Based Multimedia Maintenance Manuals. International Journal on Interactive Design and Manufacturing, 7, 183-190. https://doi.org/10.1007/s12008-013-0185-0
7. Din, R. (2010). Development and Validation of an Integrated Meaningful Hybrid E-Training (I-Met) for Computer Science: Theoretical-Empirical Based Design and Development Approach. Bangi, Malaysia: Faculty of Technology and Information Science, Universiti Kebangsaan Malaysia.
8. Din, R. (2017). Asas Pendidikan dan Kejurulatihan ICT: Integrasi Teori, Media, Teknologi dan Reka Bentuk Pembelajaran. Rangi: Penerbit UKM.
9. Felder, R. M. (2002). Designing Tests to Maximize Learning. Professional Issues in Engineering Education and Practice, 128, 1-3. https://doi.org/10.1061/(ASCE)1052-3928(2002)128:1(1)
10. Felder, R. M., & Silverman, L. K. (1988). Learning and Teaching Styles in Engineering Education. Engineering Education, 78, 674-681.

11. Gabarre, S., Gabarre, C., & Din, R. (2018). Personalizing Learning: A Critical Review of Language Learning with Mobile Phones and Social Networking Sites. Journal of Advanced Research in Dynamical and Control Systems, 10, 1782-1786.

12. Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. Instructional Science, 32, 33-58. https://doi.org/10.1023/B:TRUC.0000021809.10236.71

13. Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate Data Analysis (Vol. 6). Prentice, NJ: Prentice Hall.

14. Hoyle, R. H. (1995). Structural Equation Modeling: Concepts, Issues, and Applications, an Introduction Focusing on AMOS. Thousand Oaks, CA: Sage Publications.

15. Ibrahim, M., & Callaway, R. (2012a). Assessing the Correlations among Cognitive Overload, Online Course Design and Student Self-Efficacy. In Society for Information Technology & Teacher Education International Conference.

16. Ibrahim, M., & Callaway, R. (2012b). Implications of Designing Instructional Video Using Cognitive Theory of Multimedia Learning. Vancouver, Canada: American Educational Research Association (AERA).

17. Khalid, N. M., & Quick, D. (2016). Teaching Presence Influencing Online Students' Course Satisfaction at an Institution of Higher Education. International Education Studies, 9, 62. https://doi.org/10.5539/ies.v9n3p62

18. Mayer, R. E. (2009). Multimedia Learning (Vol. 2). New York: Cambridge University. https://doi.org/10.1017/CBO9780511811678

19. Mayer, R. E. (2010). Applying the Science of Learning. Upper Saddle River, NJ: Pearson.

20. Nasir, M. K. M., Surat, S., Maat, S. M., Abd Karim, A., & Daud, M. Y. (2018). Confirmatory Factor Analysis on the Sub-Construct of Teaching Presence's in the Community of Inquiry. Creative Education, 9, 2245-2253. https://doi.org/10.4236/ce.2018.914165

21. Romanelli, F., Bird, E., & Ryan, M. (2009). Learning Styles: A Review of Theory, Application, and Best Practices. American Journal of Pharmaceutical Education, 73, 1-5. https://doi.org/10.5688/aj730109

22. Sorden, S. (2012). The Cognitive Theory of Multimedia Learning, Handbook of Educational Theories.

23. Zainal, N. F. A., Din, R., Abd Majid, N. A., Nasrudin, M. F., & Abd Rahman, A. H. (2018). Primary and Secondary School Students Perspective on Kolb-Based STEM Module and Robotic Prototype. International Journal on Advanced Science, Engineering and Information Technology, 8, 1394-1401. https://doi.org/10.18517/ijaseit.8.4-2.6794

CHAPTER 15

A DESIGN MODEL FOR EDUCATIONAL MULTIMEDIA SOFTWARE

André Koscianski, Denise do Carmo Farago Zanotto

UTFPR, Ponta Grossa, Brazil

ABSTRACT

The design and implementation of educational software call into play two well established domains: software engineering and education. Both fields attain concrete results and are capable of making predictions in the respective spheres of action. However, at the intersection, the reports about the development of computer games and other educational software reiterate similar difficulties and an embarrassing degree of empiricism. This study aims to bring a contribution, presenting a model for the conception of

Citation: Koscianski, A. and Zanotto, D. (2014). A Design Model for Educational Multimedia Software. Creative Education, 5, 2003-2016. doi: 10.4236/ce.2014.523224.

Copyright: © 2014 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

educational multimedia software by multidisciplinary teams. It joins three elements: Ausubel's theory of meaningful learning; the theory of multimedia from Mayer; and the study of software ergonomics. The structure proposed here emerged from a theoretical study with the concurrent development of software. Observations gathered in a small scale test confirmed the expected design issues and the support provided by the model. Limitations and possible directions of study are discussed.

Keywords: Multimedia, Educational Software Design, Pedagogic Issues

INTRODUCTION

The first years of school are characterized by a ludic atmosphere that welcomes students into a warm and receptive ambience, with a careful mixture of leisure activities and serious objectives. Toys and games are practically mandatory elements from kindergarten up to K-12 classes. Some of the goals of such environments are to develop a positive attitude towards the school and cultivate reading and studying habits. Gradually there is a shift of perspective, culminating in the university, where courses are labour-intense and adopt a rather strict—if not spartan—style. Nonetheless, teachers of all educational levels tend to seek strategies to lower stress and keep students engaged and motivated.

Multimedia software can be helpful in this context for a number of reasons. These applications have native capabilities to show and interrelate textual descriptions, diagrams and photos, besides videos and sounds. All these means once combined represent a large contact surface between the previous knowledge of the student and the new information, helping to favour assimilation. The availability of images, animations and sound makes it possible to explore the presentation of material under ludic perspectives, counterbalance boredom and maintain the level of attention. In other terms, the students are presented to the same core information, although it is deliberately disguised only to change their internal disposition towards the subject. Finally, software can be designed to enhance the interaction between students and the learning materials, by means of activities that demand several inputs of different types.

The possibilities of using computers in the classroom are widely discussed in the literature, but the idiosyncrasies of educational software development are not addressed to the same extent. Using off-the-shelf products is just one facet of the introduction of technology in the classroom. Implementing such artefacts involves the design of interface and interaction which depend on a series of factors and unfold in a complex set of requirements. The project of such software should ideally be coupled to the instructional design of a discipline and not be limited to the design of a module. Aesthetic and artistic criteria also have a significant impact on the way students perceive the computer and get involved with it. Finally, the trade-off between the development cost and the utility of the product, as part of a set of comprising books, laboratories and other elements, is seldom discussed.

Several domains are interwoven in the design and construction of educational software and multimedia; three tools are considered here in order to structure the requirements and the design of such applications. The area of human-computer interaction and more particularly, the study of ergonomic interfaces helps to shape the overall design of applications and avoid mistakes that could negatively affect most users of a given audience. The Cognitive Theory of Multimedia Learning (CTML) of Richard Mayer is a landmark in the field of educational software, providing important heuristics aiming to optimize the effects of multimedia as instructional means. Finally, the Theory of Meaningful Learning (TML) of David Ausubel proposes explanations for the mechanisms behind learning and gives clues about how learning materials should be organized, so that software is integrated into a thorough instructional design. Taken together, this information can feed software engineering processes and provide references for the project and implementation of multimedia instructional programs.

This article discusses the design of multimedia software integrating these different views. An application was developed for the subject of biology, for Brazilian students with little contact with computers. Evaluations of a group of specialist and of the students are presented and discussed.

ORGANIZATIONAL ISSUES IN EDUCATIONAL SOFTWARE DEVELOPMENT

The project and implementation of any computer program is based on software engineering principles, which affect product quality and process efficiency (Pressman, 2009). The field covers a broad range of issues, varying from technical content like algorithmic complexity and hardware performance, to team management and psychological aspects involved in quality control (Weinberg, 1999). The development of educational software can be particularly complicated due to multitude of aspects to be considered, comprising cognitive and psychological effects as well as several technical problems like programming fast graphics or simulating physics to increase the realism.

Despite its relative youth, software engineering can be considered a mature field, being capable of coping with the construction and maintenance of all the complex informational infra-structure that surround us nowadays. To this end it has methodological tools to integrate specific expertises into the development of computer programs. The examples are abundant: medical tools, engineering programs, economics, chemistry and, of course, educational applications. However, working with multi-disciplinary teams may still be a daunting task. For instance, a typical programmer or software analyst is not used to the subtleties involved in the transmission of information during a teaching-learning interaction:

“Instead of guessing, designers should have access to a pool of representative users after the start of the design phase. Furthermore, users often raise questions that the development team has not even dreamed of asking” (Nielsen, 1992b).

Even if this statement is twenty years old, it remains a fresh truth in software engineering. Teachers are not acquainted with the technical barriers, nor the possibilities, that define the choice of a given set of functions or features that should be implemented in new software. Despite being well-documented in the informatics literature (Yourdon, 1989 ; Weinberg, 1999 ; Pressman, 2009), these issues are recurrent.

A minimum-team for educational software development would include a teacher, a programmer, a graphical artist and a manager, as these four persons cover the basic tasks and responsibilities associated with a project of this nature. The integration of experts in a team and the exchange of information must be constantly and actively supervised. Even small teams give rise to complex social interactions (Kirkman & Rosen, 1999). The project manager must balance different views (Weinberg, 1999).

One of the most critical aspects of software development is the initial phase of requirements elicitation (Pressman, 2009 ; Wiegers, 2003). Software behaviour, user expectations, hardware specifications are some of the characteristics that must be tailored to the application domain using information gathered from several sources, which include regular users, experts, technicians, investors and also standards and regulations. The boundaries separating the areas and the professionals are not very sharp (Weinberg, 1999 ; Flynt & Salem, 2005).

Concepts, ideas and knowledge from the domain of pedagogy enter as inputs to engineering processes. Inversely, constraints related to the computer and the resources available, may steer the project in a manner not foreseen by the teachers.

Final users and experts like teachers, psychologists and students can actively participate of the implementation, providing drawings, modifying the project of the interface, validating prototypes. Figure 1 identifies this domain of action by calling it the "teacher space", distinguishing it from the work of programmers and analysts, called here the "software engineering space". There is a mutual influence between the areas that must be balanced throughout the implementation. For instance, a choice like presenting or not videos in the interface is directly related to instructional design expectations and needs. However, it will be up to the technicians to determine the feasibility of this requirement and, if possible, to project and create a solution. Eventually a request must be turned down on the basis of technical or resource limitations, this way making it necessary to reconsider the project in the teacher space. The inverse may also happen, opening possibilities that were not considered at the beginning. The gray gradient of the figure represents the interwoven nature of this relation between experts.

There are different ways to organize the implementation of a software project. They range from the management of small applications carried out with PSP (Personal Software Process) to projects counting hundreds of developers relying on methods as the CMMI (Capability Maturity Model Integration). Giving the intrinsic complexity of multimedia, the use of a methodology for development should not be underestimated, even in modest projects. Studying the available choices is out of the scope of this article; the interested reader could refer to Pressman (2009) .

Multimedia applications share many characteristics with videogames. Both tend to contain numerous artefacts that include drawings, diagrams, sounds, music, storylines, descriptions of characters, texts and dialogues. The array of items may require specific management tools as versioning control and indexing (Flynt & Salem, 2005).

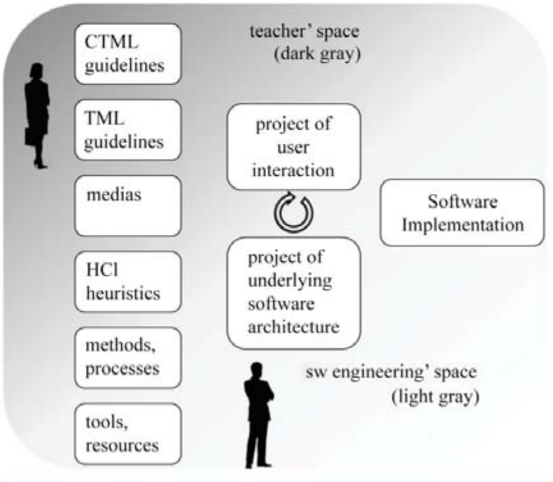

CTML guidelines
teacher' space (dark gray)
TML guidelines
project of user interaction
medias
Software Implementation
HCI heuristics
project of underlying software architecture
methods, processes
sw engineering' space (light gray)
tools, resources

Figure 1. Educational multimedia design space.

A crucial, yet simple tool, often used in management by the videogame industry, is the Game Design Document. In its original form and purpose, it is a document making a description of the software from the user view- point, without much concern for the internal architecture (Bethke, 2003 ; Rollings & Morris, 2004). It describes artefacts as the layout of screens, sequencing of tasks and possible paths of interaction that have a clear interest in an educational setting. The document functions as a focal point for the development team, helping tasks as evaluation of possible designs, negotiation of priorities and definition of deadlines. Non-ambiguous, complete and accurate design documents avoid difficulties and problems that can permeate the project up to the final product implementation (Flynt & Salem, 2005 ; Pressman, 2009).

Complementary to the Game Design Document, storyboards and prototypes (Meigs, 2003 ; Johnson & Scheleyer, 2003) are invaluable for educational multimedia. They allow discussing the project with an eye on the finished product.

INTERFACE DESIGN AND EDUCATIONAL SOFTWARE

Human Computer Interaction (HCI) can be thought of as a subset of interaction design (Preece & Rogers, 2002). Interaction design is a field of research concerned with the interaction between humans and things like cell phones, control panels used in industries, airplanes cockpits or everyday objects.

The typical computer interface devices—keyboards, mice and touch screens—are, in a certain sense, very limited artefacts: if we observe users interacting with software, we see silent people staring at screens, making movements with their hands while their bodies rest immobile for long periods. All the richness of the tasks being executed is concealed in the cognitive interaction between the users and the elements that are presented on screen. In fact, the project of computer interfaces involves a high level of subjectivity (Nielsen, 1992a ; Preece et al., 2002). As a consequence there is a rather paradoxical problem in software engineering: while the design and implementation processes are well known and reasonably controlled, the exact quality and behaviour of the final software is less certain, being highly dependent on a given context of use (Weinberg, 1999 ; Koscianski & Soares, 2006). Interfaces are, in this sense, blurry targets and the difficulty to forecast the success of a videogames is a good example of this (Bethke, 2003 ; Rollings & Morris, 2004).

The discipline of HCI studies the relationship between humans and computer interfaces from several viewpoints. For instance, the analysis of tasks offers a perspective to organize sequences of user actions, like pressing buttons and filling fields with information, in order to increase efficiency and prevent mistakes; the field of psychology gives clues on how to draw attention, avoid distractions and keep users focused; and research in physiology and cognition helps to understand how people react to stimuli and how they cope with tasks using information that is laid out on the screen. A straightforward example is to limit the branching factor and the number of items that are shown in menus. The famous work of Miller (1956) discussing short term memory is one of the firsts among a long series of research around cognitive capabilities.

Making a coherent unit out of all these pieces of information is a non-trivial task. The field of HCI has established criteria that helps guide software design and avoids most evident flaws (Nielsen, 1992a).

The international standard ISO 9241 is an important reference for interface design. It gives a series of advices concerning ergonomics, but leaves out the needs from specific domains like education. This issue has been addressed by several studies, some examples being Squirres and Preece (1999) ; Ardito, Buono, Costabile, Lanzilotti, & Piccinn (2009) ; Alsumait &

Al-Osaimi (2009) . The general character of the standard and its orthogonal organization make it possible to map virtually any set of requirements. Table 1 illustrates this point with one of the studies specific to educational software.

The first two columns of Table 1 lists the principles from ISO 9241, part 10; and the heuristics established by Nielsen (1992a) for usability of software. Both sets represent very condensed information that unfolds in more specific characteristics. For instance, in Safdari, Dargahi, Shahmoradi, & Nejad (2012) a questionnaire of 75 items was derived out of the seven ISO principles presented in the first column of the table. In the same manner, the requirement "conform to user expectations" is reworked by Alsumati and Al-Osaimi in several remarks, adapted to the universe of children education.

Table 1. A comparison between three sets of IHC heuristics.

Principles listed in ISO 9241-10	Heuristics from Nielsen (1992a)	Heuristics from Alsumait & Al-Osaimi (2009)
Suitability for the task	Aesthetic and minimalist design	Challenge the child, evoke child mental imagery, use multimedia representations
Self-descriptiveness	Recognition rather than recall	Learning content design (adequate vocabulary, illustrate abstract concepts)
Controllability	User control and freedom	Use appropriate hardware devices (match motor effort and children skills, prevent input errors)
Conformity with user expectations	Match between the system and the real world; error prevention	Attractive screen layout, support child curiosity, assessment (provide reports to instructor)
Error tolerance	Help users recognise, diagnose, and recover from errors	- (Only lists Nilsen's heuristics)
Suitability for individualization	Flexibility and efficiency of use	- (Only lists Nilsen's heuristics)
Suitability for learning (learning the software use)	Recognition rather than recall; help and documentation; consistency and standards	Design attractive screen layout (screen uncluttered, readable)

Besides covering all foundation of user interaction—a rich subject in itself—HCI should also accommodate instructional demands that are not part of most computer programs. Usual applications are designed to support tasks executed in a straightforward manner; examples are editing a letter, filtering information in databases, and buying products on Internet. ES, on the other hand, deals with the deep interaction that may (or should) exist between a person and material to be learnt, and objectives similar to those existent in class, such as:

- distribute information along of space and time according to pedagogic criteria;
- intentionally leave blanks between concepts, images and words so that links be identified during the interaction with the material or even later, during a subsequent lesson;
- ask users to solve problems, analyse their reasoning and give feedback about performance;
- entice a positive affective experience on users (Immordino-Yang & Damasio, 2011).

These ES characteristics are strongly connected to the interface organization, but they clearly go beyond ergonomics and have a complex dependency on individual cognitive differences. At this point, the software designers must call into play other conceptual references, as those discussed in this article.

A promising line of research is the development of adaptive multimedia or hypermedia systems, conceived to react to user particularities; see for example Brusilovsky & Millán (2007); De Bra & Calvi (1999) . Such approaches tend to emphasize the possibility to classify and model the interactions between user, the software and knowledge; they generally leave aside the complexities of human contact and classroom dynamics. In this sense, the ES is treated in a distance-learning perspective.

THE COGNITIVE THEORY OF MULTIMEDIA LEARNING

The field of computer ergonomics gives support to a critic part of design: the interface and immediate (short- term) user interaction. However, the studies in this area offer limited advice concerning the teaching-learning processes that are expected to occur with the support of computers. The interaction between students and the information on the screen involves internal cognitive mechanisms that are, in principle, out of the scope of HCI.

A reference for instructional multimedia is the Cognitive Theory of Multimedia Learning (CTML) developed by the psychologist Richard E.

Mayer. Undoubtedly he is best known by the investigation of the functioning of our visual and auditory channels, but he also examined other aspects of learning. This includes the effect of different sequences of presentation on retention, the influence of working with whole-part relations, or still the way information is associated and retrieved.

Mayer exploited the ability of computers to handle different information, as a means to strengthen the contact between the learner and the instructional material. A basic idea that underlies his work is the additive effect of using more than a sensory channel:

“Verbal and nonverbal systems are assumed to be functionally independent in that one system can be active without the other or both can be active in parallel. One important implication of this assumption is that verbal and nonverbal codes corresponding to the same object (e.g., pictures and their names) can have additive effects on recall” (Paivio, 1986).

It is worth pointing out that CTML present similarities with the learning styles proposed by Felder (1988) or Kolb (2005) , see also (Sankey, Birch, & Gardiner, 2011), but it has roots in a physiological comprehension of the way individuals process information. The term “multimedia learning” (Mayer, 2001) refers to the process by which an individual builds a mental representation of contents presented concurrently by means of words and pictures. Each type of information undergoes a different processing in the brain. Handling textual data requires linguistic skills to decipher a stream of symbolic data, while visual sources like a photograph, can be perceived as a whole. There is much evidence about cognitive differences associated with each type of information processing in the brain (Gardner, 1983 ; O’Reagan & Noë, 2001 ; Banich, Milham, Atchley, Cohen, Webb, & Wszalek, 2000), that support the original observations of Paivio and the subsequent work of Mayer.

CTML is based on three assumptions: verbal and visual information are treated along of different paths in the brain; each cognitive channel has a maximum processing capacity; and learning requires an active involvement of the student. These ideas are further developed into a set of principles or effects that serve to organize the design of instructional multimedia. They are listed in Table 2 according to a categorization given by Mayer himself.

The principles listed in Table 2 have been extensively tested (Mayer, 2008). They have been used as means to guide the design of educational materials (Park & Hannafin, 1993 ; Herrington & Oliver, 2000) and to predict learning results (Mayer, 2001). In order to use CTML successfully, two points should not be overlooked.

The text of the standard 9241 and the format used by Mayer to present his theory resembles a list of items. Although didactic, this can be misleading in practice, since a strict verification against a checklist can lead to a shallow evaluation (Tergan, 1998 ; Squirres & Preece, 1999). A better perspective is that of a structured walkthrough (Yourdon, 1989); the team can inspect the product according to individual expertises, and a set of guidelines remains as a reference or a reminder. As an example, a rule that determines the positioning of buttons on the screen can be broken in favour of a different layout to present a diagram, provided that the end result will be actually advantageous to students.

Videogame developers face similar difficulties: they must pay attention to clear ergonomic directives, like limiting the number of simultaneous options on the screen; and at the same time, the project must seek the wider goal of providing a joyful experience, an extremely subjective target.

CTML is also concerned with the fact that the project of software corresponds to a fraction of the instructional design. Every element—software, books, explanations, exercises, assignments and so on, should be framed by a pedagogic structure, adapted according to the teacher style and the class needs.

THE THEORY OF MEANINGFUL LEARNING

The rich variety of learning theories reflects the complexity of phenomena that take place when students engage in cognitive tasks and social interaction. Works from scientists as Piaget, Bruner or Vigotski, establish what we could call "angles of approach" to deal with the task of teaching. Each system of assertions and methods forms a coherent set that can be used as the base for instructional design of books, lessons, software.

Table 2. CTML principles.

Principle	Category	Description
Coherence effect	Reduction of extraneous load	The material presented to the student should avoid including information that is not part of the contents being studied.
Signalling effect		A presentation should give clues to students to guide their attention towards main points, by emphasizing or repeating information.
Redundancy effect		Narrated text should not be accompanied of written text, since this can distract students from observing pictorial information.
Spatial contiguity effect		The close placement between texts and pictures reduces the effort of students to inspect material and favours learning.
Temporal contiguity effect		The presentation of verbal and non-verbal pieces of information should occur simultaneously instead of sequentially.
Segmenting effect	Management of essential processing	The presentation of information should use separable units whenever possible, instead of fusing several concepts into complex texts and pictures.
Pre-training effect		Introductory material in the beginning of a presentation, may reduce the cognitive load associated to complex information that forms the core of the learning material.
Modality effect		When pictorial and verbal information are combined, the use of narrated (spoken) text is preferable over written text.
Multimedia effect	Fostering generative processing	Explanations with text and pictures are more efficient than those presenting information using only one of these possibilities.
Personalization effect		The presentation of material should preferably make students feel part of the narration, for example using second person instead of third person conjugation.

David Ausubel was a psychologist who developed research in the field of education. His work led to hypothesis concerning the functioning of the brain on an abstract, symbolic level. This view is closely related to information processing theories and cognitive approaches (Schunk, 2012).

The theory of meaningful learning was chosen in this study for two reasons. First, there is a number of links and agreements with the work of Mayer. Both authors share the viewpoint that cognitive phenomena can be explained by mechanical or information-processing metaphors, what is highly convenient in order to think about software design. Second, both theories focus the interaction between the individual and the information. In doing so, they do not preclude, neither require, adjustments to the social context. When this is necessary, the corresponding requirements elicitation must be carried out and incorporated into the project. As an example, social networking could have an effect on the sequencing of information presented on the screen. This kind of consideration falls outside the scope of this study.

Ausubel emphasized a distinction between two types of learning processes: rote memorization; and transformation of cognitive structures. In the first type of learning, individuals record facts in an arbitrary way. Examples are studying a list of items or associations, like capitals of states. In the second case new information is also stored in the brain, but this involves a network of facts that are likely to modify pre-existent knowledge. For instance, when students learn that friction is a force, their comprehension about several situations in mechanics gains new interpretations (Clement, 1982).

Learning in a meaningful way requires pre-existent knowledge that can be related to new information by the student. This process can lead to an integrative view of various concepts, or inversely to the differentiation between them. For instance, “plant” is an umbrella for things as different as “sequoias” and “roses”, while the terms “angiosperm” and “gymnosperm” can be understood as two particular types of the general concept of “plants that produce seeds”.

Not surprisingly, the anchoring mechanism does not filter or ensure the quality of information and students may acquire misconceptions (Stefani & Tsaparlis, 2009). Among the causes, teachers may propagate wrong concepts, possibly inherited in turn from their own teachers (Quílez-Pardo & Solaz-Portolés, 1995); flaws in the instructional design may cause fragmented and shallow understanding (Cooper, Grove, Underwood, & Klymkowsky, 2010); students also call into play heuristics as generalisation and simplification, which may produce distortions and false conclusions (Talanquer, 2006).

The associations between concepts can take a myriad of forms and can be represented as a network of ideas (Novak & Cañas, 2006). Some possible relations are represented in Figure 2 and labelled according to the Theory of Meaningful Learning (TML).

In subordinate learning, new ideas are connected to the cognitive structure as particular cases of a more general concept. The super-ordinate learning works in the opposite direction, by subsuming old ideas in new higher- level concepts. Finally, in combinatorial learning information get interrelated in non-hierarchical ways. In Ausubel's view, "meaning" is the result of the interaction between ideas. This way, by learning new ideas one can modify previous knowledge, up to the point of replacing and obliterating information.

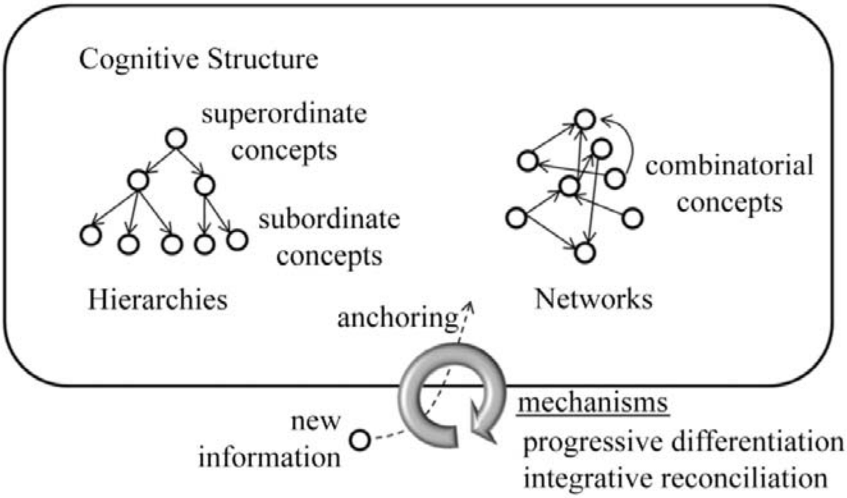


Figure 2. Overview of mechanisms from TML.

If, during a lesson, students do not remember past knowledge or fail to recognize it as relevant, the quality of assimilation is compromised. Ausubel devised a strategy to deal with this problem, by introducing students to material that can help select and activate such memories; he called them "advance organizers". They can take forms as diverse as pictures, texts, exercises or even spoken dialogues.

An advance organizer (AO) shows only information that is likely to be known, not new facts. It is presented at the beginning of a lesson, with the purpose of making students aware of their own knowledge. As an example, the concept of "couple" or "moment of a force" can be presented in a Physics lesson with the help of the formula $\tau = dF$, where d and F stand for the displacement and the force vector, respectively. Despite the formula being short and the mathematics uncomplicated, the abstract nature of the communication pose difficulties in class. A possible AO for this subject would be a video, showing someone struggling to remove a screw with spanners of different lengths (Koscianski, Ribeiro, & Silva, 2012).

Teachers know from experience that the design of a lesson must account for the level of knowledge of the students. However, the TML assigns a great weight to the detailed structure of such knowledge, which can explain subtle variations in tasks like comprehension of concepts and formulation of hypothesis. For instance, in the example of the moment of a force, the fact that the discussion about the formula is preceded by the presentation of a video is likely to make students more comfortable with the algebraic representation (Koscianski, Ribeiro, & Silva, 2012). If instead the formula is presented as the core of the lesson, then any examples can be perceived as a means to assimilate the mathematical notation and not as instances of application.

During the instructional design, it is possible to take advantage of the concepts of advance organizer and information anchoring even if the exact background of the students is not directly evaluated. Teachers may outline hypothetical cognitive structures that capture facts, ideas and situations that are expected to make part of the students' repertoire. This includes things from their quotidian, contents that have been taught previously and that are revealed by dialogues in class and textual assignments. This material will serve as an organizational basis to plan the approach in class, allowing the forecast of difficulties, potential sources of misunderstandings and different angles to attack a subject. It also clarifies potential connexions between concepts.

Hypothetical cognitive structures can be represented as concept maps that, in turn, can drive the creation of storyboards (Meigs, 2003 ; Johnson & Scheleyer, 2003). These tools can help design the different exploration paths that students may use (Ford & Chen, 2000).

CONNECTING THE TOOLS

Taken together, the areas discussed form a thorough basis for the whole engineering processes, during the implementation of educational software.

The discipline of HCI is directly related to software development and borrows material from psychology, but is has no specific links to education. One exception is the requirement that the interface of a program be easy to learn and understand, based on the premise that any computational tool should be unobtrusive.

On the other extreme, the theory of Ausubel dives into the domain of cognition but has no particular concern about the medium employed to deliver information. It makes little distinctions between the use of books, oral explanations and other vehicles. The emphasis is placed in higher cognitive tasks like comprehension and extraction of information.

Mayer's work stands as a possible bridge between both worlds. The interface is treated using principles as contiguity and multimedia, while mechanisms associated with deeper assimilation are exemplified by principles as pre-training and segmentation. CTML, however, does not offer thorough answers about the construction of interfaces as does the HCI discipline, neither predicts learning outcomes with the same amplitude as TML.

Table 3 shows a general view of the areas and their emphasis, mapped in a diagram shown in Figure 3. The layered structure of the diagram also reflects a temporal approach to the project, where the top level—the interface—corresponds to the last element to be refined.

TML provide the deepest level of organization. Its principles will justify the choice and the organization of the contents. It gives criteria to select metaphors, to draw relations, create exercises and activities, and any other aspects directly connected to understanding the actual meaning of the information. The contents of texts and images are produced and evaluated according to principles defined by TML, as the availability of previous knowledge.

Table 3. Overview of the model.

Tool	Focus	Time scale	Intervention
HCI	Symbol recognition, workflow, short-term memory	Immediate symbol and information recognition	Immediate interaction; easiness of interface comprehension and usage
CTML	Contents assimilation	Short/long term memory	Density of information and efficiency of communicating
TML	Deep, complex learning	Life-long learning	Design of the whole instructional material

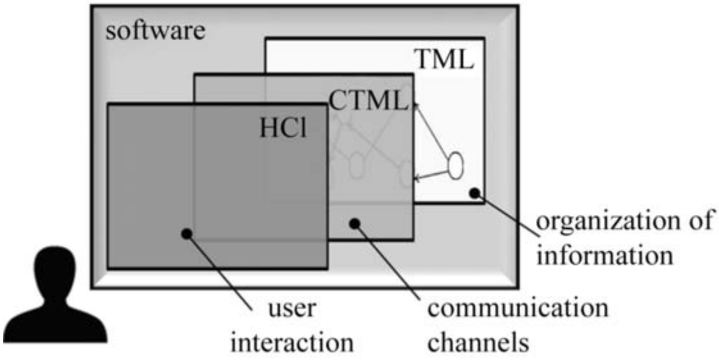


Figure 3. A layered view of the combination of the different tools for software design.

Although teachers may be eager to draw interfaces and present ideas about how students will operate the software, such considerations should be deferred in a first moment. The team should question how the teacher intends to explain each concept, idea or problem, and how that's done during a classical lesson using pencil and paper. It is only when clear strategies have been identified and understood by the team that the translation into software can take place. As a single example, during the discussion about a mathematical game showing a coordinate plan, the teacher casually mentioned that "cut the X axis" was the expression used in class to explain a certain property. That information immediately prompted the development team to change the scenery—which in principle was already validated—replacing a ball with scissors in order to strengthen the links between software, subject and verbal explanations used.

The main role of CTML is to organize the information vehicles that are assembled into the software, such as texts, videos, images. It helps to balance the use of different media and gives clues as how to combine them. In doing so, CTML may also help to consider unforeseen forms of presentation. For instance, not rarely teachers interpret the adjectives "visual" and "auditory" in a manner not compatible with CTML, and do not consider the possibility of spoken texts replacing written information on diagrams. The theory may hint teachers about potential approaches not used before, as creating a diagram adapted to the way a subject is analysed in class. Another benefit is to require an explicit control over the distribution of information. This means to quantify the number of pieces of information—diagrams, images, texts and vocabulary and to characterize their complexity. Experienced teachers may perform this task instinctively, but the same is far from being truth for programmers and graphic artists. During the project of a game, a certain pressure around the design of interface and game mechanics may build up and override other criteria for the distribution of contents.

The top level of the diagram in Figure 3 does not simply echoes the decisions made in the previous steps. The interface is a two-way street that limits both the access of students to information and the access of the software to the way students think. For instance, a program can only make approximate predictions about the knowledge acquired, by means of tests and quizzes. Interpreting natural language is still an open problem and, as a consequence, the internal layers of the software must be projected to ensure that each new piece of information is correctly grasped and assembled into the cognitive structure. This leads back to the instructional design and point the need to introduce questions and additional explanations along of the interface, using redundancy to decrease the possibility of misunderstandings. The design of the interface, on itself, follows rules that seek to make the software transparent. Mouse clicks, keyboard actions, buttons as “ok” and “cancel”, form a set of necessary objects for interaction but that should, as much as possible, be left at the edge of the attention of the student.

One last element should be considered: the affective dimension. More than a cosmetic feature, the design of characters, sceneries and storylines may have an actual influence on learning (Craig, Graesser, Sullins, & Gholson, 2004). There is evidence that a strict separation of emotions and cognition is artificial, both from an external, behavioural view (Storbeck & Clore, 2007 ; Pessoa, 2008) and from observations of physiology (Phelps, 2006). The exact mechanisms are still unclear and seem to involve interdependencies between systems of beliefs, knowledge, unconscious thoughts, feelings and other phenomena. The degree to which such relations interfere with learning and to what extent they can be controlled are subject of research, (see for instance Schunk, Pintrich, & Meece, 1996), but much remains to be explained.

For practical purposes, the sensitivity to students’ preferences and cultural traits has always been acknowledged by educators. As mentioned in the introduction, it is a common practice to make adaptations in classroom, as simple as decorating it, to favour a pleasant mood. The same can be seen with the changes in style of children's textbooks and, more recently, with the use of comics and mangas in higher-education. This artistic aspect of learning materials, not only aesthetic, is rarely evoked in the literature. There is a reasonable body of research around these issues in the industry of videogames (Bethke, 2003).

A Small Scale Study

The work described in this article was originated at an attempt to circumvent the difficulties faced by K-12 Brazilian students with science contents. The specific subject was the classification of vegetables in angiosperms and gymnosperms; the primary curriculum in Brazil, as applied in classes, is quite extensive. Textbooks include a high level of detail, what comprises the identification of anatomic structures shown in photographs and diagrams, knowledge of physiology and technical names. For instance, children aged twelve may be expected to explain the difference between xylem and phloem. This emphasis on volume of data is also illustrated by the historical undervaluing of philosophy in curricula and in exams like the vestibular (university admittance test). In recent years a new trend seems to emerge, but so far textbooks remain the same.

A multimedia software was proposed as a helping tool in this context. It would give access to texts and explanations, selected images and diagrams, without the limitations found in printed material or the potential chaos of the internet. The starting idea was a sort of encyclopaedia, but during the project, the design was steered towards a game, by means of a scenario and a storyline. Figure 4 shows one of the screens of the finished pro- duct.

The definition of the game grew out of a storyline. The plot was a conflict between good and evil, with fairies, potions and plants. Most scenes depicted a medieval castle and, although such scenery is not part of the folklore of Brazilian students, it represented nonetheless a really pleasant ambience for the public of the study. The theme was indeed selected after some input was received from children. Exercises and quizzes were included along of the software and interwoven with a background story. This ensured a continuous ambience, smoothing the rupture between "game" and "study".

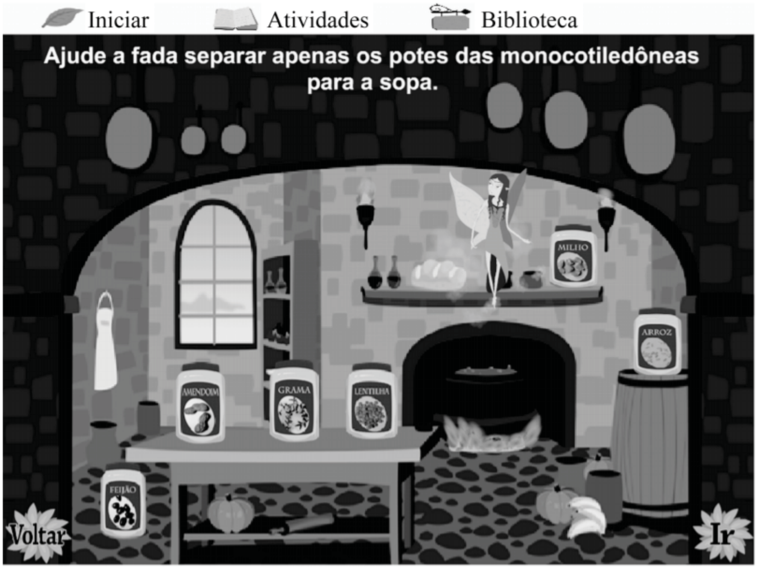


Figure 4. Screen of the implemented software (the videogame language is Portuguese).

In parallel to the product design, the definition of the layered approach depicted in Figure 3 was refined. The first issue perceived in the project was the tension between different views. The teacher and the programmer had no clues about possibilities and restrictions existent in the corresponding spheres of work. This situation made surface again during the evaluation of the product. We will return to this issue briefly.

The initial sequencing of contents was defined by the teacher, who had twenty years of experience with the subject. Two alternatives were discussed: using the software to review the subject, or to introduce the contents for the first time to the students. In order to lower the anxiety of the teacher and also to prevent any negative outcomes, a somewhat mixed approach was selected. A quick introduction to the theme would be conducted in class, followed by the use of the computer in a subsequent lesson. This way, neither of the moments was designed to fully cover the subject; the idea that the contents should be presented repeated times was part of the overall instructional design.

The contents were laid in a sequence of units, following the scheme used by the teacher and also adopted in didactic materials as textbooks. The sequence comprised a general overview, followed by roots, stem, leaf, flower, fruit and seed. During their first walk along of the interface, the students would see these themes in this strict sequence. As the sections have been visited, they were allowed to navigate freely. This restriction helped to enforce a match between software and the lessons in class and give a sense of familiarity with the discipline.

Small texts, diagrams and photos were organized in “virtual books”, present in a library of the scenery. The students followed the plot and, when presented to an exercise, could go to the library to obtain information. The choice of texts and images was based on the indications of the teacher. In this sense the software was extremely customized. Nevertheless, the final product has been used in another school with a very different public and obtained positive reviews.

In most of the screens, the texts and images were interrelated. Both the temporal and the spatial proximity between texts and images were particularly important because of the new terms that students were supposed to learn. Despite the use of storyboards, we could later spot points where this rule has not been observed. A possible explanation for this kind of mistake was the schedule pressure, since the project was part of a master thesis. Software can be compared with this respect to the preparation of other materials, as a book text. Traditional methods of peer-revision can be used to minimize this type of problem. Larger software projects, subject to financial pressures present particularities for which specific strategies have been devised (Yourdon, 1989).

Spoken narrative was never considered in the design, because the software would be operated by couples of students sharing computers in the laboratory. Nevertheless a music track was added, with a noticeable impact on the interface.

The proposed combination of viewpoints had a truly practical impact. It supported the team providing criteria to direct discussions between programmer and teacher, the realization of studies around of sketches and helping the process of decision making. The conceptual framework acted as a filter to ideas originated from personal opinions, brainstorming or simple guessing. It also gave a higher level of confidence in order to determine the sequencing of contents distributed on the software.

Once finished, the program was analysed by a group of teachers from the areas of language, biology, arts, informatics and education. The reviews were freely conducted and the reports were positive, with a few suggestions for localised improvement. In further analysis, the relative proximity between the reviewers and the researchers may have disabled negative comments. The format chosen was open-ended questions, with the objective of letting the teachers freely express their views. This decision may also have limited the depth of analysis, because the reviewers did not have any particular background on educational software. This reaffirms the cleavage between the knowledge of computer technicians and of teachers and the need to carefully administer this aspect along of software development.

FINAL POINTS

The project described in this article confirmed the classic predictions of software engineering with regard to planning, schedule and project administration. The strict division of responsibilities had a positive effect on the workflow, another result that is not new but that is never overemphasized among experts of other areas, not acquainted with software production. The departure had moments of hesitation, but in a matter of three weeks the project got a life of its own and the team dynamics was stabilized.

A "pre-design" phase to assess the cognitive structures of the students improves the requirements elicitation. The team relied on the previous experience of the teacher to define the contents, but did not systematically register this knowledge; the information remained controllable thanks to the small size of the project. A printed conceptual map would have been invaluable to guide the team through development, revealing potential gaps and acting as a reference to discuss different approaches to expose contents. We may hypothesize the development of CAD (computer aided design) tools for educational software, blending or linking concept maps into other diagrams and sources of information used in project and implementation.

Another important aspect is the complexity of designing the navigation graph of an educational application. Ausubel investigated the mechanisms underlying the storage and retrieval of networks of information, but did not address techniques for designing instructional material. Mayer studied instructional design, but with a focus on communication. Good lectures unfold complex subjects along of a clear path. It is our view that this old proven technique is still the more appropriate to lay the core design of hypermedia, as it seems to match the way we absorb information: piecewise, with clear hooks from one information to the immediately following it and, preferably, with a notion of purpose. In our study, the use of a storyline as a backbone proved to be a good solution to organize the software. According to this view, the free exploration of contents would happen backwards only; during software usage, students should follow a strict sequence. The question of self-guidance is evoked in the literature and needs to be clarified with respect to the approach proposed here.

Finally, balancing gamification with the "crude" objective of transmitting information is a problem that, to our knowledge, lacks a systematic approach. The teachers who reviewed the software missed the fact that the design underused multimedia possibilities. The non-blind nature of the review is hardly a complete explanation to the positive bias and observations about the appearance of the product. We could tentatively argue that the reviewers lacked enough familiarity with the medium in other to criticize it. However, this did not explain why they ignored points where additional concepts and contents could or should be included. In an ongoing study with a mathematical game, similar tendencies for loosing focus were observed with children, who tended to disregard instructions to focus on pure-game aspects. This seems to confirm the potential for the intentional manipulation of user perception mentioned at the origin of the text, although in this particular case it produces an undesirable, unexpected effect.

ACKNOWLEDGEMENTS

André Koscianski, Denise do Carmo Farago Zanotto The authors would like to thank Márcia Regina Carletto for her assistance and expertise, Priscila Baer Gomes Vieira for the software programming and Welington Guimarães for the illustrations.

REFERENCES

1. Alsumait , A., & Al-Osaimi, A. (2009). Usability Heuristics Evaluation for Child E-Learning Applications. Proceedings of iiWAS2009, 425-430.
2. Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R., & Piccinn, A. (2009). Enabling Interactive Exploration of Cultural Heritage: An Experience of Designing Systems for Mobile Devices. Knowledge Technology and Policy, 22, 79-86. http://dx.doi.org/10.1007/s12130-009-9079-7
3. Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., Kramer, A. F., Liang, Z. P., Wright, A., Shenker, J., & Magi, R. (2000). fMRI Studies of Stroop Tasks Reveal Unique Roles of Anterior and Posterior Brain Systems in Attentional Selection. Journal of Cognition Neuroscience, 12, 988-1000. http://dx.doi.org/10.1162/08989290051137521
4. Bethke, E. (2003). Game Development and Production. Plano, TX: Wordware Publishing, Inc.
5. Brusilovsky, P., & Millán, E. (2007). User Models for Adaptive Hypermedia and Adaptive Educational Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), The Adaptive Web (pp. 3-53). Berlin: Springer-Verlag. http://dx.doi.org/10.1007/978-3-540-72079-9_1
6. Clement, J. (1982). Student's Preconceptions in Introductory Mechanics. American Journal of Physics, 50, 66-70. http://dx.doi.org/10.1119/1.12989
7. Cooper, M. M., Grove, N., Underwood, S. M., & Klymkowsky, M. W. (2010). Lost in Lewis Structures: An Investigation of Student Difficulties in Developing Representational Competence. Journal of Chemical Education, 87, 869-874. http://dx.doi.org/10.1021/ed900004y
8. Craig, S. D., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and Learning: An Exploratory Look into the Role of Affect in Learning with Auto Tutor. Journal of Educational Media, 29, 241-250. http://dx.doi.org/10.1080/1358165042000283101
9. De Bra, P. M. E., & Calvi, L. (1999). AHA! An Open Adaptive Hypermedia Architecture. New Review of Hypermedia and Multimedia, 4, 115-139. http://dx.doi.org/10.1080/13614569808914698
10. Felder , R. M., & Silverman, L. K. (1988). Learning and Teaching

Styles in Engineering Education. Engineering Education, 78, 674-681.

11. Flynt, J. P., & Salem, O. (2005). Software Engineering for Game Developers. Boston, MA: Thomson Course Technology Ptr.
12. Ford, N., & Chen, S. Y. (2000). Individual Differences, Hypermedia Navigation, and Learning: An Empirical Study. Journal of Educational Multimedia and Hypermedia, 9, 281-311.
13. Gardner, H. (1983). Frames of Mind: The Theory of Multiple Intelligences. Philadelphia, PA: Basic Books.
14. Herrington, J., & Oliver, R. (2000). An Instructional Design Framework for Authentic Learning Environments. Educational Technology Research and Development, 48, 23-48. http://dx.doi.org/10.1007/BF02319856
15. Immordino-Yang, M. H., & Damasio, A. (2011). We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. Mind, Brain, and Education: Implications for Educators, 5, 115-132.
16. Johnson, L. A., & Schleyer, T. K. L. (2003). Developing High-Quality Educational Software. Journal of Dental Education, 67, 1209-1220.
17. Kirkman, B. L., & Rosen, B. (1999). Beyond Self-Management: Antecedents and Consequences of Team Empowerment. The Academy of Management Journal, 42, 58-74. http://dx.doi.org/10.2307/256874
18. Kolb, A. Y., & Kolb, D. A. (2005). Learning Styles and Learning Spaces: Enhancing Experiential Learning in Higher Education. Academy of Management Learning & Education, 4, 193-212. http://dx.doi.org/10.5465/AMLE.2005.17268566
19. Koscianski, A., & Soares, M. (2006). Qualidade de Software. São Paulo: Novatec.
20. Koscianski, A., Ribeiro, R. J., & Silva, S. C. R. (2012) Short Animation Movies as Advance Organizers in Physics Teaching: A Preliminary Study. Research in Science & Technological Education, 30, 1-15. http://dx.doi.org/10.1080/02635143.2012.732057
21. Mayer, R. E. (2001). Multimedia Learning. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139164603
22. Mayer, R. E. (2008). Applying the Science of Learning: Evidence-Based Principles for the Design of Multimedia Instruction. American Psychologist, 63, 760. http://dx.doi.org/10.1037/0003-066X.63.8.760
23. Meigs, T. (2003). Ultimate Game Design: Building Game Worlds.

Emeryville, CA: McGraw Hill/Osborne.

24. Miller, G. A. (1956). The Magical Number Seven, plus or minus Two: Some Limits on Our Capacity for Processing Information. Psychological Review, 101, 343-352. http://dx.doi.org/10.1037/0033-295X.101.2.343

25. Nielsen, J. (1992a). Finding Usability Problems through Heuristic Evaluation. In P. Bauersfield, J. Bennet, & G. Lynch (Eds.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 373-380). New York: Academic Press. http://dx.doi.org/10.1145/142750.142834

26. Nielsen, J. (1992b). The Usability Engineering Life Cycle. Computer, 25, 12-22. http://dx.doi.org/10.1109/2.121503

27. Novak, J. D., & Cañas, A. J. (2006). The Theory Underlying Concept Maps and How to Construct Them. Technical Report IHMC Cmap Tools, Pensacola, FL: Florida Institute for Human and Machine Cognition.

28. O'Reagan, J. K., & Noë, A. (2001). A Sensorimotor Account of Vision and Visual Consciousness. Behavioral and Brain Sciences, 24, 939-1031. http://dx.doi.org/10.1017/S0140525X01000115

29. Paivio, A. (1986). Mental Representations: A Dual Coding Approach. Oxford: Oxford University Press.

30. Park, I., & Hannafin, M. J. (1993). Empirically-Based Guidelines for the Design of Interactive Multimedia. Educational Technology Research and Development, 41, 63-85. http://dx.doi.org/10.1007/BF02297358

31. Pessoa, L. (2008). On the Relationship between Emotion and Cognition. Nature Reviews Neuroscience, 9, 148-158. http://dx.doi.org/10.1038/nrn2317

32. Phelps, E. A. (2006). Emotion and Cognition: Insights from Studies of the Human Amygdala. Annual Review of Psychology, 57, 27-53. http://dx.doi.org/10.1146/annurev.psych.56.091103.070234

33. Preece, J., Rogers, Y., & Sharp, H. (2002). Interaction Design: Beyond Human-Computer Interaction. New York: John Wiley & Sons.

34. Pressman, R., & Maxim, B. (2009). Software Engineering: A Practitioner's Approach (8th ed.). New York: McGraw-Hill Higher Education.

35. Quílez-Pardo, J., & Solaz-Portolés, J. J. (1995). Students' and Teachers' Misapplication of Le Chatelier's Principle: Implications for

the Teaching of Chemical Equilibrium. Journal of Research in Science Teaching, 32, 939-957. http://dx.doi.org/10.1002/tea.3660320906

36. Rollings, A., & Morris, D. (2004). Game Architecture and Design: A New Edition. Indianapolis, IN: New Riders Publishing.

37. Safdari, R., Dargahi, H., Shahmoradi, L., & Nejad, A. F. (2012). Comparing Four Softwares Based on ISO 9241 Part 10. Journal of Medical Systems, 36, 2787-2793. http://dx.doi.org/10.1007/s10916-011-9755-5

38. Sankey , M. D., Birch, D., & Gardiner, M. W. (2011) The Impact of Multiple Representations of Content Using Multimedia on Learning Outcomes across Learning Styles and Modal Preferences. International Journal of Education and Development Using Information and Communication Technology, 7, 18-35.

39. Schunk, D. H. (2012). Learning Theories, an Educational Perspective (6th ed.). Boston, MA: Pearson Education Inc.

40. Schunk, D. H., Meece, J., & Pintrich, P. R. (1996). Motivation in Education: Theory, Research and Applications. Englewood Cliffs, NJ: Merrill Prentice-Hall.

41. Squirres, D., & Preece, J. (1999). Predicting Quality in Educational Software: Evaluating for Learning, Usability and the Synergy between Them. Interacting with Computers, 11, 467-483. http://dx.doi.org/10.1016/S0953-5438(98)00063-0

42. Stefani, C., & Tsaparlis, G. (2009). Students' Levels of Explanations, Models, and Misconceptions in Basic Quantum Chemistry: A Phenomenographic Study. Journal of Research in Science Teaching, 46, 520-536. http://dx.doi.org/10.1002/tea.20279

43. Storbeck, J., & Clore, G. L. (2007). On the Interdependence of Cognition and Emotion. Cognition and Emotion, 21, 1212- 1237. http://dx.doi.org/10.1080/02699930701438020

44. Talanquer, V. (2006). Commonsense Chemistry: A Model for Understanding Students' Alternative Conceptions. Journal of Chemical Education, 83, 811. http://dx.doi.org/10.1021/ed083p811

45. Tergan, S. O. (1998). Checklists for the Evaluation of Educational Software: Critical Review and Prospects. Innovations in Education & Training International, 35, 9-20. http://dx.doi.org/10.1080/1355800980350103

46. Weinberg, G. M. (1999). Quality Software Management. Vol. 1: Systems Thinking. New York: Dorset House.
47. Wiegers, K. (2003). Software Requirements 2 (2nd ed.). Redmond, WA: Microsoft Press.
48. Yourdon, E. (1989). Structured Walkthroughs (4th ed.). Upper Saddle River, NJ: Yourdon Press.

CHAPTER 16

"COGNITIVE CONSTRUCTIVIST THEORY OF MULTIMEDIA: DESIGNING TEACHER-MADE INTERACTIVE DIGITAL"

Prince Hycy Bull

North Carolina Central University, Durham, USA

ABSTRACT

This paper discusses how educators could use the cognitive constructivist theory of multimedia (CCTM) to design interactive digital learning materials using Camtasia and Audacity. Camtasia allows educators to create videos that motivate students, inform parents and enhance learning. It allows educators to record live presentations or lectures and provide students with a file to review. Audacity is a free cross-platform audio editor and recorder for Windows, Mac OS X, GNU/Linux and other operating systems. CCTM

Citation: Bull, P. (2013). "Cognitive Constructivist Theory of Multimedia: Designing Teacher-Made Interactive Digital". Creative Education, 4, 614-619. doi: 10.4236/ce.2013.49088.

Copyright: © 2013 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0

advocates for the design of instruction using pictures, videos, audios and words that tap into the prior experiences of the learner, promote active learning, collaboration, personal autonomy, personal growth and alternative assessment that is aligned with multiple intelligences of learners as espoused by Gardener (1993) which are Linguistics, Logico-mathematics, Spatial, Musical, Bodily-kinesthetic, Interpersonal, Intrapersonal and Naturalist. Camtasia and Audacity promote use of CCTM because of their capabilities to construct knowledge through words, pictures, animations, videos and audio. Case studies show that use of teacher-made files could significantly impact students' learning. Use of teacher-made interactive digital learning materials could revolutionize educational presentations and enhance e-learning delivery. CCMT produced by dynamic presentations creates a balance between the learners' prior verbal and visual experiences, sensory repository, multiple intelligences and learning styles to construct new knowledge.

Keywords: Digital Interactive Materials; Multimedia; Constructivist Approach; Multiple Intelligences; Cognitive Constructivist Theory of Multimedia

INTRODUCTION

Historically, educational systems depended on teacher-made materials and activities to promote teaching and learning. The proliferation of technology, digital materials, and learning technologies have shifted the design focus from teachers and educators to profit technology solutions, the Web, and commercially designed software. There is a paradigm shift for designing materials for teaching and learning from teachers to businesses. This shift has the potential to undermine the quality of materials presented and disrupt the educational process. To promote effective integration of digital materials and learning technologies in education, the focus of designing materials and resources needs to shift back from business to classroom teachers. This paper discusses integrating an efficient and cost effective way to prepare teachers to create functional interactive digital files using Camtasia software to develop a video files and Audacity software to design audio podcasts.

CAMTASIA SOFTWARE

Camtasia allows educators to create professional finish videos that motivate and engage students, inform parents, and enhance learning. It allows educators to record live presentations, lectures, and activities and provide students with a video file to review. Camtasia digital interactive files promote flipping the classroom through use of video files that students watch as homework and apply concepts in the classroom. Camtasia has several unique features that help educators create functional digital materials to promote learning:

- Capture what you see, say, and do
- Personalize interactive videos
- Create videos with professional finish
- Allow educators to put themselves in a video by using the Chroma Key or green screen effect
- Allow easy editing of videos
- A media asset library that enhances professional finish videos
- Videos files with engaging visual effects
- Animated content to capture learners attention
- Quizzing function to assess student's understanding of videos
- Ability to share digital files in FLASH and HTML5 viewer environment
- Produce specific files for iPad, iPhone, and other devices
- Produce and share videos accessible via Youtube or Screencast. com
- Capture PowerPoint presentation using the PowerPoint addon feature

AUDACITY

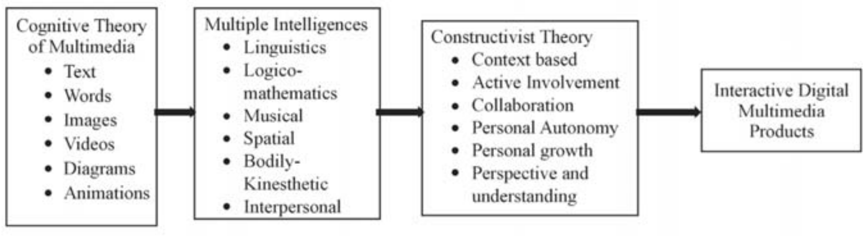
Audacity is a free cross-platform audio editor and recorder for Windows, Mac OS X, GNU/Linux and other operating systems. Audacity has several unique features that promote designing effective digital learning materials, which can be presented as audio presentations or combined with Camtasia to design effective interactive digital materials:

- Record live audio through a microphone or mixer;
- Digitize recordings from cassette tapes, records, and optical discs;
- Create multi-track recordings;
- Edit Ogg Vorbis, MP3, WAV or AIFF sound files;
- Cut, copy, splice or mix sounds together;

- Change the speed or pitch of a recording;
- Import sound files, edit and combine with other files or recordings;
- Export recording in many different file formats;
- Create MP3 files using the LAME converter file that could be uploaded to iTunes;
- Support 16-bit, 24-bit and 32-bit (floating point) samples (the latter preserves samples in excess of full scale);
- Sample rates and formats are converted using high-quality resampling and dithering;
- Track with different sample rates or formats are converted automatically in real time;
- Easy editing with Cut, Copy, Paste and Delete;
- Unlimited sequential Undo (and Redo) to go back any number of steps;
- Edit and mix large numbers of tracks;
- Multiple clips are allowed per track;
- Draw tool to alter individual sample points;
- Envelope tool to fade the volume up or down smoothly;
- Automatic Crash Recovery in the event of abnormal program termination;
- Accessibility features to manipulate tracks and selections using keyboard;
- Accessibility features to support JAWS and other screen readers on Windows, and for VoiceOver on Mac;
- Built-in sound effects for engaging and stimulating presentations.

THEORETICAL FRAMEWORK

Designing interactive digital materials with Camtasia and Audacity using the principles of cognitive constructivist theory of multimedia (CCTM) produces functional interactive teachermade digital learning materials that engages and motivates learners. See Figure 1. The cognitive constructivist theory of multimedia (CCTM) developed by Bull (2009) is the theoretical framework guiding this paper. CCTM is a theory that integrates concepts from the constructivist theory, multiple intelligences theory, and cognitive theory of multimedia learning to support designing teacher-made interactive digital materials. CCTM design is the integration of multimedia (text, image, animation, graphic, video and audio) in a systematic design structure of visual thinking combining verbal and non-verbal communication to minimize cognitive overload of the learner's memory. CCTM also utilizes the learner's multiple intelligences, learning styles, and prior knowledge of multimedia contents and content information to construct and refine knowledge through multimedia files.


Cognitive Theory of Multimedia
- Text
- Words
- Images
- Videos
- Diagrams
- Animations

Multiple Intelligences
- Linguistics
- Logico-mathematics
- Musical
- Spatial
- Bodily-Kinesthetic
- Interpersonal

Constructivist Theory
- Context based
- Active Involvement
- Collaboration
- Personal Autonomy
- Personal growth
- Perspective and understanding

Interactive Digital Multimedia Products

Figure 1. Cognitive constructivist theory of multimedia.

INTERACTIVE DIGITAL MATERIALS AND COGNITIVE THEORY OF MULTIMEDIA

CCMT is aligned with the cognitive theory of multimedia learning (CTM) as one of its supporting theories. In designing interactive digital learning materials, the learner utilizes visual and verbal information processing systems to engage in learning. All auditory information received during this process of learning goes to the verbal system and all graphics, pictures and animations goes to the visual system. This means that educators in designing interactive digital materials should create a balance between verbal and visual repository systems of the learner to fully engage them in the learning process. Mayer (2002) identified six major principles of multimedia design, which should be adhered to in designing interactive digital materials to create a balance between the verbal and visual systems for effective learning outcomes:

- Multimedia/Multiple Representation—In designing digital interactive learning materials, instructors should ensure that information is presented in at least two modes (text, video, picture, animation and audio) of representation for clarity and understanding. Using at least two modes of representation creates a multimedia learning effect, which enables learners to effectively develop verbal and visual models and build functional connections between them. This principle also taps into the multiple intelligences theory by utilizing the intelligences of each learner. For example, the presentation of pictures and text support both linguistic and spatial intelligences. A presentation of video recording supported by closed captioning supports different combination of intelligences, but predominantly addresses linguistics, spatial, reasoning, interpersonal, and bodily-kinesthetic;

- Contiguity principle—Instructors should also combine pictures and words in designing digital interactive learning materials. Using at least two modes of multimedia to present a concept allows learners to understand a concept better than when the same modes are presented separately for the same concept. Camtasia has several options to combine pictures and words on the same canvass, via PowerPoint or using over lay on videos and pictures. Files created in Audacity can be embedded in Camtasia designed videos to promote the contiguity principle, which simply states that multimedia learning combining words/text and pictures in ones verbal and visual systems constructs new knowledge. Corresponding words and pictures represented at the same time means a temporal congruity. On the other hand, words and pictures presented on same page, document, slide or screen promote spatial contiguity. This principle is aligned with the linguistics and spatial intelligences identified in multiple intelligences.
- Split-Attention principle states that when words and texts are presented as narrations they are more effective in a multimedia presentation than when presented visually. Camtasia because of its narrative and video capabilities provides more opportunities and easier creation of the split-attention principle, which simply states that on-screen text and animation may overload the visual information system, but narration aligned with the verbal information system and animation aligned with visual information system does not overload in one system. Camtasia provides opportunities for live recording of audio narrations or video presentations of instructor teaching.
- Individual Differences principle states that unique learning qualities of the learner aligned with multimedia, contiguity and split-attention principles determine how well the learner is

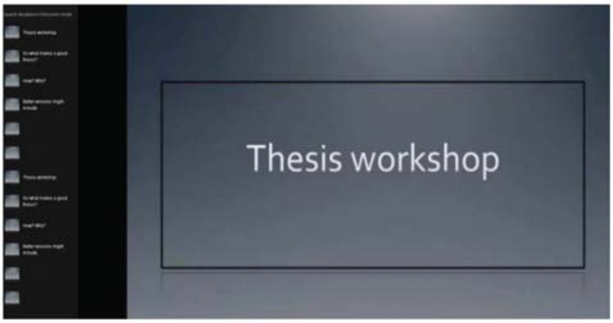
able to conceptualize learning. From the constructivist theory's perspective, learners with high prior knowledge are able to draw visual images from their image repository when presented with only an animation or reading a text. From the multiple intelligences theory's perspective, learners with high spatial ability are able to draw from their verbal and visual repositories when presented with contiguous word and picture presentations. Camtasia interactive digital videos allow learners with individual differences and different learning styles to tap into their unique learning qualities to gain knowledge.

- Coherence principle states that learners learn better in a multimedia environment when exposed to fewer rather than many words and pictures. The goal in this principle is to be concise and to the point in presenting concepts with pictures and words. For interactive digital projects teachers should minimize use of text and picture on slides, especially if video option is utilized. For digital materials with text and pictures, the design should focus on using the picture with few words to tell the story.
- The resulting effect of the coherence principle is called "redundancy effect." In designing interactive digital materials, learners gain knowledge better with animation and text, than from narration, animation and text. Text is a visual presentation, which competes with the animation for visual attention in the presentation. When video capabilities are utilized, teachers should not use animations since that will compete with the movement in the video. An animation as a short video file cannot compete with another video file in the same environment. Using CCMT as a theoretical framework for designing interactive digital materials, educators can create functional teachermade multimedia materials aligned with their curriculum, student learning outcomes and competencies. Materials created by teachers are more focused on addressing student learning outcomes. File created by teachers can easily be shared via the Web using Camtasia's web sharing location at screencast.com. This website provides 2 GBs of free space with an option to pay for additional space. The site generates a web address or URL, which is shared with students to access files anywhere and anytime.

Ms. Bender

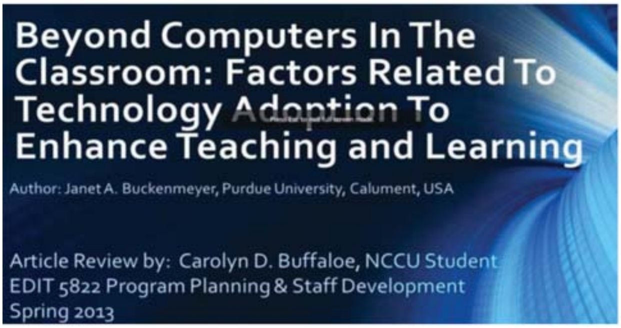
Introduction to the Scientific Method

Figure 2. Lecture presentation on the scientific method.



Thesis workshop

Figure 3. Lecture presentation on thesis workshop.

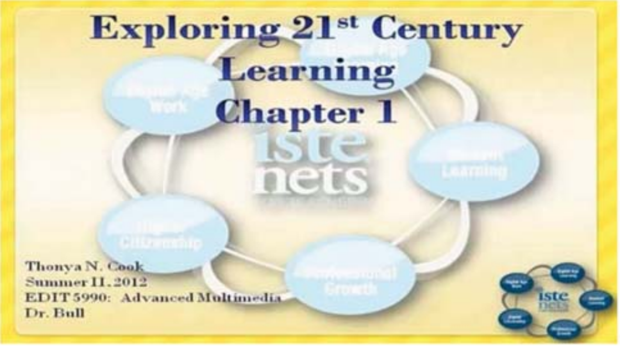


Beyond Computers In The Classroom: Factors Related To Technology Adoption To Enhance Teaching and Learning

Author: Janet A. Buckenmeyer, Purdue University, Calument, USA

Article Review by: Carolyn D. Buffaloe, NCCU Student
EDIT 5822 Program Planning & Staff Development
Spring 2013

Figure 4. Article review multimedia file.



Exploring 21st Century Learning Chapter 1

Thonya N. Cook
Summer II. 2012
EDIT 5990: Advanced Multimedia
Dr. Bull

Figure 5. Chapter presentation multimedia file.

Non-Linear PowerPoint Presentation_Part I
Prince Hycy Bull, Ph.D.

Figure 6. Lecture presentation multimedia file.

- Figures 2-6 show as variety of ways teacher made interactive digital materials can impact classroom learning:
  - Figure 2 is video presentation on the scientific method designed by a middle school science teacher; • Figure 3 is a video presentation on thesis writing by a high school language arts teacher;
  - Figure 4 is an article review done by a graduate student in my course. This exercise was done to flip class presentations;
  - Figure 5 is a chapter presentation done by a graduate student in my course. This is a shift from the tradition classroom presentation to flipping class presentations with video podcasts;

- Figure 6 is an of a class demonstration done in one of my graduate course to help students understand a concept;
- Figure 7 is an example of a mathematics lesson on numbers created by an elementary teacher.

INTERACTIVE DIGITAL MATERIALS AND MULTIPLE INTELLIGENCES THEORY

Gardner (1993) states that intelligence should not be measured as a singular entry, but by multiple entries relating intelligences possessed by all humans. Gardner's eight basic multiple entries could be addressed in different combinations in designing interactive digital materials for instruction. Multiple intelligences theory allows multimedia designers to create interactive learning materials aligned with universal design for learning and differentiated learning for students. These two principles are intricately aligned with learning and teaching styles.

The universal design for learning is a set of principles that provide all learners with equal opportunities to learn. Designing interactive digital learning materials with a focus on the multiple intelligences of learners provide learners with equal opportunities to learn based on their learning styles and combination of their intelligence entries in the areas of linguistic, logico-mathematics, spatial, bodily-kinesthetic, musical, interpersonal, intrapersonal, and naturalist intelligences.

Understanding how students demonstrate their intellectual capacity is an important factor for teachers in designing instruction to meet the specific learning needs of students who may be dominant in one or several intelligence as opposed to other forms of intelligence. On the other hand, differentiated instruction is a method of teaching that involves matching learning styles with abilities. This is best accomplished through intentional design of interactive multimedia files to address learning styles, academic levels, and intelligences of students to better facilitate the learning process.

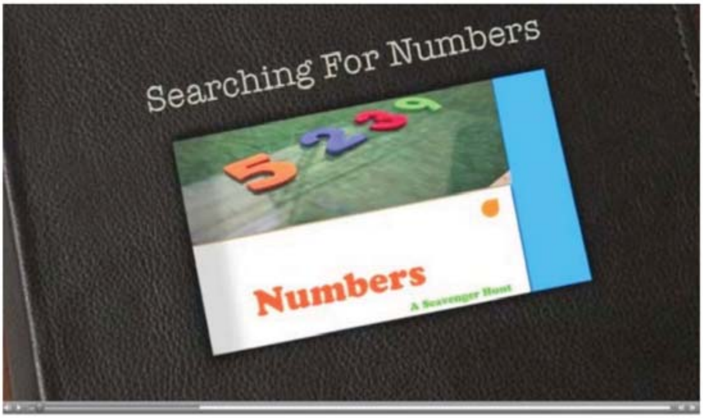


Figure 7. Lesson on numbers for kindergarten students.

The process of designing interactive digital materials to support universal learning design and differentiated learning aligned with multiple intelligences requires an understanding of how multimedia is aligned with each intelligence (Brown, 2005). The following is an overview of how multiple entries could be addressed in using CCTM as the theoretical framework to design interactive digital learning materials addressing verbal and visuals systems:

- Linguistic intelligence (Verbal system)—Interactive digital materials promotes use of this entry through a variety of options: Use of text, text overlays on videos, use of text in content area, presentation of text in videos, use of audio and audio that is part of the video (Moreno, 2006; Moreno, 2007).
- Logical-mathematical intelligence (Verbal and visual systems)—this entry is promoted in interactive digital learning materials through logical audio presentations, video presentations and mathematic/logical presentations; • Spatial intelligence (Visual system)—this entry is promoted in interactive digital learning materials through visual-spatial ideas, use of graphics, pictures, images, animations, applets, maps, and videos;
- Bodily-kinesthetic intelligence (Verbal and visual system) —this entry is promoted in interactive digital learning materials through the video components of Camtasia. Instructors can record demonstrations and movements, which students would emulate, work on or modify;

- Musical intelligence (Verbal and visual system)—this entry is promoted in interactive digital learning materials in a variety of ways. Music or sound could be embedded in the videos such as songs, musical videos, musical demonstrations and use of musical pieces as background, introductory or end of presentations. Audacity is more effective in designing audio files;
- Interpersonal intelligence (Verbal)—this entry is promoted in interactive digital learning materials through audio or video presentations. This entry deals with an individual's capacity to understand, perceive and relate to other people;
- Intrapersonal intelligence (Verbal and visual system)—this entry promotes use of interactive digital learning materials to support asynchronous learning. It provides students with opportunity to review presentations at their own pace and time and with many opportunities to do so;
- Naturalist intelligence (Visual system)—this entry promotes use videos and sounds of nature to create digital learning materials to engage and motive students. This entry also provides students with opportunities to see nature's environment and hear nature's sounds that they would experience in real life.

The combination of Gardner's multiple intelligences theory aligned with Mayer's six principles of multimedia design within CCMT promotes the design of effective interactive digital materials to engage, motivate, and achieve student learning outcomes.

INTERACTIVE DIGITAL LEARNING MATERIALS AND CONSTRUCTIVIST THEORY

The use of constructivist approach in instructional design and delivery of multimedia content of CCTM focuses on constructing knowledge through prior visual and verbal experiences. The constructivist theory of instruction is based on principles of learning that were derived from branches of cognitive science. The constructivist teaching approach theory makes effective use of students' prior knowledge and cognitive structures based on those experiences (Mayer et al., 1999; Bull, 2012). An example of what this means for designing interactive digital learning materials in that images, animations, pictures and graphics used should relate to student's prior experiences. Interactive digital learning materials support the integration of images, animations, pictures and graphic to promote learning through experiences aligned with visual and verbal repository systems of the learner (Moreno & Mayer, 2000).

To ensure that instruction presented as part of an interactive digital learning material is aligned with tenets of constructivism, teachers and educators should ensure that the content, design and presentations of interactive digital materials are aligned with following:

- Digital learning materials should be context based;
- Content presented should relate to objectives of the lesson or presentation;
- Digital learning materials should promote participation through active involvement;
- Digital learning materials should promote collaboration and engage students in learning;
- Digital learning materials presented should lay the foundation for students to develop autonomy and control over learning by creating new knowledge;
- Digital learning materials should promote personal growth of students by creating new knowledge;
- Learning outcomes should stimulates a perspective and an understanding rather than a prescribed outcome.

CASE STUDIES

Case Study #1

Anderson and Bull (2010) conducted a study on the effects of using teacher-made multimedia interactive materials on students' perceptions of learning with multimedia presentations and how multimedia promoted learning. Participants were 38 first grade students and two certified teachers.

Methodology

Researchers designed interactive digital materials on Black History Month teaching unit. Classroom teachers were trained on how to implement multimedia lessons. The study employed both qualitative and quantitative data analysis. Data for the qualitative analysis were obtained from observations, surveys, students' Black Wax Museum, and focus group discussion.

Case Study #1

Findings The case study findings yielded the following themes:

- Students were engaged in learning;
- Addressed different learning styles;
- Promoted use of prior knowledge aligned with visual and verbal knowledge;
- Promoted higher level thinking skills;
- Promoted literacy skills;
- Motivated students to learn.

Case Study #2

Researcher conducted a case study on the effects of flipping a graduate course with instructor and students' designed interactive digital materials on student learning outcomes. Participants were 10 graduate students and researcher.

Methodology

Research designed interactive learning materials for course presentations, lectures, and demonstrations of concepts. See Table 1 for samples of teacher-made interactive digital materials. Students were trained to use Camtasia and Audacity to design interactive learning materials for course presentations, article reviews, assignments, and chapter presentations. See Figures 2-6. Data for the qualitative analysis were obtained from observations, surveys, and focus group discussion.

Case Study #2

Findings Analysis of data revealed the following themes:

- Creating interactive digital materials promoted effective use of instructional time;
- Transformed students to producers of knowledge;
- Files were archived and access as needed;
- Supported synchronous and asynchronous learning;

- Students learn at their pace and time;
- Skills required to design interactive digital materials are perquisites to flip a classroom;
- Instructor's knowledge on how to design and flip was key to successful implementation;
- Promoted mobile learning as files shared via web portals were accessed with mobile devices.

BENEFITS OF INTERACTIVE DIGITAL MATERIALS

There are several benefits of using CCTM in designing teacher-made interactive digital learning materials. The following are some of the benefits:

- An effective way of delivering asynchronous learning;
- Promotes online delivery of instruction;
- Promotes flipping the classroom;
- Alternative method of instruction to face-to-face instruction;
- Easily capture and synchronize audio and video with slides from your PowerPoint presentations;
- Create rich-media presentations;
- Create an engaging rich-media presentation that can help students understand concepts better;
- Interactive digital learning materials address a variety of learning styles that tap into the multiple intelligences of students;
- Supports a wide variety of audio and video file formats, for high-quality audio and video;
- Enables capture of live audio and video using any capture device;
- Includes compelling transition and video effects.
- Saves time while offering great flexibility and high-quality output;
- Promotes anywhere and anytime learning;
- Promotes mobile learning as files can be accessed online on mobile devices.

DISCUSSION

Educational Implications

In general, when teachers create functional digital learning materials for their classrooms, learning is enhanced, students are engaged and motivated to learn. It also promotes a shift from commercially generated files to files that are more aligned with competencies and objectives. Teachers could develop training contents for tutorials, telefieldtrips, and virtual training. Interactive digital files address unique teaching and learning styles to improve how students learn. Promoting teacher-made digital learning materials would require that educational systems improve staff development training for teachers. Teachers can produce on-demand broadcast by capturing live video and audio presentations. It promotes a paperless environment in which educational institutions are able to archive events and other school activities. It lays the foundation for asynchronous learning via the web. Finally, it saves money as institutions would spend less on commercially designed materials.

CONCLUSION

Use of teacher-made interactive digital learning materials will revolutionize educational presentations and enhance elearning delivery when aligned with CCMT to produce dynamic presentations that create a balance between the learners' prior verbal and visual experiences, sensory repository, multiple intelligences and learning styles to construct new knowledge. Camtasia and Audacity are great tools with great benefits for synchronous and asynchronous learning. As is the case with designing interactive learning materials for learning, issues of copyright, fair use guidelines, access to camcorder or knowledge and use of camcorders, downloading and editing videos, time constraints, technical support and access to a server to upload final projects may pose some limitations to the effective use of Camtasia and Audacity in designing interactive digital learning materials.

ACKNOWLEDGEMENTS

The author thanks graduate students whose work are mentioned and displayed in this paper.

REFERENCES

1. Anderson, D., & Bull, P. (2010). Constructivist teaching approach: Integrating a multimedia tool to enhance student learning in a first grade classroom. In D. Gibson, & B. Dodge (Eds.), Proceedings of
2. society for information technology & teacher education international conference 2010 (pp. 3055-3059). Chesapeake, VA: AACE. Audacity. http://audacity.sourceforge.net
3. Brown, A. (2005). The challenging marriage of interaction design and learning design in multimedia educational products. In P. Kommers, & G. Richards (Eds.), Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2005 (pp. 29-932). Chesapeake, VA: AACE. http://www.editlib.org/p/20198
4. Bull, P. (2009). Cognitive constructivist theory of multimedia design: A theoretical analysis of instructional design for multimedia learning.
5. In G. Siemens, & C. Fulford (Eds.), Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2009 (pp. 735-740). Chesapeake, VA: AACE. http://www.editlib.org/p/31581
6. Bull, P. H. (2012). Using spatial constructivist thinking theory to enhance classroom instruction for students with special needs. In J.
7. Aitken, J. Fairley, & J. Carlson (Eds.), Communication technology for students in special education and gifted programs (pp. 66-81).
8. Hershey, PA: Information Science Reference. doi:10.4018/978-1-60960-878-1.ch005 Camtasia. http://www.techsmith.com/camtasia.html
9. Gardner, H. (1993). Multiple intelligences: The theory in practice. New York: Basic Books.
10. Mayer, R. E. (2002). Multimedia learning. Cambridge: Cambridge University Press.
11. Mayer, R. E., Moreno, R., Boire M., & Vagge S. (1999). Maximizing
12. constructivist learning from multimedia communications by minimizing cognitive load. Journal of Educational Psychology, 91, 638-643. doi:10.1037/0022-0663.91.4.638
13. Moreno, R., & Mayer, R. E. (2000). A learner-centered approach to multimedia explanations: Deriving instructional design principles from cognitive theory. Interactive multimedia electronic. Journal of

14. Computer-Enhanced Learning. http://imej.wfu.edu/articles/2000/2/05/index.asp

15. Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. Journal of Computer Assisted Learning, 22, 149-158. doi:10.1111/j.1365-2729.2006.00170.x

16. Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: Cognitive and affective consequences of signalling and segmentation methods. Applied Cognitive Psychology, 21, 765-781. doi:10.1002/acp.1348

CHAPTER 17

"EDUCATIONAL DIGITAL RECYCLING: DESIGN OF VIDEOGAME BASED ON INCA ABACUS"

Jorge Montalvo

University of Lima, Scientific Research Institute –IDIC Peru

INTRODUCTION

The use of multimedia applications in schools is quite common. However, according to Alfonso Gutiérrez (2003: 42), these are often attributed educational advantages, which perhaps they do not have and just like that, it is assumed it favors learning. From a creative design perspective, how can we make educational multimedia resources really contribute to an effective teaching-learning process? We believe a key aspect has to do with the possibility of digitally recreating or recycling some traditional teaching

Citation: Jorge Montalvo (March 7th 2012). "Educational Digital Recycling: Design of Videogame Based on "Inca Abacus", Interactive Multimedia, Ioannis Deliyannis, IntechOpen, DOI: 10.5772/36751.

Copyright: © 2012 by authors and IntechOpen. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

materials that have proven to be effective teaching tools. Among them, we find the so-called "Inca abacus" known in Peru and other countries of the region as "yupana".

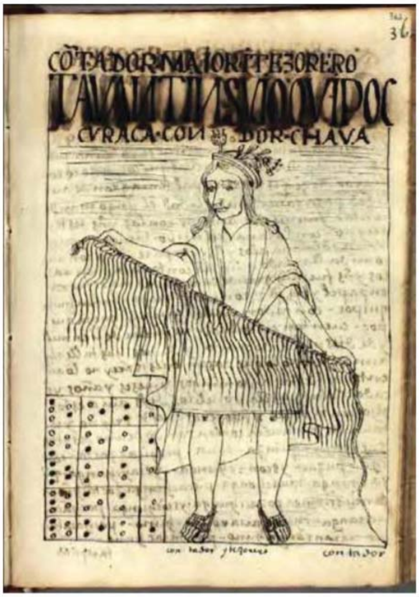


Figure 1. Drawing made by chronicler Felipe Guamán Poma

On the lower left corner of a 16th century drawing (figure 1), made by chronicler Felipe Guamán Poma, there is a sketch of a yupana next to the quipu; therefore, it is believed they were complementary calculation tools. The term yupana comes from the Quechua word "yupay" that means "to count". Originally, it consisted of a clay or stone tablet with several columns and boxes with small grooves for placing corn kernels. Researchers have not reached an agreement on how it was used in the olden days; however, for its use in schools, the interpretation made by the engineer William Burns has been chosen, which is based on the decimal system (Bousany, 2008: 18). The tablet has been turned into a horizontal position, where each column has the value of a multiple of ten and each circle has the value of one multiplied by the value in its column. The upper boxes are used as a memory or for exchange and underneath, there are ten circles, grouped in two, three and five units to facilitate counting (figure 2).

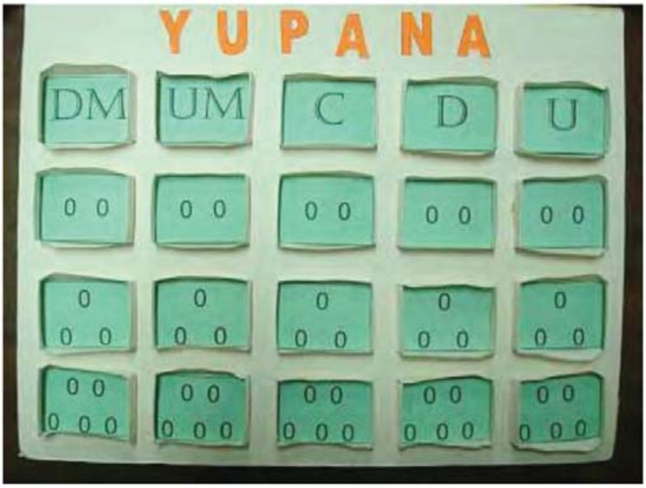


Figure 2. Representation of the decimal system in the yupana

In the eighties, the yupana was promoted as a teaching aid for the first elementary grades by the Peruvian teacher Martha Villavicencio; since then, some schools make use of traditional yupanas made from wood, cardboard or Styrofoam and numbers are represented with buttons or seeds (figure 3). There are research and testimonies certifying the educational capacity of the yupana in facilitating the understanding of the positional value of numbers in the decimal system and the execution of immediate arithmetic calculations (Vargas & López, 2000: 75). It is stated this is an educational and historic tool "[…] that helps students understand certain mathematical algorithms that are many times applied in a mechanical way without knowing the logical part." (Torres, 2009).

Taking into account the importance of mathematics in basic education and the difficulties and resistance generated in its learning, we decided to investigate the feasibility to digitally recycle the yupana with the purpose of strengthening its qualities and make it more attractive for digital natives. For this interdisciplinary work, we counted with the advisory services of a specialist in mathematics education, David Palomino, and of a multimedia application designer, David Chura.

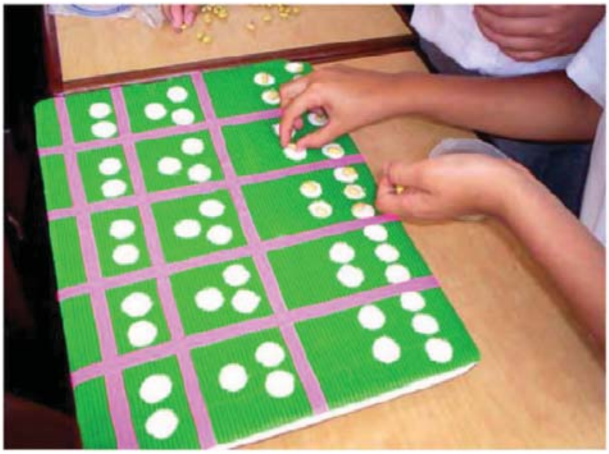


Figure 3. Use of the traditional yupana in school

EDUCATIONAL DIGITAL RECYCLING

Traditional learning and digital recycling

The origin and sense of the term "apprentice" is interesting. Professor Mariano Aral, a specialist in Spanish lexis, explains it in the following way:

> The complement of an apprentice is basically the teacher, that is, the traditional teacher (one who masters a trade). […] What was characteristic of all skills in relation to their respective apprentices was that it made them work (apprentices did not just learn, they learned by doing). […] When passing from a traditional society to an industrial one, it was discovered that the industry was not the best place for apprentices, they went straight to production and they had to know already and take a productive space in the company. Therefore, learning had to be achieved in school and not in the industry. […] It was a total revolution. Instead of masters they gave them teachers. Instead of providing them with machines, tools and working material, they gave them paper and pencils, and books and blackboards. Instead of practice, they were given explanations and more explanations and instead of turning them into apprentices, they remained as students. (Aral, 1999).

This interesting text from Mariano Aral expresses certain nostalgia for traditional education based on practical learning with specific materials. Other authors go beyond and link traditional culture with the digital era.

Juan Freire –in agreement with the expert in innovation Charles Leadbeater- believes that the industrial society was an anomaly and that the digital era is not a revolution but the recovery of ways to work, participate and share that we believed had been forgotten. The industrial revolution meant an increase in efficiency in detriment of the pleasure of learning by doing, of talking with those around us and of making collective decisions without external authorities. With the arrival of the digital era, Freire once again says, "[…] spectators become digital artisans, the profane knowledge is reappraised, and knowledge is shared once more, it is copied and remixed in a creative virtuous cycle." (Freire, 2006). A way of conceptually explaining this phenomenon is through McLuhan's laws on the media. According to Piscitelli's review (2005: 121-122), these laws state that each technology expands or amplifies a skill, by doing so, it outdates an older means and at the same time, it recovers something that was previously obsolete; and if it expands too much, it turns into something new. As we can see, the fact that a new technology rescues previous forms or means is a central part of the proposal.

We can define educational digital recycling as a way of recovering and transforming traditional learning materials to introduce them into a new life cycle more in tune with the new generations. Here, the sense of transformation involved in this process is worth pointing out. Today, there are virtual educational resources but many of them are limited to copying or simulating materials from the real world. In order to analyze this point, I would like to tell you a personal anecdote.

When my oldest son was three years old, he used to play with a Disney multimedia to learn to paint. The interactive design of the interface showed a brush, drawings for coloring, paint jars with basic colors and a pallet for mixing colors. The task or challenge consisted in painting the drawings with the same colors as in some reference pictures. My son always had a hard time obtaining green by mixing the yellow and blue paints and if he did, it was just by chance. One day, we were walking down the street and he needed to go to the bathroom. So, we went into a public restroom where they had just finished cleaning and the toilet water was blue from the disinfectant. When he started urinating, my son was surprised to see how the blue water changed color. I took the opportunity to explain that by mixing yellow with blue you get green and he never forgot this lesson. Moreover, he was capable of transferring what he had learned. Once he poured a yellow-colored soft drink into a blue glass and when he saw the glass turn green, he immediately remembered the restroom experience.

Some experiential learning processes are impossible to reproduce virtually. Therefore, the first condition for educational digital recycling is to start from a resource or material that can be subject to transformation and be enriched with digital technology. On the other hand, if we compare the Disney multimedia with the bathroom experience, we can discover several differences. The multimedia interface shows an environment that simulates an artistic surrounding, appropriate for someone who is getting ready to paint with a set of colors. On the other hand, a public restroom is an unusual and rather inappropriate place for learning to paint. But maybe, just because of that, it is capable of generating greater surprise and interest. Neuroscience has shown that emotion favors learning. According to John Medina (2010: 94- 95), when the brain detects an event with an emotional content, it releases dopamine, a neurotransmitter that aids memory and information processing. It is like if the brain placed a chemical post-it note with the statement: "¡Remember this!" so the information can be processed more thoroughly. Another basic difference between multimedia and the restroom experience is in relation to the way the green color appeared. In the Disney material, when the brush dipped in the blue paint touches the yellow paint in the palette, the green color appears automatically, as if by magic. In the toilet bowl, the effect is in real time, progressive, which fixes the attention on the mixture of colors and increases curiosity.

As we said before, many educational virtual resources simulate materials from the real world. For example, there are several digital applications of the Chinese abacus on the web (figure 4) that reproduce its physical characteristics with some interactive functions.

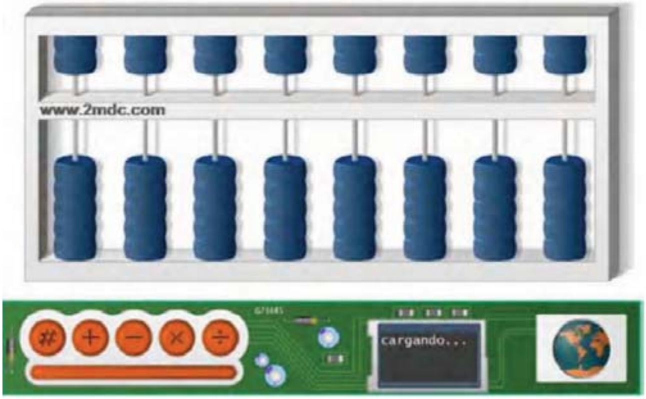
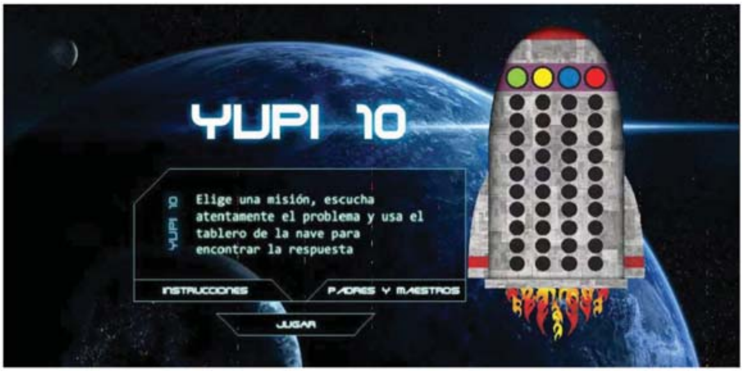


Figure 4. Digital version of the Chinese abacus

On the other hand, there are recent projects in Bolivia (Murillo, 2010) and Colombia (I.E. Once de Noviembre, 2010) of digital or virtual yupanas (figure 5). In all of these cases, they are multimedia resources that use digital technology but not the digital "culture"; that is, they are applications that do not take advantage of communication formats or styles that have developed around the new mediums.

Figure 5. Examples of virtual yupanas

In our case, we decided to creatively transform the yupana into a space videogame, which we called "Yupi 10" (figure 6). We only included four columns in the prototype (up to thousands) and we aligned all the circles vertically with the purpose of simplifying the design and facilitating its use, not only in PCs, but in mobile phones and tablets as well. The game was developed with Adobe Flash. In the case of platforms or devices that do not support Adobe Flash Player, we would have to develop versions of the game with other tools, such as Objective-C, Java or HTML5.



Elige una misión, escucha atentamente el problema y usa el tablero de la nave para encontrar la respuesta

Figure 6. Main screen of the Yupi 10 videogame

Yupi 10 basic rules

Educational videogames can be included in a category called "serious games." This concept usually includes games that have a purpose other than simple entertainment and whose application field is quite varied: ranging from military training to education in health and going through business training or artistic education. (Susi et al., 2007).

In relation to the educational potential of electronic games, it has been criticized (Buckingham, 2008: 173) that a large part of these packages are focused on the practice of out-of-context skills or factual contents of subjects, where the user just has to answer questions choosing from various options; more than teaching, the intention seems to be that of assessing skills or pre-acquired knowledge. Tejeiro & Pelegrina (2008: 135) quote studies, which alert of the possibility that educational videogames favor experiential cognition, based on reactions to successive events, but not reflexive cognition, which would enable applying what has been learned to other areas. It is stated that the context of the game causes higher motivation levels; however, this seems to interfere with reflexive cognition, so students are able to transfer their understanding of game principles to other games but not to extract the rules on which they are based. It has also been observed that without guidance from a teacher, participants in an educational videogame focus on the competitive nature of the game and not on following up of their own understanding. This has been interpreted in reference to Vygotsky's concept of the zone of proximal development, according to which individuals can pass a learning level when they are helped by a more competent person. Other authors have indicated that videogames can be useful for "[…]

teaching external abstractions such as mathematics or physics, but they have limitations when representing introspection or philosophy." (Zagalo, 2010: 65).

In reality, all these observations do not invalidate the use of electronic games in school; they just warn us about the conditions a videogame should meet to really be educational. We believe that the main condition has to do with the basic rules of the game, understood not so much as the playing instructions but rather as structural principles of the creative design. For example, a videogame in the form of a labyrinth would be appropriate to learn notions of laterality and spatial orientation (left / right, up / down). However, there are cases in which the labyrinth is only a recreational pretext to "catch" numbers, letters, animals or any object related to an allegedly educational subject. It would seem that first the format is decided upon and then content is sought for it. But this problem is not exclusive of the digital setting, in the real world there are tabletop games with a board and spaces to move, related to geography, road safety, environment, health, among others. In the case of Yupi 10, the basic rules of the game correspond to the following tasks. First the user has to choose a mission for the ship. In the prototype, we include nine missions (figure 7), with increasing levels of difficulty that correspond to second grade. All missions have an "audio-problem" form and represent acted out situations in which the captain and lieutenant of the ship take part. For example, there is an emergency and the number of oxygen tanks has to be determined through a subtraction or addition operation.

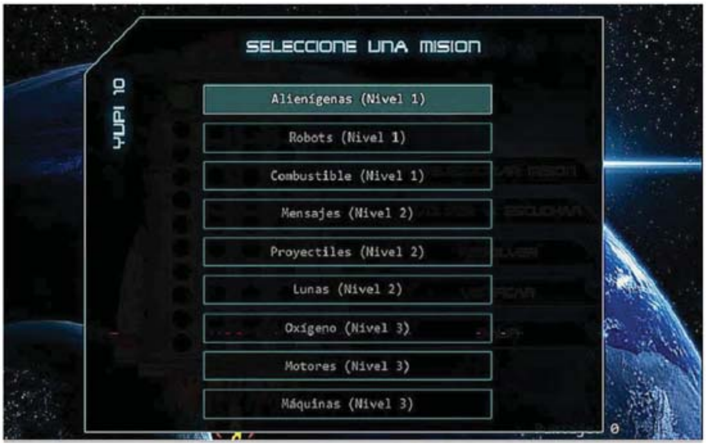


Figure 7. Missions with increasing levels of difficulty

The next step is to use the ship's color board to graphically represent the problem data (figure 8) and execute the corresponding arithmetic operation. Finally, the result is verified in the option menu in order to go on to the next mission.

Let's see how an addition operation is carried out (figure 9). Let's suppose we have to add 18 + 5. First we represent number eighteen by lighting eight units and one ten (figure 9a). Then we add the five units, but as we only have two available circles (figure 9b), we have to exchange the ten units for a ten. For this purpose, we turn off the entire red column with the upper circle and we light up a circle in the blue column (figure 9c). Then we continue adding the three remaining units and at the end the total appears represented: twenty-three (figure 9d).

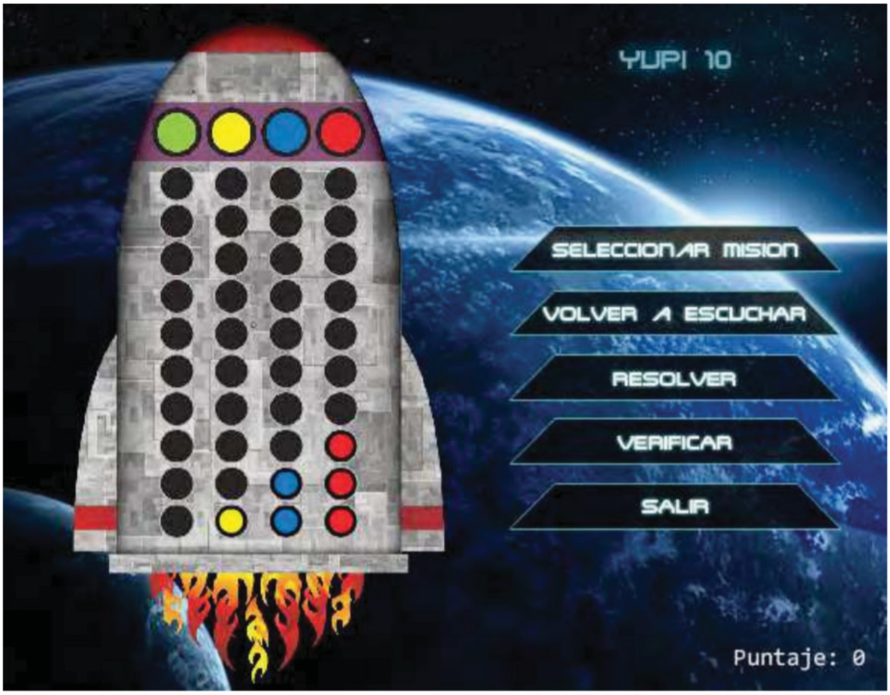


Figure 8. Color Board with data and option menu

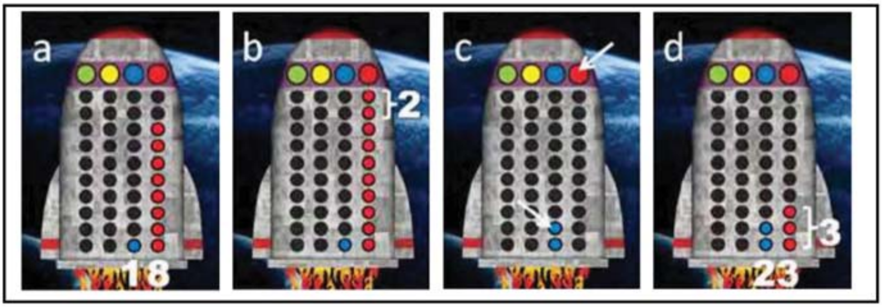


Figure 9. Addition example (18 + 5 = 23)

In a subtraction operation, the process is the other way around (figure 10). Let's suppose we want to subtract 23 – 5. First we represent number twenty-three (figure 10a). Then we subtract number five, but as there are only three circles available (figure 10b), we have to exchange a ten for ten units. For this purpose, we turn off a blue circle and we light up the entire red column with the upper circle (figure 10c). Then we continue turning off the two remaining red circles and at the end we have the subtraction result: eighteen (figure 10d).

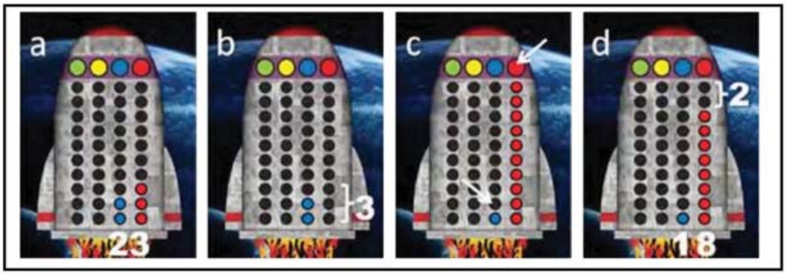


Figure 10. Subtraction example (23 – 5 = 18)

Interactivity and Sensory Elements

In relation to the educational videogame design, some authors state that unlike commercial videogames, to have very limited budgets "[…] has huge consequences in the quality of graphics and the level of interactivity and consequently, in its capacity to attract reluctant students." (Buckingham, 2008: 155). We believe that educational videogames should not be compared to commercial videogames but to school books, in face of which they are more attractive. In addition, there are studies that warn about the distracting effect of drawings that are too complex or real (Medina, 2010: 279). On the other hand, regarding the symbolic game - which has great influence in developing children's thoughts - experts affirm that playing materials "[…] the more simple and functional they are […], results obtained from the symbolic game will be positively better and greater (Licona, 2000). That is why in the case of Yupi 10 we chose a simple interface design to help give context to the game and facilitate interactivity.

Educational digital recycling of traditional material requires the identification of essential actions, which should be kept and which should be eliminated or modified. Lighting and turning off circles in Yupi 10 is an interactive way of putting and taking away seeds in the physical yupana. This is a key aspect because they are actions that represent addition and subtraction operations. Another important factor in an interactive application is to decide if certain function should be automatic or manual. For example, in the Yupi 10 exchange process, there is the technical possibility that after completing the ten circles in a column, all "automatically" turn off and one lights up in the following column; but it is preferably for the user to do it "manually" - like in the traditional yupana - so the person can assimilate decimal system equivalences better. In the case of the Disney multimedia we mentioned before, if the green color does not appear in an automatic or instant form but rather in a mechanic or progressive way, it would surely result in better learning on the blend of basic colors. Digital technology makes it possible to reproduce or strengthen both automatic processes, typical of the industrial society, as well as the manual systems, inherent to traditional society.

According to neuroscience experts, it is convenient for all educational material to be multisensorial: "When the sense of touch is combined with visual information, learning improves in almost thirty percent." (Medina, 2010: 244). It is also stated that nerve endings on our fingertips generate brain activity that helps in understanding. When you understand what is being taught, several brain areas are activated, whereas when you memorize without sense, the activity in nerve cells is much poorer (Fernández, 2010: 5). Several years ago, in relation to movies, Michel Chion (1993: 10) proposed that there are transensorial influences between what we see and what we hear. According to this author, you do not see the same when you hear and you do not hear the same when you see. At present, with the boom of touch screens and physical recognition systems, transensorial influences in the media and teaching materials will become more frequent and varied. In relation to Yupi 10, we have mentioned the exclusive audible nature of audioproblems. This decision seems to contradict the convenience for teaching materials to be multisensorial. However, we had several reasons to avoid the use of aiding images when presenting the problems. In the first place, several math teaching experts sustain that as of second grade, when children are introduced to arithmetic problems, which they have to read from their textbooks, they begin to lose the ability to listen that they developed in first grade. They also state that in daily life, the majority of arithmetic problems that arise and need to be solved are oral. Therefore, they recommend looking for an adequate balance in school between written and oral problems (Capote, 2005). On the other hand, audio stories require a higher level of attention and concentration, which helps children process and analyze information better. As Shanker (2010) indicates, the calmer, focused and alert a child is, the better will he be able to integrate sensory information received by his brain and assimilate it and organize his or her thoughts and actions.

Levels of Difficulty and Reasoning

Another important aspect related to problems was determining the way to write and classify them according to levels of difficulty. Several math teaching specialists indicate that in order to solve a problem effectively, boys and girls should make a global analysis of the text meaning, that is, to understand the problem formulation. In this sense, there is a semantic classification of verbal elemental arithmetic problems into four categories: combination, change, comparison and equating (Puig & Cerdán, 1988). Combination problems are the simplest ones and describe the relationship between two parts and a whole. For example: "There are 35 people at a meeting, 12 are men; how many of them are women?" Here, the unknown factor is one of the parts that make up a whole. Another possibility is for the unknown factor to be the whole: "There are 12 men and 23 women at a meeting; how many people are there in total?"

Change problems have a slightly higher level of difficulty. There is an initial amount, a change amount and the final amount. For example: "Ana had 12 coins, she earns 7 coins; how many coins does she have now?" Here, the unknown factor is the final amount, but it can also be the initial amount: "Ana had some coins, she earns 7 coins, now she has 19 coins; how many coins did she have at the beginning?" Another possibility is for the change amount to be the unknown factor: "Ana had 12 coins, she earns some coins, now she has 19 coins; how many coins did she earn?" These same change problems can be expressed in a decreasing way, replacing the expression "earning coins" for "losing coins."

The degree of difficulty of comparison problems is higher and shows a static relationship between two quantities: the referential one and the compared quantity. For example: "Juan is 8 years old, Ana is 13. How much older is Ana than Juan?" Here the unknown factor is the difference but it can also be the compared quantity: "Juan is 8 years old, Ana is 5 years older than Juan. How old is Ana?" Another possibility is for the unknown factor to be the referential quantity: "Ana is 13 years old and 5 years older than Juan. How old is Juan?" These same problems can be formulated in a negative way by replacing the expression "older than" for "younger than." Equating problems are usually the most difficult and also imply comparisons between two quantities but using a connector of the type "as much as" or "equal to." For example: "Juan weighs 27 kilos, Ana weighs 18 kilos. How many kilos does Ana have to gain to weigh as much as Juan?" Here, the unknown factor is the difference but it can also be the compared quantity: "Juan weighs 27 kilos, if Ana gains 9 kilos she would weigh as much as Juan. How many kilos does Ana weigh?" Another possibility is that the unknown factor is the referential quantity: "Ana weighs 18 kilos, if she gains 9 kilos she will weigh as much as Juan. How many kilos does Juan weigh?" As in the previous case, these same problems can be expressed in a negative way by replacing "gain kilos" for "lose kilos".

The degrees of difficulty of these four categories are not absolute. There are studies that suggest that some specific forms of an inferior category are more difficult that other specific forms of a higher category. For example, the change form: "Ana had some coins, she loses 7 coins, now she has 19 coins. How many coins did she have at the beginning?", is often more difficult than the comparison form: "Juan is 8 years old, Ana is13. How much older is Ana than Juan?" This semantic classification of verbal elemental arithmetic problems puts emphasis not so much on children's calculation skills but on their analysis and reasoning ability. Based on these criteria, we drew up the nine missions of the Yupi 10 prototype and we grouped them into three levels of difficulties: initial, intermediate and high.

Validation of the Prototype and Results

The challenge of every educational innovation consists in being able to turn knowledge objects into objects of desire. (Ferrés, 2008: 180). In this sense, during the qualitative validation of the prototype with three groups of children between the ages of 7 and 9, we could observe that the videogame awaken interest and curiosity, despite the simplicity of its design and the usual resistance math generates in students. However, we also noticed an inconvenience: some children listened to the audio problems too quickly and decided to do the problem - addition or subtraction - without thinking too much, almost by guessing and when they checked the answer and it said "error detected," they simply carried out the opposite operation and they got "mission accomplished". They noticed that by choosing the operation at random, they had a 50 percent chance of being successful, which encouraged a conduct similar to throwing dice to see what they got. As we mentioned before, audio problems were designed based on a semantic classification that requires analysis and reasoning. Consequently, the children's attitude contradicts the educational purpose of the multimedia. In order to correct this defect, it would be convenient to include in the game instructions a score system: winning points for each right answer and losing points for each mistake. This reward and sanction system is usual in some commercial videogames and we believe it could be useful for Yupi 10, despite the fact it is a behaviorist strategy. During the validation, we tentatively assessed this. When we warned the children: "think well, if you make a mistake, you can lose points," they listened to the audio problems again to analyze them better.

Here we need to ask ourselves about the importance of chance in playing activities. In football for example, chance is a factor that goes beyond players' ability and that in addition to adding interest to the match, it can influence the final result. Would it be appropriate to add the chance component into an educational videogame? We believe that eventually it could be used in aspects not related to those we have called basic rules of the game (unless understanding the notion of "by chance" is the educational objective of the material). Thus, this would increase the participants' interest without affecting the expected learning.

Another inconvenience we noticed during the validation was that some children, instead of using the ship board to show the problem data on a graph and carry out the respective operation, made mental calculations (or used their fingers) and only used the interface to check the result, especially the easiest problems. We got the impression children wanted to find the answer very quickly and thought that using the ship board was too slow. Probably this attitude is influenced by the customary practice of commercial videogames in which the reaction speed is a key factor in success and also by the habit of some teachers to value results of an operation more than the process followed. In any case, it is an attitude that does not contribute to the purpose of incentivizing reflection and reasoning. A way to overcome this inconvenience would be to include in the multimedia the possibility of checking not only the result of the operation but also some of the previous steps: for example, representing the problem date adequately and choosing the addition or subtraction option correctly. Thus, we would be promoting a methodic and progressive attitude in children in order to reach a goal.

Regarding teachers, a notable result of validation has to do with its mediating function between the material and the children. In a previous research on audiovisual riddles (Montalvo, 2011: 130) we maintained that such function consisted in facilitating "clues" or individual aid to students and that this type of personalized tutoring could hardly be assumed by a machine because technology tends to standardize users. The validation of Yupi 10 supports this criterion. The videogame test allowed us to experiment situations in which we had to assume the role of an aiding facilitator suited to each child. In one case, it was necessary to teach how to represent numbers on the ship board. In other cases, we had to put illustrative examples or use graphs or diagrams - made on the spot - to clarify the sense of an audio problem. In general, we verified that there are many differences and gaps in the students' previous knowledge, which implies that the mediating participation of the teacher would be essential if we want to achieve significant learning processes with Yupi 10.

CONCLUSION

In this chapter, we have systemized the experience of designing a multimedia application in the form of an educational videogame based on a traditional material of proven educational effectiveness. The study of the Yupi 10 case allows us to affirm that educational videogames are really educational when the basic rules of the game (structural principles of creative design) fully agree with the material's learning objectives. Otherwise, they only incentivize the review of previously acquired knowledge and correspond to a multimedia application category that conceive the educational process more as the transmission of information rather than the building principles of learning. On the other hand, it is worth highlighting the way math problems are formulated in Yupi 10, which seeks to privilege reasoning over calculation skills. We believe it is a very appropriate approach for current times because in a society where there is an abundance of accessible data, knowing how to reason is one of the competencies needed to analyze information and turn it into knowledge. Finally, in relation to interactivity, we must point out the educational value of manipulative actions, which are usually minimized as opposed to multimedia visual and audio aspects. Therefore, maybe we should rather call educational videogames "videotoys" so we can always keep in mind their multisensory nature.

Regarding the educational digital recycling, we believe this is a timely strategy for this transitional era, in which we can foresee the future without losing sight of the past. In order to appreciate the creative flexibility of the digital society, we have to imagine it as an inclusive culture capable of taking in and strengthening the best educational resources, both from traditional society as well as from the industrial society. Could it be that maybe we are so fascinated with new technologies that we forget about the good educational practices of the past? How many valuable traditional resources created by anonymous teachers in small schools could be digitally recycled to increase their benefits and launch them to the world? Digital culture knows no time or space boundaries. The Inca yupana was created more than 600 years ago in a long-gone society and was rescued as educational material at the end of the last century. Today, by digitally recycling it as Yupi 10, maybe it will have the opportunity of starting a new life cycle.

REFERENCES

1. Aral, M. (1999). Aprendizaje. (Access: 11 February 2011) Available from: <http://www.elalmanaque.com/julio/21-7-eti.htm>.
2. Bousany, Y. (2008). Yupanchis. La matemática inca y su incorporación a la clase. ISPCollection. Paper 1. (Access: 20February 2010) Available from: <http://digitalcollections.sit.edu/isp_collection/1>.
3. Buckingham, D. (2008). Más allá de la tecnología. Aprendizaje infantil en la era de la culturadigital. Manantial, ISBN 978-978-500-112-1, Buenos Aires, Argentina.
4. Capote, M. (2005). Planteamiento de un problema aritmético con texto en la escuela primaria. (Access: 8 October 2010) Availablefrom: <www.ucp.pr.rimed.cu/sitios/revistamendive/nanteriores/Num8/pdf/3.pdf>.
5. Chion, M. (1993). La audiovisión. Introducción a un análisis conjunto de la imagen y el sonido. Paidós. Retrieved from: <www.lapizdigital.com.ar/.../la%20audiovisión%20-%20michel%20chion.pdf>,
6. Barcelona, Spain. Fernández, J. A. (2010). Neurociencias y enseñanza de la matemática. Prólogo de algunosretos educativos. In: Revista Iberoamericana de Educación, nº 51, OEI. (Access: 22
7. August 2010) Available from: <http://fernandezbravo.ning.com/profiles/blogs/algunos-articulos-ydocumentos>.
8. Freire, J. (2006). La era industrial fue una anomalía. In: Nómada [blog]. (Access: 12 February2011) Available fro m: <http://nomada.blogs.com/jfreire/2006/12/la_era_industri.html>.
9. Ferrés, J. (2008). La educación como industria del deseo. Un nuevo estilo comunicativo. Gedisa. ISBN: 978-84-9784-288-4,Barcelona, Spain.
10. Gutiérrez, A. (2002). Alfabetización digital. Algo más que ratones y teclas. Gedisa. ISBN: 84-7432-877-2, Barcelona, Spain.
11. I.E. Once de Noviembre (2010). Proyectos pedagógicos. In: Institución Educativa Inem Once denoviembre [blog]. (Access: 20 March 2011) Available from: <http://inemoncedenoviembre.blogspot.com/>.
12. Licona, A. (2000). La importancia de los recursos materiales en el juego simbólico. Revista Pixel-Bit, n°14.(Access: 10February 2011) Available from: <http://www.sav.us.es/pixelbit/pixelbit/articulos/n14/n14art/art142.htm>.

13. Medina, J. (2010). Los 12 principios del cerebro. Una explicación sencilla de cómo funciona para obtener el máximo desempeño.Norma. ISBN: 978-958-45-2897-1, Bogotá, Colombia.
14. Montalvo, J. (2011). Adivinanzas audiovisuales para ejercitar el pensamiento creativoinfantil. In: Comunicar, nº 36,(March, 2011), pp. 123-130, Grupo Comunicar, ISSN: 1134-3478, Andalucía, Spain.
15. Murillo, J. (2010). Yupana digital. Available from: <http://es.scribd.com/doc/36501843/Yupana- Digital>.
16. Piscitelli, A. (2005). Internet, la imprenta del siglo XXI. Gedisa. ISBN: 84-9784-060-7, Barcelona, Spain.
17. Puig, L. & Cerdán, F. (1988). Problemas aritméticos escolares. Síntesis. (Access: 8 October 2010) Retrieved from:<http://www.uv.es/puigl/lpae3.pdf>. Madrid, Spain.
18. Shanker, S. (2010). Self-regulation: calm, alert and learning. Education Canada, vol. 50. (Access: 22 August 2010) Available from: <http://www.cea-ace.ca/educationcanada/article/self-regulation-calm-alert-and-learning>.
19. Susi, T.; Johannesson, M. & Backlund, P. (2007). Serious games-an overview. University of Skövde, Suecia. Available from: <http://74.125.155.132/scholar?q=cache:ttMLMX_9wRUJ:scholar.google.com/+serious+games&hl=es&as_ sdt =2000>.
20. Tejeiro, R. & Pelegrina, M. (2008). La psicología de los videojuegos. Un modelo de investigación. Aljibe. ISBN: 978-84-9700-440-4, Málaga, Spain.
21. Torres, I. (2009). Proyecto interdisciplinario la yupana para aprender matemática en el marco de la enseñanza para lacomprensión. Colegio Abraham Lincoln. Available from <http://cibem6.ulagos.cl/ponencias/ cibemPresentGimnasio/= Presentacion%20de%20la%20 conferencia%20puerto%20montt%202009.ppt>.
22. Vargas de Avella, M. & López de Castilla, M. (2000). Materiales educativos. Relato de una experiencia en Bolivia, Ecuador y Perú. Convenio Andrés Bello, Bogotá, Colombia.
23. Zagalo, N. (2010). Alfabetización creativa en los videojuegos: comunicación interactiva y alfabetización cinematográfica. In: Comunicar, nº 35, (October, 2010), pp. 61-68, Grupo Comunicar, ISSN: 1134-3478, Andalucía, Spain.

CHAPTER 18

MULTIMEDIA APPROACH IN TEACHING MATHEMATICS – EXAMPLES OF INTERACTIVE LESSONS FROM MATHEMATICAL ANALYSIS AND GEOMETRY

Marina Milovanović [1], Đurđica Takači [2] and Aleksandar Milajić [3]

[1]Faculty of Real Estate Management, Union University, Belgrade, Serbia
[2]Faculty of Natural Sciences, University of Novi Sad, Novi Sad, Serbia
[3]Faculty of Management in Civil Engeneering, Union University, Serbia

INTRODUCTION

Research on multimedia approach efficiency was carried out using lessons about the definite integral and isometric transformations (line and point reflection, translation and rotation), since these areas are among the basic ones in the fields of mathematical analysis and geometry. These topics are also important because of their presence in the mathematics programmes in the great majority of high schools and faculties, both directly and through

Citation: Marina Milovanović, Đurđica Takači and Aleksandar Milajić (March 7th 2012). Multimedia Approach in Teaching Mathematics – Examples of Interactive Lessons from Mathematical Analysis and Geometry, Interactive Multimedia, Ioannis Deliyannis, IntechOpen, DOI: 10.5772/36989.

Copyright: © 2012 by authors and IntechOpen. This paper is an open access article distributed under a Creative Commons Attribution 3.0 License

their multipurpose character. Consequently, proper approach in presenting these topics is one of the most important segments of teaching mathematics on all educational levels In teaching mathematics, it is remarkably important to avoid so-called 'knowledge/ information adoption' as the only way of work. Students often solve problems mechanically, by following the algorithm steps without real awareness of their actual meaning. For example, in case of the definite integral, one of the common problems is that students calculate its value by following steps of the algorithm, without real understanding of its definition, the meaning of upper and lower limit or relation between the definite integral and the size of an area or a volume etc. In case of isometric transformations, it was noticed that students learned what are line and point reflection, translation and rotation, but the problem appeared when they have been faced to the realistic task, where they were supposed to use the adopted knowledge. One of the most important aims of the modern approach in teaching mathematics is to combine information adoption with so-called 'knowledge transfer', i.e. learning through defining the facts and explaining their interrelations.

Mathematics teachers show great interest in visualisation of the mathematical terms and emphasize that visualised lectures are of the great help in abstract thinking in mathematics (Bishop, 1989; Tall, 1986) believes that it is of the major importance to connect the existing pictures that students have on certain terms in order to develop them further and to enable students to accept the further knowledge. Therefore, in teaching mathematics it is necessary to combine the picture method and the definition method in order to improve the existing knowledge and to enlarge it with the new facts, which is one of the points of the cognitive theory of multimedia learning (Mayer, 2001, 2005).

Nowadays, use of different kinds of multimedia is largely included in the education because it allows the wider spectrum of possibilities in teaching and learning. Visualisation is very useful in the process of explaining mathematical ideas, abstract terms, theorems, problems etc. Experience in work with students showed that they are highly interested in modern methods in learning which include all kinds of multimedia, such as educational software, internet, etc. Modern methods in multimedia approach to learning include the whole range of different possibilities applicable in mathematics lectures for different levels of education and with different levels of interactivity (Deliyiannis et al. 2008a, 2008b; D. Đ. Herceg, 2009; Milovanovic, 2005, 2009; Milovanovic et al., 2011; Takači, et al., 2003, 2004, 2006, 2008). The authors usually work on suggestions on using different kinds of software in education, especially in the field of mathematics: geometry, algebra, numerical analyses etc. (Deliyiannis et al., 2008a, 2008b; D. Đ. Herceg, 2009; Milovanovic, 2005, 2009; Milovanovic et al., 2011), as well as the definite integrals and isometric transformations.

All the above-mentioned resulted in an idea of making applicative software which would be helpful in modern and more interesting approach to the field of teaching mathematics and rising the students' knowledge from the scope of definite integrals and isometric transformation to a higher level. The aim of our research was to recognize the importance of multimedia in the teaching process as well as to examine the students' reaction to this way of learning and teaching. Therefore, we have developed experimental software with multimedia lessons about the definite integral and isometric transformations and tested them in class in order to see how they would affect teaching process and results.

MULTIMEDIA PRESENTATION OF GIVEN PROBLEMS FROM THE FIELDS OF INTEGRALS AND ISOMETRIC TRANSFORMATIONS

We would like to emphasize the importance of using computers, i.e. multimedia software in teaching and learning in the both areas, because visual presentation offers much more possibilities. Beside that, using multimedia lessons on isometric transformations enables students to actually see not only the final solution, but also the movements that have led to it.

Multimedia Presentation of the Area Calculation

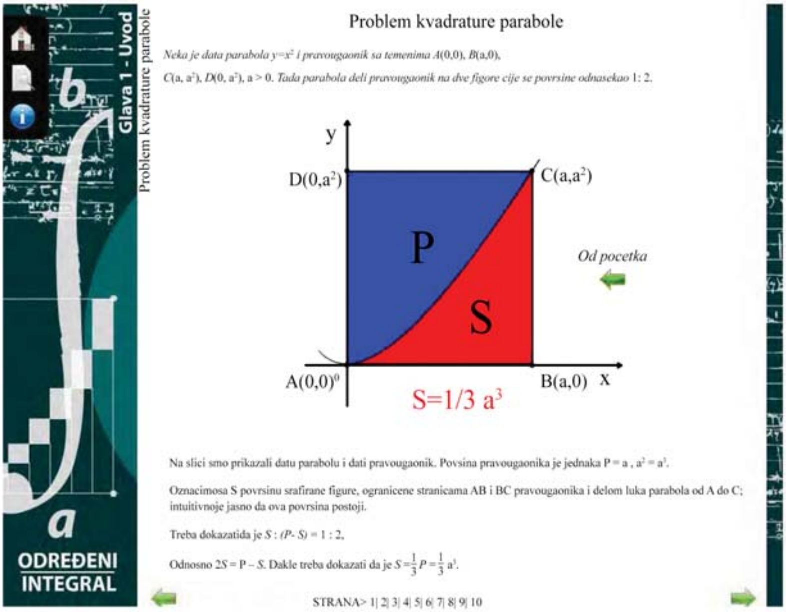
The problem of area calculation in the filed of integration calculations lead us directly to the definition of the definite integral. The basis for the defining and calculation of the definite integrals was made by Archimedes1 and his quadrature of the parabola. Because of that, we decided to use modern, multimedia approach in explaining Archimedes' quadrature of the parabola to the students. In order to use PC as a teaching aid, we were led by suggestions of Tall (Tall, 1991), who emphasized the importance of PC in the teaching because of its great possibilities in the scope of visual presentations.

The quadrature of the parabola problem is formulated as follows: For any given parabola $y = x^2$ and rectangle with nodes $A(0,0)$, $B(a, 0)$, $C(a, a^2)$ and $D(0, a^2)$, $(a > 0$)the parabola divides the rectangle in two zones of which the area of one is twice bigger than the area of the other one. Given problem is show on Figure 12. The area of the rectangle is:

$$P = a \cdot a^2 = a^3 \tag{1}$$

If we mark the zone under the parabola – limited by sides AB and BC of the rectangle and the arc AC – with S, and the other one with P, our next task is to prove the following equation:

$$S:(P-S)=1:2\text{ , i.e. } S=\frac{1}{3}P=\frac{1}{3}a^3 \tag{2}$$



Problem kvadrature parabole

Neka je data parabola $y=x^2$ i pravougaonik sa temenima $A(0,0)$, $B(a,0)$,

$C(a, a^2)$, $D(0, a^2)$, $a > 0$. *Tada parabola deli pravougaonik na dve figore cije se povrsine odnasekao* 1: 2.

Na slici smo prikazali datu parabolu i dati pravougaonik. Povsina pravougaonika je jednaka P = a , a² = a³.

Oznacimosa S povrsinu srafirane figure, ogranicene stranicama AB i BC pravougaonika i delom luka parabola od A do C; intuitivnoje jasno da ova povrsina postoji.

Treba dokazatida je $S : (P- S) = 1 : 2$.

Odnosno $2S = P – S$. Dakle treba dokazati da je $S = \frac{1}{3}P = \frac{1}{3}a^3$.

Figure 1. Illustration of given problem taken from the multimedia lesson about the definite integrals

The next step is to divide the interval [0, a] of the Ox axis in n equal parts of length a/n, where n is a natural number, which should be shown to the students by animation (step-bystep), as shown in Figure 2a. Within each of these intervals, we construct two rectangles: 'circumscribed' one, with upper right vertex on the parabola (the animation of this step is shown on Figure 2b), and 'inscribed' one, with upper left vertex on the parabola (shown on Figure 2d). It is obvious that the first part have no inscribed rectangle. The heights of these rectangles are shown on Figure 2b, and their areas are as follows:

$$\frac{a}{n}\cdot\left(\frac{a}{n}\right)^2,\frac{a}{n}\cdot\left(2\frac{a}{n}\right)^2,\dots,\frac{a}{n}\cdot\left((n-1)\frac{a}{n}\right)^2,\frac{a}{n}\cdot\left(n\frac{a}{n}\right)^2.$$

(Part of this animation is shown on Figure 2c.). Area of each inscribed rectangle is difference between the area of adjoining circumscribed rectangle and the area of 'added' rectangle (part of this animation is shown on Figures 2d and 2e)

$$P_U < S < P_O \tag{3}$$

Therefore: $\frac{a^3}{3}-\frac{a^3}{2n}+\frac{a^3}{6n^2}<S<\frac{a^3}{3}+\frac{a^3}{2n}+\frac{a^3}{6n^2}$, that is: $-\frac{a^3}{2n}+\frac{a^3}{6n^2}<S-\frac{a^3}{3}<\frac{a^3}{2n}+\frac{a^3}{6n^2}$.

These inequalities are correct for any given natural number n.

Since $\lim\limits_{n\to\infty}(-\frac{a^3}{2n}+\frac{a^3}{6n^2})=\lim\limits_{n\to\infty}(\frac{a^3}{2n}+\frac{a^3}{6n^2})=0$, we can conclude that:

$$S=\frac{1}{3}P=\frac{1}{3}a^3 \tag{4}$$

This was the solution of the given problem via numerical method, but we can offer much more by using the multimedia lessons. In animation shown on Figure 3, students can clearly see that with increasing of n, i.e. the number of circumscribed and inscribed rectangles, these areas will get closer and closer, until they, according to our intuition and visual perception, both become equal to the area S.

Led by the similar idea as in previous example, we will try to calculate the area of the curvilinear trapezium (students will see it in animation, as shown in Figure 4a and b, etc.) formed by the graph of the function $y=f(x),x\in[a,b]$, the abscissa's segment [a b,] and the two segments of the lines x = a and x = b making the figure closed (Figure 4a).

If the values $x_0,x_1,\dots,x_{n-1},x_n$ define points on x axis as follows:

$$a=x_0<x_1<\dots<x_{n-1}<x_n=b$$

these points divide interval [a, b] into n sub-intervals:

$$[x_0,x_1],[x_1,x_2],\dots,[x_{n-1},x_n].$$

Therefore, we can name the $(n+1)$-plet $(x_0,x_1,\dots,x_{n-1},x_n)$ as division of interval [a b,] . For simplification, we will mark it as

$\Pi = (x_0, x_1, ..., x_{n-1}, x_n)$.

If we choose any of these sub-intervals (Figure 4b), for example (x_{i-1}, x_i), and if ξ_i is arbitrary value within that sub-interval, the area of the rectangle whose basis is sub-interval $[x_{i-1}, x_i]$ and height is $f(\xi_i)$. can be calculated as:

$$P_i = f(\xi_i)(x_i - x_{i-1}) \tag{5}$$

If we do the same with every sub-interval $[x_{i-1}, x_i]$, $i = 1, 2, ..., n$, , we will get the series of rectangles – figure S – with total area:

$$P(S) = \sum_{i=1}^{n} P_i = f(\xi_i)(x_i - x_{i-1}) \tag{6}$$

For a given curvilinear trapezium – i.e. for given interval [a b,] and given function f(x) – the shape of figure S depends on division $\Pi = (x_0, x_1, ..., x_{n-1}, x_n)$ – and on choice of values $\xi_i \in [x_{i-1}, x_i]$, $i = 1, 2, ..., n$. Let us mark this n-plet of choices as $\xi = (\xi_1, \xi_2 ..., \xi_n)$. If all the sub-intervals $[x_{i-1}, x_i]$, $i = 1, 2, ..., n$, are 'small', shape of figure S will be 'very' similar to the curvilinear trapezium F (which are shown on Figures 5 and 6a).

If we mark value of $\Delta x_i = x_i - x_{i-1}, i - 1, 2, ... n$, than set $\{\Delta x_1, \Delta x_2, ..., \Delta x_n\}$ is finite set of positive numbers, and consequently has the largest element, which we will mark as d:

$$d = d(\Pi) = \max\{\Delta x_1, \Delta x_2, ..., \Delta x_n\}$$

If the value of d is small enough natural number, it means that sub-intervals are smaller and the division Π is 'fine'. If we introduce new breaking points, d gets smaller and smaller so the division gets finer. Consequently – and according to our intuition and multimedia presentation – figure S will get more and more similar to the curvilinear trapezium, so we can conclude that following definition of the area of the curvilinear trapezium F is valid:

Definition 1: Real number S is the area of the curvilinear trapezium F if for every ε > 0 , there exists 0 δ > , such that for every division Π for which d(Π) < δ and for any chosen set of values $\xi = (\xi_1, \xi_2 ..., \xi_n)$ in correspondent sub-intervals:

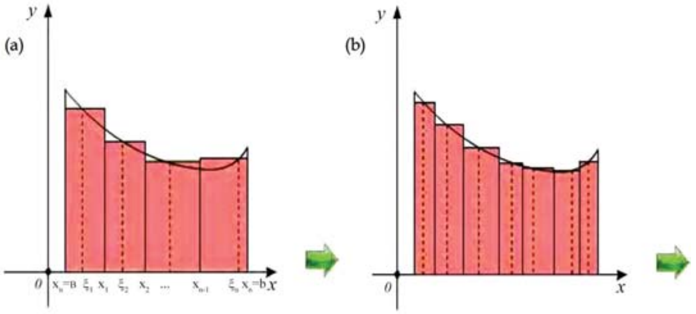
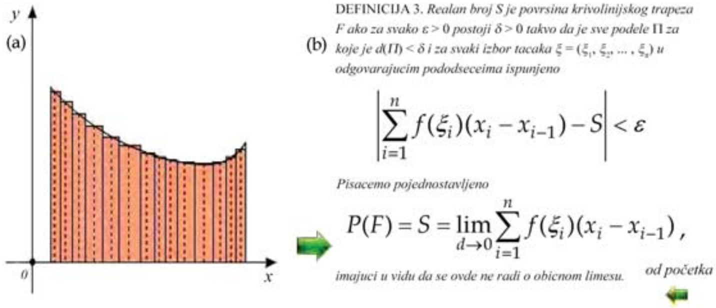


Figure 5. The second part of solution of the quadrature of the parabola problem taken from the multimedia lesson (step-by-step)



DEFINICIJA 3. *Realan broj S je povrsina krivolinijskog trapeza F ako za svako ε > 0 postoji δ > 0 takvo da je sve podele Π za koje je d(Π) < δ i za svaki izbor tacaka ξ = (ξ₁, ξ₂, ... , ξₙ) u odgovarajucim pododseceima ispunjeno*

$$\left|\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) - S\right| < \varepsilon$$

Pisacemo pojednostavljeno

$$P(F) = S = \lim_{d\to 0}\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}),$$

imajuci u vidu da se ovde ne radi o obicnom limesu.

Figure 6. Definition 1: Area of the curvilinear trapezium F – visualization method

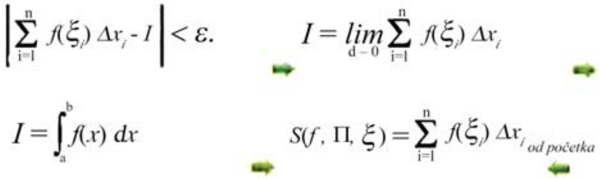
$$\left|\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) - S\right| < \varepsilon \tag{7}$$

Or it is simplified as:

$$P(F) = S = \lim_{d\to 0}\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}), \text{ i.e.} \tag{8}$$

Definition 2: Let the real-valued function f be defined on interval [a b,] . The real number I is definite integral of the function f on the interval [a b,] if for every $\varepsilon > 0$, there exists $\delta > 0$, such that for every division $\Pi = (x_0, x_1, \ldots, x_{n-1}, x_n)$, $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$ for which $d = d(\Pi) = \max\{\Delta x_1, \Delta x_2, \ldots, \Delta x_n\} < \delta$ and for any chosen set of values

$\xi = (\xi_1, \xi_2 ..., \xi_n)$, ., $\xi_i \in [x_{i-1}, x_i]$, $i = 1,2,...,n$ (animation, Figure 7 shows further development step-by-step).



$$\left|\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i - I\right| < \varepsilon. \qquad I = \lim_{d-0}\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i$$

$$I = \int_a^b f(x)\,dx \qquad S(f, \Pi, \xi) = \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i$$

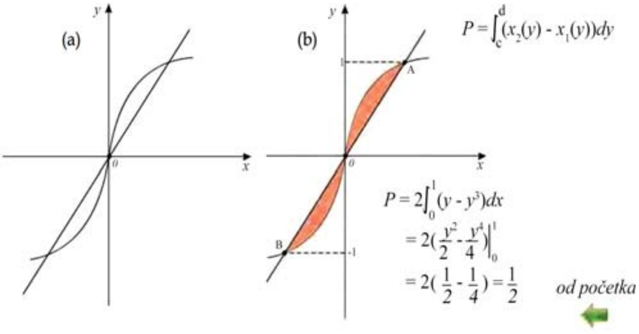
Figure 7. Definition 2: Area of the curvilinear trapezium F – visualization method

With numerous visual presentations, animations, illustrations and examples we can also introduce and explain integrability, integral sum, integrand, limits of integration, Newton-Leibniz formula, applications of integrals, etc.

Example: Determining the area of plane figure.

Task: Determine the area of the figure in the xOy plane bounded by the curves $x – y = 0$ and $x – y^3 = 0$.

Solution: Animation shows the graphs of given curves (Figure 8a) and their intersection points A(1, 1), O(0, 0) and B(-1, 1) (with numerical and graphic presentation). We can see that given figure consists of two identical parts. The next step in the animation is solving the given problem step-by-step. The final result is illustrated on Figure 8b.



$$P = \int_c^d (x_2(y) - x_1(y))dy$$

$$P = 2\int_0^1 (y - y^3)dx = 2\left(\frac{y^2}{2} - \frac{y^4}{4}\right)\Big|_0^1 = 2\left(\frac{1}{2} - \frac{1}{4}\right) = \frac{1}{2}$$

Figure 8. Animation parts which represents the graphs of the given task and solution in determining the area of a plane figure

In a similar way as shown for determining the area of a given figure, we used multimedia animations to explain application of the determined integrals for calculus of volumes of solids, as well as volumes of solids of revolution obtained by revolving a plane figure around Ox or Oy axis.

Example: Determining the volume of body by revolving. Task: Determine the volume of a right circular cone with altitude h and base radius r.

Solution: The cone is generated by revolving the right-angled triangle OAB around the Oxaxis (Figure 9a), which can be clearly shown by using animation (Figures 9b, and 9c). Animation parts which represents the given task and the triangle revolution.

Numerical solution of given problem is also shown step-by-step, by using animation.

Slant height of the cone is defined as line:

$$y = x \cdot tg\alpha = \frac{r}{h} \cdot x$$

Therefore, according to the formula for calculus of volume:

$$V = \pi \int_{0}^{h} (\frac{r}{h} \cdot x)^2 dx = \frac{\pi \cdot r^2}{h^2} \cdot \frac{x^3}{3} \Big|_{0}^{h} = \frac{\pi \cdot h \cdot r^2}{3} .$$

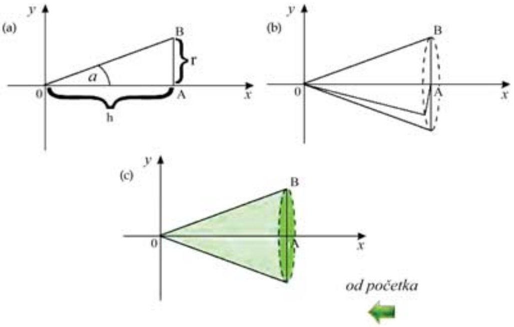


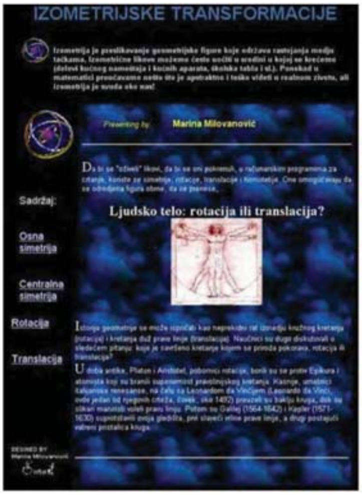
Figure 9. Animation parts which represents the graphs of the given task and solution in determine the volume of a right circular cone with altitude h and base radius r

Assorted Examples and Problems from Multimedia Lectures On Isometric Transformations

Our lectures on isometric transformations (homepage shown on Figure 10), consist of four units: line and point reflection, translation and rotation. Every transformation is presented by the following chapters:

- Basics
- Examples
- Some characteristics
- Exercises
- Problems
- Examples from everyday life

Since the field of isometric transformations is very broad, we have conducted the research in only one area – line reflection. Multimedia lessons in line reflection are presented here by characteristic examples which have enabled us to use different, multimedia approach than in classical lectures. Great advantage of multimedia lessons is particularly evident in chapter Basics, because the definitions of line reflection, axis of symmetry etc. are not only 'given' (written and drawn), but also illustrated by numerous animations which show 'movements', i.e. isometric transformation. We have also paid special attention to enabling students to find out the solutions by themselves.



IZOMETRIJSKE TRANSFORMACIJE

Izometrija je preslikavanje geometrijske figure koje održava rastojanja medju tačkama. Izometrične likove možemo često uočiti u sredini u kojoj se krećemo (delovi kućnog nameštaja i kućnih aparata, školska tabla i sl.). Ponekad u matematici proučavamo nešto što je apstraktno i teško videti u realnom životu, ali izometrija je svuda oko nas!

Presenting by Marina Milovanović

Da bi se "oživeli" likovi, da bi se oni pokrenuli, u računarskim programima za crtanje, koriste se simetrija, rotacije, translacije i homotetije. One omogućavaju da se odredjena figura okrene, da se premesti.

Sadržaj:

Osna simetrija

Centralna simetrija

Rotacija

Translacija

Ljudsko telo: rotacija ili translacija?

Istorija geometrije se može opisati kao neprekidni rat izmedju kružnog kretanja (rotacija) i kretanja duž prave linije (translacija). Naučnici su dugo diskutovali o sledećem pitanju: koje je savršeno kretanje kojem se priroda pokorava, rotacija ili translacija?

U doba antike, Platon i Aristotel, pobornici rotacije, borili su se protiv Epikura i atomista koji su branili superiornost pravolinijskog kretanja. Kasnije, umetnici italijanske renesanse, na čelu sa Leonardom da Vinčijem (Leonardo da Vinci, ovde jedan od njegovih crteža, čovek, oko 1492) preuzeli su baklju kruga, dok su slikari manirizma voleli pravu liniju. Potom su Galilej (1564-1642) i Kepler (1571-1630) suprotstavili svoja gledišta, prvi slaveći vrline prave linije, a drugi postajući vatreni pristalica kruga.

DESINED BY Marina Milovanović

Figure 10. Homepage of animation about isometric transformations

Example: Basic idea of this example is to help students to see, comprehend and implement the line reflection in different cases before giving them the exact definition. In the first task (Figure 11a), students were asked to recognize the common characteristic of given figures and to find which two of them do not belong in the group. After that, the solution was offered (for all the figures except the third and the last one) in which it was shown that there is at least one line along which we can fold the paper and every point from one side would fall on corresponding point on the other side (Figure 11b).

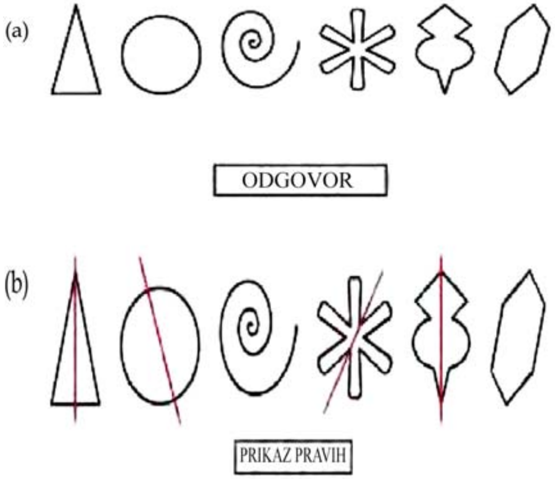


Figure 11. In next step, students were asked to look at the figures shown on Figure 12a and to find out if there is an axis of symmetry for any given pair of figures. Afther that, multimedia animation led them to the correct answer (Figure 12b)

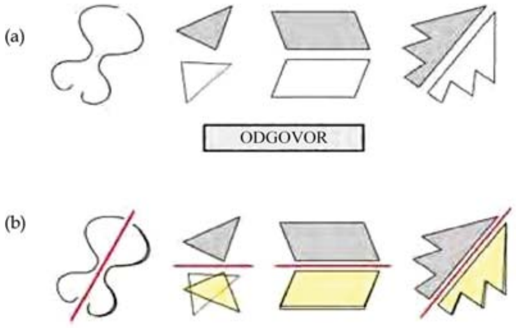


Figure 12.

In chapter ’Examples’, we have used series of animations to ask different questions, such as which of the given figures are symmetrical, how many axis of symmetry they have etc. All the answers were illustrated by complete multimedia presentations of isometric transformations. (Figure 13)

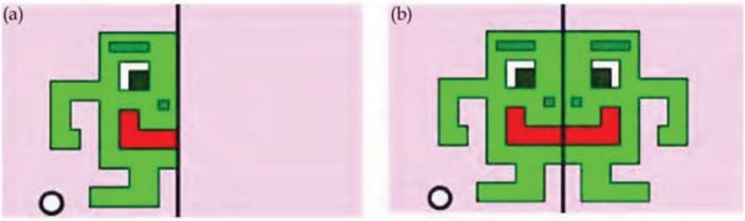


Figure 13. Example of symmetrical figure (b), which is obtained by animated isometric transformation of picture (a), point by point, by pressing the button (left)

Chapter ’Exercises’ offers variety of multimedia Q/A, quizzes and tests with purpose of resuming and exercising adopted knowledge. One of the examples is presented in Figure 14. Its purpose was to use an interesting example from the everyday life in order to enable students to resume their knowledge.

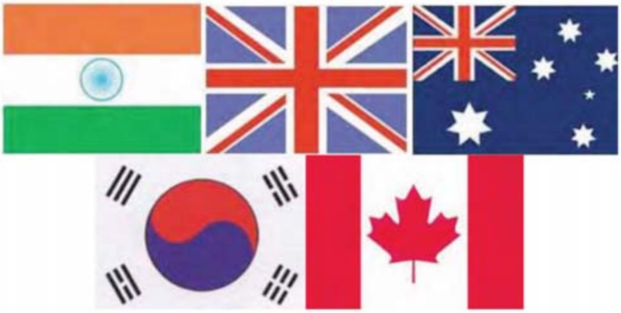


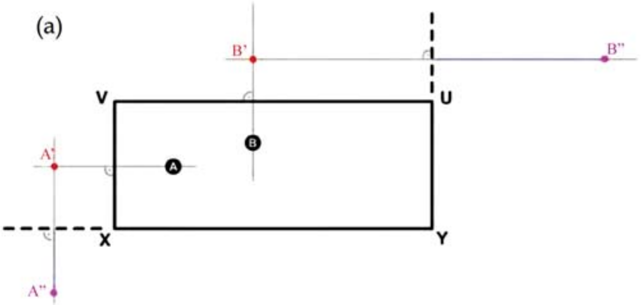
Figure 14. Task: How many axes of symmetry have these flags?

Chapter ’Problems’ include variety of different interesting tasks, ranged from easier to more complex ones. All tasks are solved, majority with complete solutions and some with instruction how to solve them. The main idea was to enable student to get to the right solution individually, before he or she see it on the screen. Animations do not offer complete solution instantly, but gradually, step-by-step. Some of tasks are typical ones, as in traditional classes, but there are also non-standard tasks taken from the mathematics competitions.

Example: Two billiard balls, A and B, are on the rectangular table, as shown on figure. How should we hit the ball A if we want it to strike all four rails before hitting the ball B?

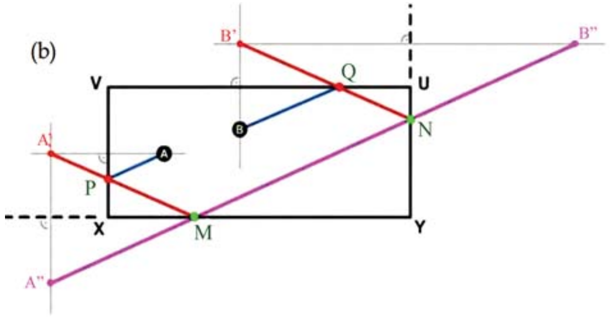
Let us mark the rectangle (billiard table) as XYUV, and $A'=I_{XV}(A)$, $A''=I_{XY}(A')$, $B'=I_{UV}(B)$, $B''=I_{UY}(B')$.

(Multimedia presentation shows transformation step by step.)



If we mark the intersection of lines A''B'' and XY as M, the intersection of lines A''B'' and UY as N, the intersection of lines A'M and XV as P, and the intersection of lines B'N and UV as Q, it can be noticed that the following angles are equal: APV=A'PV=XPM, PMX=XMA''=NMY, MNY=B''NU=UNQ, and NQU=B'QV=VQB.

(Multimedia presentation shows drawing of every line and their intersections, i.e. above mentioned points)



Therefore, ball in point A should be hit in such a way that would send it through points P, M, N and Q, and it will finally hit the ball in point B.

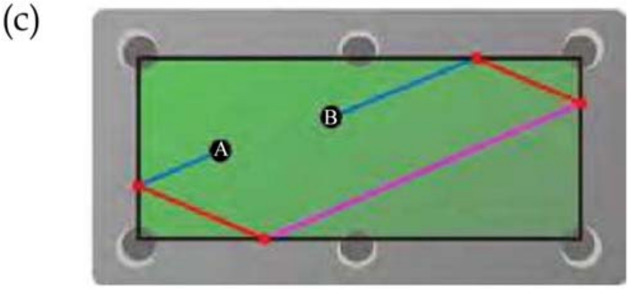


Figure 15. Solution of a given task given by the multimedia animation

In teaching mathematics, we are sometimes supposed to explain abstract terms that rarely or hardly can be seen in reality, but students often ask for proof that theory can be seen and implemented in everyday life. With help of multimedia aids, we can show numerous examples of symmetry in architecture, art, nature, psychology, religion, etc.

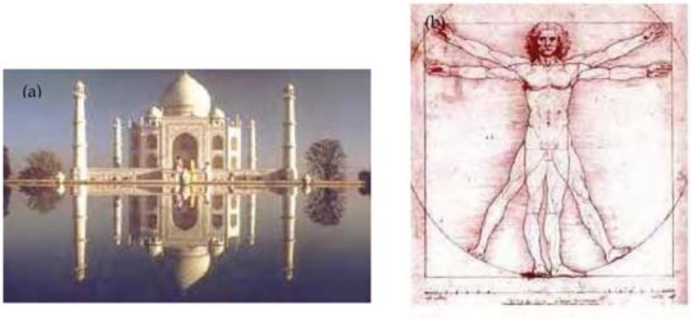


Figure 16. Examples of symmetry in everyday life: (a) This photograph of the Taj Mahal has two axes of symmetry; Beside the vertical one, there is also a horizontal one, along the water line; (b) Famous Leonardo da Vinci's drawing The Vitruvian Man is also called Canon of Proportions or Proportions of Man. It shows the symmetry of the human body

RESEARCH METHODOLOGY

Aim and Questions of the Research

Thanks to the experiences of some previous researches and results, some of the questions during this research were as follows:

- Are there any differences between results of the first group of students, who had traditional lectures (control group – traditional group) and the second group, who had multimedia lectures (experimental group – multimedia group)? Where were these differences the most obvious?
- What do students from the experimental group think about multimedia lectures? Do they prefer this or traditional way and why?
- In students' opinion, where did multimedia learning help them more, in geometry or analysis (based on lessons on the definite integrals and the isometric transformations)?
- Do students think it is easier to understand and learn the matter individually and during the classes by multimedia lectures?

Participants of the Research

The research was conducted on two groups of 50 students, divided on subgroups of 25, at two faculties: the Faculty of Architecture and the Faculty of Civil Construction Management of the UNION University, Belgrade, Serbia. In both cases, one subgroup, consisting of 25 students, had traditional lectures while the other one had multimedia lectures. Groups were formed randomly, so the previous knowledge needed for the lectures about limited integrals and isometric was practically the same, which was confirmed by pre-test. The pre-test included tasks from the area of the continuous functions (analysis and graph-drawing), as well as the tasks about the basic figures in analytic geometry (circle, ellipse, parabola etc.). Average score of this pre-test was practically equal in these groups (I: 72.35, II: 71.25 out of 100).

Methods, Techniques and Apparatus

Lectures in both groups included exactly the same information on given topics, i.e. axioms, theorems, examples and tasks. It is important to emphasize that the lecturer and the number of classes were the same, too. The main information source for the multimedia group was software created in Macromedia Flash 10.0, which is proven to be very successful and illustrative for creating multimedia applications in mathematics lectures (Bakhoum, 2008). Our multimedia lecturing material was created in accordance with methodical approach, i.e. cognitive theory of multimedia learning (Mayer, 2001, 2005), as well as with principles of multimedia teaching and design based on researches in the field of teaching mathematics

(Atkinson, 2005). This material includes a large number of dynamic and graphic presentations of definitions, theorems, characteristics, examples and tests from the area of the definite integrals based on step-by-step method with accent on visualisation. An important quality of making one's own multimedia lectures is the possibility of creating combination of traditional lecture and multimedia support in those areas we have mentioned as the 'weak links' (definite integral definition, area, volume, etc.)

After the lectures were finished, all students had the same tests consisting of tasks on definite integrals and isometric transformations. Besides that, students were interviewed after the classes and transcripts of the most characteristic opinions are also included here. In order to get as objective results as possible, participation in the interviews was voluntary and anonymous, and the interviewer was not a member of the teaching staff at any of the faculties.

Test 1 – Definite integral:

1. Use Archimedes' method to determine the area of plane figure bounded by the Ox-axis, line x a = , and part of the curve $y^3 = x$ for $0 \leq x \leq a$.
2. Write the definite integral definition.
3. Determine the definite integral $\int_0^{\pi/4} \sin x \cdot \cos^2 x dx$.
4. Determine the area of the plane figure F bounded by Oy axis, graph of the function $y = x^2$ and the tangent on this graph in point (1, 1).
5. Determine the volume of the body of revolution obtained by rotating figure F bounded by parabolas $y^2 = 8x$ and $x^2 = y$ in the xOy plane around the Oy axis.

Test 2 – Isometric transformations:

1. Which of the following figures are symmetrical:

- Ray
- Circle
- Line
- Parallelogram

- Isosceles triangle
- Isosceles trapezium
- Kite

2. Line reflection:

a. is a plane isometry

b. is not a plane isometry

How many axes of symmetry are there in the circle?

a. 2

b. 4

c. Infinite number

Immovable lines in the line reflection are:

a. parallel with axis

b. intersect axis

c. rectangular with axis

3. How many axes of symmetry are there in the following alphabet letters?

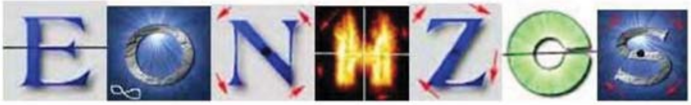


Figure 17.

4. For given sharp angle aOb and point C, find the points A and B, such that A belongs to Oa, B belongs to Ob, and the triangle ABC has the minimal possible circumference.

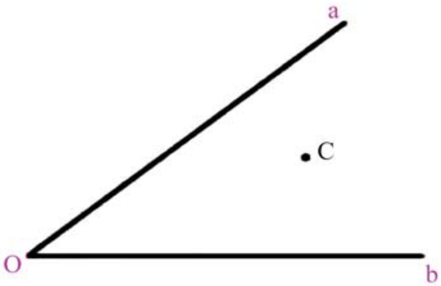


Figure 18.

5. For given line p and points A and B on the same side of p, find the point P on the line p, so the ray of light which starts in point A hits the point P and passes through the point B. (Use the fact that entering ray is symmetrical with reflected ray.)

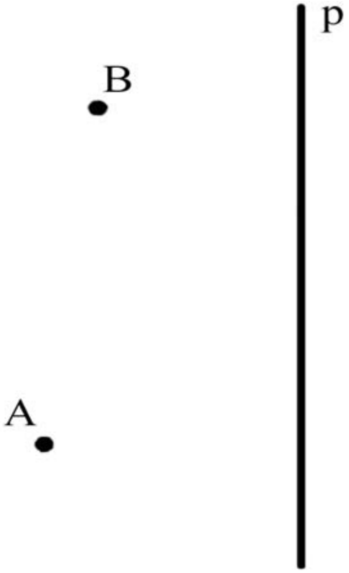


Figure 19.

Both tests were scored within the interval from 0 to 100 (20 points per task) and the average scores in both tests separately were calculated for the traditional and the multimedia group. Results were analysed with Student's t-test for independent samples using SPSS (version 10.0) software. The result was considered significant if the probability p was less than 0.05.

Results

At the Faculty of Architecture, average score on Test 1 (definite integrals) in the traditional group was 67.75 with standard deviation 14.51, and in the multimedia group, average score was 83.21 with standard deviation 15.01. Average score on Test 2 (line reflection) in the traditional group was 76.04 with standard deviation 15.25, and in the multimedia group, average score was 87.92 with standard deviation 12.5. (Figure 20)

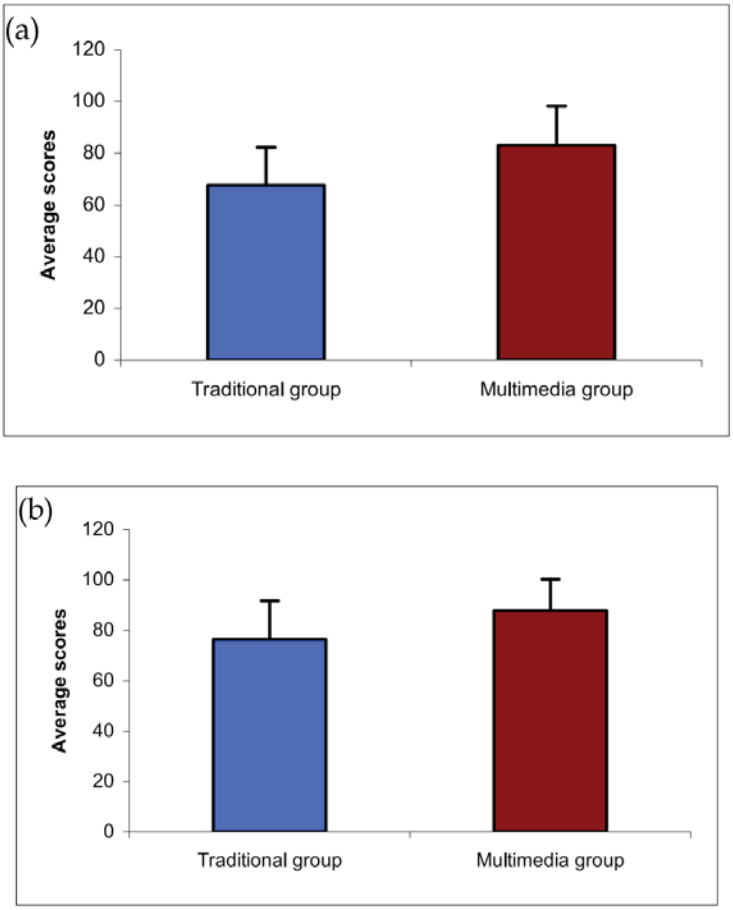


Figure 20. Total average test scores for (a) Test 1 (definite integral) and (b) Test 2 (line reflection) for traditional and multimedia groups at the faculty of Architecture

At the Faculty of Civil Construction Management, average score on Test 1 (definite integrals) in the traditional group was 60.04 with standard deviation 16.2, and in the multimedia group, average score was 76.37 with standard deviation 19.13. Average score on Test 2 (line reflection) in the traditional group was 72.21 with standard deviation 17.32, and in the multimedia group, average score was 84.37 with standard deviation 15.27. (Figure 21) In all groups, statistical comparison with t-test for two independent samples showed that multimedia groups had remarkably higher score in comparison with the traditional groups, with statistical significance $p < 0.05$.

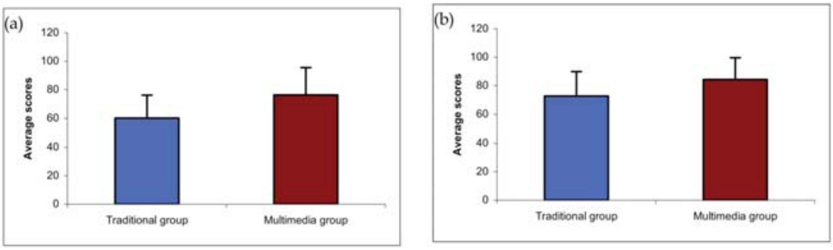


Figure 21. Total average test scores for (a) Test 1 (definite integral) and (b) Test 2 (line reflection) for traditional and multimedia groups at the Faculty of Civil Construction Management

Test scores by tasks for all groups are given in Figures 22 and 23.

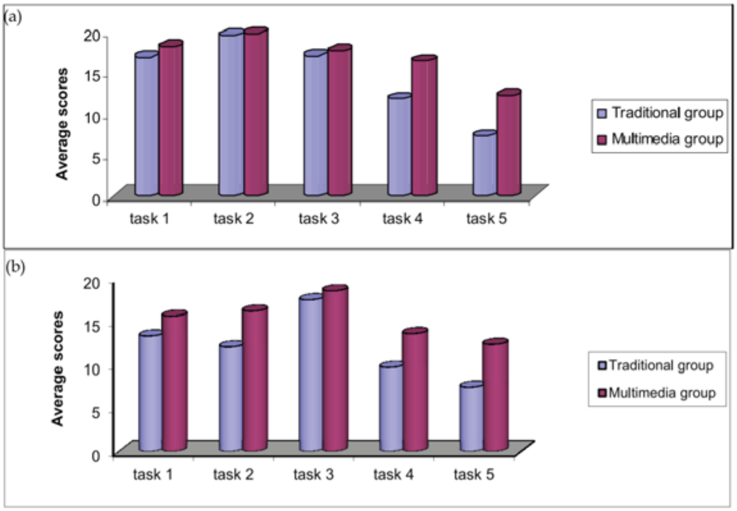


Figure 22. Average test scores by single tasks for (a) Test 1 (definite integral) and (b) Test 2 (line reflection) for traditional and multimedia groups at the faculty of Architecture

When asked whether they prefer classical or multimedia way of learning, 12% (3 students) answered classical and 82% (22 students) answered multimedia at the Faculty of Architecture, while at the Faculty of Civil Construction Management 20% (5 students) answered classical and 80% (20 students) answered multimedia, explaining it with the following reasons:

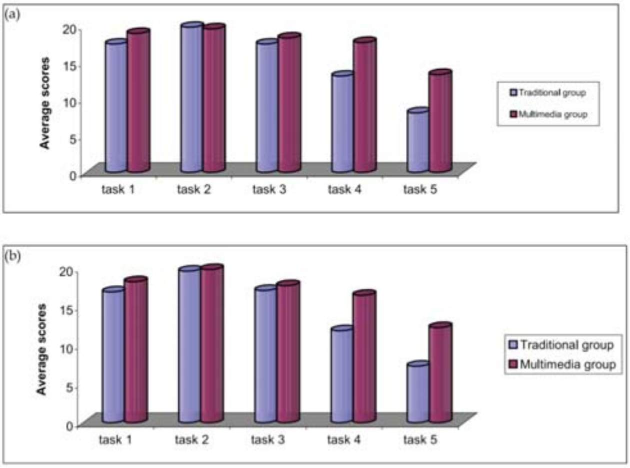


Figure 23. Average test scores by single tasks for (a) Test 1 (definite integral) and (b) Test 2 (line reflection) for traditional and multimedia groups at the Faculty of Civil Construction Management

- ‘It is much easier to see and understand some things, and much easier to comprehend with the help of step-by-step animation.’
- ‘Much more interesting and easier to follow, in opposite to traditional monotonous lectures with formulas and static graphs.’
- ‘More interesting and easier to see, understand and remember.’
- ‘I understand it much better this way and I would like to have similar lectures in other subjects, too.’
- ‘This enables me to learn faster and easier and to understand mathematical problems which demand visualisation.’
- ‘Quite interesting, although classical lectures can be interesting – depending on teacher.’

Students have also commented in which area (analyses or geometry) multimedia lessons were more helpful:

- ’Multimedia lessons certainly make learning easier, especially in the fields which are more abstract and which are better understood with help of pictures and animations, such was the one about the integrals.’
- ’I have comprehended the integrals much better now than in the high school, while in case of symmetry it was much easier to understand and solve more difficult problems with help of multimedia.’

- ’Line reflection is not very difficult to understand and learn, but it was much easier and even funny through the multimedia lessons. I have always thought that integrals are horrible, but now I understand them and know how to calculate area or volume by drawing a figure.’

When asked whether it was easier for them to learn, understand and solve problems after having lectures and individual work with multimedia approach, students answered the question as shown in Figure 24:

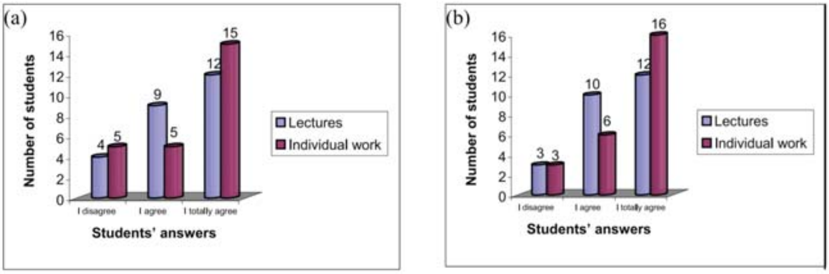


Figure 24. Students’ answers to the question: Should PC be used in lecturing and learning mathematics? (a) Architecture; (b) Civil Construction Management

DISCUSSION AND CONCLUSIONS

During past few years, multimedia learning has become very important and interesting topic in the field of teaching methodology. Researches conducted by Mayer (Mayer, 2001, 2005) and Atkinson (Atkinson, 2005) resulted in establishing the basic principles of multimedia learning and design, which were confirmed in our research, too. Multimedia lessons about the definite integrals and the isometric transformations, created in accordance with these principles, proved to be successful. According to the students’ reactions, highly understandable animations from multimedia lessons are the best proof that a picture is worth a thousand words. Their remark, and consequently one of this research’s conclusions, was that there should be much more of this kind of lessons in education, made – of course – in accordance with certain rules and created in the right way. Many research works in different scientific fields, including mathematics, have proven that multimedia makes learning process much easier.

The tests of adopted knowledge conducted during this research showed that students from multimedia group had much higher average scores in comparison with students from traditional group, which also correspondents with results of some other authors (D. Đ. Herceg, 2009; Wishart, 2000). At the Faculty of Architecture, average score on Test 1 (definite integrals) was 15.49 points greater in multimedia group than in traditional group, while the average score on Test 2 (line reflection) was 11.52 points greater in multimedia group. At the Faculty of Civil Construction Management, average score was 15.97 points greater on Test 1 and 12.6 points greater in multimedia group. Research on learning the definite integrals with software packages Mathematica and GeoGebra (D. Đ. Herceg, 2009) has shown that students who had used PC in learning process had higher scores on tests. Although this research was conducted with different multimedia teaching tools for the same subject – the definite integral as one of the most important areas in mathematical analyses – our results only proved the universality of multimedia in the process of teaching mathematics.

According to Figures 22a and 23a, which show average scores in single tasks from the field of definite integral, we can conclude that students from multimedia group were remarkably more successful in problems which demand visual comprehension (tasks 1, 2, 4 and 5), while the average score in the third task was practically the same on the both groups. Additionally, according to Figures 22b and 23b, which show average scores in single tasks from the line reflection, it can be seen that the students from multimedia groups at both faculties were remarkably more successful in solving the tasks that demanded visual comprehension, as well as in using the adopted knowledge for more complicated problems (tasks 4 and 5), while the average scores for other tasks were not significantly different. Wishart's (Wishart, 2000) research included analyses of comments on how much multimedia approach affects teaching and learning processes. Teachers emphasized that multimedia lectures have made their work easier and have proved to be motivating for students, while students said that multimedia lessons, in comparison with traditional methods, have offered better visual idea about the topic. As shown in Figure 24, a great number of them insisted that multimedia tools enabled easier understanding, learning and implementation of knowledge. Students' remark, and consequently one of this research's conclusions, was that there should be more multimedia lessons, i.e. that multimedia is an important aspect of teaching and learning process.

One of this research's conclusions can be put in the way one student did it during the survey (by answering the question: What is multimedia learning): 'Multimedia learning is use of multimedia as an addition to the traditional way of learning. Multimedia enables us to have better understanding of many mathematical problems and to experiment with them.’ Since the experimental lessons on definite integral and isometric transformations have proven to be very successful, we have decided to continue our work and to develop similar multimedia lessons for other areas of mathematics (where applicable) and to make them available for other researchers and teachers when the whole package is ready for publishing.

GUIDELINES FOR FURTHER RESEARCHES

During our research, several new questions appeared that should be solved in the future: (a) In which scientific fields does the multimedia approach give the best results? (b) How much success of the multimedia approach depends on an individual student’s ability and how much on a teacher’s skills? (c) How can we improve the understanding of lectures by using the multimedia approach, because our aim is learning and understanding, not the multimedia per se.

REFERENCES

1. Atkinson, R. (2005.). Multimedia Learning of Mathematics in The Cambridge handbook of Multimedia Learning, Mayer, R. pp. 393-408., Cambridge University Press, ISBN 0-521-54751-2, United States of America
2. Bishop, (1989.). Review of research on visualization in mathematics education, Focus on Learning Problems in Mathematics, 11 (1), pp. 7-16. ISSN 0272-8893
3. Damjanović, B. (2005). Matematička analiza, GND Produkt, ISBN 86-904989-1-5, Belgrade, Serbia.
4. Deliyannis, I., Vlamos, P., Floros, A. & Simpsiri, C. (2008). Teaching Basic Number Theory to Students with Speech and Communication Disabilities using Multimedia,
5. International Conference on Information Communication Technologies in Education (ICICTE 2008), 10-12 July, Corfu, Greece. ACON-05 I.
6. Deliyiannis, I., Floros, A., Vlamos, P., Arvanitis, M. & Tania, T. (2008). Bringing Digital Multimedia in Mathematics Education, The 7th European Conference on e-Learning, Agia Napa, Cyprus, on 6-7 November 2008
7. Herceg, D. & Herceg, Đ. (2009.). The definite integral and computer, The teaching of mathematics, Vol. 12, No.1, pp. .33-44. ISSN 0351-4463
8. Mayer, R. (2005.). The Cambridge handbook of Multimedia Learning, Cambridge University Press, ISBN 0-521-54751-2, , New York, United States of America
9. Mayer, R. (2001.). Multimedia Learning, Cambridge University Press, ISBN 0-52178-749-1, New York, United States of America
10. Miličić P.M. & Uščumlić, M.P. (2005.). Zbirka zadataka iz više matematike I, GK, ISBN 86-395- 0450-4, Belgrade, Serbia
11. Milovanovic, M. (2005.). Кそちけのろかれか すとしてけすかおけja くa とねかれか けくけすかてちけjつさけに てちaせつなそちすaぬけja,
12. Maてかすaてけчさけ なaさとしてかて Уせけうかpくけてかてa と Бかそえpaおと, master thesis Milovanovic, M. (2009.). Multimedijalni pristup nastavi matematike na primeru lekcije o osnoj simetriji, Inovacije u osnovnoškolskom obrazovanju-vrednovanje, pp. 580-588, ISBN 978-86-7849-136-8, Belgrade, Serbia, october, 2009.

13. Milovanović, M., Takaci, Đ. & Milajic, A. (2011.). Multimedia Approach in Teaching Mathemathics – Example of Lesson about the Definite Integral Application for Determining an Area, International Journal of Mathematical Education in Science and Technology, Vol. 42, No. 2, pp. 175-187, ISSN 0020-739X
14. Takači, Dj. & Pešić, D. (2004.). The Continuity of Functions in Mathematical EducationVisualization method, The Teaching of Mathematics, Belgrade, Vol. 49, No.3-4, pp. 31- 42, ISSN 0351-4463
15. Takači, Dj. , Pešić, D. & Tatar, J. (2006.). On the continuity of functions, International Journal of Mathematical Education in Science & Technology, Vol. 37, No.7, pp. 783-791, ISSN 0020-739X
16. Takači, Dj., Pešić, D. & Tatar, J. (2003.) An introduction to the Continuity of functions using Scientific Workplace, The Teaching of Mathematics, Belgrade, Vol. 6, No.2, pp. 105- 112. ISSN 0351-4463
17. Takači, Dj.. Stojković R. & Radovanovic, J. (2008.). The influence of computer on examining trigonometric functions, Teaching Mathematics and Computer Science, Debrecen, Hungary, Vol. 6, No.1, pp. 111-123. ISSN 1589-7389
18. Takači., Dj., Herceg D. & Stojković, R. (2006.). Possibilities and limitati-ons of ScientificWorkplace in studying trigonometric functions, The Teaching of Mathematics, Belgrade, Vol. 8, No.2, pp. 61-72. ISSN 0351-4463
19. Tall, D. (1991.). Advanced mathematical thinking, Springer, ISBN 978-0-7923-2812-4, New York, United States of America
20. Tall, D. (1986.). A graphical to integration and fundamental theorem, Mathematics teaching, 113, pp. 48-51, ISSN 0025-5785
21. Wishart, J. (2000.). Students' and Teachers' Perceptions of Motivation and Learning Through the Use in Schools of Multimedia Encyclopaedias on CD-ROM, Journal of Educational Multimedia and Hypermedia Vol. 9, No.4, pp. 331-345, ISSN 1055-8896
22. E. Bakhoum, (2008.). Animating an equation: a guide to using FLASH in mathematics education, International Journal of Mathematical Education in Science and Technology, Taylor & Francis, Vol. 39, No.5, pp. 637–655. ISSN 0020-739X

Index

Advances in Multimedia

Multimedia uses multiple forms of information content and processing to inform or entertain the user, or the audience. Generally speaking, we can say that multimedia refers to the use of electronic media for storage and processing of various data types. We experience the multimedia content presented by text, images, videos, animation, audio, and other media types, which stimulate our senses. There are different definitions of the term multimedia, depending on the context of use. One of the accepted definitions of multimedia is as follows: Multimedia is an area that integrates text, graphics, drawings, images and videos, animation, audio and other media, where any type of information can be displayed, stored, transmitted and processed in digital form. Despite this definition of multimedia as a scientific field, the word multimedia is also used to denote plain multimedia content, which represents the convergence of text, images, video, animation and sound (elements of multimedia) into a single form. The power of multimedia lies in the way this information is interconnected. Multimedia applications in addition to multimedia content also include the processes that execute inside them (during playback). The classification of multimedia content can be done in different ways. One of them classifies multimedia content into static and time-dependent. Text and images form static multimedia content, while audio, video and animations change over time, and the multimedia content in which they are included is called time-dependent. By multimedia structure we mean the form in which multimedia content appears. Multimedia structures are divided in two basic categories: linear and nonlinear. An example of a linear multimedia structure is a film without navigation controls. Nonlinear structures often use interactivity to control the flow, e.g. in video games or computer-assisted learning. Interactivity in multimedia means the ability to display or broadcast multimedia content according to user instructions. This edition covers several recent advances in multimedia, including: machine learning and AI methods in multimedia research, multimedia applications in health and medicine, multimedia transmission in wireless networks and multimedia applications in education.

Section 1 focuses on machine learning and AI methods in multimedia research, describing context-aware attention network for human emotion recognition in video, large-scale video retrieval via deep local convolutional features, region space guided transfer function design for nonlinear neural network augmented image visualization, data-driven methods for image and video understanding, and pretraining convolutional neural networks for image-based vehicle classification.

Section 2 focuses on multimedia applications in health and medicine, describing study of multimedia technology in posture training for the elderly, rapid extraction of target information in messy multimedia medical data, effectiveness of a multimedia messaging service reminder system in the management of knee osteoarthritis, and virtual realities in the treatment of mental disorders: a review of the current state of research.

Section 3 focuses on multimedia transmission in wireless networks, describing clustering in wireless multimedia sensor networks, a dynamic link adaptation for multimedia quality-based communications in IEEE_802.11 wireless networks, multimedia and VoIP-oriented cell search technique for the IEEE 802.11 WLANS, and ubiquitous control framework for delivering perceptual satisfaction of multimedia traffic.

Section 4 focuses on multimedia applications in education, describing online individualized multimedia instruction instrument for engineering communication skills, design model for educational multimedia software, cognitive constructivist theory of multimedia, design of videogame based on inca abacus, and multimedia approach in teaching mathematics through examples of interactive lessons from mathematical analysis and geometry.

Jovan obtained his PhD in Computer Science from RMIT University in Melbourne, Australia in 2007. His research interests include big data, business intelligence and predictive analytics, data and information science, information retrieval, XML, web services and service-oriented architectures, and relational and NoSQL database systems. He has published over 30 journal and conference papers and he also serves as a journal and conference reviewer. He is currently working as a Dean and Associate Professor at European University in Skopje, Macedonia.

AP | ARCLER
PRESS

ISBN 978-1-77469-353-7
00000
9 781774 693537