Computer Scanning and Scanners

Sean White

COMPUTER SCANNING AND SCANNERS

COMPUTER SCANNING
AND
SCANNERS

Sean White

BIBLIOTEX
Digital Library

Computer Scanning and Scanners
by Sean White

www.bibliotex.com

Ebook ISBN: 9781984664273

Published by:
Bibliotex
Canada
Website: www.bibliotex.com

Contents

1

Introduction

The process of creating art with a scanner can be as simple as arranging objects on the scanner and capturing the resulting image; in fact, some early artists in the field worked with photocopiers to capture and print in a single step, resulting in the field of Xerox art. Artist Sonia Landy Smith, artist in residence at 3M and founder of the *Generative Systems* programme at the Art Institute of Chicago was one of the first to exploit this ability in 1968, altering the variables of the photocopying process to produce artwork rather than mere copies.

Using a computer and a photo editor between the scanning and the printing process provides the artist with a greater level of control, allowing, at a minimum, the ability to "clean" the image by removing specks and other imperfections in the capture. With the increased availability of photocopiers in the 1980s and 1990s, a vernacular form of scanography developed from people making photocopies of body parts, notably their hands, faces, or their buttocks, generally humorously.

Harold Feinstein's *One Hundred Shell* and *One Hundred Flower* series contained scanned images side by side with traditional large format photography. Joseph Scheer scanned moths in *Night Visions: The Secret Designs of Moths*. A 2008 exhibition titled "Scanner as Camera" at Washington and Lee University in Lexington, Virginia drew eight artists from across the United States whose subjects ranged from scanned and digitally manipulated historic ambrotype and tintype photographs and drawings to dead birds and insects found by the artist.

Compact Digital Cameras

Compact cameras are designed to be tiny and portable and are particularly suitable for casual and "snapshot" use, thus are also called point-and-shoot cameras. The smallest, generally less than 20 mm thick, are described as *subcompacts* or "ultra-compacts" and some are nearly credit card size.

Most, apart from ruggedized or water-resistant models, incorporate a retractable lens assembly allowing a thin camera to have a moderately long focal length and thus fully exploit an image sensor larger than that on a camera phone, and a mechanized lens cap to cover the lens when retracted. The retracted and capped lens is protected from keys, coins and other hard objects, thus making a thin, pocketable package. Subcompacts commonly have one lug and a short wrist strap which aids extraction from a pocket, while thicker compacts may have two lugs for attaching a neck strap.

Compact cameras are usually designed to be easy to use, sacrificing advanced features and picture quality for compactness and simplicity; images can usually only be stored using lossy compression (JPEG). Most have a built-in flash usually of low power, sufficient for nearby subjects. Live preview is almost always used to frame the photo.

Most have limited motion picture capability. Compacts often have macro capability and zoom lenses but the zoom range is usually less than for bridge and DSLR cameras. Generally a contrast-detect autofocus system, using the image data from the live preview feed of the main imager, focuses the lens.

Typically, these cameras incorporate a nearly-silent leaf shutter into their lenses. For lower cost and smaller size, these cameras typically use image sensors with a diagonal of approximately 6 mm, corresponding to a crop factor around 6.

This gives them weaker low-light performance, greater depth of field, generally closer focusing ability, and smaller components than cameras using larger sensors. Some recent compact digital cameras can now able to take 3D still photos. These new 3D compact digital cameras can capture 3D panoramic photos for play back on a 3D TV.

Bridge Cameras

Bridge are higher-end digital cameras that physically and ergonomically resemble DSLRs and share with them some advanced features, but share with compacts the use of a fixed lens and a small sensor. Like compacts, most use live preview to frame the image. Their autofocus uses the same contrast-detect mechanism, but many bridge cameras have a manual focus mode, in some cases using a separate focus ring, for greater control.

Due to the combination of big physical size but a small sensor, many of these cameras have very highly specified lenses with large zoom range and fast aperture, partially compensating for the inability to change lenses. To compensate for the lesser sensitivity of their small sensors, these cameras almost always include an image stabilization system to enable longer handheld exposures. The highest zoom lens so far on a bridge camera is on the Nikon Coolpix P500 digital camera, which encompasses an equivalent of a super wide to ultra-telephoto 22.5-810 mm (36x).

These cameras are sometimes marketed as and confused with digital SLR cameras since the appearance is similar. Bridge cameras lack the reflex viewing system of DSLRs, are usually fitted with fixed (non-interchangeable) lenses (although some have a lens thread to attach accessory wide-angle or telephoto converters), and can usually take movies with sound. The scene is composed by viewing either the liquid crystal display or the electronic viewfinder (EVF). Most have a longer shutter lag than a true dSLR, but they are capable of good image quality (with sufficient light) while being more compact and lighter than DSLRs. High-end models of this type have comparable resolutions to low and mid-range DSLRs. Many of these cameras can store images in a Raw image format, or processed and JPEG compressed, or both. The majority have a built-in flash similar to those found in DSLRs.

In bright sun, the quality difference between a good compact camera and a digital SLR is minimal but bridgecams are more portable, cost less and have a similar zoom ability to dSLR. Thus a Bridge camera may better suit outdoor daytime activities, except when seeking professional-quality photos. In low light conditions and/or at ISO equivalents above 800, most bridge cameras (or megazooms) lack in image quality when compared to even entry level DSLRs.

Mirrorless Interchangeable Lens Camera

In late 2008 a new type of camera emerged, combining the larger sensors and interchangeable lenses of DSLRs with the live preview viewing system of compact cameras, either through an electronic viewfinder or on the rear LCD. These are simpler and more compact than DSLRs due to the removal of the mirror box, and typically emulate the handling and ergonomics of either DSLRs or compacts. The system is use by Micro Four Thirds, borrowing components from the Four Thirds DSLR system. The Ricoh GXR of 2009 puts the sensor and other electronic components in the interchangeable sensor lens unit rather than in the camera body. The first interchangeable 3D lens Lumix G 12.5mm/

F12 (H-FT012) has been announced by Panasonic. It use two lenses quite close together in one lens module adaptor and record both 3D and 2D pictures altogether. The lens module is compatible with Panasonic Lumix DMC-GH2.

Digital Single Lens Reflex Cameras

Digital single-lens reflex cameras (DSLRs) are digital cameras based on film single-lens reflex cameras (SLRs). They take their name from their unique viewing system, in which a mirror reflects light from the lens through a separate optical viewfinder.

In order to capture an image the mirror is flipped out of the way, allowing light to fall on the imager. Since no light reaches the imager during framing, autofocus is accomplished using specialized sensors in the mirror box itself. Most 21st century DSLRs also have a "live view" mode that emulates the live preview system of compact cameras, when selected.

These cameras have much larger sensors than the other types, typically 18 mm to 36 mm on the diagonal (crop factor 2, 1.6, or 1). This gives them superior low-light performance, less depth of field at a given aperture, and a larger size.

They make use of interchangeable lenses; each major DSLR manufacturer also sells a line of lenses specifically intended to be used on their cameras. This allows the user to select a lens designed for the application at hand: wide-angle, telephoto, low-light, etc. So each lens does not require its own shutter, DSLRs use a focal-plane shutter in front of the imager, behind the mirror. The mirror flipping out of the way at the moment of exposure makes a distinctive "clack" sound.

Image Sensors

Image Resolution

The resolution of a digital camera is often limited by the image sensor (typically a CCD or CMOS sensor chip) that turns light into discrete signals, replacing the job of film in traditional photography. The sensor is made up of millions of "buckets" that essentially count the number of photons that strike the sensor. This means that the brighter the image at a given point on the sensor, the larger the value that is read for that pixel.

Depending on the physical structure of the sensor, a color filter array may be used which requires a demosaicing/ interpolation algorithm. The number of resulting pixels in the image determines its "pixel count". For example, a 640x480 image would have 307,200 pixels, or approximately 307 kilopixels; a 3872x2592 image would have 10,036,224 pixels, or approximately 10 megapixels.

The pixel count alone is commonly presumed to indicate the resolution of a camera, but this simple figure of merit is a misconception. Other factors impact a sensor's resolution, including sensor size, lens quality, and the organization of the pixels (for example, a monochrome camera without a Bayer filter mosaic has a higher resolution than a typical color camera). Many digital compact cameras are criticized for having excessive pixels. Sensors can be so small that their 'buckets' can easily overfill; again, resolution of a sensor can become greater than the camera lens could possibly deliver.

As the technology has improved, costs have decreased dramatically. Counting the "pixels per dollar" as a basic measure of value for a digital camera, there has been a continuous and steady increase in the number of pixels each dollar buys in a new camera, in accord with the principles of Moore's Law. This predictability of camera prices was first presented in 1998 at the Australian PMA DIMA conference by Barry Hendy and since referred to as "Hendy's Law".

Since only a few aspect ratios are commonly used (mainly 4:3 and 3:2), the number of sensor sizes that are useful is limited. Furthermore, sensor manufacturers do not produce every possible sensor size, but take incremental steps in sizes. For example, in 2007 the three largest sensors (in terms of pixel count) used by Canon were the 21.1, 17.9, and 16.6 megapixel CMOS sensors.

Methods of Image Capture

Since the first digital backs were introduced, there have been three main methods of capturing the image, each based on the hardware configuration of the sensor and color filters. The first method is often called *single-shot*, in reference to the number of times the camera's sensor is exposed to the light passing through the camera lens. Single-shot capture systems use either one CCD with a Bayer filter mosaic, or three separate image sensors (one each for the primary additive colors red, green, and blue) which are exposed to the same image via a beam splitter.

The second method is referred to as *multi-shot* because the sensor is exposed to the image in a sequence of three or more openings of the lens aperture. There are several methods of application of the multi-shot technique. The most common originally was to use a single image sensor with three filters (once again red, green and blue) passed in front of the sensor in sequence to obtain the additive color information.

Another multiple shot method is called Microscanning. This technique utilizes a single CCD with a Bayer filter but actually moved the physical location of the sensor chip on the focus plane of the lens to “stitch” together a higher resolution image than the CCD would allow otherwise. A third version combined the two methods without a Bayer filter on the chip.

The third method is called *scanning* because the sensor moves across the focal plane much like the sensor of a desktop scanner. Their *linear* or *tri-linear* sensors utilize only a single line of photosensors, or three lines for the three colors. In some cases, scanning is accomplished by moving the sensor e.g. when using Color co-site sampling or rotate the whole camera; a digital rotating line camera offers images of very high total resolution.

The choice of method for a given capture is determined largely by the subject matter. It is usually inappropriate to attempt to capture a subject that moves with anything but a single-shot system.

However, the higher color fidelity and larger file sizes and resolutions available with multi-shot and scanning backs make them attractive for commercial photographers working with stationary subjects and large-format photographs. Dramatic improvements in single-shot cameras and raw image file processing at the beginning of the 21st century made single shot, CCD-based cameras almost completely dominant, even in high-end commercial photography. CMOS-based single shot cameras remained somewhat common.

Filter Mosaics, Interpolation, and Aliasing

Most current consumer digital cameras use a Bayer filter mosaic in combination with an optical anti-aliasing filter to reduce the aliasing due to the reduced sampling of the different primary-color images. A demosaicing algorithm is used to interpolate color information to create a full array of RGB image data. Cameras that use a beam-splitter single-shot 3CCD approach, three-filter multi-shot approach, Color co-site sampling or Foveon X3 sensor do not use anti-aliasing filters, nor demosaicing.

Firmware in the camera, or a software in a raw converter programme such as Adobe Camera Raw, interprets the raw data from the sensor to obtain a full color image, because the RGB color model requires three intensity values for each pixel: one each for the red, green, and blue (other color models, when used, also require three or more values per pixel). A single sensor element cannot simultaneously record these three intensities, and so a color filter array (CFA) must be used to selectively filter a particular color for each pixel. The Bayer filter pattern is a repeating 2×2 mosaic pattern of light filters, with green ones at opposite corners and red and blue in the other two positions. The high proportion of green takes advantage of properties of the human visual system, which determines brightness mostly from green and is far more sensitive to brightness than to hue or saturation. Sometimes a 4-color filter pattern is used, often involving two different hues of green. This provides potentially more accurate color, but requires a slightly more complicated interpolation process.

The color intensity values not captured for each pixel can be interpolated (or guessed) from the values of adjacent pixels which represent the color being calculated.

2

Computer Machine

A computer is a machine that manipulates data according to a list of instructions. Although mechanical examples of computers have existed throughout history, the first resembling a modern computer were developed in the mid-20th century (1940–1945).

The first electronic computers were the size of a large room, consuming as much power as several hundred modern personal computers (PC).

Modern computers based on tiny integrated circuits are millions to billions of times more capable than the early machines, and occupy a fraction of the space. Simple computers are small enough to fit into a wristwatch, and can be powered by a watch battery.

Personal computers in their various forms are icons of the Information Age, what most people think of as a "computer", but the embedded computers found in devices ranging from fighter aircraft to industrial robots, digital cameras, and toys are the most numerous.

The ability to store and execute lists of instructions called programmes makes computers extremely versatile, distinguishing them from calculators. The Church–Turing thesis is a mathematical statement of this versatility: any computer with a certain minimum capability is, in principle, capable of performing the same tasks that any other computer can perform.

Therefore computers ranging from a personal digital assistant to a supercomputer are all able to perform the same computational tasks, given enough time and storage capacity.

A computer is a device that accepts information (in the form of digitalized data) and manipulates it for some result based on a programme or sequence of instructions on how the data is to be processed. Complex computers also include the means for storing data (including the programme, which is also a form of data) for some necessary duration.

A programme may be invariable and built into the computer (and called *logic circuitry* as it is on microprocessors) or different programmes may be provided to the computer (loaded into its storage and then started by an administrator or user). Today's computers have both kinds of programming.

Most histories of the modern computer begin with the Analytical Engine envisioned by Charles Babbage following the mathematical ideas of George Boole, the mathematician who first stated the principles of logic inherent in today's digital computer. Babbage's assistant and collaborator, Ada Lovelace, is said to have introduced the ideas of programme loops and subroutines and is sometimes considered the first programmer.

Apart from mechanical calculators, the first really useable computers began with the vacuum tube, accelerated with the invention of the transistor, which then became embedded in large numbers in integrated circuits, ultimately making possible the relatively low-cost personal computer. Modern computers inherently follow the ideas of the stored programme laid out by John von Neumann in 1945.

Essentially, the programme is read by the computer one instruction at a time, an operation is performed, and the computer then reads in the next instruction, and so on. Recently, computers and programmes have been devised that allow multiple programmes (and computers) to work on the same problem at the same time in parallel. With the advent of the Internet and higher bandwidth data transmission, programmes and data that are part of the same overall project can be distributed over a network and embody the Sun Microsystems slogan: "The network is the computer."

Main Units

Central Processing Unit

A central processing unit (CPU) or processor is an electronic circuit that can execute computer programmes. This broad definition can easily be applied to many early computers that existed long before the term "CPU" ever came into widespread usage. The term itself and its initialism have been in use in the computer industry at least since the early 1960s (Weik 1961). The form, design and implementation of CPUs have changed dramatically since the earliest examples, but their fundamental operation has remained much the same. Early CPUs were custom-designed as a part of a larger, sometimes one-of-a-kind, computer.

However, this costly method of designing custom CPUs for a particular application has largely given way to the development of mass-produced processors that are made for one or many purposes. This standardization trend generally began in the era of discrete transistor mainframes and minicomputers and has rapidly accelerated with the popularization of the integrated circuit (IC).

The IC has allowed increasingly complex CPUs to be designed and manufactured to tolerances on the order of nanometers. Both the miniaturization and standardization of CPUs have increased the presence of these digital devices in modern life far beyond the limited application of dedicated computing machines. Modern microprocessors appear in everything from automobiles to cell phones to children's toys.

History of CPU

Prior to the advent of machines that resemble today's CPUs, computers such as the ENIAC had to be physically rewired in order to perform different tasks. These machines are often referred to as "fixed-programme computers," since they had to be physically reconfigured in order to run a different programme. Since the term "CPU" is generally defined as a software (computer programme) execution device, the earliest devices that could rightly be called CPUs came with the advent of the stored-programme computer.

The idea of a stored-programme computer was already present during ENIAC's design, but was initially omitted so the machine could be finished sooner. On June 30, 1945, before ENIAC was even completed, mathematician John von Neumann distributed the paper entitled "First Draft of a Report on the EDVAC."

It outlined the design of a stored-programme computer that would eventually be completed in August 1949 (von Neumann 1945). EDVAC was designed to perform a certain number of instructions (or operations) of various types. These instructions could be combined to create useful programmes for the EDVAC to run. Significantly, the programmes written for EDVAC were stored in high-speed computer memory rather than specified by the physical wiring of the computer.

This overcame a severe limitation of ENIAC, which was the large amount of time and effort it took to reconfigure the computer to perform a new task. With von Neumann's design, the programme, or software, that EDVAC ran could be changed simply by changing the contents of the computer's memory.

While von Neumann is most often credited with the design of the stored-programme computer because of his design of EDVAC, others before him such as Konrad Zuse had suggested similar ideas. Additionally, the so-called Harvard architecture of the Harvard Mark I, which was completed before EDVAC, also utilized a stored-programme design using punched paper tape rather than electronic memory.

The key difference between the von Neumann and Harvard architectures is that the latter separates the storage and treatment of CPU instructions and data, while the former uses the same memory space for both. Most modern CPUs are primarily von Neumann in design, but elements of the Harvard architecture are commonly seen as well.

Being digital devices, all CPUs deal with discrete states and therefore require some kind of switching elements to differentiate between and change these states. Prior to commercial acceptance of the transistor, electrical relays and vacuum tubes (thermionic valves) were commonly used as switching elements. Although these had distinct speed advantages over earlier, purely mechanical designs, they were unreliable for various reasons.

For example, building direct current sequential logic circuits out of relays requires additional hardware to cope with the problem of contact bounce. While vacuum tubes do not suffer from contact bounce, they must heat up before becoming fully operational and eventually stop functioning altogether.

Usually, when a tube failed, the CPU would have to be diagnosed to locate the failing component so it could be replaced.

Therefore, early electronic (vacuum tube based) computers were generally faster but less reliable than electromechanical (relay based) computers.

Tube computers like EDVAC tended to average eight hours between failures, whereas relay computers like the (slower, but earlier) Harvard Mark I failed very rarely (Weik 1961:238). In the end, tube based CPUs became dominant because the significant speed advantages afforded generally outweighed the reliability problems.

Most of these early synchronous CPUs ran at low clock rates compared to modern microelectronic designs . Clock signal frequencies ranging from 100 kHz to 4 MHz were very common at this time, limited largely by the speed of the switching devices they were built with.

Discrete Transistor and IC CPUs

The design complexity of CPUs increased as various technologies facilitated building smaller and more reliable electronic devices. The first such improvement came with the advent of the transistor.

Transistorized CPUs during the 1950s and 1960s no longer had to be built out of bulky, unreliable, and fragile switching elements like vacuum tubes and electrical relays. With this improvement more complex and reliable CPUs were built onto one or several printed circuit boards containing discrete (individual) components.

During this period, a method of manufacturing many transistors in a compact space gained popularity. The integrated circuit (IC) allowed a large number of transistors to be manufactured on a single semiconductor-based die, or "chip." At first only very basic non-specialized digital circuits such as NOR gates were miniaturized into ICs. CPUs based upon these "building block" ICs are generally referred to as "small-scale integration" (SSI) devices. SSI ICs, such as the ones used in the Apollo guidance computer, usually contained transistor counts numbering in multiples of ten.

To build an entire CPU out of SSI ICs required thousands of individual chips, but still consumed much less space and power than earlier discrete transistor designs. As microelectronic technology advanced, an increasing number of transistors were placed on ICs, thus decreasing the quantity of individual ICs needed for a complete CPU. MSI and LSI (medium- and large-scale integration) ICs increased transistor counts to hundreds, and then thousands.

In 1964 IBM introduced its System/360 computer architecture which was used in a series of computers that could run the same programmes with different speed and performance. This was significant at a time when most electronic computers were incompatible with one another, even those made by the same manufacturer.

To facilitate this improvement, IBM utilized the concept of a microprogram (often called "microcode"), which still sees widespread usage in modern CPUs . The System/360 architecture was so popular that it dominated the mainframe computer market for the decades and left a legacy that is still continued by similar modern computers like the IBM zSeries.

In the same year (1964), Digital Equipment Corporation (DEC) introduced another influential computer aimed at the scientific and research markets, the PDP-8. DEC would later introduce the extremely popular PDP-11 line that originally was built with SSI ICs but was eventually implemented with LSI components once these became practical. In stark contrast with its SSI and MSI predecessors, the first LSI implementation of the PDP-11 contained a CPU composed of only four LSI integrated circuits (Digital Equipment Corporation 1975).

Transistor-based computers had several distinct advantages over their predecessors. Aside from facilitating increased reliability and lower power consumption, transistors also allowed CPUs to operate at much higher speeds because of the short switching time of a transistor in comparison to a tube or relay.

Thanks to both the increased reliability as well as the dramatically increased speed of the switching elements (which were almost exclusively transistors by this time), CPU clock rates in the tens of megahertz were obtained during this period. Additionally while discrete transistor and IC CPUs were in heavy usage, new high-performance designs like SIMD (Single Instruction Multiple Data) vector processors began to appear. These early experimental designs later gave rise to the era of specialized supercomputers like those made by Cray Inc.

Microprocessors

The introduction of the microprocessor in the 1970s significantly affected the design and implementation of CPUs. Since the introduction of the first microprocessor (the Intel 4004) in 1970 and the first widely used microprocessor (the Intel 8080) in 1974, this class of CPUs has almost completely overtaken all other central processing unit implementation methods. Mainframe and minicomputer manufacturers of the time launched proprietary IC development programmes to upgrade their older computer architectures, and eventually produced instruction set compatible microprocessors that were backward-compatible with their older hardware and software.

Combined with the advent and eventual vast success of the now ubiquitous personal computer, the term "CPU" is now applied almost exclusively to microprocessors.

Previous generations of CPUs were implemented as discrete components and numerous small integrated circuits (ICs) on one or more circuit boards. Microprocessors, on the other hand, are CPUs manufactured on a very small number of ICs; usually just one. The overall smaller CPU size as a result of being implemented on a single die means faster switching time because of physical factors like decreased gate parasitic capacitance.

This has allowed synchronous microprocessors to have clock rates ranging from tens of megahertz to several gigahertz. Additionally, as the ability to construct exceedingly small transistors on an IC has increased, the complexity and number of transistors in a single CPU has increased dramatically. This widely observed trend is described by Moore's law, which has proven to be a fairly accurate predictor of the growth of CPU (and other IC) complexity to date.

While the complexity, size, construction, and general form of CPUs have changed drastically over the past sixty years, it is notable that the basic design and function has not changed much at all. Almost all common CPUs today can be very accurately described as von Neumann stored-programme machines.

As the aforementioned Moore's law continues to hold true, concerns have arisen about the limits of integrated circuit transistor technology. Extreme miniaturization of electronic gates is causing the effects of phenomena like electromigration and subthreshold leakage to become much more significant.

These newer concerns are among the many factors causing researchers to investigate new methods of computing such as the quantum computer, as well as to expand the usage of parallelism and other methods that extend the usefulness of the classical von Neumann model.

CPU Operation

The fundamental operation of most CPUs, regardless of the physical form they take, is to execute a sequence of stored instructions called a programme. The programme is represented by a series of numbers that are kept in some kind of computer memory. There are four steps that nearly all CPUs use in their operation: fetch, decode, execute, and writeback.

The first step, fetch, involves retrieving an instruction (which is represented by a number or sequence of numbers) from programme memory. The location in programme memory is determined by a programme counter (PC), which stores a number that identifies the current position in the programme. In other words, the programme counter keeps track of the CPU's place in the current programme. After an instruction is fetched, the PC is incremented by the length of the instruction word in terms of memory units. Often the instruction to be fetched must be retrieved from relatively slow memory, causing the CPU to stall while waiting for the instruction to be returned. This issue is largely addressed in modern processors by caches and pipeline architectures.

The instruction that the CPU fetches from memory is used to determine what the CPU is to do. In the decode step, the instruction is broken up into parts that have significance to other portions of the CPU. The way in which the numerical instruction value is interpreted is defined by the CPU's instruction set architecture (ISA).

Often, one group of numbers in the instruction, called the opcode, indicates which operation to perform. The remaining parts of the number usually provide information required for that instruction, such as operands for an addition operation. Such operands may be given as a constant value (called an immediate value), or as a place to locate a value: a register or a memory address, as determined by some addressing mode.

In older designs the portions of the CPU responsible for instruction decoding were unchangeable hardware devices. However, in more abstract and complicated CPUs and ISAs, a microprogram is often used to assist in translating instructions into various configuration signals for the CPU. This microprogram is sometimes rewritable so that it can be modified to change the way the CPU decodes instructions even after it has been manufactured.

After the fetch and decode steps, the execute step is performed. During this step, various portions of the CPU are connected so they can perform the desired operation. If, for instance, an addition operation was requested, an arithmetic logic unit (ALU) will be connected to a set of inputs and a set of outputs.

The inputs provide the numbers to be added, and the outputs will contain the final sum. The ALU contains the circuitry to perform simple arithmetic and logical operations on the inputs (like addition and bitwise operations). If the addition operation produces a result too large for the CPU to handle, an arithmetic overflow flag in a flags register may also be set.

The final step, writeback, simply "writes back" the results of the execute step to some form of memory. Very often the results are written to some internal CPU register for quick access by subsequent instructions. In other cases results may be written to slower, but cheaper and larger, main memory. Some types of instructions manipulate the programme counter rather than directly produce result data.

These are generally called "jumps" and facilitate behavior like loops, conditional programme execution (through the use of a conditional jump), and functions in programmes. Many instructions will also change the state of digits in a "flags" register. These flags can be used to influence how a programme behaves, since they often indicate the outcome of various operations. For example, one type of "compare" instruction considers two values and sets a number in the flags register according to which one is greater. This flag could then be used by a later jump instruction to determine programme flow.

After the execution of the instruction and writeback of the resulting data, the entire process repeats, with the next instruction cycle normally fetching the next-in-sequence instruction because of the incremented value in the programme counter. If the completed instruction was a jump, the programme counter will be modified to contain the address of the instruction that was jumped to, and programme execution continues normally.

In more complex CPUs than the one described here, multiple instructions can be fetched, decoded, and executed simultaneously. This section describes what is generally referred to as the "Classic RISC pipeline," which in fact is quite common among the simple CPUs used in many electronic devices (often called microcontroller). It largely ignores the important role of CPU cache, and therefore the access stage of the pipeline.

Evolution of Computer

- The first counting device was the abacus, originally from Asia. It worked on a place-value notion meaning that the place of a bead or rock on the apparatus determined how much it was worth.
- 1600s: John Napier discovers logarithms. Robert Bissaker invents the slide rule which will remain in popular use until 19??.

- 1642: Blaise Pascal, a French mathematician and philosopher, invents the first mechanical digital calculator using gears, called the Pascaline. Although this machine could perform addition and subtraction on whole numbers, it was too expensive and only Pascal himself could repare it.
- 1804: Joseph Marie Jacquard used punch cards to automate a weaving loom.
- 1812: Charles P. Babbage, the "father of the computer", discovered that many long calculations involved many similar, repeated operations. Therefore, he designed a machine, the difference engine which would be steam-powered, fully automatic and commanded by a fixed instruction programme. In 1833, Babbage quit working on this machine to concentrate on the analytical engine.
- 1840s: Augusta Ada. "The first programmer" suggested that a binary system shouled be used for staorage rather than a decimal system.
- 1850s: George Boole developed Boolean logic which would later be used in the design of computer circuitry.
- 1890: Dr. Herman Hollerith introduced the first electromechanical, punched-card data-processing machine which was used to compile information for the 1890 U.S. census. Hollerith's tabulator became so successful that he started his own business to market it. His company would eventually become International Business Machines (IBM).

- 1906: American physicist Lee De Forest invents the vacuum tube.
- 1939: Dr. John V. Atanasoff and his assistant Clifford Berry build the first electronic digital computer. Their machine, the Atanasoff-Berry-Computer (ABC) provided the foundation for the advances in electronic digital computers.
- 1941, Konrad Zuse (recently deceased in January of 1996), from Germany, introduced the first programmable computer designed to solve complex engineering equations. This machine, called the Z3, was also the first to work on the binary system instead of the decimal system.
- 1943: British mathematician Alan Turing developed hypothetical devices, the Turing machine which would be designed to perform logical operation and could read and write. It would presage programmable computers. He also used vacuum technology to build British Colossus, a machine used to counteract the German code-scrambling device, Enigma.
- 1944: Howard Aiken, in collaboration with engineers from IBM, constructed a large automatic digital sequence-controlled computer called the Harvard Mark I. This computer could handle all four arithmetic operations, and had special built-in programmes for logarithms and trigonometric functions.

- 1945: Dr. John von Neumann presented a paper outlining the stored-programme concept.
- 1947: The giant ENIAC (Electrical Numerical Integrator and Calculator) machine was developed by John W. Mauchly and J. Presper Eckert, Jr. at the University of Pennsylvania. It used 18, 000 vacuums, punch card input, weighed thirty tons and occupied a thirty-by-fifty-foot space. It wasn't programmable but was productive from 1946 to 1955 and was used to compute artillery-firing tables. That same year, William Shockley, John Bardeen and Walter Brattain of Bell Labs invented the transistor. It would rid computers of vacuum tubes and radios.
- 1949: Maurice V. Wilkes built the EDSAC (Electronic Delay Storage Automatic Computer), the first stored-programme computer. EDVAC (Electronic Discrete Variable Automatic Computer), the second stored-programme computer was built by Mauchly, Eckert, and von Neumann. An Wang developed magnetic-core memory which Jay Forrester would reorganize to be more efficient.
- 1950: Turing built the ACE, considered by some to be the first programmable digital computer.

Generation of Computers

The history of computer development is often referred to in reference to the different generations of computing devices A generation refers to the state of improvement in the development of a product. This term is also used in the different advancements of computer technology. With each new generation, the circuitry has gotten smaller and more advanced than the previous generation before it. As a result of the miniaturization, speed, power, and memory of computers has proportionally increased. New discoveries are constantly being developed that affect the way we live, work and play.

Each generation of computer is characterized by major technological development that fundamentally changed the way computers operate, resulting in increasingly smaller, cheaper, more powerful and more efficient and reliable devices. Read about each generation and the developments that led to the current devices that we use today.

First Generation - 1940-1956

The first computers used vacuum tubes for circuitry and magnetic drums for memory, and were often enormous, taking up entire rooms. A magnetic drum, also referred to as drum, is a metal cylinder coated with magnetic iron-oxide material on which data and programmes can be stored. Magnetic drums were once used as a primary storage device but have since been implemented as auxiliary storage devices. The tracks on a magnetic drum are assigned to channels located around the circumference of the drum, forming adjacent circular bands that wind around the drum.

Single drum can have up to 200 tracks. As the drum rotates at a speed of up to 3,000 rpm, the device's read/write heads deposit magnetized spots on the drum during the write operation and sense these spots during a read operation. This action is similar to that of a magnetic tape or disk drive. They were very expensive to operate and in addition to using a great deal of electricity, generated a lot of heat, which was often the cause of malfunctions.

First generation computers relied on machine language to perform operations, and they could only solve one problem at a time. Machine languages are the only languages understood by computers. While easily understood by computers, machine languages are almost impossible for humans to use because they consist entirely of numbers. Programmers, therefore, use either a high-level programming language or an assembly language. An assembly language contains the same instructions as a machine language, but the instructions and variables have names instead of being just numbers.

Programmes written in high-level languages retranslated into assembly language or machine language by a compiler. Assembly language programmes retranslated into machine language by a programme called an assembler. Every CPU has its own unique machine language.

Programmes must be rewritten or recompiled, therefore, to run on different types of computers. Input was based on punched cards and paper tape, and output was displayed on printouts. The UNIVAC and ENIAC computers are examples of first-generation computing devices. The UNIVAC was the first commercial computer delivered to a business client.

Acronym for Electronic Numerical Integrator And Computer, the world's first operational electronic digital computer, developed by Army Ordnance to compute World War II ballistic firing tables. The ENIAC, weighing 30 tons, using 200 kilowatts of electric power and consisting of 18,000 vacuum tubes,1,500 relays, and hundreds of thousands of resistors, capacitors, and inductors, was completed in 1945.

In addition to ballistics, the ENIAC's field of application included weather prediction, atomic-energy calculations, cosmic-ray studies, thermal ignition, random-number studies, wind-tunnel design, and other scientific uses. The ENIAC soon became obsolete as themed arose for faster computing speeds.

Second Generation - 1956-1963

Transistors replaced vacuum tubes and ushered in the second generation of computers. Transistor is a device composed of semiconductor material that amplifies a signal or opens or closes a circuit. Invented in 1947at Bell Labs, transistors have become the key ingredient of all digital circuits, including computers.

Today's microprocessors contain tens of millions of microscopic transistors. Prior to the invention of transistors, digital circuits were composed of vacuum tubes, which had many disadvantages. They were much larger, required more energy, dissipated more heat, and were more prone to failures. It's safe to say that without the invention of transistors, computing, as we know it today would not be possible.

The transistor was invented in 1947 but did not see widespread use in computers until the late 50s. The transistor was far superior to the vacuum tube, allowing computers to become smaller, faster, cheaper, more energy-efficient and more reliable than their first-generation predecessors. Though the transistor still generated a great deal of heat that subjected the computer to damage, it was a vast improvement over the vacuum tube. Second-generation computers still relied on punched cards for input and printouts for output.

Second-generation computers moved from cryptic binary machine language to symbolic, or assembly, languages, which allowed programmers to specify instructions inwards. High-level programming languages were also being developed at this time, such as early versions of COBOL and FORTRAN.

These were also the first computers that stored their instructions in their memory, which moved from a magnetic drum to magnetic core technology. The first computers of this generation were developed for the atomic energy industry.

Third Generation - 1964-1971

The development of the integrated circuit was the hallmark of the third generation of computers. Transistors were miniaturized and placed on siliconchips, called semiconductors, which drastically increased the speed and efficiency of computers.

A nonmetallic chemical element in the carbon family of elements. Silicon - atomic symbol "Si" - is the second most abundant element in the earth's crust, surpassed only by oxygen. Silicon does not occur uncombined in nature. Sand

and almost all rocks contain silicon combined with oxygen, forming silica. When silicon combines with other elements, such as iron, aluminum or potassium, a silicate is formed. Compounds of silicon also occur in the atmosphere, natural waters, and many plants and in the bodies of some animals.

Silicon is the basic material used to make computer chips, transistors, silicon diodes and other electronic circuits and switching devices because its atomic structure makes the element an ideal semiconductor. Silicon is commonly doped, or mixed, with other elements, such as boron, phosphorous and arsenic, to alter its conductive properties. A chip is a small piece of semi conducting material (usually silicon) on which an integrated circuit is embedded.

A typical chip is less than ¼-square inches and can contain millions of electronic components(transistors). Computers consist of many chips placed on electronic boards called printed circuit boards. There are different types of chips. For example, CPU chips (also called microprocessors) contain an entire processing unit, whereas memory chips contain blank memory.

Semiconductor is a material that is neither a good conductor of electricity (like copper) nor a good insulator (like rubber). The most common semiconductor materials are silicon and germanium. These materials are then doped to create an excess or lack of electrons.

Computer chips, both for CPU and memory, are composed of semiconductor materials. Semiconductors make it possible to miniaturize electronic components, such as transistors. Not only does miniaturization mean that the components take up less space, it also means that they are faster and require less energy. Instead of punched cards and printouts, users interacted with third generation computers through keyboards and monitors and interfaced with an operating system, which allowed the device to run many different applications at one time with a central programme that monitored the memory.

Computers for the first time became accessible to a mass audience because they were smaller and cheaper than their predecessors.

Fourth Generation - 1971-Present

The microprocessor brought the fourth generation of computers, as thousands of integrated circuits were built onto a single silicon chip. A silicon chip that contains a CPU.

In the world of personal computers, the terms microprocessor and CPU are used interchangeably. At the heart of all personal computers and most workstations sits a microprocessor. Microprocessors also control the logic of almost all digital devices, from clock radios to fuel-injection systems for automobiles.

Three basic characteristics differentiate microprocessors:

- Instruction Set: The set of instructions that the microprocessor can execute.
- Bandwidth: The number of bits processed in a single instruction.
- Clock Speed: Given in megahertz (MHz), the clock speed determines how many instructions per second the processor can execute.

In both cases, the higher the value, the more powerful the CPU. For example, a 32-bit microprocessor that runs at 50MHz is more powerful than a 16-bitmicroprocessor that runs at 25MHz.

What in the first generation filled an entire room could now fit in the palm of the hand. The Intel 4004chip, developed in 1971, located all the components of the computer - from the central processing unit and memory to input/output controls - on a single chip.

Abbreviation of central processing unit, and pronounced as separate letters. The CPU is the brains of the computer.

Sometimes referred to simply as the processor or central processor, the CPU is where most calculations take place. In terms of computing power, the CPU is the most important element of a computer system. On large machines, CPUs require one or more printed circuit boards. On personal computers and small workstations, the CPU is housed in a single chip called a microprocessor.

Two typical components of a CPU are:

- The arithmetic logic unit (ALU), which performs arithmetic and logical operations.
- The control unit, which extracts instructions from memory and decodes and executes them, calling on the ALU when necessary.

In 1981 IBM introduced its first computer for the home user, and in 1984 Apple introduced the Macintosh. Microprocessors also moved out of the realm of desktop computers and into many areas of life as more and more everyday products began to use microprocessors.

As these small computers became more powerful, they could be linked together to form networks, which eventually led to the development of the Internet. Fourth generation computers also saw the development of GUIs, the mouse and handheld devices

Fifth Generation

Fifth generation computing devices, based on artificial intelligence, are still in development, though there are some applications, such as voice recognition, that are being used today.

Artificial Intelligence is the branch of computer science concerned with making computers behave like humans. John McCarthy at the Massachusetts Institute of Technology.

Artificial intelligence includes coined the term in 1956:

- *Games Playing*: Programming computers to play games such as chess and checkers
- *Expert Systems*: Programming computers to make decisions in real-life situations (for example, some expert systems help doctors diagnose diseases based on symptoms)
- *Natural Language*: Programming computers to understand natural human languages
- Neural Networks: Systems that simulate intelligence by attempting to reproduce the types of physical connections that occur in animal brains

- *Robotics*: Programming computers to see and hear and react to other sensory stimuli

Currently, no computers exhibit full artificial intelligence (that is, are able to simulate human behaviour). The greatest advances have occurred in the field of games playing. The best computer chess programmes are now capable of beating humans.

In May,1997, an IBM super-computer called Deep Blue defeated world chess champion Gary Kasparov in a chess match. In the area of robotics, computers are now widely used in assembly plants, but they are capable only of very limited tasks. Robots have great difficulty identifying objects based on appearance or feel, and they still move and handle objects clumsily.

Natural-language processing offers the greatest potential rewards because it would allow people to interact with computers without needing any specialized knowledge. You could simply walk up to a computer and talk to it. Unfortunately, programming computers to understand natural languages has proved to be more difficult than originally thought. Some rudimentary translation systems that translate from one human language to another are in existence, but they are not nearly as good as human translators.

There are also voice recognition systems that can convert spoken sounds into written words, but they do not understand what they are writing; they simply take dictation. Even these systems are quite limited — you must speak slowly and distinctly.

In the early 1980s, expert systems were believed tore present the future of artificial intelligence and of computers in general. To date, however, they have not lived up to expectations.

Many expert systems help human experts in such fields as medicine and engineering, but they are very expensive to produce and are helpful only in special situations. Today, the hottest area of artificial intelligence is a neural network, which is proving successful in a number of disciplines such as voice recognition and natural language processing.

There are several programming languages that are known as AI languages because they are used almost exclusively for AI applications. The two most common are LISP and Prolog.

Supercomputer and Mainframe

Supercomputer is a broad term for one of the fastest computers currently available. Supercomputers are very expensive and are employed for specialized applications that require immense amounts of mathematical calculations. For example, weather forecasting requires a supercomputer.

Other uses of supercomputers scientific simulations, (animated) graphics, fluid dynamic calculations, nuclear energy research, electronic design, and analysis of geological data (*e.g.* in petrochemical prospecting). Mainframe was a term originally referring to the cabinet containing the central processor unit or "main frame" of a room-filling Stone Age batch machine. After the emergence of smaller "minicomputer" designs in the early 1970s, the traditional big iron machines were described as "mainframe computers" and eventually just as mainframes. Nowadays a Mainframe is a very large and expensive computer capable of supporting hundreds, or even thousands, of users simultaneously.

The chief difference between a supercomputer and a mainframe is that a supercomputer channels all its power into executing a few programmes as fast as possible, whereas a mainframe uses its power to execute many programmes concurrently. In some ways, mainframes are more powerful than supercomputers because they support more simultaneous programmes.

But supercomputers can execute a single programme faster than a mainframe. The distinction between small mainframes and minicomputers is vague, depending really on how the manufacturer wants to market its machines.

It is a midsize computer. In the past decade, the distinction between large minicomputers and small mainframes has blurred, however, as has the distinction between small minicomputers and workstations. But in general, a minicomputer is a multiprocessing system capable of supporting from up to 200 users simultaneously.

It is a type of computer used for engineering applications (CAD/CAM), desktop publishing, software development, and other types of applications that require a moderate amount of computing power and relatively high quality graphics capabilities. Workstations generally come with a large, high-resolution graphics screen, at large amount of RAM, built-in network support, and a graphical user interface. Most workstations also have a mass storage device such as a disk drive, but a special type of workstation, called a diskless workstation, comes without a disk drive.

The most common operating systems for workstations are UNIX and Windows NT. Like personal computers, most workstations are single-user computers. However, workstations are typically linked together to form a local-area network, although they can also be used as stand-alone systems.

Personal Computer

It can be defined as a small, relatively inexpensive computer designed for an individual user. In price, personal computers range anywhere from a few hundred pounds to over five thousand pounds.

All are based on the microprocessor technology that enables manufacturers to put an entire CPU on one chip. Businesses use personal computers for word processing, accounting, desktop publishing, and for running spreadsheet and database management applications. At home, the most popular use for personal computers is for playing games and recently for surfing the Internet.

Personal computers first appeared in the late 1970s. One of the first and most popular personal computers was the Apple II, introduced in 1977 by Apple Computer. During the late 1970s and early 1980s, new models and competing operating systems seemed to appear daily. Then, in 1981, IBM entered the fray with its first personal computer, known as the IBM PC. The IBM PC quickly became the personal computer of choice, and most other personal computer manufacturers fell by the wayside. P.C. is short for personal computer or IBM PC.

One of the few companies to survive IBM's onslaught was Apple Computer, which remains a major player in the personal computer marketplace. Other companies adjusted to IBM's dominance by building IBM clones, computers that were internally almost the same as the IBM PC, but that cost less. Because IBM clones used the same microprocessors as IBM PCs, they were capable of running the same software. Over the years, IBM has lost much of its influence in directing the evolution of PCs.

Therefore after the release of the first PC by IBM the term PC increasingly came to mean IBM or IBM-compatible personal computers, to the exclusion of other types of personal computers, such as Macintoshes. In recent years, the term PC has become more and more difficult to pin down. In general, though, it applies to any personal computer based on an Intel microprocessor, or on an Intel-compatible microprocessor.

For nearly every other component, including the operating system, there are several options, all of which fall under the rubric of PC Today, the world of personal computers is basically divided between Apple Macintoshes and PCs. The principal characteristics of personal computers are that they are single-user systems and are based on microprocessors.

However, although personal computers are designed as single-user systems, it is common to link them together to form a network. In terms of power, there is great variety. At the high end, the distinction between personal computers and workstations has faded. High-end models of the Macintosh and PC offer the same computing power and graphics capability as low-end workstations by Sun Microsystems, Hewlett-Packard, and DEC.

Personal Computer Types

Actual personal computers can be generally classified by size and chassis/case. The chassis or case is the metal frame that serves as the structural support for electronic components. Every computer system requires at least one chassis to house the circuit boards and wiring.

The chassis also contains slots for expansion boards. If you want to insert more boards than there are slots, you will need an expansion chassis, which provides additional slots. There are two basic flavours of chassis designs–desktop models and tower models–but there are many variations on these two basic types. Then come the portable computers that are computers small enough to carry. Portable computers include notebook and subnotebook computers, hand-held computers, palmtops, and PDAs.

Tower Model

The term refers to a computer in which the power supply, motherboard, and mass storage devices are stacked on top of each other in a cabinet. This is in contrast to desktop models, in which these components are housed in a more compact box. The main advantage of tower models is that there are fewer space constraints, which makes installation of additional storage devices easier.

Desktop Model

A computer designed to fit comfortably on top of a desk, typically with the monitor sitting on top of the computer. Desktop model computers are broad and low, whereas tower model computers are narrow and tall.

Because of their shape, desktop model computers are generally limited to three internal mass storage devices. Desktop models designed to be very small are sometimes referred to as slim line models.

Notebook Computer

An extremely lightweight personal computer. Notebook computers typically weigh less than 6 pounds and are small enough to fit easily in a briefcase. Aside from size, the principal difference between a notebook computer and a personal computer is the display screen.

Notebook computers use a variety of techniques, known as flat-panel technologies, to produce a lightweight and non-bulky display screen. The quality of notebook display screens varies considerably.

In terms of computing power, modern notebook computers are nearly equivalent to personal computers. They have the same CPUs, memory capacity, and disk drives. However, all this power in a small package is expensive.

Notebook computers cost about twice as much as equivalent regular-sized computers. Notebook computers come with battery packs that enable you to run them without plugging them in. However, the batteries need to be recharged every few hours.

Laptop Computer

A small, portable computer — small enough that it can sit on your lap. Nowadays, laptop computers are more frequently called notebook computers.

Sub Notebook Computer

A portable computer that is slightly lighter and smaller than a full-sized notebook computer. Typically, sub notebook computers have a smaller keyboard and screen, but are otherwise equivalent to notebook computers.

Hand-held Computer

A portable computer that is small enough to be held in one's hand. Although extremely convenient to carry, handheld computers have not replaced notebook computers because of their small keyboards and screens. The most popular hand-held computers are those that are specifically designed to provide PIM (personal information manager) functions, such as a calendar and address book. Some manufacturers are trying to solve the small keyboard problem by replacing the keyboard with an electronic pen.

However, these pen-based devices rely on handwriting recognition technologies, which are still in their infancy. Hand-held computers are also called PDAs, palmtops and pocket computers.

Palmtop

A small computer that literally fits in your palm. Compared to full-size computers, palmtops are severely limited, but they are practical for certain functions such as phone books and calendars. Palmtops that use a pen rather than a keyboard for input are often called hand-held computers or PDAs. Because of their small size, most palmtop computers do not include disk drives. However, many contain PCMCIA slots in which you can insert disk drives, modems, memory, and other devices. Palmtops are also called PDAs, hand-held computers and pocket computers.

PDA

Short for personal digital assistant, a handheld device that combines computing, telephone/fax, and networking features. A typical PDA can function as a cellular phone, fax sender, and personal organizer.

Unlike portable computers, most PDAs are pen-based, using a stylus rather than a keyboard for input. This means that they also incorporate handwriting recognition features.

Some PDAs can also react to voice input by using voice recognition technologies. The field of PDA was pioneered by Apple Computer, which introduced the Newton MessagePad in 1993. Shortly thereafter, several other manufacturers offered similar products.

To date, PDAs have had only modest success in the marketplace, due to their high price tags and limited applications. However, many experts believe that PDAs will eventually become common gadgets.

Classification of Computers according to Size

Microcomputers are the most common type of computers in existence today, whether at work in school or on the desk at home. The term “microcomputer” was introduced with the advent of single chip microprocessors. The term "microcomputer" itself is now practically an anachronism. These computers include:

- Desktop computers – A case and a display, put under and on a desk.
- In-car computers ("carputers") – Built into a car, for entertainment, navigation, etc.

A separate class is that of mobile devices:

- Laptops, notebook computers and Palmtop computers – Portable and all in one case. Varying sizes, but other than smartbooks expected to be "full" computers without limitations.
- Tablet PC – Like laptops, but with only a touch-screen instead of a physical keyboard.
- Smartphones, smartbooks and PDAs (personal digital assistants) – Small handheld computers with limited hardware.
- Programmable calculator– Like small handhelds, but specialised on mathematical work.
- Game consoles – Fixed computers specialized for entertainment purposes (computer games).
- Handheld game consoles – Ditto, but small and portable.

Minicomputers (Midrange Computers)

A minicomputer (colloquially, mini) is a class of multi-user computers that lies in the middle range of the computing spectrum, in between the smallest multi-user systems (mainframe computers) and the largest single-user systems (microcomputers or personal computers). The contemporary term for this class of system is midrange computer, such as the higher-end SPARC, POWER and Itanium-based systems from Sun Microsystems, IBM and Hewlett-Packard.

Mainframe Computers

The term mainframe computer was created to distinguish the traditional, large, institutional computer intended to service multiple users from the smaller, single user machines. These computers are capable of handling and processing very large amounts of data quickly. Mainframe computers are used in large institutions such as government, banks and large corporations.

These institutions were early adopters of computer use, long before personal computers were available to individuals. "Mainframe" often refers to computers compatible with the computer architectures established in the 1960s. Thus, the origin of the architecture also affects the classification, not just processing power.

Mainframes are measured in millions of instructions per second or MIPS. An example of integer operation is moving data around in memory or I/O devices. A more useful industrial benchmark is transaction processing as defined by the Transaction Processing Performance Council.

Mainframes are built to be reliable for transaction processing as it is commonly understood in the business world: a commercial exchange of goods, services, or money. A typical transaction, as defined by the Transaction

Processing Performance Council, would include the updating to a database system for such things as inventory control (goods), airline reservations (services), or banking (money). A transaction could refer to a set of operations including disk read/writes, operating system calls, or some form of data transfer from one subsystem to another.

Supercomputer

A supercomputer is focused on performing tasks involving intense numerical calculations such as weather forecasting, fluid dynamics, nuclear simulations, theoretical astrophysics, and complex scientific computations. A supercomputer is a computer that is at the frontline of current processing capacity, particularly speed of calculation. The term supercomputer itself is rather fluid, and today's supercomputer tends to become tomorrow's ordinary computer. Supercomputer processing speeds are measured in floating point operations per second or FLOPS. Example of floating point operation is the calculation of mathematical equations in real numbers. In terms of computational capability, memory size and speed, I/O technology, and topological issues such as bandwidth and latency, Supercomputers are the most powerful. Supercomputers are very expensive and not cost-effective just to perform batch or transaction processing. Transaction processing is handled by less powerful computer such as server computer or mainframe.

Classes by Function

Server usually refers to a computer that is dedicated to providing a service. For example, a computer dedicated to a database may be called a "database server". "File servers" manage a large collection of computer files. "Web servers" process web pages and web applications. Many smaller servers are actually personal computers that have been dedicated to providing services for other computers.

Information Appliances

Information appliances are computers specially designed to perform a specific user-friendly function —such as playing music, photography, or editing text. The term is most commonly applied to mobile devices, though there are also portable and desktop devices of this class.

Embedded Computers

Embedded computers are computers that are a part of a machine or device. Embedded computers generally execute a program that is stored in non-volatile memory and is only intended to operate a specific machine or device. Embedded computers are very common.

Embedded computers are typically required to operate continuously without being reset or rebooted, and once employed in their task the software usually cannot be modified. An automobile may contain a number of embedded computers; however, a washing machine and a DVD player would contain only one.

The central processing units (CPUs) used in embedded computers are often sufficient only for the computational requirements of the specific application and may be slower and cheaper than CPUs found in a personal computer.

Classification of Computers

Computers are classified according to their data processing speed, amount of data that they can hold and price. Generally, a computer with high processing speed and large internal storage is called a big computer. Due to rapidly improving technology, we are always confused among the categories of computers.

Depending upon their speed and memory size, computers are classified into following four main groups.

- Supercomputer.
- Mainframe computer.
- Mini computer.
- Microcomputer.

Supercomputer: powerful and fastest

Supercomputer is the most powerful and fastest, and also very expensive. It was developed in 1980s. It is used to process large amount of data and to solve the complicated scientific problems. It can perform more than one trillions calculations per second. It has large number of processors connected parallel. So parallel processing is done in this computer. In a single supercomputer thousands of users can be connected at the same time and the supercomputer handles the work of each user separately.

Supercomputer are mainly used for:

- Weather forecasting.
- Nuclear energy research.
- Aircraft design.

- Automotive design.
- Online banking.
- To control industrial units.

The supercomputers are used in large organizations, research laboratories, aerospace centres, large industrial units etc. Nuclear scientists use supercomputers to create and analyse models of nuclear fission and fusions, predicting the actions and reactions of millions of atoms as they interact. The examples of supercomputers are CRAY-1, CRAY-2, Control Data CYBER 205 and ETA A-10 etc.

Mainframe Computers: large-scale computers

Mainframe computers are also large-scale computers but supercomputers are larger than mainframe. These are also very expensive. The mainframe computer specially requires a very large clean room with air-conditioner. This makes it very expensive to buy and operate.

It can support a large number of various equipments. It also has multiple processors. Large mainframe systems can handle the input and output requirements of several thousand of users.

For example, IBM, S/390 mainframe can support 50,000 users simultaneously. The users often access then mainframe with terminals or personal computers. Tere are basically two types of terminals used with mainframe systems. These are:

Dumb Terminal

Dumb terminal does not have its own CPU and storage devices. This type of terminal uses the CPU and storage devices of mainframe system. Typically, a dumb terminal consists of monitor and a keyboard (or mouse).

Intelligent Terminal

Intelligent terminal has its own processor and can perform some processing operations. Usually, this type of terminal does not have its own storage. Typically, personal computrers are used as intelligent terminals.

A personal computer as an intelligent terminal gives facility to access data and other services from mainframe system. It also enables to store and process data locally. The mainframe computers are specially used as servers on the World Wide Web. The mainframe computers are used in large organizations such as Banks, Airlines and Universities etc. where many people (users) need frequent access to the same data, which is usually organized into one or more huge databases. IBM is the major manufacturer of mainframe computers. The examples of mainframes are IBM S/390, Control Data CYBER 176 and Amdahl 580 etc.

Minicomputers: smaller in size

These are smaller in size, have lower processing speed and also have lower cost than mainframe. These computers are known as minicomputers because of their small size as compared to other computers at that time. The capabilities of a minicomputer are between mainframe and personal computer. These computers are also known as midrange computers.

The minicomputers are used in business, education and many other government departments. Although some minicomputers are designed for a single user but most are designed to handle multiple terminals.

Minicomputers are commonly used as servers in network environment and hundreds of personal computers can be connected to the network with a minicomputer acting as server like mainframes, minicomputers are used as web servers. Single user minicomputers are used for sophisticated design tasks.

The first minicomputer was introduced in the mid-1960s by Digital Equipment Corporation (DEC). After this IBM Corporation (AS/400 computers) Data General Corporation and Prime Computer also designed the mini computers.

Microcomputer: personal computers

The microcomputers are also known as personal computers or simply PCs. Microprocessor is used in this type of computer.

These are very small in size and cost. The IBM's first microcomputer was designed in 1981 and was named as IBM-PC. After this many computer hardware companies copied the design of IBM-PC. The term "PC-compatible" refers any personal computer based on the original IBM personal computer design.

The most popular types of personal computers are the PC and the Apple. PC and PC-compatible computers have processors with different architectures than processors in Apple computers. These two types of computers also use different operating systems. PC and PC-compatible computers use the Windows operating system while Apple computers use the Macintosh operating system (MacOS).

The majority of microcomputers sold today are part of IBM-compatible. However the Apple computer is neither an IBM nor a compatible. It is another family of computers made by Apple computer. Personal computers are available in two models.

These are:

- Desktop PCs
- Tower PCs.

A desktop personal computer is most popular model of personal computer. The system unit of the desktop personal computer can lie flat on the desk or table. In desktop personal computer, the monitor is usually placed on the system unit.

Microcomputer are further divided into following categories:

- Laptop computer
- Workstation
- Network computer
- Handheld computer.

Laptop Computer: Notebook Computer

Laptop computer is also known as notebook computer. It is small size (85-by-11 inch notebook computer and can fit inside a briefcase. The laptop computer is operated on a special battery and it does not have to be plugged in like desktop computer. The laptop computer is portable and fully functional microcomputer. It is mostly used during journey. It can be used on your lap in an airplane. It is because it is referred to as laptop computer. The memory and storage capacity of laptop computer is almost equivalent to the PC or desktop computer. It also has the hard dist, floppy disk drive, Zip disk drive, CD-ROM drive, CD-writer etc. it has built-in keyboard and built-in trackball as pointing device.

Laptop computer is also available with the same processing speed as the most powerful personal computer. It means that laptop computer has same features as personal computer. Laptop computers are more expensive than desktop computers. Normally these computers are frequently used in business travellers.

Network Computers: Personal Computers

Network computers are also version of personal computers having less processing power, memory and storage. These are specially designed as terminals for network environment. Some types of network computers have no storage.

The network computers are designed for network, Internet or Intranet for data entry or to access data on the network.

The network computers depend upon the network's server for data storage and to use software. These computers also use the network's server to perform some processing tasks. Network computers are cheaper to purchase and to maintain than personal computers.

Workstations: Single User Computers

Workstations are special single user computers having the same features as personal computer but have the processing speed equivalent to minicomputer or mainframe computer. A workstation computer can be fitted on a desktop.

Scientists, engineers, architects and graphic designers mostly use these computers. Workstation computers are expensive and powerful computers. These have advanced processors, more RAM and storage capacity than personal computers. These are usually used as single-user applications but these are used as servers on computer network and web servers as well.

Handheld Computer

In the mid 1990s, many new types of small personal computing devices have been introduced and these are referred to as handheld computers. These computers are also referred to as Palmtop Computers.

The handheld computers sometimes called Mini-Notebook Computers.

The type of computer is named as handheld computer because it can fit in one hand while you can operate it with the other hand. Because of its reduced size, the screen of handheld computer is quite small.

Similarly it also has small keyboard. The handheld computers are preferred by business traveller. Some handheld computers have a specialized keyboard.

These computers are used by mobile employees, such as meter readers and parcel delivery people, whose jobs require them to move from place to place.

The examples of handheld computers are:

- Personal Digital Assistance
- Cellular telephones
- H/PC Pro devices.

Types of Computers

A computer is a machine that can be programmed to manipulate symbols. Its principal characteristics are:

- It responds to a specific set of instructions in a well-defined manner.
- It can execute a prerecorded list of instructions (a programme).
- It can quickly store and retrieve large amounts of data.

Therefore computers can perform complex and repetitive procedures quickly, precisely and reliably. Modern computers are electronic and digital.

The actual machinery (wires, transistors, and circuits) is called hardware; the instructions and data are called software. All general-purpose computers require the following hardware components:

- Central processing unit (CPU): The heart of the computer, this is the component that actually executes instructions organized in programmes ("software") which tell the computer what to do.
- Memory (fast, expensive, short-term memory): Enables a computer to store, at least temporarily, data, programmes, and intermediate results.
- Mass storage device (slower, cheaper, long-term memory): Allows a computer to permanently retain large amounts of data and programmes between jobs.

Common mass storage devices include disk drives and tape drives.

- Input device: Usually a keyboard and mouse, the input device is the conduit through which data and instructions enter a computer.
- Output device: A display screen, printer, or other device that lets you see what the computer has accomplished.

In addition to these components, many others make it possible for the basic components to work together efficiently. For example, every computer requires a bus that transmits data from one part of the computer to another.

Computer Sizes and Power

Computers can be generally classified by size and power as follows, though there is considerable overlap:

- Personal computer: A small, single-user computer based on a microprocessor.
- Workstation: A powerful, single-user computer. A workstation is like a personal computer, but it has a more powerful microprocessor and, in general, a higher-quality monitor.
- Minicomputer: A multi-user computer capable of supporting up to hundreds of users simultaneously.
- Mainframe: A powerful multi-user computer capable of supporting many hundreds or thousands of users simultaneously.

- Supercomputer: An extremely fast computer that can perform hundreds of millions of instructions per second.

3

The Scanners Capture Process

Scanners differ significantly from digital camera in many areas. First, the optical resolution of a flatbed scanner can exceed 5000 pixels per inch (200 pixels per mm). Even at a relatively low resolution of 1200 pixels per inch (47 p/mm) a letter sized image would be 134 megapixels in size.

The depth of field of most scanners is very limited, usually no more than half an inch (12 mm), but the built-in light source provides excellent sharpness, color saturation, and unique shadow effects.

The time it takes the scanning head to traverse the bed means that scanners can only be used to capture still objects, and common items used are flowers, leaves, and other suitable "still life" subjects.

Equipment

Using a flatbed scanner to scan items other than paper documents exceeds the original purpose of the scanner, so special care must be taken with the process. The bed of the scanner is typically made of glass, and care needs to be taken that the glass not be scratched or cracked when placing or removing items on the bed.

Since the items to be captured are often placed directly on the bed, dust and other particles will often land on the glass, and care must be taken to keep the glass clean. Scanners will also hold only a limited amount of weight, and items that may damage the scanner, such a liquids or items that might scratch the glass, should be placed on a plastic barrier to protect the bed.

The larger the scanner bed, the larger an image may be captured, so scanners with large beds provide the artist with more flexibility than smaller document scanners. Many scanners advertise two resolutions, an optical resolution and a higher resolution that is achieved by interpolation. A higher optical resolution is desirable, since that captures more data, while interpolation can actually result in reduced quality.

Flatbed scanners typically have a hinged cover that covers the bed, and reflects light back into the scan head. This cover is usually removed or propped open when scanning 3-D objects, to prevent damage or compression of the subject. Removal of the cover also allows the artist to use additional light sources positioned above the bed, which can be used to enhance the depth captured by the scanner. Scanners can also be modified to provide additional capture abilities. For example, the scanner, with the illumination removed or disabled, can be used as a giant CCD replacement, producing a large format digital camera back at a fraction of the cost of professional large format systems.

Techniques

The simplest use of the scanner, which also most closely matches its use for document capture, is as a specialized tool for macro photography. As long as the subject can be placed on the scanner bed, the scanner is excellent for capturing very high resolution images, within its limitations.

A common artistic use of the scanner is to capture collages of objects. The objects are arranged by the artist on the scanner bed, and then captured. Since the artist is working from the back of the image, it can be difficult to get the desired arrangement. Scanning software with the ability to generate a low resolution preview scan can help in obtaining the desired arrangement before the final, high resolution scan is made. Since the subjects are often placed in contact with the scanner, there is a high potential for damage to the scanner from objects scratching or cracking the surface of the bed, or from liquids that might seep from the subject into the interior of the scanner. These risks can be mitigated by placing a layer of transparent protective material, such as clear plastic film, onto the scanner bed. Another approach is to invert the scanner, so the bed is above the subject and not quite in contact with it.

Capturing a moving subject with the scanner can be viewed as a problem, or as an opportunity for artistic effect. As the subject moves during the scan, distortions are caused along the axis of the scan head's movement, as it captures different periods of the subject's movement line by line in a manner similar to slit-scan photography. The artist can use this by aligning the direction of the scan head's movement to deliberately caused the desired distortion.

Further Manipulation

While the result of a scanner capture provides a work of digital art, just as a digital photograph does, further manipulation of the captured image are possible as well. This may be as simple as flattening the background to enhance the "floating" effect provided by the scanner to complete reworking of the image.

4

Barcode Reader

A barcode reader (or barcode scanner) is an electronic device for reading printed barcodes. Like a flatbed scanner, it consists of a light source, a lens and a light sensor translating optical impulses into electrical ones. Additionally, nearly all barcode readers contain *decoder* circuitry analyzing the barcode's image data provided by the sensor and sending the barcode's content to the scanner's output port.

Types of Barcode Readers

Methods

Types of Technology

The reader types can be distinguished as follows:

Pen-type Readers

Pen-type readers consist of a light source and a photodiode that are placed next to each other in the tip of a pen or wand. To read a bar code, the tip of the pen moves across the bars in a steady motion. The photodiode measures the intensity of the light reflected back from the light source and generates a waveform that is used to measure the widths of the bars and spaces in the bar code. Dark bars in the bar code absorb light and white spaces reflect light so that the voltage waveform generated by the photo diode is a representation of the bar and space pattern in the bar code. This waveform is decoded by the scanner in a manner similar to the way Morse code dots and dashes are decoded.

Laser Scanners

Laser scanners work the same way as pen type readers except that they use a laser beam as the light source and typically employ either a reciprocating mirror or a rotating prism to scan the laser beam back and forth across the bar code. As with the pen type reader, a photodiode is used to measure the intensity of the light reflected back from the bar code. In both pen readers and laser scanners, the light emitted by the reader is rapidly varied in brightness with a data pattern and the photodiode receive circuitry is designed to detect only signals with the same modulated pattern.

CCD Readers

CCD readers use an array of hundreds of tiny light sensors lined up in a row in the head of the reader. Each sensor measures the intensity of the light immediately in front of it. Each individual light sensor in the CCD reader is extremely small and because there are hundreds of sensors lined up in a row, a voltage pattern identical to the pattern in a bar code is generated in the reader by sequentially measuring the voltages across each sensor in the row. The important difference between a CCD reader and a pen or laser scanner is that the CCD reader is measuring emitted ambient light from the bar code whereas pen or laser scanners are measuring reflected light of a specific frequency originating from the scanner itself.

Camera-based Readers

Two-dimensional imaging scanners are the fourth and newest type of bar code reader. They use a camera and image processing techniques to decode the bar code. Video camera readers use small video cameras with the same CCD technology as in a CCD bar code reader except that instead of having a single row of sensors, a video camera has hundreds of rows of sensors arranged in a two dimensional array so that they can generate an image. Large field-of-view readers use high resolution industrial cameras to capture multiple bar codes simultaneously. All the bar codes appearing in the photo are decoded instantly (ImageID patents 6801245 & 6922208). There are a number of open source libraries for barcode reading from images. These include the ZXing project, which reads one- and two-dimensional barcodes using Android and JavaME, the JJIL project, which includes code for reading EAN-13 barcodes from cellphone cameras using Java ME, and ZBar, which reads various one-dimensional barcodes in C.

Even web site integration, either by image uploads (e.g. Folke Ashberg: EAN-13 Image-Scanning and code creation tools) or by use of plugins (e.g. the Barcodepedia uses a flash application and some web cam for querying a database), have been realized options for resolving the given tasks. Finally, there are free closed source libraries including Barcode Scanner, which read one- and two-dimensional barcodes on iPhone, Windows Mobile, Android, BlackBerry and Symbian smartphones.

Omni-directional Barcode Scanners

Omni-directional scanning uses "series of straight or curved scanning lines of varying directions in the form of a starburst, a lissajous pattern, or other multiangle arrangement are projected at the symbol and one or more of them will be able to cross all of the symbol's bars and spaces, no matter what the orientation." Omni-directional scanners almost all use a laser.

Unlike the simpler single-line laser scanners, they produce a pattern of beams in varying orientations allowing them to read barcodes presented to it at different angles. Most of them use a single rotating polygonal mirror and an arrangement of several fixed mirrors to generate their complex scan patterns. Omni-directional scanners are most familiar through the horizontal scanners in supermarkets, where packages are slid across a glass or sapphire window. There are a range of different omni-directional units available which can be used for differing scanning applications, ranging from retail type applications with the barcodes read only a few centimetres away from the scanner to industrial conveyor scanning where the unit can be a couple of metres away or more from the code. Omni-directional scanners are also better at reading poorly printed, wrinkled, or even torn barcodes.

Cell Phone Cameras

While cell phone cameras without auto-focus are not ideal for reading some common barcode formats, there are 2D barcodes (such as Semacode) which are optimized for cell phones, as well as QR Codes and Data Matrix codes which can be read quickly and accurately with or without auto-focus. These open up a number of applications for consumers:

- Movies: DVD/VHS movie catalogs
- Music: CD catalogs, play MP3 when scanned
- Book catalogs
- Groceries, nutrition information, making shopping lists when the last of an item is used, etc.
- Personal Property inventory (for insurance and other purposes)
- Calling cards: 2D barcodes can store contact information for importing.
- Brick and mortar shopping: Portable scanners can be used to record items of interest for looking up online at home.
- Coupon management: weeding expired coupons.
- Personal finance. Receipts can be tagged with a barcode label and the barcode scanned into personal

finance software when entering. Later, scanned receipt images can then be automatically associated with the appropriate entries. Later, the bar codes can be used to rapidly weed out paper copies not required to be retained for tax or asset inventory purposes.

- If retailers put barcodes on receipts that allowed downloading an electronic copy or encoded the entire receipt in a 2D barcode, consumers could easily import data into personal finance, property inventory, and grocery management software. Receipts scanned on a scanner could be automatically identified and associated with the appropriate entries in finance and property inventory software.

A number of enterprise applications using cell phones are appearing:

- Access control (ex. ticket validation at venues), inventory reporting (ex. tracking deliveries), asset tracking (ex. anti-counterfeiting).

ZXing Barcode Scanner, an Android platform application, is one example of a barcode reader that has been optimized to work with a cellphone camera. Big in Japan's ShopSavvy and Occipital's RedLaser, among others, are examples of barcode reader apps for the iPhone optimized to work with a cellphone camera. There are also a few other iPhone apps, which integrates barcode scanner into Google shopping web site.

Housing types

The reader packaging can be distinguished as follows :

Handheld scanner with a handle and typically a trigger button for switching on the light source.

Pen scanner (or wand scanner) a pen-shaped scanner that is swiped.

Stationary scanner; wall- or table-mounted scanners that the barcode is passed under or beside. These are commonly found at the checkout counters of supermarkets and other retailers.

Fixed-position scanner: an industrial barcode reader used to identify products during manufacture or logistics. Often used on conveyor tracks to identify cartons or pallets which need to be routed to another process or shipping location.

Another application joins holographic scanners with a checkweigher to read bar codes of any orientation or placement, and weighs the package. Systems like this are used in factory and farm automation for quality management and shipping.

PDA scanner (or Auto-ID PDA) a PDA with a built-in barcode reader or attached barcode scanner.

Automatic reader a back office equipment to read barcoded documents at high speed (50,000/hour).

Cordless scanner (or Wireless scanner) a cordless barcode scanner is operated by a battery fitted inside it and is not connected to the electricity mains

Methods of Connection

Early Serial Interfaces

Early barcode scanners, of all formats, almost universally used the then-common RS-232 serial interface. This was an electrically simple means of connection and the software to access it is also relatively simple, although needing to be written for specific computers and their serial ports with abdelilah.

Proprietary Interfaces

There are a few other less common interfaces. These were used in large EPOS systems with dedicated hardware, rather than attaching to existing commodity computers. In some of these interfaces, the scanning device returned a "raw" signal proportional to the intensities seen while scanning the barcode.

This was then decoded by the host device. In some cases the scanning device would convert the symbology of the barcode to one that could be recognized by the host device, such as Code 39.

Keyboard Wedges

With the popularity of the PC and its standard keyboard interface, it became ever easier to connect physical hardware to a PC and so there was commercial demand similarly to reduce the complexity of the associated software.

"Keyboard wedge" hardware plugged between the PC and its normal keyboard, with characters from the barcode scanner appearing exactly as if they had been typed at the keyboard. This made the addition of simple barcode reading abilities to existing programmes very easy, without any need to change them, although it did require some care by the user and could be restrictive in the content of the barcodes that could be handled.

USB

Later barcode readers began to use USB connectors rather than the keyboard port, as this became a more convenient hardware option. To retain the easy integration with existing programmes, a device driver called a "software wedge" could be used, to emulate the keyboard-impersonating behaviour of the old "keyboard wedge" hardware.

In many cases a choice of USB interface types (HID, CDC) are provided. Some have Powered USB.

Wireless Networking

Modern handheld barcode readers are operated in wireless networks according to IEEE 802.11g (WLAN) or IEEE 802.15.1 (Bluetooth). However, such configuration limits the time of operation from battery or rechargeable battery and required recharging at least after a shift of operation.

Resolution

The scanner resolution is measured by the size of the dot of light emitted by the reader. If this dot of light is wider than any bar or space in the bar code, then it will overlap two elements (two spaces or two bars) and it may produce wrong output. On the other hand, if a too small dot of light is used, then it can misinterpret any spot on the bar code making the final output wrong.

The most commonly used dimension is 13 mils (0.33 mm), although some scanners can read codes with dimensions as small as 3 mils. Smaller bar codes must be printed at high resolution to be read accurately.

5

Desktop Digital Camera Scanner

One printer manufacturer (Lexmark) has introduced all-in-one printer which is provided with a desktop digital camera scanner that has 10 megapixel image sensors. For scanning a business card or a full 8.5x11 inch image, it takes not more than 3 seconds including the processing time.

Smartphone Scanner Apps

Cameras in smartphones have reached a resolution and quality that reasonable quality scans can be achieved by taking a photo with the phone and using a scanning app for post-processing (such as whitening the background of a page, correcting perspective distortion so that a document is output as a correct rectangle, conversion to black-and-white, etc.) Most smartphone platforms now have a range of scanner apps available (the largest range being for Apple iPhone). These apps can typically scan multiple page documents through the use of multiple camera exposures, and output them to a PDF document or as separate JPEG images.

Some smartphone scanning apps can also save documents directly to online storage locations such as Dropbox, Evernote, send via email or fax documents via email-to-fax gateways.

Quality

Scanners typically read red-green-blue color (RGB) data from the array. This data is then processed with some proprietary algorithm to correct for different exposure conditions, and sent to the computer via the device's input/output interface (usually USB, previous to which was SCSI or bidirectional parallel port in older units). Color depth varies depending on the scanning array characteristics, but is usually at least 24 bits.

High quality models have 48 bits or more color depth. Another qualifying parameter for a scanner is its *resolution*, measured in pixels per inch (ppi), sometimes more accurately referred to as Samples per inch (spi). Instead of using the scanner's true optical resolution, the only meaningful parameter, manufacturers like to refer to the *interpolated resolution*, which is much higher thanks to software interpolation. As of 2009, a high-end flatbed scanner can scan up to 5400 ppi and a good drum scanner has an optical resolution of 12,000 ppi.

Manufacturers often claim interpolated resolutions as high as 19,200 ppi; but such numbers carry little meaningful value, because the number of possible interpolated pixels is unlimited and doing so does not increase the level of captured detail.

The size of the file created increases with the square of the resolution; doubling the resolution quadruples the file size. A resolution must be chosen that is within the capabilities of the equipment, preserves sufficient detail, and does not produce a file of excessive size. The file size can be reduced for a given resolution by using "lossy" compression methods such as JPEG, at some cost in quality.

If the best possible quality is required lossless compression should be used; reduced-quality files of smaller size can be produced from such an image when required (e.g., image designed to be printed on a full page, and a much smaller file to be displayed as part of a fast-loading web page).

Purity can be diminished be scanner noise, optical flare, poor analog to digital conversion, color interpolation (from scanners that use CCD chips with bayer grids) and poor software. Drum scanners are said to produce the purist digital representations of the film, followed by high end film scanners that use the larger Kodak Tri-Linear sensors. The third important parameter for a scanner is its density range. A high-density range means that the scanner is able to reproduce shadow details and brightness details in one scan. By combining full-color imagery with 3D models, modern hand-held scanners are able to completely reproduce objects electronically. The addition of 3D color printers enables accurate miniaturization of these objects, with applications across many industries and professions.

Computer Connection

Scanning the document is only one part of the process. For the scanned image to be useful, it must be transferred from the scanner to an application running on the computer. There are two basic issues: (1) how the scanner is physically connected to the computer and (2) how the application retrieves the information from the scanner.

Direct Physical Connection to a Computer

The amount of data generated by a scanner can be very large: a 600 DPI 23 x 28 cm (9"x11") (slightly larger than A4 paper) uncompressed 24-bit image is about 100 megabytes of data which must be transferred and stored. Recent scanners can generate this volume of data in a matter of seconds, making a fast connection desirable. Scanners communicate to their host computer using one of the following physical interfaces, listing from slow to fast:

- Parallel port - Connecting through a parallel port is the slowest common transfer method. Early scanners had parallel port connections that could not transfer data faster than 70 kilobytes/second. The primary advantage of the parallel port connection was economic and user skill level: it avoided adding an interface card to the computer.

- GPIB - General Purpose Interface Bus. Certain drumscanners like the Howtek D4000 featured both a SCSI and GPIB interface. The latter conforms to the IEEE-488 standard, introduced in the mid '70's. The GPIB-interface has only been used by a few scanner manufactures, mostly serving the DOS/ Windows environment. For Apple Macintosh systems, National Instruments provided a NuBus GPIB interface card.
- Small Computer System Interface (SCSI), which is supported by most computers only via an additional SCSI interface card. Some SCSI scanners are supplied together with a dedicated SCSI card for a PC, although any SCSI controller can be used. During the evolution of the SCSI standard speeds increased, with backwards compatibility; a SCSI connection can transfer data at the highest speed which both the controller and the device support. SCSI has been largely replaced by USB and Firewire, one or both of which are directly supported by most computers, and which are easier to set up than SCSI.
- Universal Serial Bus (USB) scanners can transfer data quickly, and they are easier to use and cheaper than SCSI devices. The early USB 1.1 standard could transfer data at only 1.5 megabytes per second (slower than SCSI), but the later USB 2.0 standard can theoretically transfer up to 60 megabytes per second

(although everyday rates are much lower), resulting in faster operation.

- FireWire is an interface that is much faster than USB 1.1 and comparable to USB 2.0. FireWire speeds are 25, 50, and 100, 400 and 800 megabits per second (but a device may not support all speeds). Also known as: IEEE-1394.
- Proprietary interfaces were used on some early scanners that used a proprietary interface card rather than a standard interface.

Indirect (Network) Connection to a Computer

During the early nineties, professional flatbed scanners were targeted to professional users. Some vendors (like Umax) allowed a single scanner connected to a host computer to function as a scanner accessible by all users within a local computer network.

This proved to be very handy to e.g. publishers, print shops, etc. This functionality gradually disappeared after the mid-'90's as flatbed scanners became more affordable each year. However, as of 2000 and later, all-in-one multi-purpose devices targeted to serve both (small) offices and consumers usually combine a printer, scanner, copier and fax into a single apparatus available to a whole workgroup, providing each individual fax, scan, copy and print functionality.

Applications Programming Interface

A paint application such as GIMP or Adobe Photoshop must communicate with the scanner. There are many different scanners, and many of those scanners use different protocols. In order to simplify applications programming, some Applications Programming Interfaces ("API") were developed. The API presents a uniform interface to the scanner.

This means that the application does not need to know the specific details of the scanner in order to access it directly. For example, Adobe Photoshop supports the TWAIN standard; therefore in theory Photoshop can acquire an image from any scanner that also supports TWAIN. In practice, there are often problems with an application communicating with a scanner. Either the application or the scanner manufacturer (or both) may have faults in their implementation of the API. Typically, the API is implemented as a dynamically linked library. Each scanner manufacturer provides software that translates the API procedure calls into primitive commands that are issued to a hardware controller (such as the SCSI, USB, or FireWire controller). The manufacturer's part of the API is commonly called a device driver, but that designation is not strictly accurate: the API does not run in kernel mode and does not directly access the device. Rather the scanner API library translates application requests into hardware requests.

Common Scanner Software API Interfaces

SANE (Scanner Access Now Easy) is a free/open source API for accessing scanners. Originally developed for Unix and Linux operating systems, it has been ported to OS/2, Mac OS X, and Microsoft Windows. Unlike TWAIN, SANE does not handle the user interface. This allows batch scans and transparent network access without any special support from the device driver.

TWAIN is used by most scanners. Originally used for low-end and home-use equipment, it is now widely used for large-volume scanning.

ISIS (Image and Scanner Interface Specification) created by Pixel Translations, which still uses SCSI-II for performance reasons, is used by large, departmental-scale, machines.

WIA (Windows Image Acquisition) is an API provided by Microsoft.

Bundled Applications

Although no software beyond a scanning utility is a feature of any scanner, many scanners come bundled with software. Typically, in addition to the scanning utility, some type of image-editing application (such as Photoshop), and optical character recognition (OCR) software are supplied. OCR software converts graphical images of text into standard text that can be edited using common word-processing and text-editing software; accuracy is rarely perfect.

6

Charge-Coupled Device

A charge-coupled device (CCD) is a device for the movement of electrical charge, usually from within the device to an area where the charge can be manipulated, for example conversion into a digital value. This is achieved by “shifting” the signals between stages within the device one at a time. CCDs move charge between capacitive *bins* in the device, with the shift allowing for the transfer of charge between bins.

Often the device is integrated with an image sensor, such as a photoelectric device to produce the charge that is being read, thus making the CCD a major technology for digital imaging. Although CCDs are not the only technology to allow for light detection, CCDs are widely used in professional, medical, and scientific applications where high-quality image data is required.

History

The charge-coupled device was invented in 1969 at AT&T Bell Labs by Willard Boyle and George E. Smith. The lab was working on semiconductor bubble memory when Boyle and Smith conceived of the design of what they termed, in their notebook, "Charge 'Bubble' Devices".

A description of how the device could be used as a shift register and as a linear and area imaging devices was described in this first entry. The essence of the design was the ability to transfer charge along the surface of a semiconductor from one storage capacitor to the next. The concept was similar in principle to the bucket-brigade device (BBD), which was developed at Philips Research Labs during the late 1960's.

The initial paper describing the concept listed possible uses as a memory, a delay line, and an imaging device. The first experimental device demonstrating the principle was a row of closely spaced metal squares on an oxidized silicon surface electrically accessed by wire bonds.

The first working CCD made with integrated circuit technology was a simple 8-bit shift register. This device had input and output circuits and was used to demonstrate its use as a shift register and as a crude eight pixel linear imaging device. Development of the device progressed at a rapid rate. By 1971, Bell researchers Michael F. Tompsett et al. were able to capture images with simple linear devices.

Several companies, including Fairchild Semiconductor, RCA and Texas Instruments, picked up on the invention and began development programmes. Fairchild's effort, led by ex-Bell researcher Gil Amelio, was the first with commercial devices, and by 1974 had a linear 500-element device and a 2-D 100 x 100 pixel device.

The first KH-11 KENNAN reconnaissance satellite equipped with charge-coupled device array technology for imaging was launched in December 1976. Under the leadership of Kazuo Iwama, Sony also started a big development effort on CCDs involving a significant investment. Eventually, Sony managed to mass produce CCDs for their camcorders. Before this happened, Iwama died in August 1982; subsequently, a CCD chip was placed on his tombstone to acknowledge his contribution.

In January 2006, Boyle and Smith were awarded the National Academy of Engineering Charles Stark Draper Prize, and in 2009 they were awarded the Nobel Prize for Physics, for their work on the CCD.

Basics of Operation

In a CCD for capturing images, there is a photoactive region (an epitaxial layer of silicon), and a transmission region made out of a shift register (the CCD, properly speaking).

An image is projected through a lens onto the capacitor array (the photoactive region), causing each capacitor to accumulate an electric charge proportional to the light intensity at that location. A one-dimensional array, used in line-scan cameras, captures a single slice of the image, while a two-dimensional array, used in video and still cameras, captures a two-dimensional picture corresponding to the scene projected onto the focal plane of the sensor. Once the array has been exposed to the image, a control circuit causes each capacitor to transfer its contents to its neighbor (operating as a shift register).

The last capacitor in the array dumps its charge into a charge amplifier, which converts the charge into a voltage. By repeating this process, the controlling circuit converts the entire contents of the array in the semiconductor to a sequence of voltages.

In a digital device, these voltages are then sampled, digitized, and usually stored in memory; in an analog device (such as an analog video camera), they are processed into a continuous analog signal (e.g. by feeding the output of the charge amplifier into a low-pass filter) which is then processed and fed out to other circuits for transmission, recording, or other processing.

Detailed Physics of Operation

The photoactive region of the CCD is, generally, an epitaxial layer of silicon. It has a doping of p+ (Boron) and is grown upon a substrate material, often p++. In buried channel devices, the type of design utilized in most modern CCDs, certain areas of the surface of the silicon are ion implanted with phosphorus, giving them an n-doped designation. This region defines the channel in which the photogenerated charge packets will travel. The gate oxide, i.e. the capacitor dielectric, is grown on top of the epitaxial layer and substrate.

Later on in the process polysilicon gates are deposited by chemical vapor deposition, patterned with photolithography, and etched in such a way that the separately phased gates lie perpendicular to the channels. The channels are further defined by utilization of the LOCOS process to produce the channel stop region. Channel stops are thermally grown oxides that serve to isolate the charge packets in one column from those in another.

These channel stops are produced before the polysilicon gates are, as the LOCOS process utilizes a high temperature step that would destroy the gate material. The channels stops are parallel to, and exclusive of, the channel, or “charge carrying”, regions. Channel stops often have a p+ doped region underlying them, providing a further barrier to the electrons in the charge packets (this discussion of the physics of CCD devices assumes an electron transfer device, though hole transfer is possible). The clocking of the gates, alternately high and low, will forward and reverse bias to the diode that is provided by the buried channel (n-doped) and the epitaxial layer (p-doped).

This will cause the CCD to deplete, near the p-n junction and will collect and move the charge packets beneath the gates—and within the channels—of the device. CCD manufacturing and operation can be optimized for different uses. The above process describes a frame transfer CCD.

While CCDs may be manufactured on a heavily doped p++ wafer it is also possible to manufacture a device inside p-wells that have been placed on an n-wafer.

This second method, reportedly, reduces smear, dark current, and infrared and red response. This method of manufacture is used in the construction of interline transfer devices. Another version of CCD is called a peristaltic CCD. In a peristaltic charge-coupled device, the charge packet transfer operation is analogous to the peristaltic contraction and dilation of the digestive system.

The peristaltic CCD has an additional implant that keeps the charge away from the silicon/silicon dioxide interface and generates a large lateral electric field from one gate to the next. This provides an additional driving force to aid in transfer of the charge packets.

Architecture

The CCD image sensors can be implemented in several different architectures. The most common are full-frame, frame-transfer, and interline. The distinguishing characteristic of each of these architectures is their approach to the problem of shuttering. In a full-frame device, all of the image area is active, and there is no electronic shutter.

A mechanical shutter must be added to this type of sensor or the image smears as the device is clocked or read out. With a frame-transfer CCD, half of the silicon area is covered by an opaque mask (typically aluminum). The image can be quickly transferred from the image area to the opaque area or storage region with acceptable smear of a few percent.

That image can then be read out slowly from the storage region while a new image is integrating or exposing in the active area. Frame-transfer devices typically do not require a mechanical shutter and were a common architecture for early solid-state broadcast cameras. The downside to the frame-transfer architecture is that it requires twice the silicon real estate of an equivalent full-frame device; hence, it costs roughly twice as much. The interline architecture extends this concept one step further and masks every other column of the image sensor for storage.

In this device, only one pixel shift has to occur to transfer from image area to storage area; thus, shutter times can be less than a microsecond and smear is essentially eliminated. The advantage is not free, however, as the imaging area is now covered by opaque strips dropping the fill factor to approximately 50 percent and the effective quantum efficiency by an equivalent amount.

Modern designs have addressed this deleterious characteristic by adding microlenses on the surface of the device to direct light away from the opaque regions and on the active area. Microlenses can bring the fill factor back up to 90 percent or more depending on pixel size and the overall system's optical design. The choice of architecture comes down to one of utility. If the application cannot tolerate an expensive, failure-prone, power-intensive mechanical shutter, an interline device is the right choice. Consumer snap-shot cameras have used interline devices. On the other hand, for those applications that require the best possible light collection and issues of money, power and time are less important, the full-frame device is the right choice.

Astronomers tend to prefer full-frame devices. The frame-transfer falls in between and was a common choice before the fill-factor issue of interline devices was addressed. Today, frame-transfer is usually chosen when an interline architecture is not available, such as in a back-illuminated device. CCDs containing grids of pixels are used in digital cameras, optical scanners, and video cameras as light-sensing devices.

They commonly respond to 70 percent of the incident light (meaning a quantum efficiency of about 70 percent) making them far more efficient than photographic film, which captures only about 2 percent of the incident light. Most common types of CCDs are sensitive to near-infrared light, which allows infrared photography, night-vision devices, and zero lux (or near zero lux) video-recording/ photography.

For normal silicon-based detectors, the sensitivity is limited to 1.1 ìm. One other consequence of their sensitivity to infrared is that infrared from remote controls often appears on CCD-based digital cameras or camcorders if they do not have infrared blockers. Cooling reduces the array's dark current, improving the sensitivity of the CCD to low light intensities, even for ultraviolet and visible wavelengths. Professional observatories often cool their detectors with liquid nitrogen to reduce the dark current, and therefore the thermal noise, to negligible levels.

Use in Astronomy

Due to the high quantum efficiencies of CCDs, linearity of their outputs (one count for one photon of light), ease of use compared to photographic plates, and a variety of other reasons, CCDs were very rapidly adopted by astronomers for nearly all UV-to-infrared applications. Thermal noise and cosmic rays may alter the pixels in the CCD array.

To counter such effects, astronomers take several exposures with the CCD shutter closed and opened. The average of images taken with the shutter closed is necessary to lower the random noise.

Once developed, the *dark frame* average image is then subtracted from the open-shutter image to remove the dark current and other systematic defects (dead pixels, hot pixels, etc.) in the CCD. The Hubble Space Telescope, in particular, has a highly developed series of steps ("data reduction pipeline") to convert the raw CCD data to useful images. See the references for a more in-depth description of the steps in astronomical CCD image-data correction and processing. CCD cameras used in astrophotography often require sturdy mounts to cope with vibrations from wind and other sources, along with the tremendous weight of most imaging platforms.

To take long exposures of galaxies and nebulae, many astronomers use a technique known as auto-guiding. Most autoguiders use a second CCD chip to monitor deviations during imaging.

This chip can rapidly detect errors in tracking and command the mount motors to correct for them. An interesting unusual astronomical application of CCDs, called *drift-scanning*, uses a CCD to make a fixed telescope behave like a tracking telescope and follow the motion of the sky. The charges in the CCD are transferred and read in a direction parallel to the motion of the sky, and at the same speed. In this way, the telescope can image a larger region of the sky than its normal field of view.

The Sloan Digital Sky Survey is the most famous example of this, using the technique to produce the largest uniform survey of the sky yet accomplished. In addition to astronomy, CCDs are also used in laboratory analytical instrumentation such as monochromators, spectrometers, and N-slit laser interferometers.

Color Cameras

Digital color cameras generally use a Bayer mask over the CCD. Each square of four pixels has one filtered red, one blue, and two green (the human eye is more sensitive to green than either red or blue). The result of this is that luminance information is collected at every pixel, but the color resolution is lower than the luminance resolution.

Better color separation can be reached by three-CCD devices (3CCD) and a dichroic beam splitter prism, that splits the image into red, green and blue components. Each of the three CCDs is arranged to respond to a particular color. Most professional video camcorders, and some semi-professional camcorders, use this technique. Another advantage of 3CCD over a Bayer mask device is higher quantum efficiency (and therefore higher light sensitivity for a given aperture size). This is because in a 3CCD device most of the light entering the aperture is captured by a sensor, while a Bayer mask absorbs a high proportion (about 2/3) of the light falling on each CCD pixel.

For still scenes, for instance in microscopy, the resolution of a Bayer mask device can be enhanced by Microscanning technology. During the process of color co-site sampling, several frames of the scene are produced. Between acquisitions, the sensor is moved in pixel dimensions, so that each point in the visual field is acquired consecutively by elements of the mask that are sensitive to the red, green and blue components of its color. Eventually every pixel in the image has been scanned at least once in each color and the resolution of the three channels become equivalent (the resolutions of red and blue channels are quadrupled while the green channel is doubled).

Sensor Sizes

Sensors (CCD / CMOS) are often referred to with an inch fraction designation such as 1/1.8" or 2/3" called the optical format. This measurement actually originates back in the

1950s and the time of Vidicon tubes. Compact digital cameras and Digicams typically have much smaller sensors than a digital SLR and are thus less sensitive to light and inherently more prone to noise.

Frame Transfer CCD

A frame transfer CCD is a specialized CCD, often used in astronomy and some professional video cameras, designed for high exposure efficiency and correctness.

The normal functioning of a CCD, astronomical or otherwise, can be divided into two phases: exposure and readout. During the first phase, the CCD passively collects incoming photons, storing electrons in its cells. After the exposure time is passed, the cells are read out one line at a time. During the readout phase, cells are shifted down the entire area of the CCD. While they are shifted, they continue to collect light.

Thus, if the shifting is not fast enough, errors can result from light that falls on a cell holding charge during the transfer. These errors are referred to as "vertical smear" and cause a strong light source to create a vertical line above and below its exact location. In addition, the CCD cannot be used to collect light while it is being read out. Unfortunately, a faster shifting requires a faster readout, and a faster readout can introduce errors in the cell charge measurement, leading to a higher noise level. A frame transfer CCD solves both problems: it has a hidden, not normally used, area containing as many cells as the area exposed to light. Typically, this area is covered by a reflective material such as aluminium. When the exposure time is up, the cells are transferred very rapidly to the hidden area. Here, safe from any incoming light, cells can be read out at any speed one deems necessary to correctly measure the cells' charge. At the same time, the exposed part of the CCD is collecting light again, so no delay occurs between successive exposures.

The disadvantage of such a CCD is the higher cost: the cell area is basically doubled, and more complex control electronics are needed.

Intensified Charge-coupled Device

An intensified charge-coupled device (ICCD) is a CCD that is optically connected to an image intensifier that is mounted in front of the CCD.

An image intensifier includes three functional elements: a photocathode, a micro-channel plate (MCP) and a phosphor screen. These three elements are mounted one close behind the other in the mentioned sequence.

The photons which are coming from the light source fall onto the photocathode, thereby generating photoelectrons. The photoelectrons are accelerated towards the MCP by an electrical control voltage, applied between photocathode and MCP. The electrons are multiplied inside of the MCP and thereafter accelerated towards the phosphor screen. The phosphor screen finally converts the multiplied electrons back to photons which are guided to the CCD by a fiber optic or a lens. An image intensifier inherently includes a shutter functionality: If the control voltage between the photocathode and the MCP is reversed, the emitted photoelectrons are not accelerated towards the MCP but return to the photocathode. Thus, no electrons are multiplied and emitted by the MCP, no electrons are going to the phosphor screen and no light is emitted from the image intensifier. In this case no light falls onto the CCD, which means that the shutter is closed. The process of reversing the control voltage at the photocathode is called gating and therefore ICCDs are also called gateable CCD cameras.

Beside of the extremely high sensitivity of ICCD cameras, which enable single photon detection, the gateability is one of the major advantages of the ICCD over the EMCCD cameras. The highest performing ICCD cameras enable shutter times as short as 200 picoseconds. ICCD cameras are in general somewhat higher in price than EMCCD cameras because they need the expensive image intensifier. On the other hand EMCCD cameras need a cooling system to cool the EMCCD chip down to temperatures around 170 K. This cooling system adds additional costs to the EMCCD camera and often yields heavy condensation problems in the application. ICCDs are used in night vision devices and in a large variety of scientific applications.

7

Digital Image

A digital image is a representation of a two-dimensional image using ones and zeros (binary). Depending on whether or not the image resolution is fixed, it may be of vector or raster type. Without qualifications, the term "digital image" usually refers to raster images also called bitmap images.

Raster

Raster images have a finite set of digital values, called *picture elements* or pixels. The digital image contains a fixed number of rows and columns of pixels. Pixels are the smallest individual element in an image, holding quantized values that represent the brightness of a given color at any specific point.

Typically, the pixels are stored in computer memory as a raster image or raster map, a two-dimensional array of small integers. These values are often transmitted or stored in a compressed form. Raster images can be created by a variety of input devices and techniques, such as digital cameras, scanners, coordinate-measuring machines, seismographic profiling, airborne radar, and more.

They can also be synthesized from arbitrary non-image data, such as mathematical functions or three-dimensional geometric models; the latter being a major sub-area of computer graphics. The field of digital image processing is the study of algorithms for their transformation.

Raster Image Types

Each pixel of a raster image is typically associated to a specific 'position' in some 2D region, and has a *value* consisting of one or more quantities (samples) related to that position. Digital images can be classified according to the number and nature of those samples:

- binary
- grayscale
- color
- false-color
- multi-spectral
- thematic
- picture function

The term *digital image* is also applied to data associated to points scattered over a three-dimensional region, such as produced by tomographic equipment. In that case, each datum is called a voxel.

Raster File Formats

Most users come into contact with raster images through digital cameras. Some digital cameras give access to almost all the data captured by the camera, using a raw image format. *The Universal Photographic Imaging Guidelines (UPDIG)* suggests this format be used when possible since raw files produce the best quality images.

These file formats allow the photographer and the processing agent the greatest level of control and accuracy for output. Unfortunately, there is an issue of proprietary information [trade secrets] for some camera makers, but organizations are attempting to influence the manufacturers of them to avail these records publicly. An alternative may be a Digital Negative (DNG) a proprietary Adobe product described as "the public, archival format for digital camera raw data". Although this format is not yet universally accepted, support for the product is growing and archival confidence is building.

Vector

Vector images resulted from mathematical geometry (vector). In mathematical term, a vector consist of point that have direction and length. Often, both raster and vector elements will be combined in one image, for example, in the case of a billboard with text (vector) and photographs (raster).

Image Viewing

The user can utilize different programme to see the image. The GIF, JPEG, and PNG images can be seen simply using a web browser because they are the standard internet image formats. The SVG format is more and more used in the web and is a standard W3C format. Some viewers offer a slideshow utility, to see the images in a certain folder one after the other automatically.

History

The first picture to be scanned, stored, and recreated in digital pixels was displayed on the Standards Eastern Automatic Computer at NIST. The advancement of digital imagery continued in the early 1960s, alongside development of the space programme and in medical research. Projects at the Jet Propulsion Laboratory, MIT, Bell Labs and the University of Maryland, among others, used digital images to advance satellite imagery, wirephoto standards conversion, medical imaging, videophone technology, character recognition, and photo enhancement.

Rapid advances in digital imaging began with the introduction of microprocessors in the early 1970s, alongside progress in related storage and display technologies. The invention of computerised axial tomography (CAT scanning), using x-rays to produce a digital image of a “slice” through a three-dimensional object, was of great importance to medical diagnostics.

As well as origination of digital images, digitization of analog images allowed the enhancement and restoration of archaeolgical artefacts and began to be used in fields as diverse as nuclear medicine, astronomy, law enforcement, defence and industry. Advances in microprocessor technology paved the way for the development and marketing of charge-coupled devices (CCDs) for use in a wide range of image capture devices and gradually displaced the use of analog film and tape in photography and videography towards the end of the 20th century. The computing power necessary to process digital image capture also allowed computer-generated digital images to achieve a level of refinement close to photorealism.

Computer Animation

Computer animation is the process used for generating animated images by using computer graphics. The more general term computer generated imagery encompasses both static scenes and dynamic images, while *computer animation* only refers to moving images produced by exploiting the persistence of vision to make a series of images look animated.

Given that images last for about one twenty-fifth of a second on the retina fast image replacement creates the illusion of movement. Modern computer animation usually uses 3D computer graphics, although 2D computer graphics are still used for stylistic, low bandwidth, and faster real-time renderings. Sometimes the target of the animation is the computer itself, but sometimes the target is another medium, such as film. Computer animation is essentially a digital successor to the stop motion techniques used in traditional animation with 3D models and frame-by-frame animation of 2D illustrations. Computer generated animations are more controllable than other more physically based processes, such as constructing miniatures for effects shots or hiring extras for crowd scenes, and because it allows the creation of images that would not be feasible using any other technology. It can also allow a single graphic artist to produce such content without the use of actors, expensive set pieces, or props.

To create the illusion of movement, an image is displayed on the computer screen and repeatedly replaced by a new image that is similar to the previous image, but advanced slightly in the time domain (usually at a rate of 24 or 30 frames/second). This technique is identical to how the illusion of movement is achieved with television and motion pictures. For 3D animations, objects (models) are built on the computer monitor (modeled) and 3D figures are rigged with a virtual skeleton.

For 2D figure animations, separate objects (illustrations) and separate transparent layers are used, with or without a virtual skeleton. Then the limbs, eyes, mouth, clothes, etc. of the figure are moved by the animator on key frames. The differences in appearance between key frames are automatically calculated by the computer in a process known as tweening or morphing. Finally, the animation is rendered. For 3D animations, all frames must be rendered after modeling is complete. For 2D vector animations, the rendering process is the key frame illustration process, while tweened frames are rendered as needed. For pre-recorded presentations, the rendered frames are transferred to a different format or medium such as film or digital video. The frames may also be rendered in real time as they are presented to the end-user audience. Low bandwidth animations transmitted via the internet (e.g. 2D Flash, X3D) often use software on the end-users computer to render in real time as an alternative to streaming or pre-loaded high bandwidth animations.

A Simple Example

The screen is blanked to a background color, such as black. Then, a goat is drawn on the right of the screen. Next, the screen is blanked, but the goat is re-drawn or duplicated slightly to the left of its original position. This process is repeated, each time moving the goat a bit to the left. If this process is repeated fast enough, the goat will appear to move smoothly to the left.

This basic procedure is used for all moving pictures in films and television. The moving goat is an example of shifting the location of an object. More complex transformations of object properties such as size, shape, lighting effects often require calculations and computer rendering instead of simple re-drawing or duplication.

Explanation

To trick the eye and brain into thinking they are seeing a smoothly moving object, the pictures should be drawn at around 12 frames per second (frame/s) or faster (a frame is one complete image). With rates above 70 frames/s no improvement in realism or smoothness is perceivable due to the way the eye and brain process images. At rates below 12 frame/s most people can detect jerkiness associated with the drawing of new images which detracts from the illusion of realistic movement.

Conventional hand-drawn cartoon animation often uses 15 frames/s in order to save on the number of drawings needed, but this is usually accepted because of the stylized nature of cartoons. Because it produces more realistic imagery computer animation demands higher frame rates to reinforce this realism.

The reason no jerkiness is seen at higher speeds is due to "persistence of vision." From moment to moment, the eye and brain working together actually store whatever one looks at for a fraction of a second, and automatically "smooth out" minor jumps. Movie film seen in theaters in the United States runs at 24 frames per second, which is sufficient to create this illusion of continuous movement.

One of the earliest steps in the history of computer animation was the 1973 movie *Westworld,* a science-fiction film about a society in which robots live and work among humans, though the first use of 3D Wireframe imagery was in its sequel, *Futureworld* (1976), which featured a computer-generated hand and face created by then University of Utah graduate students Edwin Catmull and Fred Parke. Developments in CGI technologies are reported each year at SIGGRAPH, an annual conference on computer graphics and interactive techniques, attended each year by tens of thousands of computer professionals. Developers of computer games and 3D video cards strive to achieve the same visual quality on personal computers in real-time as is possible for CGI films and animation. With the rapid advancement of real-time rendering quality, artists began to use game engines to render non-interactive movies. This art form is called *machinima*.

Methods of Animating Virtual Characters

In most 3D computer animation systems, an animator creates a simplified representation of a character's anatomy, analogous to a skeleton or stick figure. The position of each segment of the skeletal model is defined by *animation variables*, or Avars. In human and animal characters, many parts of the skeletal model correspond to actual bones, but skeletal animation is also used to animate other things, such as facial features (though other methods for facial animation exist).

The character "Woody" in *Toy Story*, for example, uses 700 Avars, including 100 Avars in the face. The computer does not usually render the skeletal model directly (it is invisible), but uses the skeletal model to compute the exact position and orientation of the character, which is eventually rendered into an image. Thus by changing the values of Avars over time, the animator creates motion by making the character move from frame to frame. There are several methods for generating the Avar values to obtain realistic motion. Traditionally, animators manipulate the Avars directly. Rather than set Avars for every frame, they usually set Avars at strategic points (frames) in time and let the computer interpolate or 'tween' between them, a process called keyframing. Keyframing puts control in the hands of the animator, and has roots in hand-drawn traditional animation.

In contrast, a newer method called motion capture makes use of live action. When computer animation is driven by motion capture, a real performer acts out the scene as if they were the character to be animated. His or her motion is recorded to a computer using video cameras and markers, and that performance is then applied to the animated character. Each method has its advantages, and as of 2007, games and films are using either or both of these methods in productions.

Keyframe animation can produce motions that would be difficult or impossible to act out, while motion capture can reproduce the subtleties of a particular actor. For example, in the 2006 film Pirates of the Caribbean: Dead Man's Chest, actor Bill Nighy provided the performance for the character Davy Jones. Even though Nighy himself doesn't appear in the film, the movie benefited from his performance by recording the nuances of his body language, posture, facial expressions, etc. Thus motion capture is appropriate in situations where believable, realistic behaviour and action is required, but the types of characters required exceed what can be done through conventional costuming.

Creating Characters and Objects on a Computer

3D computer animation combines 3D models of objects and programmed or hand "keyframed" movement. Models are constructed out of geometrical vertices, faces, and edges in a 3D coordinate system. Objects are sculpted much like real clay or plaster, working from general forms to specific details with various sculpting tools. A bone/joint animation system is set up to deform the CGI model (e.g., to make a humanoid model walk). In a process called rigging, the virtual marionette is given various controllers and handles for controlling movement. Animation data can be created using motion capture, or keyframing by a human animator, or a combination of the two. 3D models rigged for animation may contain thousands of control points - for example, the character "Woody" in Pixar's movie *Toy Story*, uses 700 specialized animation controllers. Rhythm and Hues Studios labored for two years to create Aslan in the movie *The Chronicles of Narnia: The Lion, the Witch and the Wardrobe* which had about 1851 controllers, 742 in just the face alone. In the 2004 film *The Day After Tomorrow*, designers had to design forces of extreme weather with the help of video references and accurate meteorological facts. For the 2005 remake of *King Kong*, actor Andy Serkis was used to help designers pinpoint the gorilla's prime location in the shots and used his expressions to model "human" characteristics onto the creature. Serkis had earlier provided the voice and performance for Gollum in J. R. R. Tolkien's *The Lord of the Rings* trilogy.

Computer Animation Development Equipment

Computer animation can be created with a computer and animation software. Some impressive animation can be achieved even with basic programmes; however, the rendering can take a lot of time on an ordinary home computer. Because of this, video game animators tend to use low resolution, low polygon count renders, such that the graphics can be rendered in real time on a home computer. Photorealistic animation would be impractical in this context. Professional animators of movies, television, and video sequences on computer games make photorealistic animation with high detail.

This level of quality for movie animation would take tens to hundreds of years to create on a home computer. Many powerful workstation computers are used instead. Graphics workstation computers use two to four processors, and thus are a lot more powerful than a home computer, and are specialized for rendering. A large number of workstations (known as a render farm) are networked together to effectively act as a giant computer.

The result is a computer-animated movie that can be completed in about one to five years (this process is not comprised solely of rendering, however). A workstation typically costs $2,000 to $16,000, with the more expensive stations being able to render much faster, due to the more technologically advanced hardware that they contain. Pixar's Renderman is rendering software which is widely used as the movie animation industry standard, in competition with Mental Ray. It can be bought at the official Pixar website for about $3,500. It will work on Linux, Mac OS X, and Microsoft Windows based graphics workstations along with an animation programme such as Maya and Softimage XSI. Professionals also use digital movie cameras, motion capture or performance capture, bluescreens, film editing software, props, and other tools for movie animation.

Modeling Human Faces

The modeling of human facial features is both one of the most challenging and sought after elements in computer-generated imagery. Computer facial animation is a highly complex field where models typically include a very large number of animation variables.

Historically speaking, the first SIGGRAPH tutorials on *State of the art in Facial Animation* in 1989 and 1990 proved to be a turning point in the field by bringing together and consolidating multiple research elements, and sparked interest among a number of researchers. The Facial Action Coding System (with 46 *action units* such as "lip bite" or "squint") which had been developed in 1976 became a popular basis for many systems.

As early as 2001 MPEG-4 included 68 facial animation parameters for lips, jaws, etc., and the field has made significant progress since then and the use of facial microexpression has increased. In some cases, an affective space such as the PAD emotional state model can be used to assign specific emotions to the faces of avatars. In this approach the PAD model is used as a high level emotional space, and the lower level space is the MPEG-4 Facial Animation Parameters (FAP). A mid-level Partial Expression

Parameters (PEP) space is then used to in a two level structure: the PAD-PEP mapping and the PEP-FAP translation model.

The Future

One open challenge in computer animation is a photorealistic animation of humans. Currently, most computer-animated movies show animal characters (*A Bug's Life*, *Finding Nemo*, *Ratatouille*, *Ice Age*, *Over the Hedge*), fantasy characters (*Monsters Inc.*, *Shrek*, *Teenage Mutant Ninja Turtles 4*, *Monsters vs. Aliens*), anthropomorphic machines (*Cars*, *WALL-E*, *Robots*) or cartoon-like humans (*The Incredibles*, *Despicable Me*, *Up*).

The movie *Final Fantasy: The Spirits Within* is often cited as the first computer-generated movie to attempt to show realistic-looking humans. However, due to the enormous complexity of the human body, human motion, and human biomechanics, realistic simulation of humans remains largely an open problem. Another problem is the distasteful psychological response to viewing nearly perfect animation of humans, known as "the uncanny valley." It is one of the "holy grails" of computer animation.

Eventually, the goal is to create software where the animator can generate a movie sequence showing a photorealistic human character, undergoing physically-plausible motion, together with clothes, photorealistic hair, a complicated natural background, and possibly interacting with other simulated human characters. This could be done in a way that the viewer is no longer able to tell if a particular movie sequence is computer-generated, or created using real actors in front of movie cameras.

Complete human realism is not likely to happen very soon, but when it does it may have major repercussions for the film industry. For the moment it looks like three dimensional computer animation can be divided into two main directions; photorealistic and non-photorealistic rendering. Photorealistic computer animation can itself be divided into two subcategories; real photorealism (where performance capture is used in the creation of the virtual human characters) and stylized photorealism.

Real photorealism is what Final Fantasy tried to achieve and will in the future most likely have the ability to give us live action fantasy features as *The Dark Crystal* without having to use advanced puppetry and animatronics, while *Antz* is an example on stylistic photorealism (in the future stylized photorealism will be able to replace traditional stop motion animation as in *Corpse Bride*). None of these mentioned are perfected as of yet, but the progress continues. The non-photorealistic/cartoonish direction is more like an extension of traditional animation, an attempt to make the animation look like a three dimensional version of a cartoon, still using and perfecting the main principles of animation articulated by the Nine Old Men, such as squash and stretch. While a single frame from a photorealistic computer-animated feature will look like a photo if done right, a single frame vector from a cartoonish computer-animated feature will look like a painting (not to be confused with cel shading, which produces an even simpler look).

Detailed Examples and Pseudocode

In 2D computer animation, moving objects are often referred to as "sprites." A sprite is an image that has a location associated with it. The location of the sprite is changed slightly, between each displayed frame, to make the sprite appear to move. The following pseudocode makes a sprite move from left to right:

```
var int x := 0, y := screenHeight / 2;
while x < screenWidth
drawBackground()
drawSpriteAtXY (x, y) // draw on top of the background
x := x + 5 // move to the right
```

Computer animation uses different techniques to produce animations. Most frequently, sophisticated mathematics is used to manipulate complex three dimensional polygons, apply "textures", lighting and other effects to the polygons and finally rendering the complete image. A sophisticated graphical user interface may be used to create the animation and arrange its choreography.

Another technique called constructive solid geometry defines objects by conducting boolean operations on regular shapes, and has the advantage that animations may be accurately produced at any resolution. Let's step through the rendering of a simple image of a room with flat wood walls with a grey pyramid in the center of the room. The pyramid will have a spotlight shining on it. Each wall, the floor and the ceiling is a simple polygon, in this case, a rectangle. Each corner of the rectangles is defined by three values referred to as X, Y and Z. X is how far left and right the point is. Y is how far up and down the point is, and Z is far in and out of the screen the point is. The wall nearest us would be defined by four points: (in the order x, y, z). Below is a representation of how the wall is defined

(0, 10, 0)   (10, 10, 0)

(0,0,0)   (10, 0, 0)

The far wall would be:

(0, 10, 20)   (10, 10, 20)

(0, 0, 20)   (10, 0, 20)

The pyramid is made up of five polygons: the rectangular base, and four triangular sides. To draw this image the computer uses math to calculate how to project this image, defined by three dimensional data, onto a two dimensional computer screen. First we must also define where our view point is, that is, from what vantage point will the scene be drawn. Our view point is inside the room a bit above the floor, directly in front of the pyramid.

First the computer will calculate which polygons are visible. The near wall will not be displayed at all, as it is behind our view point. The far side of the pyramid will also not be drawn as it is hidden by the front of the pyramid. Next each point is perspective projected onto the screen. The portions of the walls 'furthest' from the view point will appear to be shorter than the nearer areas due to perspective. To make the walls look like wood, a wood pattern, called a texture, will be drawn on them. To accomplish this, a technique called "texture mapping" is often used. A small drawing of wood that can be repeatedly drawn in a matching tiled pattern (like wallpaper) is stretched and drawn onto the walls' final shape.

The pyramid is solid grey so its surfaces can just be rendered as grey. But we also have a spotlight. Where its light falls we lighten colors, where objects blocks the light we darken colors. Next we render the complete scene on the computer screen. If the numbers describing the position of the pyramid were changed and this process repeated, the pyramid would appear to move.

Movies

CGI short films have been produced as independent animation since 1976, though the popularity of computer animation (especially in the field of special effects) skyrocketed during the modern era of U.S. animation. The first completely computer-generated television series was *ReBoot*, in 1994, and the first completely computer-generated animated movie was *Toy Story* (1995). See List of computer-animated films for more.

Amateur Animation

The popularity of websites which allows members to upload their own movies for others to view has created a growing community of amateur computer animators. With utilities and programmes often included free with modern operating systems, many users can make their own animated movies and shorts. Several free and open source animation software applications exist as well. A popular amateur approach to animation is via the animated GIF format, which can be uploaded and seen on the web easily.

Computer-Generated Imagery

Computer-generated imagery (CGI) is the application of the field of computer graphics or, more specifically, 3D computer graphics to special effects in art, films, television programmes, commercials, simulators and simulation generally, and printed media. The visual scenes may be either dynamic or static.

The term *computer animation* refers to dynamic CGI rendered as a movie. The term *virtual world* refers to agent-based, interactive environments. 3D computer graphics software is used to make computer-generated imagery for movies, etc. Recent availability of CGI software and increased computer speeds have allowed individual artists and small companies to produce professional grade films, games, and fine art from their home computers. This has brought about an Internet subculture with its own set of global celebrities, clichés, and technical vocabulary.

Static Images and Landscapes

Not only do animated images form part of computer-generated imagery, natural looking landscapes, such as fractal landscapes are also generated via computer algorithms. A simple way to generate fractal surfaces is to use an extension of the triangular mesh method, relying on the construction of some special case of a de Rham curve, e.g. midpoint displacement. For instance, the algorithm may start with a large triangle, then recursively zoom in by dividing it into 4 smaller Sierpinski triangles, then interpolate the height of each point from its nearest neighbors. The creation of a Brownian surface may be achieved not only by adding noise as new nodes are created, but by adding additional noise at multiple levels of the mesh.

Thus a topographical map with varying levels of height can be created using relatively straightforward fractal algorithms. Some typical, and easy to programme fractals used in CGI are the *plasma fractal* and the more dramatic *fault fractal*. A large number of specific techniques have been researched and developed to produce highly focused computer-generated effects, e.g. the use of specific models to represent the chemical weathering of stones to model erosion and produce an "aged appearance" for a given stone-based surface.

Architectural Scenes

Modern architects use services from computer graphic firms to create 3-dimensional models for both customers and builders. These computer generated models can be more accurate than traditional drawings. Architectural animation (which provides animated movies of buildings, rather than interactive images) can also be used to see the possible relationship a building will have in relation to the environment and its surrounding buildings. The rendering of architectural spaces without the use of paper and pencil tools is now a widely accepted practice with a number of computer-assisted architectural design systems. Architectural modeling tools allow an architect to visualize a space and perform "walk throughs" in an interactive manner, thus providing "interactive environments" both at the urban and building levels.

Specific applications in architecture not only include the specification of building structures such as walls and windows, and walk-throughs, but the effects of light and how sunlight will affect a specific design at different times of the day. Architectural modeling tools have now become increasingly internet-based. However, the quality of internet-based systems still lags those of sophisticated inhouse modeling systems. In some applications, computer-generated images are used to "reverse engineer" historical buildings. For instance, a computer-generated reconstruction of the monastery at Georgenthal in Germany was derived from the ruins of the monastery, yet provides the viewer with a "look and feel" of what the building would have looked like in its day.

Anatomical Models

Computer generated models used in skeletal animation are not always anatomically correct, however, organizations such as the Scientific Computing and Imaging Institute have developed anatomically correct computer-based models. Computer generated anatomical models can be used both for instructional and operational purposes. To date, a large body of artist produced medical images continue to be used by medical students, such as images by Frank Netter, e.g. Cardiac images. However, a number of online anatomical models are becoming available. A single patient X-ray is not a computer generated image, even in the case of digitized x-rays. However, in applications which involve CT scans a three dimensional model is automatically produced from a large number of single slice x-rays, producing “computer generated image”.

Applications involving magnetic resonance imaging also bring together a number of “snapshots” (in this case via magnetic pulses) to produce a composite, internal image. In modern medical applications, patient specific models are constructed in “computer assisted surgery”. For instance, in total knee replacement, the construction of a detailed patient specific model can be used to carefully plan the surgery.

These three dimensional models are usually extracted from multiple CT scans of the appropriate parts of the patient’s own anatomy. Such models can also be used for planning aortic valve implantations, one of the common procedures for treating heart disease.

Given that the shape, diameter and position of the coronary openings can vary greatly from patient to patient, the extraction (from CT scans) of a model that closely resembles a patient’s valve anatomy can be highly beneficial in planning the procedure.

Generating Cloth and Skin Images

Models of cloth generally fall into three groups: the geometric-mechanical structure at yarn crossings, secondly the mechanics of continuous elastic sheets and thirdly the geometric macroscopic features of cloth. To date, making the clothing of a digital character automatically fold in a natural way remains a challenge for many animators.

In addition to their use in film, advertising and other modes of public display, computer generated images of clothing are now routinely used by top fashion design firms. The challenge in rendering human skin images involves three levels of realism: *photo realism* in that it should look like real skin at the static level; *physical realism* in that it should closely simulate real skin's movements and *functional realism* in that it should act like real skin in response to actions.

Interactive Simulation and Visualization

Interactive visualization is a general term that applies to the rendering of data that may vary dynamically and allowing a user to view the data from multiple perspectives. The applications areas may vary significantly, ranging from the visualization of the flow patterns in fluid dynamics to specific computer aided design applications.

The data rendered may correspond to specific visual scenes that change as the user interacts with the system, e.g. simulators such as flight simulators make extensive use of CGI techniques for representing the world. At the abstract level an interactive visualization process involves a '*data pipeline* in which the raw data is managed and filtered to a form that makes it suitable for rendering. This is often called the "visualization data".

The visualization data is then mapped to a "visualization representation" that can be fed to a rendering system. This is usually called a "renderable representation". This representation is then rendered as a displayable image. As the user interacts with the system, e.g. by using joystick controls to change their position within the virtual world, the raw data is fed through the pipeline to create a new rendered image, often making real-time computational efficiency a key consideration in such applications.

Computer Animation

While computer generated images of landscapes may be static, the term computer animation only applies to dynamic images that resemble a movie. However, in general the term computer animation refers to dynamic images that do not allow user interaction, and the term virtual world is used for the interactive animated environments. Computer animation is essentially a digital successor to the art of stop motion animation of 3D models and frame-by-frame animation of 2D illustrations.

Computer generated animations are more controllable than other more physically based processes, such as constructing miniatures for effects shots or hiring extras for crowd scenes, and because it allows the creation of images that would not be feasible using any other technology. It can also allow a single graphic artist to produce such content without the use of actors, expensive set pieces, or props.

To create the illusion of movement, an image is displayed on the computer screen and repeatedly replaced by a new image that is similar to the previous image, but advanced slightly in the time domain (usually at a rate of 24 or 30 frames/second).

This technique is identical to how the illusion of movement is achieved with television and motion pictures.

Virtual Worlds

A *virtual world* is a simulated environment, which allows user to interact with animated characters, or interact with other users through the use of animated characters known as avatars.

Virtual worlds are intended for its users to inhabit and interact, and the term today has become largely synonymous with interactive 3D virtual environments, where the users take the form of avatars visible to others graphically.

These avatars are usually depicted as textual, two-dimensional, or three-dimensional graphical representations, although other forms are possible (auditory and touch sensations for example).

Some, but not all, virtual worlds allow for multiple users.

3D Modeling

In 3D computer graphics, 3D modeling (also known as meshing) is the process of developing a mathematical representation of any three-dimensional surface of object (either inanimate or living) via specialized software. The product is called a 3D model.

It can be displayed as a two-dimensional image through a process called *3D rendering* or used in a computer simulation of physical phenomena. The model can also be physically created using 3D Printing devices. Models may be created automatically or manually. The manual modeling process of preparing geometric data for 3D computer graphics is similar to plastic arts such as sculpting.

Models

3D models represent a 3D object using a collection of points in 3D space, connected by various geometric entities such as triangles, lines, curved surfaces, etc. Being a collection of data (points and other information), 3D models can be created by hand, algorithmically (procedural modeling), or scanned. 3D models are widely used anywhere in 3D graphics.

Actually, their use predates the widespread use of 3D graphics on personal computers. Many computer games used pre-rendered images of 3D models as sprites before computers could render them in real-time. Today, 3D models are used in a wide variety of fields. The medical industry uses detailed models of organs. The movie industry uses them as characters and objects for animated and real-life motion pictures. The video game industry uses them as assets for computer and video games.

The science sector uses them as highly detailed models of chemical compounds. The architecture industry uses them to demonstrate proposed buildings and landscapes through Software Architectural Models. The engineering community uses them as designs of new devices, vehicles and structures as well as a host of other uses. In recent decades the earth science community has started to construct 3D geological models as a standard practice.

Representation

Almost all 3D models can be divided into two categories.

- Solid - These models define the volume of the object they represent (like a rock). These are more realistic, but more difficult to build. Solid models are mostly used for nonvisual simulations such as medical and engineering simulations, for CAD and specialized visual applications such as ray tracing and constructive solid geometry
- Shell/boundary - these models represent the surface, e.g. the boundary of the object, not its volume (like an infinitesimally thin eggshell). These are easier to work with than solid models. Almost all visual models used in games and film are shell models.

Because the appearance of an object depends largely on the exterior of the object, boundary representations are common in computer graphics. Two dimensional surfaces are a good analogy for the objects used in graphics, though quite often these objects are non-manifold. Since surfaces are not finite, a discrete digital approximation is required: polygonal meshes (and to a lesser extent subdivision surfaces) are by far the most common representation, although point-based representations have been gaining some popularity in recent years.

Level sets are a useful representation for deforming surfaces which undergo many topological changes such as fluids. The process of transforming representations of objects, such as the middle point coordinate of a sphere and a point on its circumference into a polygon representation of a sphere, is called tessellation.

This step is used in polygon-based rendering, where objects are broken down from abstract representations ("primitives") such as spheres, cones etc., to so-called *meshes*, which are nets of interconnected triangles. Meshes of triangles (instead of e.g. squares) are popular as they have proven to be easy to render using scanline rendering. Polygon representations are not used in all rendering techniques, and in these cases the tessellation step is not included in the transition from abstract representation to rendered scene.

Modeling Processes

There are five popular ways to represent a model:

- Polygonal modeling - Points in 3D space, called vertices, are connected by line segments to form a polygonal mesh. Used, for example, by Blender. The vast majority of 3D models today are built as textured polygonal models, because they are flexible and because computers can render them so quickly. However, polygons are planar and can only approximate curved surfaces using many polygons.

- NURBS modeling - NURBS Surfaces are defined by spline curves, which are influenced by weighted control points. The curve follows (but does not necessarily interpolate) the points. Increasing the weight for a point will pull the curve closer to that point. NURBS are truly smooth surfaces, not approximations using small flat surfaces, and so are particularly suitable for organic modeling. Maya, Rhino 3d and solidThinking are the most well-known commercial programmes which use NURBS natively.
- Splines & Patches modeling - Like NURBS, Splines and Patches depend on curved lines to define the visible surface. Patches fall somewhere between NURBS and polygons in terms of flexibility and ease of use.
- Primitives modeling - This procedure takes geometric primitives like balls, cylinders, cones or cubes as building blocks for more complex models. Benefits are quick and easy construction and that the forms are mathematically defined and thus absolutely precise, also the definition language can be much simpler. Primitives modeling is well suited for technical

applications and less for organic shapes. Some 3D software can directly render from primitives (like POV-Ray), others use primitives only for modeling and convert them to meshes for further operations and rendering.

- Sculpt modeling - Still fairly new method of modeling 3D sculpting has become very popular in the few short years it has been around. There are 2 types of this currently, Displacement which is the most widely used among applications at this moment, and volumetric. Displacement uses a dense model (often generated by Subdivision surfaces of a polygon control mesh) and stores new locations for the vertex positions through use of a 32bit image map that stores the adjusted locations. Volumetric which is based loosely on Voxels has similar capabilities as displacement but does not suffer from polygon stretching when there are not enough polygons in a region to achieve a deformation. Both of these methods allow for very artistic exploration as the model will have a new topology created over it once the models form and possibly details have been sculpted. The new mesh will usually have the original high resolution mesh information transferred into displacement data or normal map data if for a game engine.

The modeling stage consists of shaping individual objects that are later used in the scene. There are a number of modeling techniques, including:

- constructive solid geometry
- implicit surfaces
- subdivision surfaces

Modeling can be performed by means of a dedicated programme (e.g., form•Z, Maya, 3DS Max, Blender, Lightwave, Modo, solidThinking) or an application component (Shaper, Lofter in 3DS Max) or some scene description language (as in POV-Ray). In some cases, there is no strict distinction between these phases; in such cases modeling is just part of the scene creation process (this is the case, for example, with Caligari trueSpace and Realsoft 3D).

Complex materials such as blowing sand, clouds, and liquid sprays are modeled with particle systems, and are a mass of 3D coordinates which have either points, polygons, texture splats, or sprites assigned to them. Sculpt

Scene Setup

Scene setup involves arranging virtual objects, lights, cameras and other entities on a scene which will later be used to produce a still image or an animation. Lighting is an important aspect of scene setup. As is the case in real-world scene arrangement, lighting is a significant contributing factor to the resulting aesthetic and visual quality of the finished work.

As such, it can be a difficult art to master. Lighting effects can contribute greatly to the mood and emotional response effected by a scene, a fact which is well-known to photographers and theatrical lighting technicians. It is usually desirable to add color to a model's surface in a user controlled way prior to rendering. Most 3D modeling software allows the user to color the model's vertices, and that color is then interpolated across the model's surface during rendering. This is often how models are colored by the modeling software while the model is being created.

The most common method of adding color information to a 3D model is by applying a 2D texture image to the model's surface through a process called texture mapping. Texture images are no different than any other digital image, but during the texture mapping process, special pieces of information (called texture coordinates or UV coordinates) are added to the model that indicate which parts of the texture image map to which parts of the 3D model's surface. Textures allow 3D models to look significantly more detailed and realistic than they would otherwise.

Other effects, beyond texturing and lighting, can be done to 3D models to add to their realism. For example, the surface normals can be tweaked to affect how they are lit, certain surfaces can have bump mapping applied and any other number of 3D rendering tricks can be applied. 3D models are often animated for some uses. They can sometimes be animated from within the 3D modeler that created them or else exported to another programme.

If used for animation, this phase usually makes use of a technique called "keyframing", which facilitates creation of complicated movement in the scene. With the aid of keyframing, one needs only to choose where an object stops or changes its direction of movement, rotation, or scale, between which states in every frame are interpolated. These moments of change are known as keyframes. Often extra data is added to the model to make it easier to animate. For example, some 3D models of humans and animals have entire bone systems so they will look realistic when they move and can be manipulated via joints and bones, in a process known as skeletal animation.

Compared to 2D Methods

3D photorealistic effects are often achieved without wireframe modeling and are sometimes indistinguishable in the final form. Some graphic art software includes filters that can be applied to 2D vector graphics or 2D raster graphics on transparent layers. Advantages of wireframe 3D modeling over exclusively 2D methods include:

- *Flexibility,* ability to change angles or animate images with quicker rendering of the changes;
- *Ease of rendering,* automatic calculation and rendering photorealistic effects rather than mentally visualizing or estimating;
- *Accurate photorealism,* less chance of human error in misplacing, overdoing, or forgetting to include a visual effect.

Disadvantages compare to 2D photorealistic rendering may include a software learning curve and difficulty achieving certain photorealistic effects. Some photorealistic effects may be achieved with special rendering filters included in the 3D modeling software. For the best of both worlds, some artists use a combination of 3D modeling followed by editing the 2D computer-rendered images from the 3D model.

3D Model Market

3CT (3D Catalog Technology) has revolutionized the 3D model market by offering quality 3D model libraries free of charge for professionals using various CAD programmes. Some believe that this uprising technology is gradually eroding the traditional "buy and sell" or "object for object exchange" markets although the quality of the products do not match those sold on specialized 3d marketplaces. A large market for 3D models (as well as 3D-related content, such as textures, scripts, etc.) still exists - either for individual models or large collections. Online marketplaces for 3D content allow individual artists to sell content that they have created. Often, the artists' goal is to get additional value out of assets they have previously created for projects. By doing so, artists can earn more money out of their old content, and companies can save money by buying pre-made models instead of paying an employee to create one from scratch. These marketplaces typically split the sale between themselves and the artist that created the asset, often in a roughly 50-50 split. In most cases, the artist retains ownership of the 3d model; the customer only buys the right to use and present the model.

Human Models

The first widely available commercial application of human Virtual Models appeared in 1998 on the Lands' End web site. The human Virtual Models were created by the company My Virtual Model Inc. and enabled users to create a model of themselves and try on 3D clothing. There are several modern programmes that allow for the creation of virtual human models (Poser being one example).

8

Computer Scanography

Scanography, also spelled scannography more commonly referred to as scanner photography, is the process of capturing digitized images of objects for the purpose of creating printable art using a flatbed "photo" scanner with a CCD (charge-coupled device) array capturing device. The term scanography formerly referred to *medical* scanning, but is unrelated in either purpose or technique. Fine art scanography differs from traditional document scanning by using atypical objects, often three dimensional, as well as from photography, due to the nature of the scanner's operation.

Image Scanner

In computing, an image scanner—often abbreviated to just scanner— is a device that optically scans images, printed text, handwriting, or an object, and converts it to a digital image. Common examples found in offices are variations of the *desktop (or flatbed) scanner* where the document is placed on a glass window for scanning. *Hand-held scanners*, where the device is moved by hand, have evolved from text scanning "wands" to 3D scanners used for industrial design, reverse engineering, test and measurement, orthotics, gaming and other applications. Mechanically driven scanners that move the document are typically used for large-format documents, where a flatbed design would be impractical.

Modern scanners typically use a charge-coupled device (CCD) or a Contact Image Sensor (CIS) as the image sensor, whereas older *drum scanners* use a photomultiplier tube as the image sensor. A *rotary scanner,* used for high-speed document scanning, is another type of drum scanner, using a CCD array instead of a photomultiplier. Other types of scanners are planetary scanners, which take photographs of books and documents, and 3D scanners, for producing three-dimensional models of objects.

Another category of scanner is digital camera scanners, which are based on the concept of reprographic cameras. Due to increasing resolution and new features such as anti-shake, digital cameras have become an attractive alternative to regular scanners. While still having disadvantages compared to traditional scanners (such as distortion, reflections, shadows, low contrast), digital cameras offer advantages such as speed, portability and gentle digitizing of thick documents without damaging the book spine. New scanning technologies are combining 3D scanners with digital cameras to create full-color, photo-realistic 3D models of objects.

In the biomedical research area, detection devices for DNA microarrays are called scanners as well. These scanners are high-resolution systems (up to 1 μm/ pixel), similar to microscopes. The detection is done via CCD or a photomultiplier tube (PMT).

Historical Precedent

Modern scanners may be considered the successors of early telephotography and fax input devices, consisting of a rotating drum with a single photodetector at a standard speed of 60 or 120 rpm (later models up to 240 rpm). They send a linear analog AM signal through standard telephone voice lines to receptors, which synchronously print the proportional intensity on special paper. This system was in use in press from the 1920s to the mid-1990s. Color photos were sent as three separated RGB filtered images consecutively, but only for special events due to transmission costs.

Types

Drum

Drum scanners capture image information with photomultiplier tubes (PMT), rather than the charge-coupled device (CCD) arrays found in flatbed scanners and inexpensive film scanners. Reflective and transmissive originals are mounted on an acrylic cylinder, the scanner drum, which rotates at high speed while it passes the object being scanned in front of precision optics that deliver image information to the PMTs. Most modern color drum scanners use three matched PMTs, which read red, blue, and green light, respectively. Light from the original artwork is split into separate red, blue, and green beams in the optical bench of the scanner.

The drum scanner gets its name from the clear acrylic cylinder, the drum, on which the original artwork is mounted for scanning. Depending on size, it is possible to mount originals up to 11"x17", but maximum size varies by manufacturer. One of the unique features of drum scanners is the ability to control sample area and aperture size independently. The sample size is the area that the scanner encoder reads to create an individual pixel. The aperture is the actual opening that allows light into the optical bench of the scanner. The ability to control aperture and sample size separately is particularly useful for smoothing film grain when scanning black-and white and color negative originals.

While drum scanners are capable of scanning both reflective and transmissive artwork, a good-quality flatbed scanner can produce good scans from reflective artwork. As a result, drum scanners are rarely used to scan prints now that high-quality, inexpensive flatbed scanners are readily available. Film, however, is where drum scanners continue to be the tool of choice for high-end applications. Because film can be wet-mounted to the scanner drum and because of the exceptional sensitivity of the PMTs, drum scanners are capable of capturing very subtle details in film originals.

Only a few companies continue to manufacture drum scanners. While prices of both new and used units have come down over the last decade, they still require a considerable monetary investment when compared to CCD flatbed and film scanners. However, drum scanners remain in demand due to their capacity to produce scans that are superior in resolution, color gradation, and value structure. Also, because drum scanners are capable of resolutions up to 12,000 PPI, their use is generally recommended when a scanned image is going to be enlarged.

In most graphic-arts operations, very-high-quality flatbed scanners have replaced drum scanners, being both less expensive and faster. However, drum scanners continue to be used in high-end applications, such as museum-quality archiving of photographs and print production of high-quality books and magazine advertisements. In addition, due to the greater availability of pre-owned units, many fine-art photographers are acquiring drum scanners, which has created a new niche market for the machines.

The first image scanner developed for use with a computer, was a drum scanner. It was built in 1957 at the US National Bureau of Standards by a team led by Russell A. Kirsch. The first image ever scanned on this machine was a 5 cm square photograph of Kirsch's then-three-month-old son, Walden. The black and white image had a resolution of 176 pixels on a side.

Flatbed

CCD Scanner

A flatbed scanner is usually composed of a glass pane (or platen), under which there is a bright light (often xenon or cold cathode fluorescent) which illuminates the pane, and a moving optical array in CCD scanning. CCD-type scanners typically contain three rows (arrays) of sensors with red, green, and blue filters.

CIS Scanner

CIS scanning consists of a moving set of red, green and blue LEDs strobed for illumination and a connected monochromatic photodiode array for light collection. Images to be scanned are placed face down on the glass, an opaque cover is lowered over it to exclude ambient light, and the sensor array and light source move across the pane, reading the entire area. An image is therefore visible to the detector only because of the light it reflects. Transparent images do not work in this way, and require special accessories that illuminate them from the upper side. Many scanners offer this as an option.

Film

Slide" (positive) or negative film can be scanned in equipment specially manufactured for this purpose. Usually, uncut film strips of up to six frames, or four mounted slides, are inserted in a carrier, which is moved by a stepper motor across a lens and CCD sensor inside the scanner. Some models mainly used for same-size scans. Film scanners vary a great deal in price and quality. Consumer scanners are relatively inexpensive while the most expensive professional CCD based film scanning system was around 120,000 USD. More expensive solutions are said to produce better results.

Hand

Hand scanners come in two forms: document and 3D scanners. Hand held document scanners are manual devices that are dragged across the surface of the image to be scanned. Scanning documents in this manner requires a steady hand, as an uneven scanning rate would produce distorted images - a little light on the scanner would indicate if the motion was too fast. They typically have a "start" button, which is held by the user for the duration of the scan; some switches to set the optical resolution; and a roller, which generates a clock pulse for synchronization with the computer. Most hand scanners were monochrome, and produced light from an array of green LEDs to illuminate the image. A typical hand scanner also had a small window through which the document being scanned could be viewed. They were popular during the early 1990s and usually had a proprietary interface module specific to a particular type of computer, usually an Atari ST or Commodore Amiga.

While popularity for document scanning has waned, use of hand held 3D scanners remains popular for many applications, including industrial design, reverse engineering, inspection & analysis, digital manufacturing and medical applications. To compensate for the uneven motion of the human hand, most 3D scanning systems rely on the placement of reference markers – typically adhesive reflective tabs that the scanner uses to align elements and mark positions in space.

Desktop Digital Camera Scanner

One printer manufacturer (Lexmark) has introduced all-in-one printer which is provided with a desktop digital camera scanner that has 10 megapixel image sensors. For scanning a business card or a full 8.5x11 inch image, it takes not more than 3 seconds including the processing time.

Smartphone Scanner Apps

Cameras in smartphones have reached a resolution and quality that reasonable quality scans can be achieved by taking a photo with the phone and using a scanning app for post-processing (such as whitening the background of a page, correcting perspective distortion so that a document is output as a correct rectangle, conversion to black-and-white, etc.) Most smartphone platforms now have a range of scanner apps available (the largest range being for Apple iPhone). These apps can typically scan multiple page documents through the use of multiple camera exposures, and output them to a PDF document or as separate JPEG images. Some smartphone scanning apps can also save documents directly to online storage locations such as Dropbox, Evernote, send via email or fax documents via email-to-fax gateways.

Quality

Scanners typically read red-green-blue color (RGB) data from the array. This data is then processed with some proprietary algorithm to correct for different exposure conditions, and sent to the computer via the device's input/output interface (usually USB, previous to which was SCSI or bidirectional parallel port in older units). Color depth varies depending on the scanning array characteristics, but is usually at least 24 bits. High quality models have 48 bits or more color depth. Another qualifying parameter for a scanner is its *resolution*, measured in pixels per inch (ppi), sometimes more accurately referred to as Samples per inch (spi). Instead of using the scanner's true optical resolution, the only meaningful parameter, manufacturers like to refer to the *interpolated resolution*, which is much higher thanks to software interpolation. As of 2009, a high-end flatbed scanner can scan up to 5400 ppi and a good drum scanner has an optical resolution of 12,000 ppi.

Manufacturers often claim interpolated resolutions as high as 19,200 ppi; but such numbers carry little meaningful value, because the number of possible interpolated pixels is unlimited and doing so does not increase the level of captured detail.

The size of the file created increases with the square of the resolution; doubling the resolution quadruples the file size. A resolution must be chosen that is within the capabilities of the equipment, preserves sufficient detail, and does not produce a file of excessive size. The file size can be reduced for a given resolution by using "lossy" compression methods such as JPEG, at some cost in quality. If the best possible quality is required lossless compression should be used; reduced-quality files of smaller size can be produced from such an image when required (e.g., image designed to be printed on a full page, and a much smaller file to be displayed as part of a fast-loading web page).

Purity can be diminished be scanner noise, optical flare, poor analog to digital conversion, color interpolation (from scanners that use CCD chips with bayer grids) and poor software. Drum scanners are said to produce the purist digital representations of the film, followed by high end film scanners that use the larger Kodak Tri-Linear sensors. The third important parameter for a scanner is its density range. A high-density range means that the scanner is able to reproduce shadow details and brightness details in one scan. By combining full-color imagery with 3D models, modern hand-held scanners are able to completely reproduce objects electronically. The addition of 3D color printers enables accurate miniaturization of these objects, with applications across many industries and professions.

Computer Connection

Scanning the document is only one part of the process. For the scanned image to be useful, it must be transferred from the scanner to an application running on the computer. There are two basic issues: (1) how the scanner is physically connected to the computer and (2) how the application retrieves the information from the scanner.

Direct Physical Connection to a Computer

The amount of data generated by a scanner can be very large: a 600 DPI 23 x 28 cm (9"x11") (slightly larger than A4 paper) uncompressed 24-bit image is about 100 megabytes of data which must be transferred and stored. Recent scanners can generate this volume of data in a matter of seconds, making a fast connection desirable. Scanners communicate to their host computer using one of the following physical interfaces, listing from slow to fast:

- Parallel port - Connecting through a parallel port is the slowest common transfer method. Early scanners had parallel port connections that could not transfer data faster than 70 kilobytes/second. The primary advantage of the parallel port connection was economic and user skill level: it avoided adding an interface card to the computer.
- GPIB - General Purpose Interface Bus. Certain drumscanners like the Howtek D4000 featured both a SCSI and GPIB interface. The latter conforms to the IEEE-488 standard, introduced in the mid '70's. The GPIB-interface has only been used by a few scanner manufactures, mostly serving the DOS/ Windows environment. For Apple Macintosh systems, National Instruments provided a NuBus GPIB interface card.
- Small Computer System Interface (SCSI), which is supported by most computers only via an additional SCSI interface card. Some SCSI scanners are supplied

together with a dedicated SCSI card for a PC, although any SCSI controller can be used. During the evolution of the SCSI standard speeds increased, with backwards compatibility; a SCSI connection can transfer data at the highest speed which both the controller and the device support. SCSI has been largely replaced by USB and Firewire, one or both of which are directly supported by most computers, and which are easier to set up than SCSI.

- Universal Serial Bus (USB) scanners can transfer data quickly, and they are easier to use and cheaper than SCSI devices. The early USB 1.1 standard could transfer data at only 1.5 megabytes per second (slower than SCSI), but the later USB 2.0 standard can theoretically transfer up to 60 megabytes per second (although everyday rates are much lower), resulting in faster operation.
- FireWire is an interface that is much faster than USB 1.1 and comparable to USB 2.0. FireWire speeds are 25, 50, and 100, 400 and 800 megabits per second (but a device may not support all speeds). Also known as: IEEE-1394.
- Proprietary interfaces were used on some early scanners that used a proprietary interface card rather than a standard interface.

Indirect (Network) Connection to a Computer

During the early nineties, professional flatbed scanners were targeted to professional users. Some vendors (like

Umax) allowed a single scanner connected to a host computer to function as a scanner accessible by all users within a local computer network. This proved to be very handy to e.g. publishers, print shops, etc. This functionality gradually disappeared after the mid-'90's as flatbed scanners became more affordable each year. However, as of 2000 and later, all-in-one multi-purpose devices targeted to serve both (small) offices and consumers usually combine a printer, scanner, copier and fax into a single apparatus available to a whole workgroup, providing each individual fax, scan, copy and print functionality.

Applications Programming Interface

A paint application such as GIMP or Adobe Photoshop must communicate with the scanner. There are many different scanners, and many of those scanners use different protocols. In order to simplify applications programming, some Applications Programming Interfaces ("API") were developed. The API presents a uniform interface to the scanner. This means that the application does not need to know the specific details of the scanner in order to access it directly. For example, Adobe Photoshop supports the TWAIN standard; therefore in theory Photoshop can acquire an image from any scanner that also supports TWAIN. In practice, there are often problems with an application communicating with a scanner. Either the application or the scanner manufacturer (or both) may have faults in their implementation of the API. Typically, the API is implemented as a dynamically linked library. Each scanner manufacturer provides software that translates the API procedure calls into primitive commands that are issued to a hardware controller (such as the SCSI, USB, or FireWire controller). The manufacturer's part of the API is commonly called a device driver, but that designation is not strictly accurate: the API does not run in kernel mode and does not directly access the device. Rather the scanner API library translates application requests into hardware requests.

Common Scanner Software API Interfaces

SANE (Scanner Access Now Easy) is a free/open source API for accessing scanners. Originally developed for Unix and Linux operating systems, it has been ported to OS/2, Mac OS X, and Microsoft Windows. Unlike TWAIN, SANE does not handle the user interface. This allows batch scans and transparent network access without any special support from the device driver.

TWAIN is used by most scanners. Originally used for low-end and home-use equipment, it is now widely used for large-volume scanning.

ISIS (Image and Scanner Interface Specification) created by Pixel Translations, which still uses SCSI-II for performance reasons, is used by large, departmental-scale, machines.

WIA (Windows Image Acquisition) is an API provided by Microsoft.

Bundled Applications

Although no software beyond a scanning utility is a feature of any scanner, many scanners come bundled with software. Typically, in addition to the scanning utility, some type of image-editing application (such as Photoshop), and optical character recognition (OCR) software are supplied. OCR software converts graphical images of text into standard text that can be edited using common word-processing and text-editing software; accuracy is rarely perfect.

Output Data

The scanned result is a non-compressed RGB image, which can be transferred to a computer's memory. Some scanners compress and clean up the image using embedded firmware. Once on the computer, the image can be processed with a raster graphics programme (such as Photoshop or the GIMP) and saved on a storage device (such as a hard disk). Images are usually stored on a hard disk. Pictures are normally stored in image formats such as uncompressed Bitmap, "non-lossy" (lossless) compressed TIFF and PNG, and "lossy" compressed JPEG. Documents are best stored in TIFF or PDF format; JPEG is particularly unsuitable for text. Optical character recognition (OCR) software allows a scanned image of text to be converted into editable text with reasonable accuracy, so long as the text is cleanly printed and in a typeface and size that can be read by the software. OCR capability may be integrated into the scanning software, or the scanned image file can be processed with a separate OCR programme.

Document Processing

The scanning or digitization of paper documents for storage makes different requirements of the scanning equipment used than scanning of pictures for reproduction. While documents can be scanned on general-purpose scanners, it is more efficiently performed on dedicated document scanners. When scanning large quantities of documents, speed and paper-handling is very important, but the resolution of the scan will normally be much lower than for good reproduction of pictures. Document scanners have document feeders, usually larger than those sometimes found on copiers or all-purpose scanners. Scans are made at high speed, perhaps 20 to 150 pages per minute, often in grayscale, although many scanners support color. Many scanners can scan both sides of double-sided originals (duplex operation). Sophisticated document scanners have firmware or software that cleans up scans of text as they are produced, eliminating accidental marks and sharpening type; this would be unacceptable for photographic work, where marks cannot reliably be distinguished from desired fine detail. Files created are compressed as they are made.

The resolution used is usually from 150 to 300 dpi, although the hardware may be capable of somewhat higher resolution; this produces images of text good enough to read and for optical character recognition (OCR), without the higher demands on storage space required by higher-resolution images.

Document scans are often processed using OCR technology to create editable and searchable files. Most scanners use ISIS or TWAIN device drivers to scan documents into TIFF format so that the scanned pages can be fed into a document management system that will handle the archiving and retrieval of the scanned pages. Lossy JPEG compression, which is very efficient for pictures, is undesirable for text documents, as slanted straight edges take on a jagged appearance, and solid black (or other color) text on a light background compresses well with lossless compression formats.

While paper feeding and scanning can be done automatically and quickly, preparation and indexing are necessary and require much work by humans. Preparation involves manually inspecting the papers to be scanned and making sure that they are in order, unfolded, without staples or anything else that might jam the scanner. Additionally, some industries such as legal and medical may require documents to have Bates Numbering or some other mark giving a document identification number and date/time of the document scan.

Indexing involves associating relevant keywords to files so that they can be retrieved by content. This process can sometimes be automated to some extent, but it often requires manual labour performed by data-entry clerks. One common practice is the use of barcode-recognition technology: during preparation, barcode sheets with folder names or index information are inserted into the document files, folders, and document groups. Using automatic batch scanning, the documents are saved into appropriate folders, and an index is created for integration into document-management systems.

A specialized form of document scanning is book scanning. Technical difficulties arise from the books usually being bound and sometimes fragile and irreplaceable, but some manufacturers have developed specialized machinery to deal with this. For instance, Atiz DIY scanner uses a V-shaped cradle and a V-shaped transparent platen to handle brittle books. Often special robotic mechanisms are used to automate the page turning and scanning process.

Infrared Cleaning

Infrared cleaning is a technique used to remove dust and scratches from film, and most modern scanners incorporate this feature. It works by scanning the film with infrared light. From this, it is possible to detect dust and scratches that cut off the infrared light; and they can then be automatically removed, by considering their position, size, shape, and surroundings. Scanner manufacturers usually have their own name attached to this technique. For example, Epson, Nikon, Microtek, and others use Digital ICE, while Canon uses its own system FARE (Film Automatic Retouching and Enhancement system). Some independent software developers are designing their own infrared cleaning tools.

Trivia

Scanner Music

Flatbed scanners are capable of synthesising simple musical scores, due to the variable speed (and tone) of their stepper motors. This property can be applied for hardware diagnostics: for example the HP Scanjet 5 plays Ode to Joy if powered on with the Scan button held down and the SCSI ID set to zero. Windows- and Linux-based software is available for several brands and types of flatbed scanners to play MIDI files for fun purposes.

Scanner Art

Scanner art is art made by placing objects on a flatbed scanner and scanning them. There has been some debate as to whether scanner art is a form of digital photography. Images made with a scanner differ from those made with a camera, as the scanner has very little depth of field and a constant light all over the surface.

Planetary Scanner

A planetary scanner (also called an orbital scanner) is a type of image scanner for making scans of rare books and other easily damaged documents. In essence, such a scanner is a mounted camera taking photos of a well-lit environment. Originally, such scanners were expensive and could only be found in archives and museums, but with the availability of cheap, high-resolution digital cameras, DIY planetary scanners have become affordable, and for instance are being used by volunteer scan providers for Project Gutenberg. Flatbed scanners often come in contact with at least part of the object to be scanned. They also require books to be fully opened most of the time (there are some exceptions where the scanning surface ends at the edge of the flatbed scanner, so that a book can be opened partially). Both practices can damage rare books; For example, opening a book 180 degrees can be damaging to its spine. These scanners are also implemented to scan other fragile documents such as old maps.

Planetary scanners tend to touch fewer parts of a book, and provide an option of only opening a book partially.

Film Scanner

A film scanner is a device made for scanning photographic film directly into a computer without the use of any intermediate printmaking. It provides several benefits over using a flatbed scanner to scan in a print of any size: the photographer has direct control over cropping and aspect ratio from the original, unmolested image on film; and many film scanners have special software or hardware that removes scratches and film grain and improves color reproduction from film. Film scanners can accept either strips of 35 mm or 120 film, or individual slides. Low-end scanners typically only take 35mm film strips, while medium- and high-end film scanners often have interchangeable film loaders. This allows the one scanning platform to be used for different sizes and packaging. For example, some allow microscope slides to be loaded for scanning, while mechanised slide loaders allow many individual slides to be batch scanned unattended.

Dust and Scratches

Dust and scratches on the film can be a big problem for scanning. Because of their reduced size (compared to prints) the scanners are capable of resolutions much higher than a regular flatbed scanner; typically at least 2000 samples per inch (spi), up to 4000 spi or more. At these resolutions dust and scratches take on gigantic proportions. Even small specks of dust, invisible to the naked eye, can obscure a cluster of several pixels. For this reason, techniques have been developed to remove their appearance from a scan, see film restoration.

The simplest is the median filter, often called *despeckle* in many graphic manipulation programmes, e.g. in Adobe Photoshop and the GIMP. It works by examining a pixel in relation to the pixels surrounding it; if it is too different from the surrounding pixels then it is replaced with one set to their median value. This and other methods can be quite effective, but have the disadvantage that the filter cannot know what actually is dust or noise. It will also obliterate fine detail in the scan.

Infrared Cleaning

Infrared cleaning works by collecting an infrared channel from the scan at the same time as the visible colour channels (red, green, and blue). This is done by using a light source that also produces infrared radiation, and having a fourth row of sensors on the linear CCD sensor. Photographic film is mostly transparent to infrared radiation (no matter what the visible image contains) but dust and scratches aren't, so they show up in the IR channel. This information can then be used to automatically remove the appearance of dust and scratches in the visible channels and replace them with something similar to their surroundings. A major limitation of this technique is that it can only be used on dye-based (color and chromogenic black-and-white) films; the image-forming silver particles in most black-and-white film stocks are opaque to infrared radiation.

Scanner manufacturers usually have their own name attached to this technique. Kodak developed Digital ICE at their Austin development centre, and is licensed by Epson, Nikon, Microtek and some others. Canon developed its own FARE (Film Automatic Retouching and Enhancement) system. LaserSoft Imaging is developing the iSRD Dust and Scratch Removal, on which among others Plustek is relying.

Book Scanning

Book scanning is the process of converting physical books into digital media such as images, electronic text, or electronic books (e-books) by using an image scanner. Digital books can be easily distributed, reproduced, and read on-screen. Common file formats are DjVu, Portable Document Format (PDF), and Tagged Image File Format (TIFF). To convert the raw images Optical character recognition (OCR) is used to turn book pages into a digital text format like ASCII or other similar format, which reduces the file size and allows the text to be reformatted, searched, or processed by other applications.

Image scanners may be manual or automated. In an ordinary commercial image scanner, the book is placed on a flat glass plate (or platen), and a light and optical array moves across the book underneath the glass. In manual book scanners, the glass plate extends to the edge of the scanner, making it easier to line up the book's spine. Other book scanners place the book face up in a v-shaped frame, and photograph the pages from above. Pages may be turned by hand or by automated paper transport devices. Glass or plastic sheets are usually pressed against the page to flatten it. After scanning, software adjusts the document images by lining it up, cropping it, picture-editing it, and converting it to text and final e-book form. Human proofreaders usually check the output for errors.

Scanning at 118 dots/centimeter (300 dpi) is adequate for conversion to digital text output, but for archival reproduction of rare, elaborate or illustrated books, much higher resolution is used. High-end scanners capable of thousands of pages per hour can cost thousands of dollars, but do-it-yourself manual book scanners capable of 1200 pages per hour have been built for 300 USD.

Commercial Book Scanners

Commercial book scanners are not like normal scanners; these book scanners are usually a high quality digital camera with light sources on either side of the camera mounted on some sort of frame to provide easy access for a person or machine to flip the pages of the book. Some models involve V-shaped book cradles, which provide support for book spines and also center book position automatically. The advantage of this type of scanner is that it is very fast, compared to the productivity of overhead scanners. Compared with traditional overhead scanners whose prices normally start from 10 000 USD upwards, this type of digital camera-based book scanner is much more cost-effective.

Book Scanning by Organizations on a Large Scale

One of the main challenges to this is the sheer volume of books that must be scanned, expected to be in the tens of millions.

All of these must be scanned and then made searchable online for the public to use as a universal library. Currently, there are 3 main ways that large organizations are relying on: outsourcing, scanning in-house using commercial book scanners, and scanning in-house using robotic scanning solutions.

As for outsourcing, books are often shipped to be scanned by low-cost sources such as India or China. Alternatively, due to convenience, safety and technology improvement, many organizations choose to scan in-house by using either overhead scanners which are time-consuming, or digital camera-based scanning solutions which are substantially faster, and is a method employed by Internet Archive as well as Google.

Traditional methods have included cutting off the book's spine and scanning the pages in a scanner with automatic page-feeding capability, with rebinding of the loose pages occurring afterwards. Once the page is scanned, the data is either entered manually or via OCR, another major cost of the book scanning projects. Due to copyright issues, most scanned books are those that are out of copyright; however, Google Book Search is known to scan books still protected under copyright unless the publisher specifically excludes them.

Destructive Scanning

For book scanning on a low budget, the least expensive method to scan a book or magazine is to cut off the binding. This converts the book or magazine into a sheaf of looseleaf papers, which can then be loaded into a standard automatic document feeder and scanned using inexpensive and common scanning technology. While this is definitely not a desirable solution for very old and uncommon books, it is a useful tool for book and magazine scanning where the book is not an expensive collector's item and replacement of the scanned content is easy.

There are two technical difficulties with this process, first with the cutting and second with the scanning.

Cutting

One method of cutting a stack of 500 to 1000 pages in one pass is accomplished with a guillotine paper cutter. This is a large steel table with a paper vise that screws down onto the stack and firmly secures it before cutting. The cut is accomplished with a large sharpened steel blade which moves straight down and cuts the entire length of each sheet all at once.

A lever on the blade permits several hundred pounds of force to be applied to the blade for a quick one-pass cut.

A clean cut through a thick stack of paper cannot be made with a traditional inexpensive sickle-shaped hinged paper cutter.

These cutters are only intended for a few sheets, with up to ten sheets being the practical cutting limit. A large stack of paper applies torsional forces on the hinge, pulling the blade away from the cutting edge on the table. The cut becomes more inaccurate as the cut moves away from the hinge, and the force required to hold the blade against the cutting edge increases as the cut moves away from the hinge.

The guillotine cutting process dulls the blade over time, requiring that it be resharpened. Coated paper such as slick magazine paper dulls the blade more quickly than plain book paper, due to the kaolinite clay coating. Additionally, removing the binding of an entire hardcover book causes excessive wear due to cutting through the cover's stiff backing material. Instead the outer cover can be removed and only interior pages need be cut.

Scanning

Once the paper is liberated from the spine, it can be scanned one sheet at a time using a traditional flatbed scanner or automatic document feeder (ADF). Pages with a decorative riffled edging or curving in an arc due to a non-flat binding can be difficult to scan using an ADF. An ADF is designed to scan pages of uniform shape and size, and variably sized or shaped pages can lead to improper scanning.

The riffled edges or curved edge can be guillotined off to render the outer edges flat and smooth before the binding is cut.

The coated paper of magazines and bound textbooks can make them difficult for the rollers in an ADF to pick up and guide along the paper path. An ADF which uses a series of rollers and channels to flip sheets over may jam or misfeed when fed coated paper.

Generally there are fewer problems by using as straight of a paper path as is possible, with few bends and curves. The clay can also rub off the paper over time and coat sticky pickup rollers, causing them to loosely grip the paper. The ADF rollers may need periodic cleaning to prevent this slipping.

Magazines can pose a bulk-scanning challenge due to small nonuniform sheets of paper in the stack, such as magazine subscription cards and fold out pages. These need to be removed before the bulk scan begins, and are either scanned separately if they include worthwhile content, or are simply left out of the scan process.

Non-destructive Scanning

In recent years, software driven machines and robots have been developed to scan books without the need of disbinding them in order to preserve both the contents of the document and create a digital image archive of its current state. This recent trend has been due in part to ever improving imaging technologies that allow a high quality digital archive image to be captured with little or no damage to a rare or fragile book in a reasonably short period of time. Some high-end scanning systems employ vacuum and air and static charges to turn pages while imaging is performed automatically, usually from a high resolution camera located over an adjustable v-shaped cradle. Images are then shuttled from the imaging device into various editing suites which can further process the images for either an archival-quality file such as TIFF or JPEG 2000, or a web-friendly output such as JPEG or PDF.