Chhattisgarh Swami Vivekananda Technical University, Bhilai

Program / Semester: **B.Tech (VIII Sem)**	Branch: **Computer Science & Engineering**
Subject: **Multimedia & Computer Vision**	Course Code: **D022833(022)**
Total / Minimum-Pass Marks (End Semester Exam): **100 / 35**	L: 2 T: 1 P: 0 Credits: **3**
Class Tests & Assignments to be conducted: **2 each**	Duration (End Semester Exam): **03 Hours**

Course Objective:

1. To understand the fundamental issues and problems in the representation, manipulation, and deliveryof multimedia content particularly in a networked environment.
2. To understand the concepts of multimedia components.
3. To understand the basic concepts of Computer vision.

UNIT-I: Introduction

Concept of Multimedia, media & data stream, Main properties of multimedia system, Data stream characteristics of continuous media, multimedia Applications, Hardware and software requirements, Multimedia Products & its evolution.

UNIT-II: Components Of Multimedia

Text, Basic sound concepts, MIDI, Speech, Basic concept of Images, Graphics format, Overview of image processing, Basic concepts of Video & animation, Conventional system, Transmission, Enhanced system, High-Definition system, Computer based animation, Design & authoring Tools, Categories of Authority Tools, Types of products

UNIT-III: Data Compression

Coding requirement, Source, entropy, hybrid coding, JPEG, MPEG, Text compression using static Huffmann technique, Dynamic Huffmann Technique, Statistical coding techniques.

UNIT-IV: Optical Storage Media

Videodisk and other WORMS, Compact Disk digital audio, Advantage of CD-DA Frames tracks blocks of CD-DA, CD-ROM, and Further CD-ROM based developments, Principles of CDWO, Prospects of CD technologies.

UNIT-V: Introduction To Computer Vision

Overview, computer imaging systems, lenses, Image formation and sensing, Image analysis, pre-processing and Binary image analysis, feature detection, image classification.

Text Books:

1. Multimedia System Design, Andleigh and Thakarar , PHI, 2003.
2. Multimedia Technology & Application, David Hillman, Galgotia Publications.
3. Computer Vision: A modern approach, Forsyth & Ponce, 2nd Ed., Pearson 2011

Reference Books:

1. Multimedia Computing Communication and Application, Steinmetz, Pearson Edn.
2. Fundamentals of Computer Graphics and Multimedia, D.P. Mukherjee, PHI

Course Outcomes [After completion of this course the students will be able to:]

1. To Know the fundamental video, audio, image, text processing techniques
2. Acquire the basic skill of designing video compression, audio compression, image compression, text compression.
3. To Know the basic techniques in designing video transmission systems: error control and rate control
4. To Identify basic concepts, terminology, theories, models and methods in the field of computer vision.

Unit-1

What is Multimedia?

The word multi and media are combined to form the word multimedia. The word "multi" signifies "many." Multimedia is a type of medium that allows information to be easily transferred from one location to another.

Multimedia is the presentation of text, pictures, audio, and video with links and tools that allow the user to navigate, engage, create, and communicate using a computer. Multimedia refers to the computer-assisted integration of text, drawings, still and moving images(videos) graphics, audio, animation, and any other media in which any type of information can be expressed, stored, communicated, and processed digitally. To begin, a computer must be present to coordinate what you see and hear, as well as to interact with. Second, there must be interconnections between the various pieces of information. Third, you'll need navigational tools to get around the web of interconnected data.
Multimedia is being employed in a variety of disciplines, including education, training, and business.

Categories of Multimedia

Linear Multimedia:

It is also called Non-interactive multimedia. In the case of linear multimedia, the end-user cannot control the content of the application. It has literally no interactivity of any kind. Some multimedia projects like movies in which material is thrown in a linear fashion from beginning to end. A linear multimedia application lacks all the features with the help of which, a user can interact with the application such as the ability to choose different options, click on icons, control the flow of the media, or change the pace at which the media is displayed. Linear multimedia works very well for providing information to a large group of people such as at training sessions, seminars, workplace meetings, etc.

Non-Linear Multimedia:

In Non-Linear multimedia, the end-user is allowed the navigational control to rove through multimedia content at his own desire. The user can control the access of the application. Non-linear offers user interactivity to control the movement of data. For example computer games, websites, self-paced computer-based training packages, etc.

Applications of Multimedia

Multimedia indicates that, in addition to text, graphics/drawings, and photographs, computer information can be represented using audio, video, and animation. Multimedia is used in:

Education
In the subject of education, multimedia is becoming increasingly popular. It is often used to produce study materials for pupils and to ensure that they have a thorough comprehension of various disciplines. Edutainment, which combines education and entertainment, has become highly popular in recent years. This system gives learning in the form of enjoyment to the user.

Entertainment
The usage of multimedia in films creates a unique auditory and video impression. Today, multimedia has completely transformed the art of filmmaking around the world. Multimedia is the only way to achieve difficult effects and actions. The entertainment sector makes extensive use of multimedia. It's particularly useful for creating special effects in films and video games. The most visible illustration of the emergence of multimedia in entertainment is music and video apps. Interactive games become possible thanks to the use of multimedia in the gaming business. Video games are more interesting because of the integrated audio and visual effects.

Business
Marketing, advertising, product demos, presentation, training, networked communication, etc. are applications of multimedia that are helpful in many businesses. The audience can quickly understand an idea when multimedia presentations are used. It gives a simple and effective technique to attract visitors' attention and effectively conveys information about numerous products. It's also utilized to encourage clients to buy things in business marketing.

Technology & Science
In the sphere of science and technology, multimedia has a wide range of applications. It can communicate audio, films, and other multimedia documents in a variety of formats. Only multimedia can make live broadcasting from one location to another possible.
It is beneficial to surgeons because they can rehearse intricate procedures such as brain removal and reconstructive surgery using images made from imaging scans of the human body. Plans can be produced more efficiently to cut expenses and problems.

Fine Arts
Multimedia artists work in the fine arts, combining approaches employing many media and incorporating viewer involvement in some form. For example, a variety of digital mediums can be used to combine movies and operas. Digital artist is a new word for these types of artists. Digital painters make digital paintings, matte paintings, and vector graphics of many varieties using computer applications.

Engineering
Multimedia is frequently used by software engineers in computer simulations for military or industrial training. It's also used for software interfaces created by creative experts and software engineers in partnership. Only multimedia is used to perform all the minute calculations.

Components of Multimedia

Multimedia consists of the following 5 components:

Text

Characters are used to form words, phrases, and paragraphs in the text. Text appears in all multimedia creations of some kind. The text can be in a variety of fonts and sizes to match the multimedia software's professional presentation. Text in multimedia systems can communicate specific information or serve as a supplement to the information provided by the other media.

Graphics

Non-text information, such as a sketch, chart, or photograph, is represented digitally. Graphics add to the appeal of the multimedia application. In many circumstances, people dislike reading big amounts of material on computers. As a result, pictures are more frequently used than words to clarify concepts, offer background information, and so on. Graphics are at the heart of any multimedia presentation. The use of visuals in multimedia enhances the effectiveness and presentation of the concept. Windows Picture, Internet Explorer, and other similar programs are often used to see visuals. Adobe Photoshop is a popular graphics editing program that allows you to effortlessly change graphics and make them more effective and appealing.

Animations

A sequence of still photographs is being flipped through. It's a set of visuals that give the impression of movement. Animation is the process of making a still image appear to move. A presentation can also be made lighter and more appealing by using animation. In multimedia applications, the animation is quite popular. The following are some of the most regularly used animation viewing programs: Fax Viewer, Internet Explorer, etc.

Video

Photographic images that appear to be in full motion and are played back at speeds of 15 to 30 frames per second. The term video refers to a moving image that is accompanied by sound, such as a television picture. Of course, text can be included in videos, either as captioning for spoken words or as text embedded in an image, as in a slide presentation. The following programs are widely used to view videos: Real Player, Window Media Player, etc.

Audio

Any sound, whether it's music, conversation, or something else. Sound is the most serious aspect of multimedia, delivering the joy of music, special effects, and other forms of entertainment. Decibels are a unit of measurement for volume and sound pressure level. Audio files are used as part of the application context as well as to enhance interaction. Audio files must occasionally be distributed using plug-in media players when they appear within online applications and webpages. MP3, WMA, Wave, MIDI, and RealAudio are examples of audio formats. The following programs are widely used to view videos: Real Player, Window Media Player, etc.

Multimedia Systems with features or characteristics

What are Multimedia Systems:

A multimedia system is responsible for developing a multimedia application. A multimedia application is a bundle of different kinds of data. A multimedia computer system is one that can create, integrate, store, retrieve delete two or more types of media materials in digital form, such as audio, image, video, and text information.

Following are some major characteristics or features of a Multimedia System:

Very High Processing Power:

To deal with large amount of data, very high processing power is used.

File System:

File system must be efficient to meet the requirements of continuous media. These media files requires very high-disk bandwidth rates. Disks usually have low transfer rates and high latency rates. To satisfy the requirements for multimedia data, disk schedulers must reduce the latency time to ensure high bandwidth.

File formats that support multimedia:

Multimedia data consists of a variety of media formats or file representation including ,JPEG, MPEG, AVI, MID, WAV, DOC, GIF,PNG, etc. AVI files can contain both audio and video data in a file container that allows synchronous audio-with-video playback. Like the DVD video format, AVI files support multiple streaming audio and video. Because of restrictions on the conversion from one format to the other, the use of the data in a specific format has been limited as well.

Input/Output:

In multimedia applications, the input and output should be continuous and fast. Real-time recording as well as playback of data are common in most of the multimedia applications which need efficient I/O.

Operating System:

The operating system must provide a fast response time for interactive applications. High throughput for batch applications, and real-time scheduling,

Storage and Memory:

Multimedia systems require storage for large capacity objects such as video, audio, animation and images. Depending on the compression scheme and reliability video and audio require large amount of memory.

Network Support:

It includes internet, intranet, LAN, WAN, ATM, Mobile telephony and others. In recent years, there has been a tremendous growth of multimedia applications on the internet like streaming video, IP telephony, interactive games, teleconferencing, virtual world, distance learning and so on. These multimedia networking applications are referred as continuous-media applications and require high communication latency. Communication Latency is the time it takes for a data packet to be received by the remote computer.

Software Tools:
For the development of multimedia applications, the various software tools like programming languages, graphics software's, multimedia editing software's scripting languages: authoring tools, design software's etc are required. In addition to these the device drivers are required for interfacing the multimedia peripherals.

Data stream characteristics for continuous media:

The data stream characteristics in transmission are associated with any audio and video data transfer. Also the info stream characteristics are influenced by the compression during the information transfer. Its characteristics apply for distributed furthermore as local environment. Hence the data stream characteristics can be discussed on the basic of three given factors or properties .

1.According to the time intervals between consecutive packets.

2.According to the variation of the amount between consecutive packets.

3.According to the continuity or connection between consecutive packets.

1.According to time intervals between consecutive packets:

On the basic of this factor we find out three properties they are:

a) Strongly periodic data stream:

If time intervals are of the identical length between two consecutive t packets that's a continuing, then the stream is named strongly periodic and within the ideal case the jitter has the worth zero. For eg: PCM coded speech in traditional telephone switching.

b) Weakly periodic data stream:

If time intervals between two consecutive packets is not constant but are of periodic nature with finite period then the data stream is called weakly periodic.

c) A-periodic data stream:

If the sequence of time intervals is neither strongly nor weakly periodic, instead the time period or time gap various between packets to packets during transmission then such data stream is called A-periodic data stream.

2.According to variation of consecutive packet amounts:

On the basic of these factors there are three types of data stream:

a) Strongly regular data stream:

If the number of information stays constant during the life time of a knowledge stream, this feature is specially found in uncompressed digital data transmission, as an example audio stream of CD, video stream of camera in uncompressed form.

b) Weakley regular data stream:

If the amount of data stream varies periodically with time and not shows the behaviors of strongly regular data stream then it is called Weakley regular data stream, For example compressed video stream.

c) Irregular data stream:

If the number of information is neither constant nor changes in keeping with a periodic function, then the information streams are called irregular data stream. Transmission and processing of this category data stream is complicated. Since data stream includes a variable (bit) rate after applying compression methods.

3.According to continuity or connection between consecutive packets.:

On the basic of this factor their can be also 2 types or characters:

a) Continuous data stream:

If consecutive packets are directly transmitted one after another without any time gap then such data streams are called continuous data stream, For example audio data use for B channel of Isdn with transmission rate for 64 kbps.

b) Unconnected data stream:

A data stream with gaps between information units is named and unconnected data stream. The transmission of a connected data stream through a channel with the next capacity treads gaps between individual packets, as an example the information stream coded with JPEG method with 1.2 mbps on a FDDI network.

MULTIMEDIA HARDWARE REQUIREMENTS

CPU

Central Processing Unit (CPU) is an essential part in any computer. It is considered as the brain of computer, where processing and synchronization of all activities takes place. The efficiency of a computer is judged by the speed of the CPU in processing of data. For a multimedia computer a Pentium processor is preferred because of higher efficiency.

Monitor

The monitor is used to see the computer output. Generally, it displays 25 rows and 80 columns of text. The text or graphics in a monitor is created as a result of an arrangement of tiny dots, called pixels. Resolution is the amount of details the monitor can render. Resolution is defined in terms of horizontal and vertical pixel (picture elements) displayed on the screen.

Video Grabbing Card

We need to convert the analog video signal to digital signal for processing in a computer. Normal computer will not be able to do it alone. It requires special equipment called video grabbing card and software to this conversion process. This card translates the analog signal it receives from conventional sources such as a VCR or a video camera, and converts them into digital format.

Sound Card

Today’s computers are capable of creating the professional multimedia needs. Not only you can use computer to compose your own music, but it can also be used for recognition of speech and synthesis. It can even read back the entire document for you. But before all this happens, we need to convert the conventional sound signal to computer understandable digital signals. This is done using a special component added to the system called sound card.

CD-Rom

CD-ROM is a magnetic disk of 4.7 inches diameter and it can contain data up to 680 Megabytes. It has become a standard by itself basically for its massive storage capacity, faster data transfer rate. To access CD-ROM a very special d1rive is required and it is known as CD-ROM drive.

MULTIMEDIA SOFTWARE REQUIREMENTS

For the creation of multimedia on the PC there are hundreds of software packages that are available from manufacturers all over the world.

These software packages can cost anything from being absolutely free (normally this software is called freeware or shareware) to anything upwards of £500.

Here is a summary of just a few of these programs.

Adobe CS4

Adobe CS4 is a collection of graphic design, video editing, and web development applications made by Adobe Systems many of which are the industry standard that includes

Adobe Dreamweaver

WYSIWYG WYSIWYG HTML Although a hybrid and code-based web design and development application, Dreamweaver's mode can hide the code details of pages from the user, making it possible for non-coders to create web pages and sites.WYSIWYG (What You See Is What You Get) web development software that allows users to create websites with out using Html, everything can be done visually.

Adobe Fireworks

A graphics package that allows users to create bitmap and vector graphics editor with features such as: slices, the ability to add hotspots etc.) for rapidly creating website prototypes and application interfaces.

Gimp

Is an alternative to Photoshop and cheaper but not quite as good.

Google Sketchup

3D modeling program architectscivil engineersSketchUp is a designed for , , filmmakers, game developers, and related professions.

Microsoft Frontpage

WYSIWYG web pages sitesAs a editor, FrontPage is designed to hide the details of pages' HTML code from the user, making it possible for novices to easily create and .

Apple Quicktime

proprietary multimedia frameworkAppledigital video3D modelsmusicpanoramic imagesinteractivityQuickTime is an extensible developed by , capable of handling various formats of , , sound, text, animation, , , and .

Photoshop Pro

graphics editing programAdobe Systemsbitmap Adobe Photoshop, or simply Photoshop, is a developed and published by . It is the current market leader for commercial and image manipulation software, and is the flagship product of Adobe Systems. It has been described as "an industry standard for graphics professionals"

Microsoft Powerpoint

Powerpoint Presentations are generally made up of slides may contain text, graphics, movies, and other objects, which may be arranged freely on the slide.

Adobe Flash Player

multimedia platformanimation interactivity Macromedia1996Adobe SystemsAdobe Flash (formerly Macromedia Flash) is a that is popular for adding and to web pages. Originally acquired by , Flash was introduced in , and is currently developed and distributed by .

advertisementsFlash componentsrich Internet applicationsFlash is commonly used to create animation, , and various web page , to integrate video into web pages, and more recently, to develop .

Adobe Shockwave

multimedia player MacromediaAdobe Systems Adobe Director web browser Adobe Shockwave (formerly Macromedia Shockwave) is a program, first developed by , acquired by in 2005. It allows applications to be published on the Internet and viewed in a on any computer which has the Shockwave plug-in installed.

Multimedia Product

A multimedia product is any form of content that combines different types of media, such as text, images, audio, video, and animations, to convey information, ideas, or messages. Examples of multimedia products include websites, e-books, online courses, interactive presentations, video games, and mobile apps.

Multimedia products are designed to be engaging, interactive, and immersive, providing users with a more dynamic and personalized experience than traditional forms of media. They can be used for a variety of purposes, such as education, entertainment, marketing, and communication.

Creating multimedia products requires specialized skills and tools, such as graphic design software, video editing software, and programming languages. Content creators must also consider factors such as usability, accessibility, and interactivity when developing multimedia products.

The evolution of technology has played a significant role in the development of multimedia products, enabling new possibilities for creating and distributing content. Today, multimedia products are an essential part of our daily lives, and their influence can be seen in everything from online advertising to education and training.

Evolution of Multimedia

The evolution of multimedia can be traced back to the 1950s when the first interactive multimedia system was developed for the United States military. However, it wasn't until the introduction of personal computers in the 1980s that multimedia began to take off.

Here are some key milestones in the evolution of multimedia:

- The first computer-based multimedia system, called Aspen Movie Map, was developed in the late 1970s by researchers at MIT.
- The introduction of CD-ROMs in the 1980s allowed for the storage and distribution of large amounts of multimedia content.
- In the late 1980s and early 1990s, the development of multimedia authoring tools such as HyperCard and Director made it easier for individuals and organizations to create multimedia content.
- The introduction of the World Wide Web in 1991 provided a platform for the distribution of multimedia content on a global scale.
- The development of HTML and web browsers in the mid-1990s enabled the creation of interactive web pages with multimedia elements such as images, audio, and video.
- The emergence of Flash technology in the late 1990s and early 2000s allowed for the creation of more sophisticated multimedia products such as interactive animations and games.
- The widespread adoption of smartphones and tablets in the 2010s has led to the development of mobile apps and mobile-optimized multimedia content.

- The rise of social media and video sharing platforms such as YouTube, Instagram, and TikTok has facilitated the creation and distribution of multimedia content by individuals and organizations.
- The increasing use of virtual and augmented reality technologies is creating new possibilities for immersive and interactive multimedia experiences.

Overall, the evolution of multimedia has been driven by advances in technology, and has enabled more engaging, interactive, and immersive experiences for users, and has opened up new opportunities for content creators, marketers, and businesses.

Unit-2

Component Of Multimedia

Multimedia Component

Multimedia consists of different components that work together to create a rich and engaging experience for users. Here are the main components of multimedia:

1. Text: Text is the most basic component of multimedia and is used to convey information, instructions, and messages. Text can be presented in different fonts, sizes, colors, and styles to enhance its visual appeal and readability.

2. Images: Images are a visual component of multimedia that can be used to convey information, evoke emotions, and enhance the overall aesthetic of the content. Images can be static or dynamic and can be presented in different formats such as JPEG, PNG, or GIF.

3. Audio: Audio is a component of multimedia that includes sounds, music, and voiceovers. Audio can be used to convey information, create atmosphere, and evoke emotions. Audio can be presented in different formats such as MP3 or WAV.

4. Video: Video is a dynamic component of multimedia that includes moving images, animations, and special effects. Video can be used to convey complex information, tell stories, and entertain. Video can be presented in different formats such as MP4 or AVI.

5. Animation: Animation is a dynamic component of multimedia that creates the illusion of movement or action. Animation can be used to explain complex concepts, entertain, and create engaging interactive experiences.

6. Interactivity: Interactivity is a component of multimedia that allows users to engage with the content and control their experience. Interactivity can include elements such as buttons, menus, and links, as well as more advanced features such as simulations and games.

Overall, the combination of these components creates a rich and engaging multimedia experience for users, providing them with a variety of sensory stimuli that enhance their understanding and retention of information.

Text in Multimedia

Text is an essential component of multimedia that is used to convey information, instructions, and messages. Text can be presented in different fonts, sizes, colors, and styles to enhance its visual appeal and readability.

In multimedia, text can be used in different ways, such as:

1. Titles and headings: Titles and headings are used to introduce the content and provide a summary of what the user can expect. They can be presented in larger font sizes, bold, or different colors to attract attention.

2. Captions: Captions are used to provide additional information or context about an image or video. They can be presented below or beside the media element and can be presented in a smaller font size.

3. Descriptions: Descriptions are used to provide detailed information about a product, service, or concept. They can be presented in paragraphs or bullet points and can be presented in a font size that is easy to read.

4. Instructions: Instructions are used to guide the user on how to interact with the multimedia content. They can be presented as a step-by-step guide, in bullet points or in short sentences.

5. Interactive elements: Interactive elements such as buttons, menus, and links may include text that explains what the user can do or where the link will take them.

When using text in multimedia, it is essential to consider its readability, legibility, and accessibility. The font size, style, and color should be chosen to ensure that the text is easy to read, and the contrast between the text and the background should be sufficient. Additionally, text should be accessible to users with disabilities, such as those who use screen readers, and should conform to accessibility guidelines such as WCAG 2.0.

Basic Sound Concept Of Multimedia

Sound is an essential component of multimedia that includes sounds, music, and voiceovers. Sound can be used to convey information, create atmosphere, and evoke emotions. In multimedia, sound is typically presented as an audio file that can be played back on a computer, mobile device, or other media player.

Here are some basic sound concepts of multimedia:

1. Frequency: Frequency is the number of sound waves per second and is measured in Hertz (Hz). Higher frequencies create higher-pitched sounds, while lower frequencies create lower-pitched sounds.

2. Amplitude: Amplitude is the intensity or loudness of a sound and is measured in decibels (dB). Higher amplitudes create louder sounds, while lower amplitudes create quieter sounds.

3. Sound quality: Sound quality refers to the clarity and fidelity of a sound. Factors that affect sound quality include the type of microphone used to record the sound, the environment in which the sound was recorded, and the compression used to create the audio file.

4. Stereo sound: Stereo sound refers to the use of two or more audio channels to create a sense of depth and space in the sound. Stereo sound is typically used in music and movies to create a more immersive listening experience.

5. Voiceovers: Voiceovers are recorded spoken words that are synchronized with visual content. Voiceovers can be used to narrate a story, explain a concept, or provide instructions.

6. Sound effects: Sound effects are audio clips that are used to create a particular sound or atmosphere. Sound effects can be used to create a sense of realism or to evoke an emotional response from the user.

When using sound in multimedia, it is essential to consider its quality, clarity, and volume. Sound should be adjusted to ensure that it is audible but not too loud, and any background noise or distortion should be minimized. Additionally, it is essential to consider the accessibility of sound for users with disabilities, such as those who are hard of hearing, and to provide alternatives such as captions or transcripts when necessary.

MIDI in Multimedia

MIDI (Musical Instrument Digital Interface) is a protocol that allows digital musical instruments, computers, and other devices to communicate with each other. MIDI is often used in multimedia applications to synchronize music, sound effects, and other audio elements with video, animation, and other multimedia content.

In multimedia production, MIDI can be used in several ways, including:

1. Controlling virtual instruments: MIDI allows multimedia producers to control virtual instruments, such as software synthesizers and drum machines, in real-time. This allows them to create music and sound effects that are synchronized with the multimedia content.
2. Triggering sound effects: MIDI can be used to trigger sound effects, such as explosions, footsteps, and other environmental sounds, in response to specific events in the multimedia content.
3. Timecode synchronization: MIDI can be used to synchronize multimedia content with external timecode sources, such as SMPTE timecode, which is commonly used in film and television production.
4. Controlling lighting and other effects: MIDI can be used to control lighting and other effects in live performances, such as concerts and theater productions.

Overall, MIDI is a powerful tool for multimedia producers who want to create dynamic and immersive audiovisual experiences.

Regenerate response

Speech in multimedia

Speech is a key component of multimedia, as it enables effective communication of ideas, emotions, and information. In multimedia production, speech can be used in several ways to enhance the user experience and convey messages more effectively.

Here are some examples of how speech can be used in multimedia:

1. Narration: Narration is a common use of speech in multimedia, where a voiceover artist reads a script to provide context and information about the content. This is often used in documentaries, educational videos, and promotional materials.
2. Dialogue: Dialogue is used in multimedia to create engaging characters and tell stories. It can be used in animation, video games, and films to give characters a voice and make them more relatable to the audience.
3. Automated speech recognition: Automated speech recognition (ASR) is a technology that enables computers to recognize and transcribe spoken words. ASR can be used in

multimedia to provide closed captions, subtitles, and transcripts, making the content more accessible to people with hearing impairments or those who prefer to read along.

4. Voice commands: Voice commands are becoming more common in multimedia, especially in virtual assistants and smart home devices. Users can use their voice to control the multimedia content, such as playing a video or adjusting the volume.

Basic concept of images in multimedia

Images are a fundamental component of multimedia, as they help to convey information and enhance the user experience. In multimedia production, images can be used in several ways to create engaging and dynamic content.

Here are some basic concepts of images in multimedia:

1. **Resolution:** Resolution refers to the number of pixels in an image. The higher the resolution, the more detail the image can contain. However, high-resolution images also require more storage space and processing power.
2. **Color depth:** Color depth refers to the number of colors that can be displayed in an image. The higher the color depth, the more colors the image can contain. This is important for creating realistic and vibrant images, especially in photographs and illustrations.

3. **File format:** There are several different file formats for images, including JPEG, PNG, and GIF. Each format has its own advantages and disadvantages in terms of file size, quality, and compatibility with different devices and software.

4. **Compression:** Compression is the process of reducing the file size of an image without significantly reducing its quality. This is important for multimedia production, as large image files can slow down the performance of the content.

5. **Transparency:** Transparency is the ability of an image to have parts of it be transparent, meaning that the background can be seen through those parts. This is useful for creating overlays and combining images with other multimedia elements.

Graphic format in multimedia

In multimedia production, there are several different graphic formats that are commonly used to store and display images. Each format has its own advantages and disadvantages in terms of file size, quality, and compatibility with different devices and software.

Here are some of the most common graphic formats used in multimedia:

1. JPEG (Joint Photographic Experts Group): JPEG is a popular format for storing photographs and other images with complex color gradients. It uses lossy compression, which means that some image data is lost during compression, resulting in a smaller file size but potentially lower image quality.

2. PNG (Portable Network Graphics): PNG is a format that supports transparency and lossless compression, which means that the image quality is not compromised during compression. PNG is commonly used for web graphics and images with simple color gradients.

3. GIF (Graphics Interchange Format): GIF is a format that supports animation and transparency, making it popular for creating simple animated graphics and memes. GIFs use lossless compression, but they are limited to a maximum of 256 colors, which can result in a lower quality image.

4. BMP (Bitmap): BMP is a format that stores images pixel by pixel, without compression. This results in large file sizes, but it also preserves image quality. BMP is commonly used for simple graphics and icons.

5. SVG (Scalable Vector Graphics): SVG is a format that uses vector graphics, which means that the images can be scaled to any size without losing quality. SVG is commonly used for web graphics and logos.

Overview of Image Processing

Image processing is the use of algorithms and techniques to manipulate digital images in order to enhance their quality or extract useful information from them. In multimedia production, image processing is often used to improve the visual quality of images and videos, and to create special effects and visualizations.

Here are some common techniques used in image processing:

1. Image filtering: Image filtering is the process of applying a filter to an image in order to enhance or modify certain features. Filters can be used to remove noise, sharpen edges, blur or smoothen images, and more.

2. Color correction: Color correction is the process of adjusting the color balance of an image in order to achieve a desired look or to correct for color cast. This can involve adjusting the brightness, contrast, saturation, and hue of the image.

3. Image segmentation: Image segmentation is the process of dividing an image into regions or objects in order to extract useful information or separate foreground from background. This is commonly used in computer vision and object recognition applications.

4. Image compression: Image compression is the process of reducing the file size of an image without significantly reducing its quality. This can be achieved through lossy or lossless compression techniques.

5. Image restoration: Image restoration is the process of recovering a degraded or damaged image to its original quality. This can involve removing scratches, dust, or other artifacts, as well as restoring color and brightness.

Overall, image processing plays an important role in multimedia production, allowing producers to enhance the quality of images and videos, as well as to create special effects and visualizations that enhance the user experience.

ANIMATION and VIDEO

Introduction

Animation makes static presentations come alive. It is visual change over time and can add great power to our multimedia projects. Carefully planned, well-executed

video clips can make a dramatic difference in a multimedia project. Animation is created from drawn pictures and video is created using real time visuals.

Principles of Animation

Animation is the rapid display of a sequence of images of 2-D artwork or model positions in order to create an illusion of movement. It is an optical illusion of motion due to the phenomenon of persistence of vision, and can be created and demonstrated in a number of ways. The most common method of presenting animation is as a motion picture or video program, although several other forms of presenting animation also exist Animation is possible because of a biological phenomenon known as *persistence of vision* and a psychological phenomenon called *phi.* An object seen by the human eye remains chemically mapped on the eye's retina for a brief time after viewing. Combined with the human mind's need to conceptually complete a perceived action, this makes it possible for a series of images that are changed very slightly and very rapidly, one after the other, to seemingly blend together into a visual illusion of movement. The following shows a few cells or frames of a rotating logo.

When the images are progressively and rapidly changed, the arrow of the compass is perceived to be

spinning.

Television video builds entire frames or pictures every second; the speed with which each

frame is replaced by the next one makes the images appear to blend smoothly into movement.

To make an object travel across the screen while it changes its shape, just change the shape and also move or *translate* it a few pixels for each frame.

Animation Techniques

When you create an animation, organize its execution into a series of logical steps. First,

gather up in your mind all the activities you wish to provide in the animation; if it is

complicated, you may wish to create a written script with a list of activities and required objects.

Choose the animation tool best suited for the job.

Then build and tweak your sequences; experiment with lighting effects. Allow plenty of time for this phase when

you are experimenting and testing. Finally, post-process your animation, doing any special rendering and adding sound effects.

Cel Animation

The term *cel* derives from the clear celluloid sheets that were used for drawing each frame, which have been replaced today by acetate or plastic. Cels of famous animated cartoons have become sought-after, suitable-for-framing collector's items.

Cel animation artwork begins with *keyframes* (the first and last frame of an action). For example, when an animated figure of a man walks across the screen, he balances the weight of his entire body on one foot and then the other in a series of falls and recoveries, with the opposite foot and leg catching up to support the body.

- The animation techniques made famous by Disney use a series of

progressively different on each frame of movie film which plays at 24 frames

per second.

- A minute of animation may thus require as many as 1,440 separate frames.
- The term cel derives from the clear celluloid sheets that were used for drawing

each frame, which is been replaced today by acetate or plastic.

- Cel animation artwork begins with keyframes.

Computer Animation

Computer animation programs typically employ the same logic and procedural concepts as cel animation, using layer, keyframe, and tweening techniques, and even borrowing from the vocabulary of classic animators. On the computer, paint is most often filled or drawn with tools using features such as gradients and antialiasing. The word *links*, in computer animation terminology, usually means special methods for computing RGB pixel values, providing edge detection, and layering so that images can blend or otherwise mix their colors to produce special transparencies, inversions, and effects.

- Computer Animation is same as that of the logic and procedural concepts as cel animation and use the vocabulary of classic cel animation – terms such as layer, Keyframe, and tweening.
- The primary difference between the animation software program is in how much must be drawn by the animator and how much is automatically generated by the software
- In 2D animation the animator creates an object and describes a path for the object to follow. The software takes over, actually creating the animation on the fly as the program is being viewed by your user.
- In 3D animation the animator puts his effort in creating the models of individual and designing the characteristic of their shapes and surfaces.
- Paint is most often filled or drawn with tools using features such as gradients and anti- aliasing.

Kinematics

- It is the study of the movement and motion of structures that have joints, such as a walking man.
- Inverse Kinematics is in high-end 3D programs, it is the process by which you link objects such as hands to arms and define their relationships and limits.
- Once those relationships are set you can drag these parts around and let the computer calculate the result.

Morphing

- Morphing is popular effect in which one image transforms into another. Morphing application and other modeling tools that offer this effect can perform transition not only between still images but often between moving images as well.
- The morphed images were built at a rate of 8 frames per second, with each transition taking a total of 4 seconds.
- Some product that uses the morphing features are as follows
- o Black Belt's EasyMorph and WinImages,

o Human Software's Squizz

o Valis Group's Flo , MetaFlo, and MovieFlo.

Animation File Formats

Some file formats are designed specifically to contain animations and the can be ported among application and platforms with the proper translators.

- Director *.dir, *.dcr
- AnimationPro *.fli, *.flc
- 3D Studio Max *.max
- SuperCard and Director *.pics
- CompuServe *.gif
- Flash *.fla, *.swf

Following is the list of few Software used for computerized animation:

- 3D Studio Max
- Flash
- AnimationPro

Video

Analog versus Digital

Digital video has supplanted analog video as the method of choice for making video for multimedia use. While broadcast stations and professional production and postproduction houses remain greatly invested in analog video hardware (according to Sony, there are more than 350,000 Betacam SP devices in use today), digital video gear produces excellent finished products at a fraction of the cost of analog. A digital camcorder directly connected to a computer workstation eliminates the image-degrading analog-to-digital conversion step typically performed by expensive video capture cards, and brings the power of nonlinear video editing and production to everyday users.

Broadcast Video Standards

Four broadcast and video standards and recording formats are commonly in use around the world: NTSC, PAL, SECAM, and HDTV. Because these standards and formats are not easily interchangeable, it is important to know where your multimedia project will be used.

NTSC

The United States, Japan, and many other countries use a system for broadcasting and displaying video that is based upon the specifications set forth by the 1952 National Television Standards Committee. These standards define a method for encoding information into the electronic signal that ultimately creates a television

picture. As specified by the NTSC standard, a single frame of video is made up of 525 horizontal scan lines drawn onto the inside face of a phosphor-coated picture tube every 1/30th of a second by a fast-moving electron beam.

PAL

The Phase Alternate Line (PAL) system is used in the United Kingdom, Europe, Australia, and South Africa. PAL is an integrated method of adding color to a black-and-white television signal that paints 625 lines at a frame rate 25 frames per second.

SECAM

The Sequential Color and Memory (SECAM) system is used in France, Russia, and few other countries. Although SECAM is a 625-line, 50 Hz system, it differs greatly from both the NTSC and the PAL color systems in its basic technology and broadcast method.

HDTV

High Definition Television (HDTV) provides high resolution in a 16:9 aspect ratio (see following Figure). This aspect ratio allows the viewing of Cinemascope and Panavision movies. There is contention between the broadcast and computer industries about whether to use interlacing or progressive-scan technologies.

Shooting and Editing Video

To add full-screen, full-motion video to your multimedia project, you will need to invest in specialized hardware and software or purchase the services of a professional video production studio. In many cases, a professional studio will also provide editing tools and post-production capabilities that you cannot duplicate with your Macintosh or PC.

Video Tips

A useful tool easily implemented in most digital video editing applications is "blue screen," "Ultimate," or "chromo key" editing. Blue screen is a popular technique for making multimedia titles because expensive sets are not required. Incredible backgrounds can be generated using 3-D modeling and graphic software, and one or more actors, vehicles, or other objects can be neatly layered onto that background. Applications such as VideoShop, Premiere, Final Cut Pro, and iMovie provide this capability.

Recording Formats

S-VHS video

In S-VHS video, color and luminance information are kept on two separate tracks. The result is a definite improvement in picture quality. This standard is also used in Hi-8. still, if your ultimate goal is to have your project accepted by broadcast stations, this would not be the best choice.

Component (YUV)

In the early 1980s, Sony began to experiment with a new portable professional video format based on Betamax. Panasonic has developed their own standard based on a similar technology, called "MII," Betacam SP has become the industry standard for professional video field recording. This format may soon be eclipsed by a new digital version called "Digital Betacam."

Digital Video

Full integration of motion video on computers eliminates the analog television form of video from the multimedia delivery platform. If a video clip is stored as data on a hard disk, CD-ROM, or other mass-storage device, that clip can be played back on the computer's monitor without overlay boards, videodisk players, or second monitors. This playback of digital video is accomplished using software architecture such as QuickTime or AVI, a multimedia producer or developer; you may need to convert video source material from its still common analog form (videotape) to a digital form manageable by the end user's computer system. So an understanding of analog video and some special hardware must remain in your multimedia toolbox. Analog to digital conversion of video can be accomplished using the video overlay hardware described above, or it can be delivered direct to disk using FireWire cables. To repetitively digitize a full-screen color video image every 1/30 second and store it to disk or RAM severely taxes both Macintosh and PC processing capabilities–special hardware, compression firmware, and massive amounts of digital storage space are required.

Video Compression

To digitize and store a 10-second clip of full-motion video in your computer requires transfer of an enormous amount of data in a very short amount of time. Reproducing just one frame of digital video component video at 24 bits requires almost 1MB of computer data; 30 seconds of video will fill a gigabyte hard disk. Full-size, full-motion video requires that the computer deliver data at about 30MB per second. This overwhelming technological bottleneck is overcome using digital video compression schemes or *codecs* (coders/decoders). A codec is the algorithm used to compress a video for delivery and then decode it in real-time for fast playback. Real-time video compression algorithms such as MPEG, P*64, DVI/Indeo, JPEG, Cinepak, Sorenson, ClearVideo, RealVideo, and VDOwave are available to compress digital video information. Compression schemes use Discrete Cosine Transform (DCT), an encoding algorithm that quantifies the human eye's ability to detect color and image

distortion. All of these codecs employ lossy compression algorithms.

In addition to compressing video data, *streaming* technologies are being implemented to provide reasonable quality low-bandwidth video on the Web. Microsoft, RealNetworks, VXtreme, VDOnet, Xing, Precept, Cubic, Motorola, Viva, Vosaic, and Oracle are actively pursuing the commercialization of streaming technology on the Web.

QuickTime, Apple's software-based architecture for seamlessly integrating sound, animation, text, and video (data that changes over time), is often thought of as a compression standard, but it is really much more than that.

MPEG

The MPEG standard has been developed by the Moving Picture Experts Group, a working group convened by the International Standards Organization (ISO) and the International Electro-technical Commission (IEC) to create standards for digital representation of moving pictures and associated audio and other data. MPEG1 and

MPEG2 are the current standards. Using MPEG1, you can deliver 1.2 Mbps of video and 250 Kbps of two-channel stereo audio using CD-ROM technology. MPEG2, a completely different system from MPEG1, requires higher data rates (3 to 15 Mbps) but delivers higher image resolution, picture quality, interlaced video formats, multiresolution scalability, and multichannel audio features.

DVI/Indeo

DVI is a property, programmable compression/decompression technology based on the Intel i750 chip set. This hardware consists of two VLSI (Very Large Scale Integrated) chips to separate the image processing and display functions. Two levels of compression and decompression are provided by DVI: Production Level

Video (PLV) and Real Time Video (RTV). PLV and RTV both use variable compression rates. DVI's algorithms can compress video images at ratios between 80:1 and 160:1. DVI will play back video in full-frame size and in full color at 30 frames per second.

Optimizing Video Files for CD-ROM

CD-ROMs provide an excellent distribution medium for computer-based video: they are inexpensive to mass produce, and they can store great quantities of information. CDROM players offer slow data transfer rates, but adequate video transfer can be achieved by taking care to properly prepare your digital video files.

- Limit the amount of synchronization required between the video and audio. With Microsoft's AVI files, the audio and video data are already interleaved, so this is not a necessity, but with QuickTime files, you should "flatten" your movie. *Flattening* means you interleave the audio and video segments together.
- Use regularly spaced key frames, 10 to 15 frames apart, and temporal compression can correct for seek time delays. *Seek time* is how long it takes the CD-ROM player to locate specific data on the CD-ROM disc. Even fast 56x drives must spin up, causing some delay (and occasionally substantial noise).
- The size of the video window and the frame rate you specify dramatically affect

performance. In QuickTime, 20 frames per second played in a 160X120-pixel window is equivalent to playing 10 frames per second in a 320X240 window.

The more data that has to be decompressed and transferred from the CD-ROM to the screen, the slower the playback.

Multimedia Authoring

Multimedia authoring is a process of assembling different types of media contents like text, audio, image, animations and video as a single stream of information with the help of various software tools available in the market. Multimedia authoring tools give an integrated environment for joining together the different elements of a multimedia production. It gives the framework for organizing and editing the components of a multimedia project. It enables the developer to create interactive presentation by combining text, audio, video, graphics and animation.

Features of Authoring Tools

- **Editing Features**- Most authoring environment and packages exhibit capabilities to create edit and transform different kinds of media that they support. For example, Macromedia Flash comes bundled with its own sound editor. This eliminates the need for buying dedicated software to edit sound data. So authoring systems include editing tools to create, edit and convert multimedia components such as animation and video clips.
- **Organizing Features**- The process of organization, design and production of multimedia involve navigation diagrams or storyboarding and flowcharting. Some of the authoring tools provide a system of visual flowcharting or overview facility to showcase your project's structure at a macro level. Navigation diagrams help to organize a project. Many web-authoring programs like Dreamweaver include tools that create helpful diagrams and links among the pages of a website.
- **Visual programming with icons or objects**- It is simplest and easiest authoring process. For example, if you want to play a sound then just clicks on its icon.
- **Programming with a scripting language**- Authoring software offers the ability to write scripts for software to build features that are not supported by the software itself. With script you can perform computational tasks - sense user input and respond, character creation, animation, launching other application and to control external multimedia devices.
- **Document Development tools**- Some authoring tools offers direct importing of pre-formatted text, to index facilities, to use complex text search mechanism and to use hypertext link-ing tools.
- **Interactivity Features**- Interactivity empowers the end users to control the content and flow of information of the project. Authoring tools may provide one or more levels of interactivity.
- **Simple branching**- Offers the ability to go to another section of the multimedia production.
- **Conditional branching**- Supports a go to base on the result of IF-THEN decision or events.

- **Playback Features**- When you are developing multimedia project, you will continousally assembling elements and testing to see how the assembly looks and performs. Therefore authoring system should have playback facility.
- **Supporting CD-ROM or Laser Disc Sources**- This software allows over all control of CD-drives and Laser disc to integrate audio, video and computer files. CD-ROM drives, video and laserdisc sources are directly controlled by authoring programs.
- **Supporting Video for Windows**- Videos are the right media for your project which are stored on the hard disk. Authoring software has the ability to support more multimedia elements like video for windows.
- **Hypertext**- Hypertext capabilities can be used to link graphics, some animation and other text. The help system of window is an example of hypertext. Such systems are very useful when a large amount of textual information is to be represented or referenced.
- **Cross-Platform Capability**- Some authoring programs are available on several platforms and provide tools for transforming and converting files and programs from one to the other.
- **Run-time Player for Distribution**- Run time software is often included in authoring software to explain the distribution of your final product by packaging playback software with content. Some advanced authoring programs provide special packaging and run-time distribution for use with devices such as CD-ROM.
- **Internet Playability**- Due to Web has become a significant delivery medium for multimedia, authoring systems typically provide a means to convert their output so that it can be delivered within the context of HTML or DHTML.

Authoring Tools Classification

Card or Page based authoring tools

In these authoring systems, elements are organized as pages of a book or a stack of cards. In the book or stack there are thousand of pages or cards available. These tools are best used when the bulk of your content consists of elements that can be viewed individually, for example the pages of a book or file cards in card file. You can jump from page to page because all pages can be interrelated. In the authoring system you can organize pages or cards in the sequences manner. Every page of the book may contain many media elements like sounds, videos and animations.

One page may have a hyperlink to another page that comes at a much later stage and by clicking on the same you might have effectively skipped several pages in between. Some examples of card or page tools are:

- Hypercard (Mac)
- Tool book (Windows)
- PowerPoint (Windows)
- Supercard (Mac)

Advantages

Following are the advantages of card based authoring tools.

- Easy to understand.
- One screen is equal to 1card or 1page.

- Easy to use as these tools provide template.
- Short development time.

Disadvantages

Following are the disadvantages of card based authoring tools.

- Some run only on one platform.
- Tools not as powerful as equivalent stand alones.

Icon based or Event driven authoring tools

Icon-based tools give a visual programming approach to organizing and presenting multimedia. First you build a structure or flowchart of events, tasks and decisions by dragging appropriate icons from a library. Each icon does a specific task, for example- plays a sound, open an image etc. The flowchart graphically displays the project's logic. When the structure is built you can add your content text, graphics, animation, video movies and sounds. A nontechnical multimedia author can also build sophisticated applications without scripting using icon based authoring tools. Some examples of icon based tools are:

- Authorware Professional (Mac/Windows)
- Icon Author (Windows)

Advantages:

Following are the advantages of icon/event based authoring tools.

- Clear Structure.
- Easy editing and updating

Disadvantages:

Following are the disadvantages of icon/event based authoring tools.

- Difficult to learn.
- Expensive.

Time based authoring tools

Time based authoring tools allow the designer to arrange various elements and events of the multimedia project along a well defined time line. By time line, we simply mean the passage of time. As the time advances from starting point of the project, the events begin to occur, one after another. The events may include media files playback as well as transition from one portion of the project to another. The speed at which these transitions occur can also be accurately controlled. These tools are best to use for those projects, wherein the information flow can be directed from beginning to end much like the movies. Some example of Time based tools are:

- Macromedia's Director
- Macromedia Flash

Advantages

Following are the advantages of time based authoring tools.

- Good for creating animation.
- Branching, user control, interactivity facilities.

Disadvantages

Following are the disadvantages of time based authoring tools.

- Expensive
- Large file size
- Steep learning curve to understand various features.

Object-Oriented authoring tools:

Object oriented authoring tools support environment based on object. Each object has the following two characteristics:

1. **State or Attributes** - The state or attributes refers to the built in characteristics of an object. For example, a color T.V has the following attributes:
    - Color receiver
    - Volume control
    - Picture control
    - 128 channels
    - Remote control unit
2. **Behavior or Operations** - The behavior or operations of an object refers to its action. For example, a T.V can behave in any of the following manner at a given point of time:
    - Switched on
    - Switched off
    - Displays picture and sound from
        - A TV cable connection
        - A TV transmitter
        - A DVD
        - A VCR

In these systems, multimedia elements events are often treated as objects that live in a hierarchical order of parent and child relationships. These objects use messages passed among them to do things according to the properties assigned to them. For example, a video object will likely have a duration property i.e how long the video plays and a source property that is the location of the video file. This video object will likely accept commands from the system such as play and stop. Some examples of the object oriented tools are:

- mTropolis (Mac/Windows)
- Apple Media Tool (Mac/Windows)
- Media Forge (Windows)

Types of Multimedia products

It is efficient and helpful to categorise Multimedia products. The main categories are: Education, entertainment and information.

Educational Products

Educational multimedia products can have an impact on learning that is greater than that of a lecture or talk. Some people argue that they inhibit a person's creative thoughts, as they are focused on many senses at the same time. These products should be seen as a helpful supplementary resource, instead of a substitute for the interaction between a teacher and a student.

Interactive CD-ROMs

Multimedia resources have been used in classrooms for a number of years. One of the first multimedia CD-ROMs was Microsoft's encyclopaedia, Encarta®. This resource let student's access text, video clips and audio from significant instances in history (with a heavy American bias!).

Many products produced on behalf of the NSW board of studies are acclaimed for their innovation. Foe example, the Nardoo River CD-ROM allows students to investigate the effects of human development and activity on a river, spanning decades. Students can sample water quality, research natural history, operate a water usage simulator and listen to a number of video and audio broadcasts.

Multimedia Presentations

Microsoft PowerPoint® allows users to create slides with interactive elements (eg. animations, web links, movies). PowerPoint® can be used effectively to add a visual interest to talks, but there is an issue that many people choose the default templates for their presentation, resulting in many similar displays. Presentation software is the most widely used multimedia application.

Computer Based Training

Computer based training (CBT) uses multimedia to assist user in learning about a topic, or teaching skills to others in the workplace. Someone can receive training on the operation of machinery, or office procedures from CBF. An advantage of CBT over other types of training is that users can retrace their steps as many times as they want, to reaffirm their understanding. CBT is used to assist learning in a specific topic by asking timed multiple-choice questions. For example, students can answer their computing test by selecting or dragging their response or answers on screen. Once a template is created, material can be modified or added to easily. These pre-designed shells can reduce time and cost when developing future material.

Entertainment products

Entertainment drives advancements in computing, and multimedia is an example of this. Games have become one of the most popular applications of multimedia. Early games consisted of either 2-dimensional platform games or computerised copies of board games.

Multimedia games

Today, games have entered a 3rd dimension, where the user can control the camera angle as well, as well as the characters direction and speed. These types of games can have dramatic and use realistic sound effects, and can have complex puzzles that require solving and can immerse the player in a real-world environment. As gaming consoles become more sophisticated, the realism of their graphics increase.

Interactive DVD movies

DVD's do not only offer additional video to that seen on the big screen, but they also allow the viewer to interact with the plot, actors, writers and producers. These changes are encouraged by improved higher capacity of DVD's and CD's that are constantly being advanced.

Interactive Digital Television

With interactive digital television, the audience is involved and it becomes part of the "home entertainment" experience. The viewer can change the camera angle for sports broadcasters, and even choose alternate endings for soap operas!

Information products

Information Kiosks

Many large institutions have replaced human guides with multimedia touch-screen information kiosks. These productions are designed for public places and are usually used at museums, hospital, exhibitions and businesses.

These information kiosks save money on helps desks and allow people to search and inquire for information at their own pace. It also helps stop embarrassment in seeking answers to questions.

Electronic books and magazines

Electronic books are among the first example of computer-based multimedia. These were designed to see if people would opt to read books using a laptop computer screen.

There are many electronic magazines (e-zines) that combine text based stories with animation, audio and video, and can be supplied by a CD or on the Internet. Not many e-zines have been successful. The earliest challenge faced by these was the lack of a cross-platform CD standard format. Another challenge was the difficulty of distribution and attracting paid advertisement. People usually like to browse before they buy, and it is not possible to browse the contents of a CD in a busy newsagency.

Multimedia Databases

Multimedia databases are used to catalogue media. The designer of a multimedia product can assign key words and ideas to video/graphical elements, so they are easily identifiable. These databases can be used for advertisement, catalogues and information systems. Some schools have established digital photographic registries of their enrolled students, which they use for attendance systems. The students swipe their identity card, recording their arrival and departure times.

Other multimedia Products

Communication has become a significant part of the multimedia category recently, with the development of true multimedia mobile phones. Multimedia is associated in so many areas that it can be hard to decide which category something should be placed.

Sometimes, a label can cross over more than a single area, and a new word is coined by the combined categories, eg. Edutainment (education + entertainment) and infotainment (information + entertainment).

Edutainment Products

Edutainment is about the multimedia blend of educational and entertainment elements. Teachers have used this throughout history as an effective teaching method. Edutainment is designed to educate the user, whilst capturing their interest with the entertainment within.

Infotainment Products

Infotainment refers to the multimedia blend of information and entertainment. This is also known as docutainment (documentary + entertainment), as it refers to documentaries that cover issues in an entertaining way. The entertainment used in infotainment products have been known to reduce the seriousness of a subject, and doing the opposite of what it was designed to achieve.

Unit-III
Data Compression

Coding requirement source

In data compression, the coding requirement source refers to the method used to encode the compressed data. This can include various coding techniques such as Huffman coding, arithmetic coding, or run-length encoding.

The coding requirement source is an essential aspect of data compression because it determines how efficiently the compressed data can be stored and transmitted. The goal of coding techniques is to reduce the number of bits required to represent the compressed data while minimizing loss of information.

Different data types and compression algorithms may require different coding techniques. For example, Huffman coding is often used for text data, while run-length encoding may be more effective for images or audio data.

Entropy

Entropy is a fundamental concept in information theory and plays a critical role in data compression. In simple terms, entropy measures the amount of uncertainty or randomness in a given set of data. It is calculated as the average amount of information contained in each symbol of the data.

In data compression, entropy is often used as a measure of the minimum number of bits required to represent the data without loss of information. The entropy of a data source is a theoretical limit on the compression ratio that can be achieved by any compression algorithm. In other words, the lower the entropy, the more efficiently the data can be compressed.

Huffman coding is a popular compression algorithm that uses entropy as a basis for encoding data. The algorithm works by constructing a tree-based codebook, where symbols with lower probability of occurrence are assigned shorter codes. This is possible because symbols with lower probability of occurrence contribute less to the entropy of the data and therefore require fewer bits to represent.

Arithmetic coding is another compression algorithm that makes use of entropy. It works by dividing the data into subintervals based on their probabilities, then encoding each subinterval with a unique code. The length of each code is determined by the size of the subinterval, which is proportional to its probability.

In both of these algorithms, entropy is used as a measure of the inherent redundancy in the data. By removing this redundancy through encoding, the data can be compressed without loss of information. However, it's worth noting that entropy-based compression algorithms are only effective if the data contains some form of statistical structure, such as repeated patterns or symbols.

In summary, entropy is a key concept in data compression that helps measure the amount of information contained in a data source. It's used as a basis for many compression algorithms, such as Huffman coding and arithmetic coding, and can significantly impact the efficiency and effectiveness of the compression process.

Entropy

Shannon borrowed the definition of *entropy* from statistical physics, where entropy represents the randomness or disorder of a system. In particular a system is assumed to have a set of possible states it can be in, and at a given time there is a probability distribution over those states. Entropy

is then defined as:

$$H(S) = \sum_{s \in S} p(s) \log_2 \frac{1}{p(s)}$$

where S is the set of possible states, and $p(s)$ is the probability of state $s \in S$. This definition indicates that the more even the probabilities the higher the entropy (disorder) and the more biased the probabilities the lower the entropy—e.g. if we know exactly what state the system is in then $H(S) = 0$. One might remember that the second law of thermodynamics basically says that the entropy of a closed system can only increase.

In the context of information theory Shannon simply replaced "state" with "message", so S is a set of possible messages, and $p(s)$ is the probability of message $s \in S$. Shannon also defined the notion of the *self information* of a message as

$$i(s) = \log_2 \frac{1}{p(s)} .$$

This self information represents the number of bits of information contained in it and, roughly speaking, the number of bits we should use to encode that message. The definition of self informa- tion indicates that messages with higher probability will contain less information (*e.g.*, a message saying that it will be sunny out in LA tomorrow is less informative than one saying that it is going to snow).

The entropy is then simply a probability weighted average of the self information of each mes- sage. It is therefore the average number of bits of information contained in a message picked at random from the probability distribution. Larger entropies represent larger average information, and perhaps counter-intuitively, the more random a set of messages (the more even the probabili- ties) the more information they contain on average.

Here are some examples of entropies for different probability distributions over five messages.

$$
\begin{aligned}
p(S) &= \{0.25, 0.25, 0.25, 0.125, 0.125\} \\
H &= 3 \times 0.25 \times \log2\ 4 + 2 \times 0.125 \times \log2\ 8 \\
&= 1.5 + 0.75 \\
&= 2.25 \\
p(s) &= \{0.5, 0.125, 0.125, 0.125, 0.125\} \\
H &= 0.5 \times \log2\ 2 + 4 \times 0.125 \times \log2\ 8 \\
&= 0.5 + 1.5 \\
&= 2 \\
p(s) &= \{0.75, 0.0625, 0.0625, 0.0625, 0.0625\}
\end{aligned}
$$

4

$H \quad = 0.75 \times \log2(3\ ) + 4 \times 0.0625 \times \log2\ 16\ -$

$= 0.3 + 1$

$= 1.3$

Note that the more uneven the distribution, the lower the Entropy.

Why is the logarithm of the inverse probability the right measure for self information of a mes- sage? Although we will relate the self information and entropy to message length more formally in Section 3 lets try to get some intuition here. First, for a set of $n = 2^i$ equal probability messages, the probability of each is $1/n$. We also know that if all are the same length, then $\log_2 n$ bits are required

to encode each message. Well this is exactly the self information since $i(S_i) = \log2 \frac{1}{p_i} = \log2\ n$.

Another property of information we would like, is that the information given by two independent messages should be the sum of the information given by each. In particular if messages A and B are independent, the probability of sending one after the other is $p(A)p(B)$ and the information contained is them is

$$i(AB) = \lg \frac{1}{p(A)p(B)} = \lg \frac{1}{p(A)} + \lg \frac{1}{p(A)} = i(A) + i(B)\ .$$

The logarithm is the “simplest” function that has this property.

The Entropy of the English Language

We might be interested in how much information the English Language contains. This could be used as a bound on how much we can compress English, and could also allow us to compare the density (information content) of different languages.

One way to measure the information content is in terms of the average number of bits per character. Table 1 shows a few ways to measure the information of English in terms of bits-per- character. If we assume equal probabilities for all characters, a separate code for each character, and that there are 96 printable characters (the number on a standard keyboard) then each character

	bits/char
bits $\lceil \log(96) \rceil$	7
entropy	4.5
Huffman Code (avg.)	4.7
Entropy (Groups of 8)	2.4
Asymptotically approaches:	1.3
Compress	3.7
Gzip	2.7
BOA	2.0

Table 1: Information Content of the English Language

would take $\lceil \log 96 \rceil = 7$ bits. The entropy assuming even probabilities is $\log 96 = 6.6$ bits/char. If we give the characters a probability distribution (based on a corpus of English text) the entropy is reduced to about 4.5 bits/char. If we assume a separate code for each character (for which the Huffman code is optimal) the number is slightly larger 4.7 bits/char.

Note that so far we have not taken any advantage of relationships among adjacent or nearby characters. If you break text into blocks of 8 characters, measure the entropy of those blocks (based on measuring their frequency in an English corpus) you get an entropy of about 19 bits. When we divide this by the fact we are coding 8 characters at a time, the entropy (bits) per character is 2.4. If we group larger and larger blocks people have estimated that the entropy would approach 1.3 (or lower). It is impossible to actually measure this because there are too many possible strings to run statistics on, and no corpus large enough.

This value 1.3 bits/char is an estimate of the information content of the English language. As- suming it is approximately correct, this bounds how much we can expect to compress English text if we want lossless compression. Table 1 also shows the compression rate of various compressors. All these, however, are general purpose and not designed specifically for the English language. The last one, BOA, is the current state-of-the-art for general-purpose compressors. To reach the

1.3 bits/char the compressor would surely have to "know" about English grammar, standard id- ioms, etc..

A more complete set of compression ratios for the Calgary corpus for a variety of compressors is shown in Table 2. The Calgary corpus is a standard benchmark for measuring compression ratios and mostly consists of English text. In particular it consists of 2 books, 5 papers, 1 bibliography, 1 collection of news articles, 3 programs, 1 terminal session, 2 object files, 1 geophysical data, and 1 bit-map b/w image. The table shows how the state of the art has improved over the years.

JPEG

JPEG is a lossy compression scheme for color and gray-scale images. It works on full 24-bit color, and was designed to be used with photographic material and naturalistic artwork. It is not the ideal format for line-drawings, textual images, or other images with large areas of solid color or a very limited number of distinct colors. The lossless techniques, such as JBIG, work better for such images.

JPEG is designed so that the loss factor can be tuned by the user to tradeoff image size and image quality, and is designed so that the loss has the least effect on human perception. It however does have some anomalies when the compression ratio gets high, such as odd effects across the boundaries of 8x8 blocks. For high compression ratios, other techniques such as wavelet compression appear to give more satisfactory results.

An overview of the JPEG compression process is given in Figure 19. We will cover each of the steps in this process.

The input to JPEG are three color planes of 8-bits per-pixel each representing Red, Blue and Green (RGB). These are the colors used by hardware to generate images. The first step of JPEG compression, which is optional, is to convert these into YIQ color planes. The YIQ color planes are

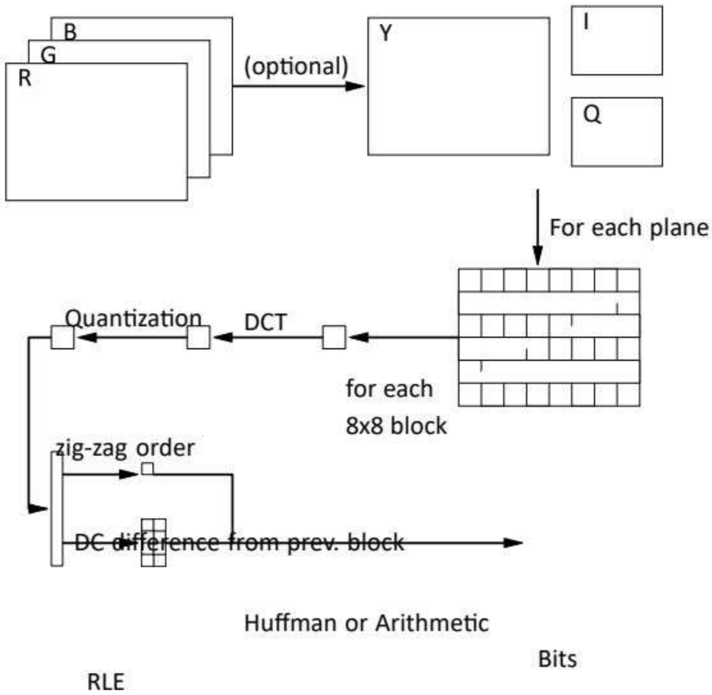


Figure 19: Steps in JPEG compression.

designed to better represent human perception and are what are used on analog TVs in the US (the NTSC standard). The Y plane is designed to represent the brightness (luminance) of the image. It is a weighted average of red, blue and green (0.59 Green + 0.30 Red + 0.11 Blue). The weights are not balanced since the human eye is more responsive to green than to red, and more to red thanto blue. The I (interphase) and Q (quadrature) components represent the color hue (chrominance).If you have an old black-and-white television, it uses only the Y signal and drops the I and Q components, which are carried on a sub-carrier signal. The reason for converting to YIQ is that it is more important in terms of perception to get the intensity right than the hue. Therefore JPEG keeps all pixels for the intensity, but typically down samples the two color planes by a factor of 2 in each dimension (a total factor of 4). This is the first lossy component of JPEG and gives a factor of 2 compression: $(1 + 2 * .25)/3 = .5$.

The next step of the JPEG algorithm is to partition each of the color planes into 8x8 blocks. Each of these blocks is then coded separately. The first step in coding a block is to apply a cosine transform across both dimensions. This returns an 8x8 block of 8-bit frequency terms. So far this does not introduce any loss, or compression. The block-size is motivated by wanting it to be large enough to capture some frequency components but not so large that it causes “frequency spilling”. In particular if we cosine-transformed the whole image, a sharp boundary anywhere in a line would cause high values across all frequency components in that line.

After the cosine transform, the next step applied to the blocks is to use uniform scalar quantization on each of the frequency terms. This quantization is controllable based on user parameters and is the main source of information loss in JPEG compression. Since the human eye is more perceptive to certain frequency components than to others, JPEG allows the quantization scaling factor to be different for each frequency component. The scaling factors are specified using an 8x8 table that simply is used to element-wise divide the 8x8 table of frequency components. JPEG

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 6: JPEG default quantization table, luminance plane.

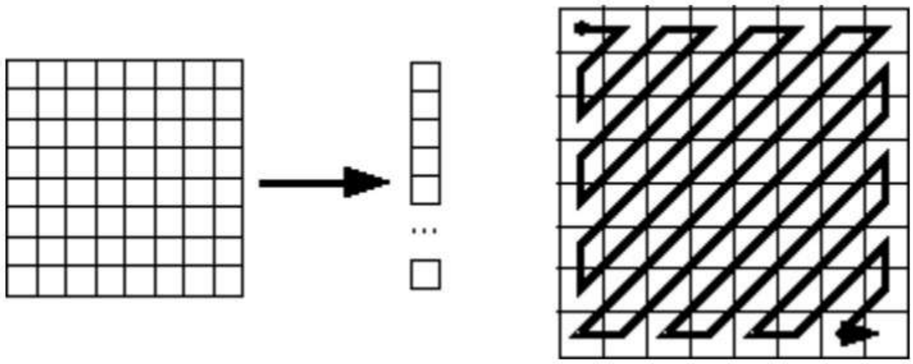


Figure 20: Zig-zag scanning of JPEG blocks.

defines standard quantization tables for both the Y and I-Q components. The table for Y is shown in Table 6. In this table the largest components are in the lower-right corner. This is because these are the highest frequency components which humans are less sensitive to than the lower-frequency components in the upper-left corner. The selection of the particular numbers in the table seems magic, for example the table is not even symmetric, but it is based on studies of human perception. If desired, the coder can use a different quantization table and send the table in the head of the message. To further compress the image, the whole resulting table can be divided by a constant, which is a scalar "quality control" given to the user. The result of the quantization will often drop most of the terms in the lower left to zero.

JPEG compression then compresses the DC component (upper-leftmost) separately from the other components. In particular it uses a difference coding by subtracting the value given by the DC component of the previous block from the DC component of this block. It then Huffman or arithmetic codes this difference. The motivation for this method is that the DC component is often similar from block-to-block so that difference coding it will give better compression.

The other components (the AC components) are now compressed. They are first converted into a linear order by traversing the frequency table in a zig-zag order (see Figure 20). The motiva- tion for this order is that it keeps frequencies of approximately equal length close to each other

Playback order:	0	1	2	3	4	5	6	7	8	9
Frame type:	I	B	B	P	B	B	P	B	B	I
Data stream order:	0	2	3	1	5	6	4	8	9	7

Figure 21: MPEG B-frames postponed in data stream.

in the linear-order. In particular most of the zeros will appear as one large contiguous block at the end of the order. A form of run-length coding is used to compress the linear-order. It is coded as a sequence of (skip,value) pairs, where skip is the number of zeros before a value, and value is the value. The special pair (0,0) specifies the end of block. For example, the sequence [4,3,0,0,1,0,0,0,1,0,0,0,...] is represented as [(0,4),(0,3),(2,1),(3,1),(0,0)]. This sequence is then compressed using either arithmetic or Huffman coding. Which of the two coding schemes used is specified on a per-image basis in the header.

MPEG

Correlation improves compression. This is a recurring theme in all of the approaches we have seen; the more effectively a technique is able to exploit correlations in the data, the more effectively it will be able to compress that data.

This principle is most evident in MPEG encoding. MPEG compresses video streams. In theory, a video stream is a sequence of discrete images. In practice, successive images are highly interrelated. Barring cut shots or scene changes, any given video frame is likely to bear a close resemblance to neighboring frames. MPEG exploits this strong correlation to achieve far better compression rates than would be possible with isolated images.

Each frame in an MPEG image stream is encoded using one of three schemes:

I-frame , or intra-frame, are coded as isolated images.

P-frame , or predictive coded frame, are based on the previous I- or P-frame.

B-frame , or bidirectionally predictive coded frame, are based on either or both the previous and next I- or P-frame.

Figure 21 shows an MPEG stream containing all three types of frames. I-frames and P-frames appear in an MPEG stream in simple, chronological order. However, B-frames are moved so that they appear *after* their neighboring I- and P-frames. This guarantees that each frame appears after any frame upon which it may depend. An MPEG encoder can decode any frame by buffering the two most recent I- or P-frames encountered in the data stream. Figure 21 shows how B-frames are postponed in the data stream so as to simplify decoder buffering. MPEG encoders are free to mix the frame types in any order. When the scene is relatively static, P- and B-frames could be used, while major scene changes could be encoded using I-frames. In practice, most encoders use some fixed pattern.

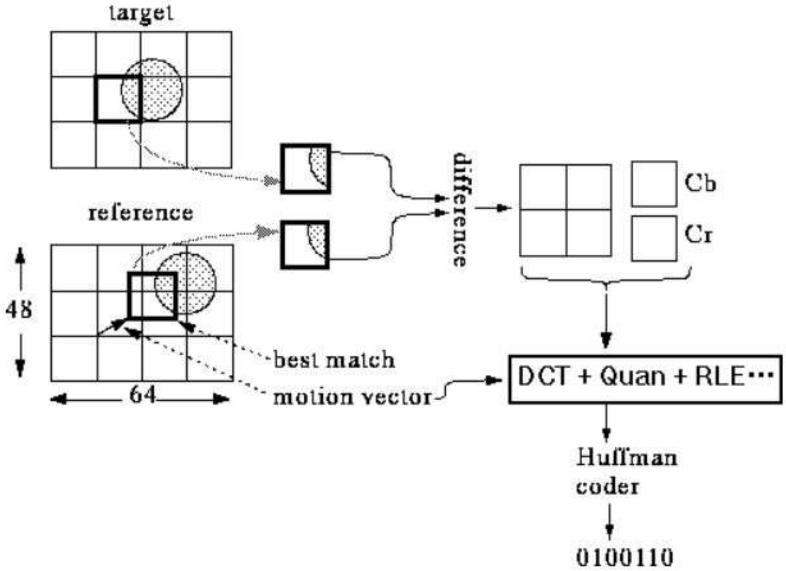


Figure 22: P-frame encoding.

Since I-frames are independent images, they can be encoded as if they were still images. The particular technique used by MPEG is a variant of the JPEG technique (the color transformation and quantization steps are slightly different). I-frames are very important for use as anchor points so that the frames in the video can be accessed randomly without requiring one to decode all previous frames. To decode any frame we need only find its closest previous I-frame and go from there. This is important for allowing reverse playback, skip-ahead, or error-recovery.

The intuition behind encoding P-frames is to find matches, *i.e.*, groups of pixels with similar patterns, in the previous reference frame and then coding the difference between the P-frame and its match. To find these "matches" the MPEG algorithm partitions the P-frame into 16x16 blocks. The process by which each of these blocks is encoded is illustrated in Figure 22. For each *target* block in the P-frame the encoder finds a *reference* block in the previous P- or I-frame that most closely matches it. The reference block need not be aligned on a 16-pixel boundary and can potentially be anywhere in the image. In practice, however, the x-y offset is typically small. The offset is called the *motion vector*. Once the match is found, the pixels of the reference block are subtracted from the corresponding pixels in the target block. This gives a residual which ideally is close to zero everywhere. This residual is coded using a scheme similar to JPEG encoding, but will ideally get a much better compression ratio because of the low intensities. In addition to sending the coded residual, the coder also needs to send the motion vector. This vector is Huffman coded. The motivation for searching other locations in the reference image for a match is to allow for the efficient encoding of motion. In particular if there is a moving object in the sequence of images (*e.g.*, a car or a ball), or if the whole video is panning, then the best match will not be in the same location in the image. It should be noted that if no good match is found, then the block is coded as if it were from an I-frame.

In practice, the search for good matches for each target block is the most computationally expensive part of MPEG encoding. With current technology, real-time MPEG encoding is only possible with the help of custom hardware. Note, however, that while the *search* for a match is expensive, regenerating the image as part of the decoder is cheap since the decoder is given the motion vector and only needs to look up the block from the previous image.

B-frames were not present in MPEG's predecessor, H.261. They were added in an effort to address the following situation: portions of an intermediate P-frame may be completely absent from all previous frames, but may be present in future frames. For example, consider a car entering a shot from the side. Suppose an I-frame encodes the shot before the car has started to appear, and another I-frame appears when the car is completely visible. We would like to use P-frames for the intermediate scenes. However, since no portion of the car is visible in the first I-frame, the P-frames will not be able to "reuse" that information. The fact that the car is visible in a later I-frame does not help us, as P-frames can only look *back* in time, not forward.

B-frames look for reusable data in both directions. The overall technique is very similar to that used in P-frames, but instead of just searching in the previous I- or P-frame for a match, it also searches in the next I- or P-frame. Assuming a good match is found in each, the two reference frames are averaged and subtracted from the target frame. If only one good match is found, then it is used as the reference. The coder needs to send some information on which reference(s) is (are) used, and potentially needs to send two motion vectors.

How effective is MPEG compression? We can examine typical compression ratios for each frame type, and form an average weighted by the ratios in which the frames are typically interleaved.

Starting with a 356×260 pixel, 24-bit color image, typical compression ratios for MPEG-I are:

Type	Size	Ratio
I	18 Kb	7:1
P	6 Kb	20:1
B	2.5 Kb	50:1
Avg	4.8 Kb	27:1

If one 356×260 frame requires 4.8 Kb, how much bandwidth does MPEG require in order to provide a reasonable video feed at thirty frames per second?

$$30 frames/sec \cdot 4.8 Kb/frame \cdot 8 b/bit = 1.2 Mbits/sec$$

Thus far, we have been concentrating on the visual component of MPEG. Adding a stereo audio stream will require roughly another 0.25 Mbits/sec, for a grand total bandwidth of 1.45 Mbits/sec.

This fits nicely within the 1.5 Mbit/sec capacity of a T1 line. In fact, this specific limit was a design goal in the formation of MPEG. Real-life MPEG encoders track bit rate as they encode, and will dynamically adjust compression qualities to keep the bit rate within some user-selected bound. This bit-rate control can also be important in other contexts. For example, video on a multimedia CD-ROM must fit within the relatively poor bandwidth of a typical CD-ROM drive.

MPEG in the Real World

MPEG has found a number of applications in the real world, including:

1. Direct Broadcast Satellite. MPEG video streams are received by a dish/decoder, which unpacks the data and synthesizes a standard NTSC television signal.
2. Cable Television. Trial systems are sending MPEG-II programming over cable television lines.
3. Media Vaults. Silicon Graphics, Storage Tech, and other vendors are producing on-demand video systems, with twenty file thousand MPEG-encoded films on a single installation.
4. Real-Time Encoding. This is still the exclusive province of professionals. Incorporating special-purpose parallel hardware, real-time encoders can cost twenty to fifty thousand dollars.

<u>Text compression</u>

Text compression should be strictly lossless. For example a text file consisting of some financial information is to be compressed. If single digit is changed due to compression it will change the complete meaning. Hence text compression cannot be lossy. Normally statistical encoding techniques are more suitable for text compression due to higher compression ratios.

Static Huffmann Technique

Static Huffman coding is a lossless data compression technique that assigns variable-length codes to each character in the input text based on its frequency of occurrence in the text. The most frequent characters are assigned shorter codes, while the least frequent characters are assigned longer codes. Here is a step-by-step guide to compress text using static Huffman coding.

1. Count the frequency of each character in the input text.

Suppose we have the following input text: "ABBCCCDDDDEEEEE". We can count the frequency of each character as follows:

Character	**Frequency**
A	1
B	2
C	3

Character	Frequency
D	4
E	5

2. Build a Huffman tree based on the character frequencies.

Starting with the leaf nodes (i.e., the characters), we combine the two nodes with the lowest frequencies into a single parent node until we have a single root node. We also assign 0 to the left branch and 1 to the right branch. The resulting Huffman tree for our example is shown below:

mathematicaCopy code

```
        Root
        /  \
      15/    \
      /      \
     E(5)     10/
               / \
             4/D 6/
               / \
              3/C 3/
              / \
           2/B 1/A
```

3. Generate the Huffman code for each character by traversing the Huffman tree.

To generate the Huffman code for a character, we traverse the Huffman tree from the root to the corresponding leaf node, recording 0s and 1s along the way. The resulting Huffman codes for our example are:

Character	Frequency	Huffman Code
A	1	111
B	2	110
C	3	10
D	4	0
E	5	11

4. Replace each character in the input text with its corresponding Huffman code.

We replace each character in the input text with its corresponding Huffman code. The compressed output for our example is:

Copy code

1101101100000000000001111111111111111111111111111111

5. Concatenate all the Huffman codes to form the compressed output.

Finally, we concatenate all the Huffman codes to form the compressed output.

To decompress the compressed text, the same Huffman tree used for compression is needed. The compressed text can be decoded by reading the compressed bits one by one, traversing the Huffman tree from the root to a leaf node until a character is reached.

Dynamic Huffmann Technique

Dynamic Huffman coding is a lossless data compression technique that assigns variable-length codes to each character in the input text based on its frequency of occurrence in the text. Unlike static Huffman coding, dynamic Huffman coding does not require knowledge of the character frequencies in advance, making it suitable for dynamic or streaming data. Here is a step-by-step guide to compress text using dynamic Huffman coding.

1. Initialize the Huffman tree.

We start with an initial Huffman tree that contains only the EOF (end-of-file) symbol.

markdownCopy code

```
     0
    / \
  EOF  -
```

2. Read the input text one character at a time.

As we read each character, we check if it is already in the Huffman tree. If it is, we increment its frequency count and update the Huffman tree accordingly. If it is not, we add a new leaf node to the tree for the character and update the tree.

Suppose we have the following input text: "ABBCCCDDDDEEEEE". Here is how the Huffman tree is updated for each character:

Character	Huffman Tree
A	
B	

Character	Huffman Tree
B	
C	
C	
C	
D	
D	
D	
D	
E	
E	
E	
E	
E	

After processing each character, the Huffman tree is updated as follows:

```
     0
    / \
 EOF   -
        / \
        E   -
       / \
     D   -
    / \
   C   -
  / \
  B   A
```

3. Generate the Huffman code for each character by traversing the Huffman tree.

To generate the Huffman code for a character, we traverse the Huffman tree from the root to the corresponding leaf node, recording 0s and 1s along the way. The resulting Huffman codes for our example are:

Character	Frequency	Huffman Code
A	1	1110
B	2	110
C	3	10
D	4	0
E	5	11

4. Replace each character in the input text with its corresponding Huffman code.

We replace each character in the input text with its corresponding Huffman code. The compressed output for our example is:

Copy code

```
110110110000000000001111111111111111111111111111111111
```

5. Concatenate all the Huffman codes to form the compressed output.

Finally, we concatenate all the Huffman codes to form the compressed output.

To decompress the compressed text, we need to maintain the same Huffman tree used for compression. The compressed text can be decoded by reading the compressed bits one by one, traversing the Huffman tree from the root to a leaf node until a character is reached. Whenever we encounter a leaf node, we output the corresponding character and update the Huffman tree accordingly by adding a new leaf node for the character and incrementing its frequency count. If we encounter the EOF symbol, we stop decoding.

Statistical coding technique

Statistical coding is a lossless data compression technique that compresses text based on the probability distribution of the characters in the input text. It uses fewer bits to represent characters that occur more frequently in the input text and more bits to represent characters that occur less frequently. Here is a step-by-step guide to compress text using statistical coding.

1. Calculate the frequency of each character in the input text.

We first count the number of occurrences of each character in the input text. This information is used to compute the probability distribution of the characters.

2. Generate a probability distribution of the characters.

We compute the probability of each character by dividing its frequency by the total number of characters in the input text. For example, suppose we have the following input text: "ABBCCCDDDDEEEEE". Here is the frequency and probability distribution of the characters:

Character	Frequency	Probability
A	1	0.056
B	2	0.111
C	3	0.167
D	4	0.222
E	5	0.444

3. Generate a code for each character based on its probability.

We assign shorter codes to characters that have higher probability and longer codes to characters that have lower probability. One popular method for generating codes is the Shannon-Fano coding algorithm.

The Shannon-Fano algorithm works by recursively dividing the set of characters into two subsets with approximately equal probabilities until each subset contains only one character. The code for each character is generated by concatenating the bits assigned to it at each level of the tree. Here is the Shannon-Fano tree for our example:

```
      1
     /\
    /  \
   /    \
  /      \
 0        1
/\       /\
/ \     / \
A   B  C   0.5
          /\
         D E
```

The codes for our example are:

Character	Probability	Code
A	0.056	1011

Character	Probability	Code
B	0.111	11
C	0.167	010
D	0.222	00
E	0.444	001

4. Replace each character in the input text with its corresponding code.

We replace each character in the input text with its corresponding code. The compressed output for our example is:

Copy code

110111001000010010010001001001001

5. Concatenate all the codes to form the compressed output.

Finally, we concatenate all the codes to form the compressed output.

To decompress the compressed text, we need to have the same probability distribution and code assignments used for compression. The compressed text can be decoded by reading the compressed bits one by one and tracing a path down the Shannon-Fano tree until a character is reached. Whenever we encounter a leaf node, we output the corresponding character and start over at the root of the tree. If we encounter the EOF symbol, we stop decoding.

Unit-IV

Videodiscs were one of the earliest forms of optical storage, and paved the way for the development of other optical storage devices such as CDs, DVDs, and Blu-ray discs. Here are some notes on videodiscs and their role in the history of optical storage:

1. Videodiscs were first introduced in the 1970s, and were a popular means of storing and playing back video content. They used an analog format, which meant that video and audio were stored as continuous waves on the disc's surface.
2. The two main types of videodiscs were CAV (constant angular velocity) and CLV (constant linear velocity). CAV discs stored one video frame per track, whereas CLV discs stored multiple frames per track.
3. Laserdiscs were a type of videodisc that used a laser to read and write data. They were first introduced in the 1970s and were popular in the home entertainment market until the introduction of DVD in the late 1990s.
4. CD (Compact Disc) was introduced in the 1980s and became one of the most popular forms of optical storage. CDs use a digital format, which means that data is stored as a series of 1s and 0s on the disc's surface. CDs can hold up to 700 MB of data and are commonly used for music, software, and data storage.
5. DVD (Digital Versatile Disc) was introduced in the late 1990s as an upgrade to CDs. DVDs use a digital format and can hold up to 4.7 GB of data. They are commonly used for movies, software, and data storage.
6. Blu-ray Disc was introduced in the mid-2000s and is the latest form of optical storage. Blu-ray discs use a digital format and can hold up to 50 GB of data. They are commonly used for high-definition movies and data storage.

Compact Disc Digital Audio (CD-DA)

Compact Disc Digital Audio (CD-DA) is a format used for storing digital audio on CDs. Here are some notes on CD-DA and a diagram explaining its structure:

1. CD-DA was first introduced in the 1980s and quickly became a popular format for storing and playing back digital audio.
2. CD-DA discs are 12 cm in diameter and can hold up to 74 minutes of audio. They use a spiral track that starts at the center of the disc and moves outward, with each revolution consisting of a single audio frame.
3. CD-DA uses a sampling rate of 44.1 kHz and a bit depth of 16 bits, which provides a high level of fidelity and ensures that the original analog waveform is accurately represented in digital form.

4. CD-DA discs are organized into a series of sectors, each of which contains a fixed number of data bytes. The standard sector size for CD-DA is 2,352 bytes, which corresponds to a single audio frame.
5. CD-DA discs are structured in a way that allows for error correction and detection. The disc includes a table of contents (TOC) that describes the location of each track on the disc, as well as subcode information that provides additional information about each track, such as its start and end times and the track number.
6. CD-DA uses a disc format called Red Book, which defines the technical specifications for CD-DA discs. The Red Book standard specifies the sampling rate, bit depth, and other parameters for CD-DA, as well as the physical and logical layout of the disc.

Compact Disc Digital Audio (CD-DA) has several advantages over other audio formats. Here are some brief notes on the advantages of CD-DA:

1. High sound quality: CD-DA provides high-quality sound with a sampling rate of 44.1 kHz and a bit depth of 16 bits. This provides a high level of fidelity and ensures that the original analog waveform is accurately represented in digital form.
2. Durability: CD-DA discs are highly durable and resistant to scratches, dust, and other environmental factors. This makes them ideal for archival purposes and ensures that the audio on the disc can be played back reliably for many years.
3. Large storage capacity: CD-DA discs can hold up to 74 minutes of audio, which is much more than other audio formats such as cassette tapes or vinyl records. This makes them ideal for storing entire albums or compilations on a single disc.
4. Compatibility: CD-DA is a widely accepted standard that can be played back on most CD and DVD players, as well as personal computers and other digital audio devices. This makes it easy to share and distribute audio recordings in CD-DA format.
5. Error correction: CD-DA includes error correction and detection capabilities that help to ensure accurate playback of the audio data. This ensures that even if the disc is scratched or damaged, the audio can still be played back without any significant loss of quality.

In Compact Disc Digital Audio (CD-DA), the audio data is organized into frames, tracks, and blocks. Here are some notes on each of these components:

1. Frame: A CD-DA frame is the basic unit of audio data on a CD-DA disc. Each frame contains 2,352 bytes of data, which corresponds to 1/75th of a second of audio. Each frame includes subcode information that provides additional information about the audio data, such as the track number and the start and end times of the frame.

2. Track: A CD-DA track is a sequence of frames that contains a single audio recording. Each track is identified by a unique track number and contains a fixed number of frames. The start of each track is marked by a lead-in area, and the end of each track is marked by a lead-out area.
3. Block: A CD-DA block is a group of frames that are read from the disc in a single operation. Each block contains 98 frames, which corresponds to 1.5 seconds of audio. Blocks are used to facilitate error detection and correction during playback.

The organization of CD-DA into frames, tracks, and blocks allows for efficient storage and playback of digital audio. Frames provide a precise timing mechanism that ensures accurate playback of the audio data, while tracks allow for easy navigation between different recordings on the disc. Blocks provide error correction capabilities that help to ensure accurate playback of the audio data, even in the presence of scratches or other damage to the disc.

CDROM

CD-ROM (Compact Disc Read-Only Memory) is a type of optical disc that is used to store digital data. Unlike CD-DA (Compact Disc Digital Audio), which is used to store audio recordings, CD-ROM is designed for storing computer data, such as software programs, multimedia files, and other types of digital information.

Here are some key features and characteristics of CD-ROM:

1. Storage capacity: A standard CD-ROM has a storage capacity of 650-700 MB, which is equivalent to approximately 450 floppy disks or 70 minutes of uncompressed audio.
2. Data access: CD-ROM data is accessed using a laser beam that reads pits and lands on the disc's reflective surface. The data is read in a continuous spiral track from the inside to the outside of the disc.
3. File system: CD-ROMs use a file system called ISO 9660, which is designed to be compatible with multiple operating systems, including Windows, Mac OS, and Unix.
4. Read-only: CD-ROMs are read-only, which means that once data has been recorded onto the disc, it cannot be changed or erased.
5. Durability: CD-ROMs are generally more durable than floppy disks and other forms of magnetic media, and can withstand scratches and other types of damage that might render other types of media unreadable.

CD-ROMs have been widely used for software distribution, multimedia presentations, and other types of digital information storage. While they have largely been replaced by more advanced storage technologies such as DVDs and USB drives, CD-ROMs continue to be used in some applications where their compatibility with legacy systems is important.

CD-ROMs have been used as a medium for software development and distribution for many years. Here are some key ways in which CD-ROMs have been used for software development:

1. Software development kits (SDKs): SDKs are collections of tools, documentation, and sample code that are used by developers to create software applications. CD-ROMs have been used to distribute SDKs for a wide variety of programming languages and platforms, including Java, .NET, and various operating systems.
2. Libraries and frameworks: CD-ROMs have been used to distribute libraries and frameworks that provide common functionality to developers. These include graphics libraries, database access libraries, and user interface frameworks.
3. Documentation: CD-ROMs have been used to distribute documentation for software development, including programming guides, reference manuals, and user guides.
4. Training materials: CD-ROMs have been used to distribute training materials for software development, including video tutorials, interactive lessons, and sample projects.
5. Demos and samples: CD-ROMs have been used to distribute demos and samples of software applications, allowing developers to see how specific features and functionality can be implemented.

Overall, CD-ROMs have been a popular medium for software development and distribution, providing a convenient way to distribute large amounts of data and software tools to developers. While CD-ROMs have largely been replaced by other storage technologies, such as DVDs, USB drives, and cloud-based solutions, they continue to be used in some development contexts, particularly where compatibility with legacy systems is important.

Principle of CD-WO brief notes

CD-WO (Compact Disc-Write Once) is a type of optical disc that is similar to CD-ROMs but can be recorded on once. Here are some key principles of CD-WO:

1. Physical structure: CD-WO discs are made of a polycarbonate substrate that is coated with a reflective layer of metal, such as aluminum. A dye layer is then applied on top of the reflective layer. This dye layer is what allows the disc to be recorded on once.
2. Recording process: CD-WO discs are recorded using a laser beam that heats the dye layer, causing it to change color. This change in color represents the data that is being recorded onto the disc. Once the dye layer has been heated and recorded on, it cannot be changed or erased.
3. Compatibility: CD-WO discs are designed to be compatible with standard CD-ROM drives, meaning that they can be read by any CD-ROM drive. However, they require a special recorder that is capable of writing data onto the disc.

4. Storage capacity: CD-WO discs have a similar storage capacity to CD-ROMs, with a standard capacity of 650-700 MB.
5. Data integrity: CD-WO discs are designed to be resistant to scratches and other forms of damage, which helps to ensure the integrity of the data stored on the disc.
6. Longevity: CD-WO discs are expected to have a lifespan of approximately 50-100 years, making them a reliable storage medium for long-term archival purposes.
7. Applications: CD-WO discs have been used in a variety of applications where data needs to be recorded once and then distributed, such as for archival purposes or software distribution.
8. Cost-effectiveness: CD-WO is a relatively inexpensive storage medium, which makes it an attractive option for applications where large quantities of data need to be stored and distributed.
9. Write speed: CD-WO discs can be recorded at relatively slow speeds, typically ranging from 1x to 4x, which means that it can take some time to record large amounts of data onto a disc.
10. Limitations: CD-WO discs are not suitable for applications where data needs to be updated or modified on a regular basis, as they can only be written to once. Additionally, CD-WO discs are not as durable or reliable as some other storage technologies, such as solid-state drives or cloud-based storage solutions.

CD technology has revolutionized the way we store and distribute digital information. Here are some of the key prospects and benefits of CD technology:

1. High capacity: CDs have a large storage capacity, typically ranging from 650 MB to 700 MB, which makes them ideal for storing large amounts of data, such as music, video, software, and other digital content.
2. Compatibility: CDs are designed to be compatible with a wide range of devices, including CD-ROM drives, CD players, and other types of optical disc drives.
3. Durability: CDs are designed to be durable and resistant to scratches and other forms of damage, which helps to ensure the integrity of the data stored on the disc.
4. Longevity: CDs have a lifespan of approximately 50-100 years, making them a reliable storage medium for long-term archival purposes.
5. Portability: CDs are compact and lightweight, making them easy to transport and distribute. This makes them an ideal medium for distributing music, software, and other digital content.
6. Low cost: CDs are a relatively inexpensive storage medium, which makes them an attractive option for applications where large quantities of data need to be stored and distributed.

7. Easy to duplicate: CDs can be easily duplicated using CD burners, which makes them an ideal medium for creating backup copies of important data.
8. High quality: CDs offer high-quality digital sound and video, making them ideal for storing and distributing music and movies.
9. Versatility: CDs can be used for a wide range of applications, including data storage, music, video, software distribution, and more.

Overall, the prospects of CD technology are vast, and the benefits it provides make it a popular and reliable medium for storing and distributing digital information. Despite the emergence of other storage technologies, such as flash drives and cloud-based storage solutions, CDs remain a popular choice for many applications.

Unit-V

Computer Vision

Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Machines can accurately identify and locate objects then react to what they **"see"** using digital images from cameras, videos, and deep learning models.

Starting in the late 1950s and early 1960s, the goal of image analysis was to mimic human vision systems and to ask computers what they see. Prior to this, image analysis had been completed manually using x-rays, MPIs or hi-res space photography. Nasa's map of the moon took the lead with **digital image processing,** but wasn't fully accepted until 1969.

As computer vision evolved, programming algorithms were created to solve individual challenges. Machines became better at doing the job of vision recognition with repetition. Over the years, there has been a huge improvement of deep learning techniques and technology. We now have the ability to program supercomputers to train themselves, self-improve over time and provide capabilities to businesses as online applications.

Computer imaging

It is a fascinating and exciting area to be involved today. Visual information, transmitted in the form of digital image, is becoming a major method of communication in the modern age.

Computer imaging Can be defined a acquisition and processing of visual information by computer. **The importance of computer imaging** is derived from the fact that our primary sense is our visual sense, and the information that be conveyed in images has been known throughout the centuries to be extraordinary (one picture is worth a thousand words). Computer representation of an image requires the equivalent of many thousands of words of data, so the massive amount of data required for image is a primary reason for the development of many sub areas with field of computer imaging, such as image compression and segmentation

.Another important aspect of computer imaging involves the ultimate "receiver" of visual information in some case the human visual system and in some cases the human visual system and in others the computer itself.

Computer imaging can be separate into two primary categories:

1. Computer Vision.
2. Image Processing.

In computer vision application the processed images output for use by a computer, whereas in image processing applications the output images are for human consumption.

These two categories are not totally separate and distinct. The boundaries that separate the two are fuzzy, but this definition allows us to explore the differences between the two and to explore the difference between the two and to understand how they fit together (Figure 1.1). Historically, the field of image processing grew from electrical engineering as an extension of the signal processing branch, whereas are the computer science discipline was largely responsible for developments in computer vision.

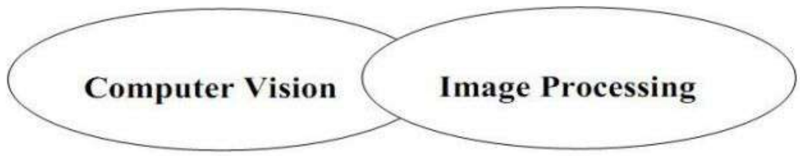


Fig.(1.1) computer imaging

Computer vision is computer imaging where the application doses not involve a human being in visual loop. One of the major topics within this field of computer vision is image analysis.

Image Analysis: involves the examination of the image data to facilitate solving vision problem.

Computer vision systems are used in many and **various types of environments, such as:**

1. Manufacturing Systems
2. Medical Community
3. Law Enforcement
4. Infrared Imaging
5. Satellites Orbiting.

Image processing is computer imaging where application involves a human being in the visual loop. In other words the image are to be examined and a acted upon by people.

There are two categories of the steps involved in the image processing

–

(1) Methods whose **outputs are input** are **images**.

(2) Methods whose outputs are attributes extracted from those images.

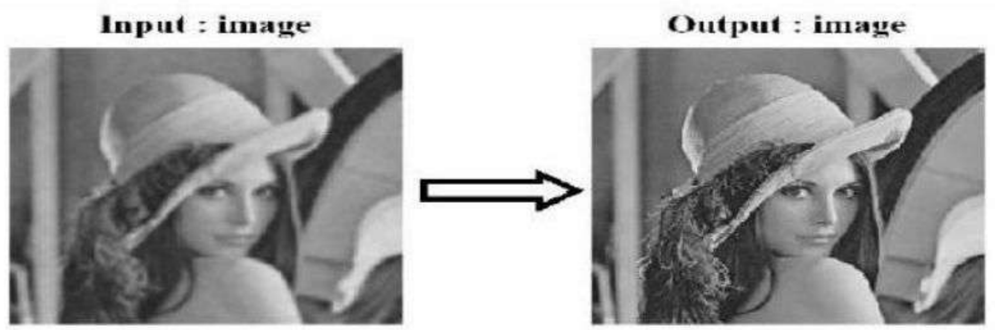
The major topics within the **field of image processing include**:

1. Image restoration.
2. Image enhancement.
3. Image compression.

1. *Image restoration.*

Is the process of taking an image with some known, or estimated

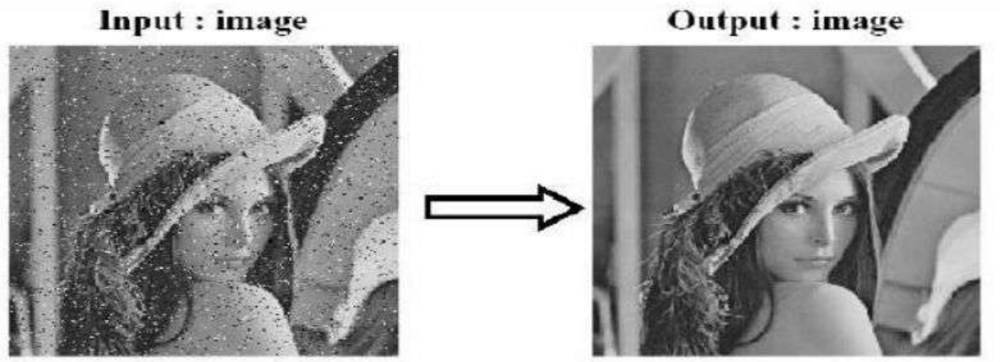
degradation, and restoring it to its original appearance. Image restoration is often used in the field of **photography or publishing** where an image was somehow degraded but needs to be improved before it can be printed.



2. *Image enhancement.*

Involves taking an image and improving it visually, typically by taking advantages of human Visual Systems responses. One of the simplest enhancement techniques is to simply **stretch the contrast of an image**. Enhancement methods tend to be problem specific. For example, a method that is used to enhance satellite images may not suitable for enhancing medical images.

Although enhancement and restoration are similar in aim, to make an **image look better**. They differ in how they approach the problem. **Restoration method** attempt to model the distortion to the image and reverse the degradation, where **enhancement methods** use knowledge of the human visual systems responses to improve an image visually.



3. *Image compression.*

Involves reducing the typically massive amount of data needed to represent an image. This done by eliminating data that are visually

unnecessary and by taking advantage of the redundancy that is inherent in most images. **It has to major approaches**

a) Lossless Compression b) Lossy Compression

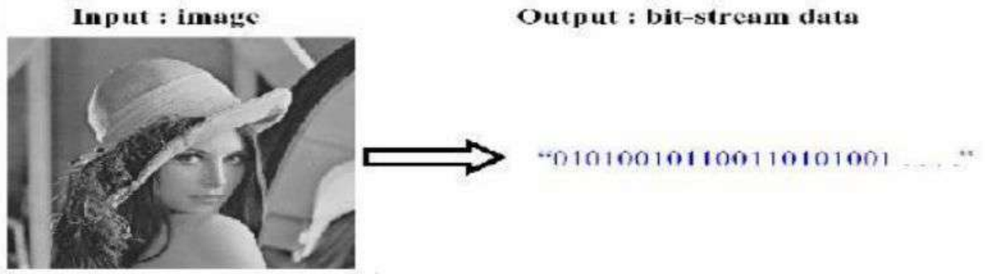


Image processing systems are used in many and various types of environments, such as:

1. Medical community
2. Computer – Aided Design
3. Virtual Reality
4. Image Processing.

Image Formation

In modeling any image formation process, geometric primitives and transformations are crucial to project 3-D geometric features into 2-D features. However, apart from geometric features, image formation also depends on discrete color and intensity values. It needs to know the lighting of the environment, camera optics, sensor properties, etc. Therefore, while talking about image formation in Computer Vision, the article will be focussing on **photometric image formation.**

2.1 Photometric Image Formation

Fig. 1 gives a simple explanation of image formation. The light from a source is reflected on a particular surface. A part of that reflected light goes through an image plane that reaches a sensor plane via optics.

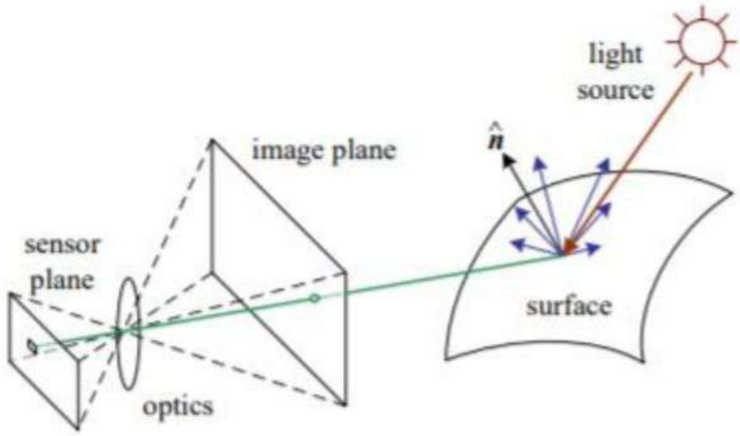


Fig: 1. Photometric Image Formation (Image cr*edit: Szeliski, Computer Vision: Algorithms and Applications 2010)*

Some factors that affect image formation are:

- The strength and direction of the light emitted from the source.
- The material and surface geometry along with other nearby surfaces.
- Sensor Capture properties

2.1.1 Reflection and Scattering

Images cannot exist without light. Light sources can be a point or an area light source. When the light hits a surface, three major reactions might occur-

1. Some light is absorbed. That depends on the factor called ρ (albedo). Low ρ of the surface means more light will get absorbed.
2. Some light gets reflected diffusively, which is independent of viewing direction. It follows **Lambert's cosine law** that the amount of reflected light is proportional to cos(θ). E.g., cloth, brick.

3. Some light is reflected specularly, which depends on the viewing direction. E.g., mirror.

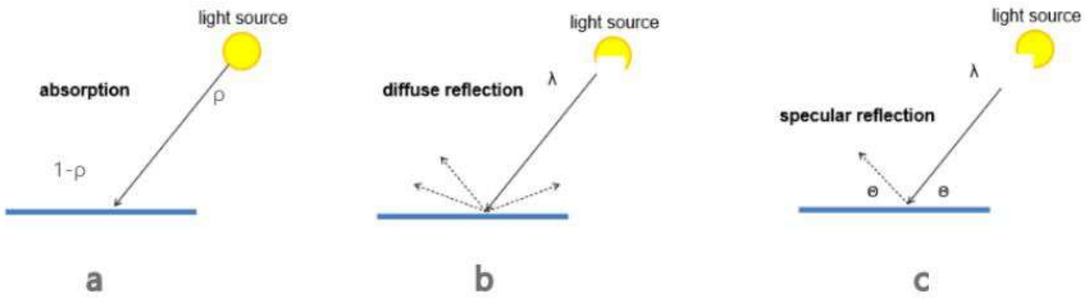


Fig. 2: Models of reflection (Image credits: Derek Hoiem, University of Illinois)

Apart from the above models of reflection, the most common model of light scattering is the **Bidirectional Reflectance Distribution Function (BRDF).** It gives the measure of light scattered by a medium from one direction into another. The scattering of the light can determine the topography of the surface — smooth surfaces reflect almost entirely in the specular direction, while with increasing roughness the light tends to diffract into all possible directions. Eventually, an object will appear equally bright throughout the outgoing hemisphere if its surface is perfectly diffuse (i.e., Lambertian). Owing to this, BRDF can give valuable information about the nature of the target sample.

There are multiple other shading models and ray tracing approaches that are used in unison to properly understand the environment by evaluating the appearance of the scene.

2.1.2 Color

From a viewpoint of color, we know visible light is only a small portion of a large electromagnetic spectrum.

Two factors are noticed when a colored light arrives at a sensor:

- Colour of the light
- Colour of the surface

Bayer Grid/Filter is an important development to capture the color of the light. In a camera, not every sensor captures all the three components (RGB) of light. Inspired by human visual preceptors, Bayers proposed a grid in which there are 50% green, 25 % red, and 25% blue sensors.

Demosaicing algorithm is then used to obtain a full-color image where the surrounding pixels are used to estimate the values for a particular pixel.

There are many such color filters that have been developed to sense colors apart from Bayer Filter.

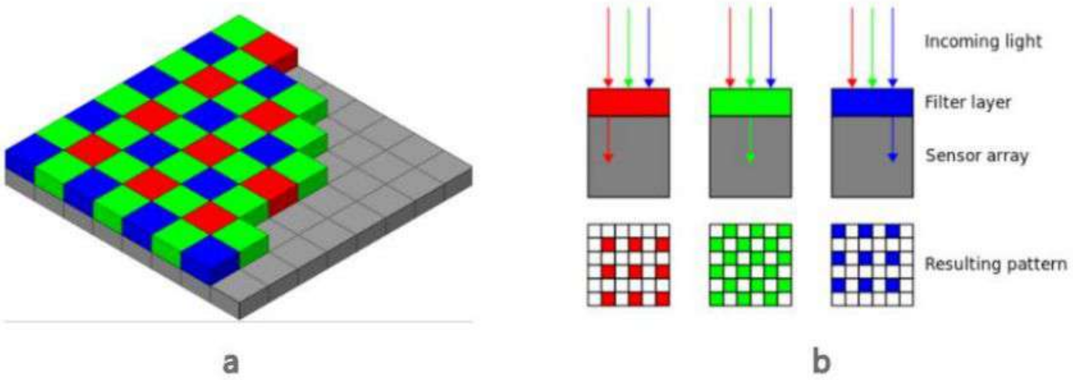


Fig. 3: (a) Bayer arrangement of filters on an image sensor. (b) The cross-section of the sensor. (Image credits: https://en.wikipedia.org/wiki/Bayer_filter)

2.2 Image sensing Pipeline (The digital camera)

The light originates from multiple light sources, gets reflected on multiple surfaces, and finally enters the camera where the photons are converted into the (R, G, B) values that we see while looking at a digital image.

An image sensing pipeline in the camera follows the flowchart that is given in fig. 4.

In a camera, the light first falls on the lens (optics). Following that is the aperture and shutter which can be specified or adjusted. Then the light falls on sensors which can be CCD or CMOS (discussed below), then the image is obtained in an analog or digital form and we get the raw image.

Typically cameras do not stop here. They use demosaic algorithms mentioned in the above topic. Image is sharpened if required or any other important processing algorithms are applied. Post this, white balancing and other digital signal processing tasks are done and the image is finally compressed to a suitable format and stored.

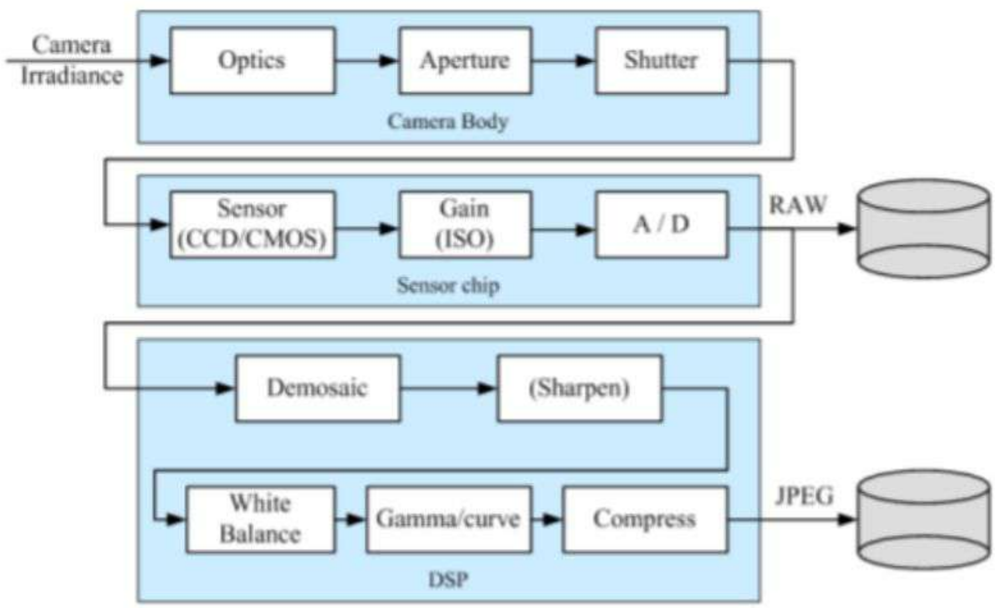


Fig. 4: Image sensing pipeline in a camera (Image credit: *Szeliski, Computer Vision: Algorithms and Applications 2010)*

2.2.1 CCD vs CMOS

The camera sensor can be CCD or CMOS. In charged coupled device (CCD). A charge is generated at each sensing element and this photogenerated charge is moved from pixel to pixel and is converted into a voltage at the output node. Then an analog to digital converter (ADC) converts the value of each pixel to a digital value.

The complementary metal-oxide-semiconductor (CMOS) sensors work by converting charge to voltage inside **each** element as opposed to CCD which accumulates the charge. CMOS signal is digital and therefore does not need ADC. CMOS is widely used in cameras in the current times.

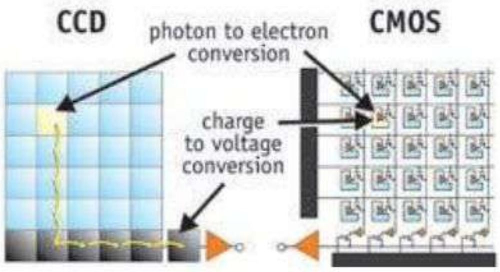


Fig. 5: CCD vs CMOS (Image credit: D. Litwiller, CMOS vs. CCD: Maturing technologies, maturing markets)

2.2.2 Properties of Digital Image Sensor

Let us look at some properties that you may see while clicking a picture on a camera.

Shutter Speed: It controls the amount of light reaching the sensor

Sampling Pitch: It defines the physical space between adjacent sensor cells on the imaging chip.

Fill Factor: It is the ratio of active sensing area size with respect to the theoretically available sensing area (product of horizontal and vertical sampling pitches)

Chip Size: Entire size of the chip

Sensor Noise: Noise from various sources in the sensing process

Resolution: It tells you how many bits are specified for each pixel.

Post-processing: Digital image enhancement methods used before compression and storage.

3.0 Image Representation

After getting an image, it is important to devise ways to represent the image. There are various ways by which an image can be represented. Let's look at the most common ways to represent an image.

3.1 Image as a matrix

The simplest way to represent the image is in the form of a matrix.

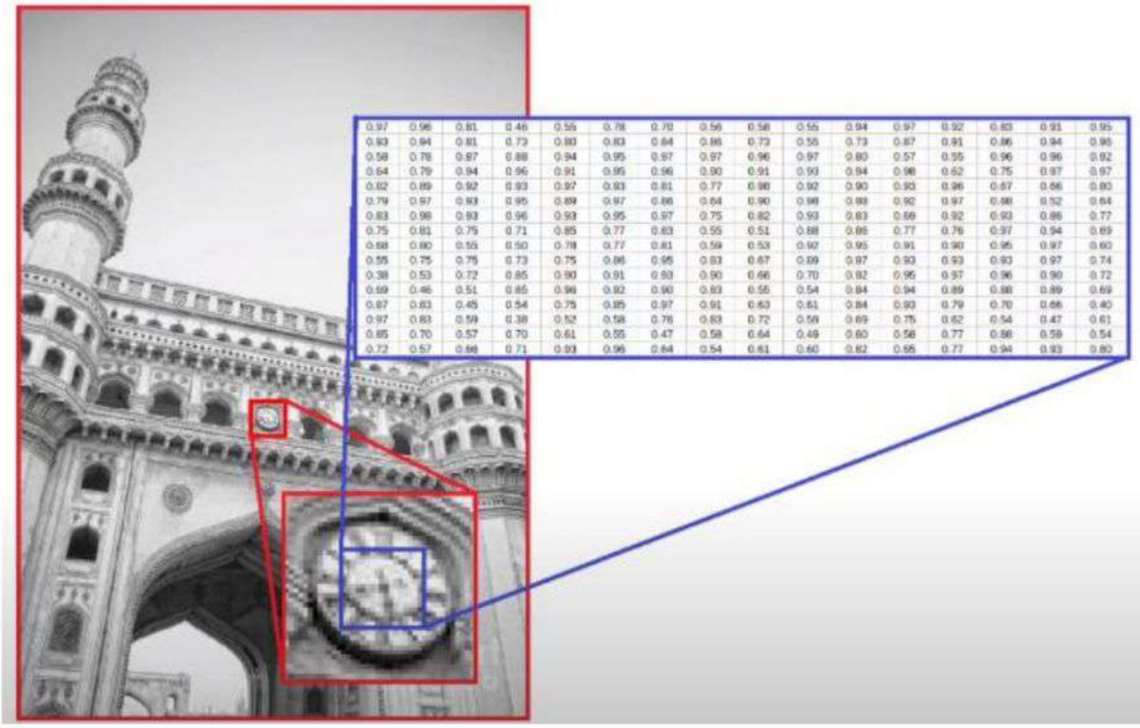


Fig. 6: Representing a part of the image as a matrix (Image credit: IIT, Madras, NPTEL Deep Learning for Computer Vision)

In fig. 6, we can see that a part of the image, i.e., the clock, has been represented as a matrix. A similar matrix will represent the rest of the image too.

It is commonly seen that people use up to a byte to represent every pixel of the image. This means that values between 0 to 255 represent the intensity for each pixel in the image where 0 is black and 255 is white. For every color channel in the image, one such matrix is generated. In practice, it is also common to normalize the values between 0 and 1 (as done in the example in the figure above).

3.2 Image as a function

An image can also be represented as a function. An image (grayscale) can be thought of as a function that takes in a pixel coordinate and gives the intensity at that pixel.

It can be written as function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ that outputs the intensity at any input point (x,y). The value of intensity can be between 0 to 255 or 0 to 1 if values are normalized.

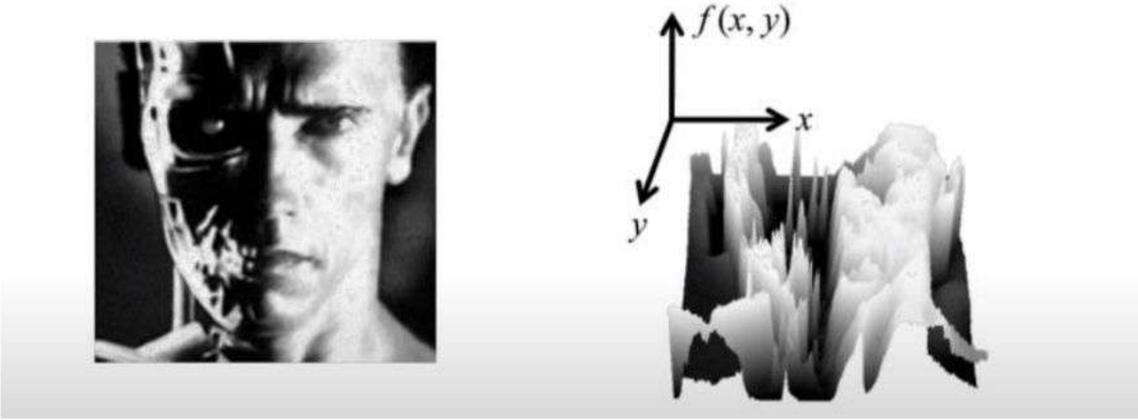


Fig. 7: An image represented as a function (Image credits: Noah Snavely, Cornell University)

3.2.1 Image Transformation

Images can be transformed when they are looked upon as functions. A change in the function can result in changes in the pixel values of the image. Given below are a few examples of the same.

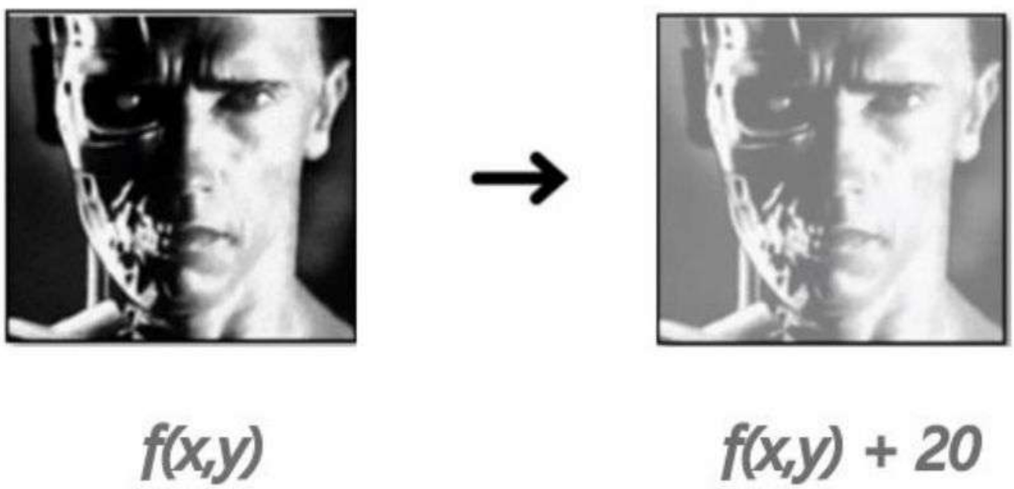


Fig. 8 Image transformation — Lightening the image (Source used: Noah Snavely, Cornell University)

In fig. 8, we wish to make the image lighter. Therefore, for every pixel in the image, we increase the corresponding intensity value. Here we are assuming that the values lie between 0 and 255.

Fig. 9: Image Transformation — Flipping the image (Source used: Noah Snavely, Cornell University)

Similarly, fig. 9 shows the change in the function to flip the image around the vertical axis.

In the above examples, transformation takes place at a pixel level. There are other ways too by which we can perform image transformation.

3.2.2 Image Processing Operations

Essentially, there are three main operations that can be performed on an image.

- Point Operations
- Local Operations
- Global Operations

Given below is an explanation of each of these operations.

3.2.2.1 Point Operation

The examples of image transformations shown above are point operations. In this, the output value depends only on the input value at that particular coordinate.

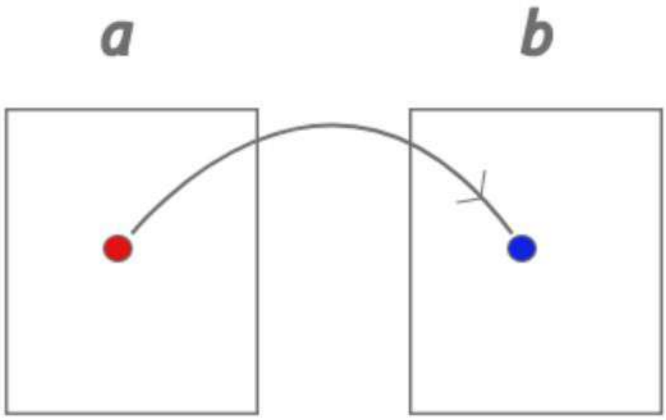


Fig. 10: Point Operation. Output coordinate in ‘b’ only depends on the corresponding input coordinate in ‘a’ (Image credit: Author)

A very famous point operation example that one uses a lot while editing images is reversing the contrast. In the most simple terms, it flips the dark pixels into light pixels and vice versa.

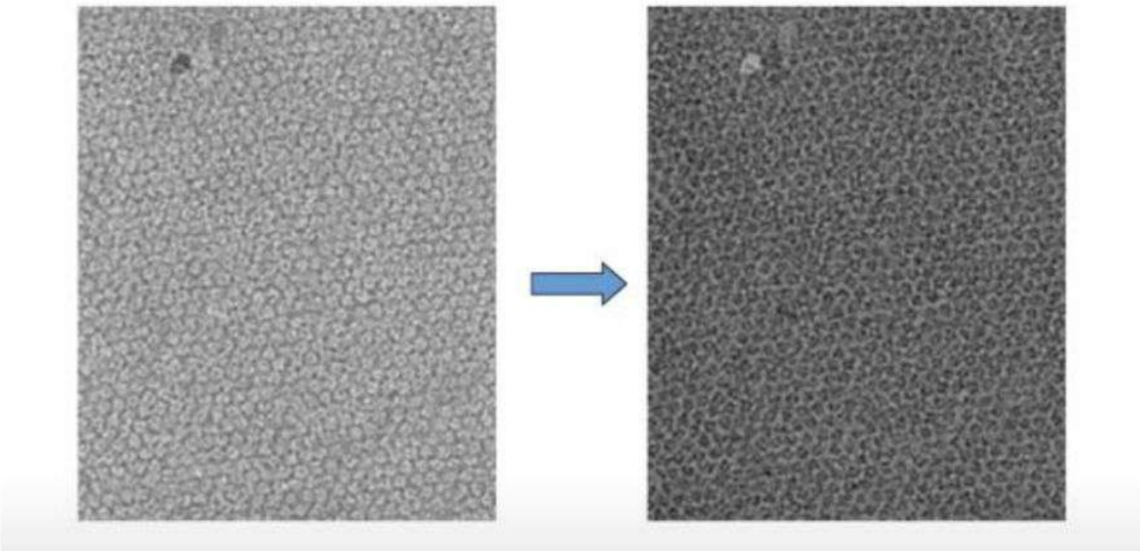


Fig. 11: Reversing the contrast (Image credit: IIT, Madras, NPTEL Deep Learning for Computer Vision)

Fig. 11 displays the application of reversing the contrast. The point operation that helps us to achieve this is stated below.

$$\hat{I}\left(x, y\right) = I_{\max} - I\left(x, y\right) + I_{\min}$$

Fig.12: Point operation to reverse contrast (Image credit: Author)

Here, **I(x,y)** stands for the intensity value at coordinate (x,y) of an image **I**. $\mathbf{I_{max}}$ and $\mathbf{I_{min}}$ refer to the maximum and minimum intensity value of image **I**. For example, say that an image **I** has an intensity between 0 and 255. Therefore, $\mathbf{I_{max}}$ and $\mathbf{I_{min}}$ become 255 and 0 respectively. You wish to flip the intensity value at a coordinate say (x,y) where the current intensity value is 5. By using the above operation, you get the output as : (255) — 5 + 0 = 250 which will be the new value of intensity at coordinate (x,y).

Let's say you clicked a still scene using a camera. But there can be noise in the image due to many reasons like dust particles on the lens, damage in a sensor, and many more. **Noise reduction** using point operations can be very tedious. One way is to take multiple still scenes and average the value at every pixel and hope that the noise gets removed. But at times, it is not possible to get multiple images of a scene and the stillness of a scene can not be guaranteed every time. To do this, we need to move from point operation to local operation.

3.2.2.2 Local Operation

In local operation, as shown in fig. 13, the output value is dependent on the input value and its neighbors.

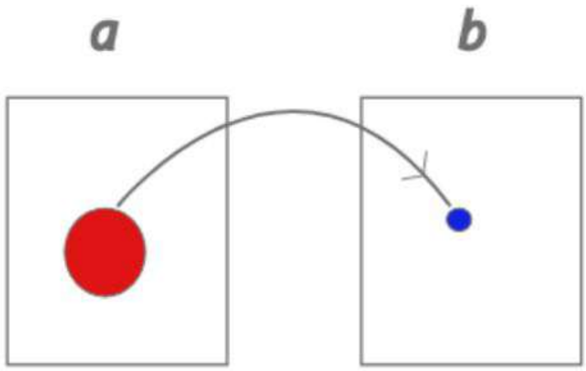


Fig. 13: Local Operation (Image credit: Author)

A simple example to understand local operation is the **moving average**. Suppose an image I as shown in fig. 14. It is clear by looking at the image that it is a white box placed in the dark background. However, we see noise in the picture as a couple of pixels seem to be misplaced (circled in the figure).

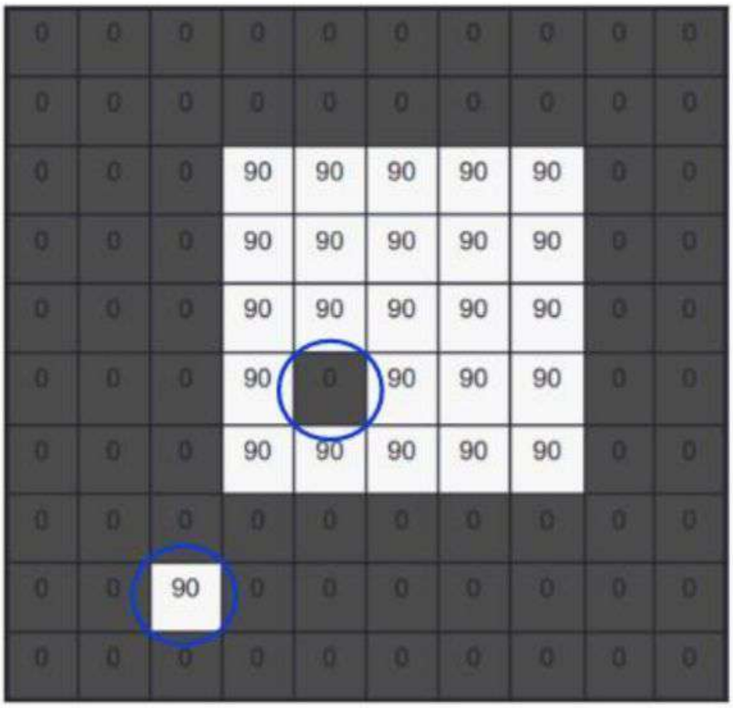


Fig. 14: Intensity values of corresponding pixels of an image (Source used: Steve Seitz, University of Washington)

How do you remove this noise from the image? Assume a 3 X 3 window in the image (any size of the window can be chosen). Move the window across the image and take the average of all the pixels falling within the window. A demonstration of this can be seen below. The final output of the operation can be seen in fig. 15.

Moving Average 2D (Source used to make the visual description: Steve Seitz, Unversity of Washington)

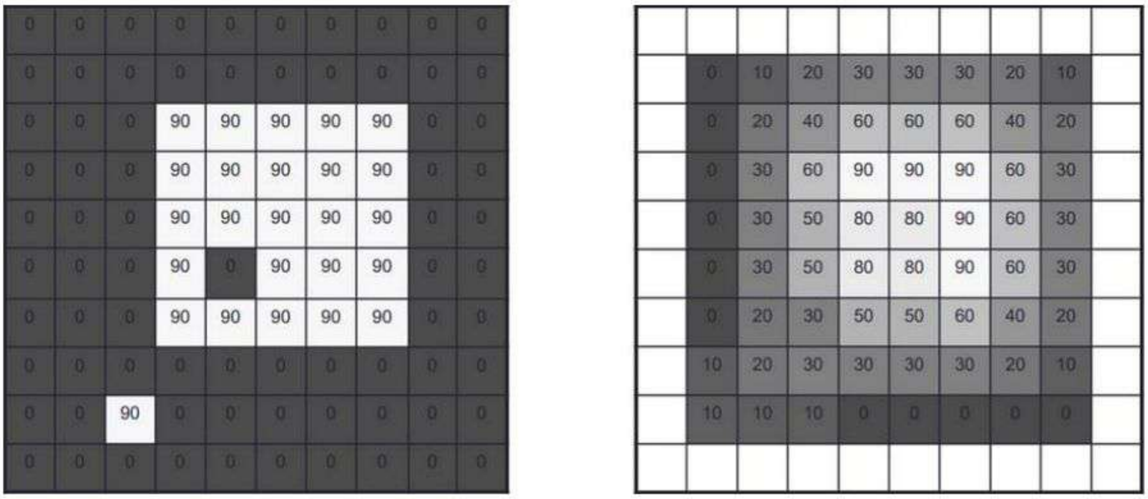


Fig. 15: Final output of the moving average (Image credit: Steve Seitz, University of Washington)

The above operation is a local operation as the output is dependent on the input pixel and its neighbors. Due to the operation, noise pixels in the image are smoothened out in the output.

3.2.2.3 Global Operation

As the name suggests, in global operation, the value at the output pixel is dependent on the entire input image. An example of the global operation is the Fourier transformation which is shown in fig.17.

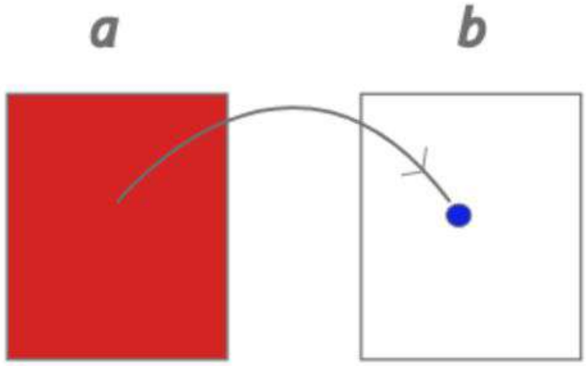


Fig. 16: Global Operation (Image credit: Author)

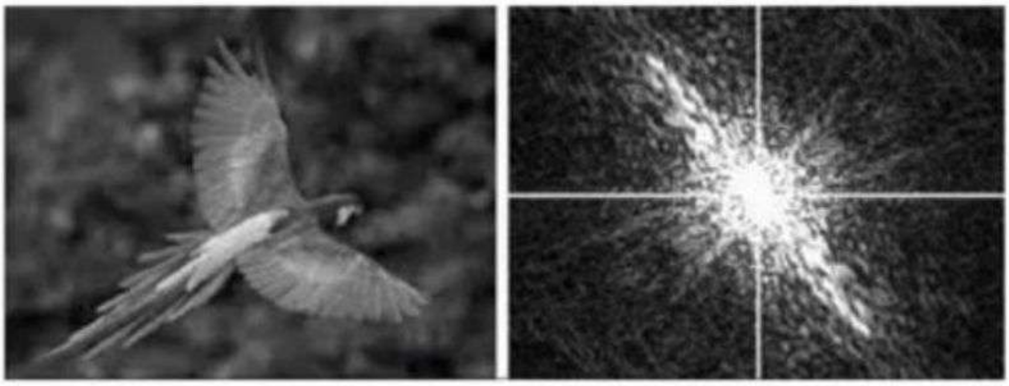


Image Pre-Processing

As a Machine Learning Engineer, data pre-processing or data cleansing is a crucial step and most of the ML engineers spend a good amount of time in data pre-processing before building the model. Some examples for data pre-processing includes outlier detection, missing value treatments and remove the unwanted or noisy data.

Similarly, Image pre-processing is the term for operations on images at the lowest level of abstraction. These operations do not increase image information content but they decrease it if entropy is an information measure. The aim of pre-processing is an improvement of the image data that suppresses undesired distortions or enhances some image features relevant for further processing and analysis task.

There are 4 different types of Image Pre-Processing techniques and they are listed below.

1. Pixel brightness transformations/ Brightness corrections
2. Geometric Transformations
3. Image Filtering and Segmentation
4. Fourier transform and Image restauration

Let's discuss each type in detail.

Pixel brightness transformations(PBT)

Brightness transformations modify pixel brightness and the transformation depends on the properties of a pixel itself. In PBT, output pixel's value depends only on the corresponding input pixel value. Examples of such operators include brightness and contrast adjustments as well as colour correction and transformations.

Contrast enhancement is an important area in image processing for both human and computer vision. It is widely used for medical image processing and as a pre-processing step in speech recognition, texture synthesis, and many other image/video processing applications

There are two types of Brightness transformations and they are below.

1. Brightness corrections
2. Gray scale transformation

The most common Pixel brightness transforms operations are

1. Gamma correction or Power Law Transform
2. Sigmoid stretching
3. Histogram equalization

Two commonly used point processes are multiplication and addition with a constant.

$g(x)=\alpha f(x)+\beta$

The parameters $\alpha>0$ and β are called the gain and bias parameters and sometimes these parameters are said to control contrast and brightness respectively.

cv.convertScaleAbs(image, alpha=alpha, beta=beta)

for different values of alpha and beta, the image brightness and contrast varies.

Gamma Correction

Gamma correction is a non-linear adjustment to individual pixel values. While in image normalization we carried out linear operations on individual pixels, such as scalar multiplication and addition/subtraction, gamma correction carries out a non-linear operation on the source image pixels, and can cause saturation of the image being altered.

Histogram equalization

Histogram equalization is a well-known contrast enhancement technique due to its performance on almost all types of image. Histogram equalization provides a sophisticated method for modifying the dynamic range and contrast of an image by altering that image such that its intensity histogram has the desired shape. Unlike contrast stretching, histogram modelling operators may employ non-linear and non-monotonic transfer functions to map between pixel intensity values in the input and output images.

The normalized histogram.

P(n) = number of pixels with intensity n/ total number of pixels

Sigmoid stretching

Sigmoid function is a continuous nonlinear activation function. The name, sigmoid, is obtained from the fact that the function is "S" shaped. Statisticians call this function the logistic function.

g (x,y) is Enhanced pixel value

c is Contrast factor

th is Threshold value

fs(x,y) is original image

By adjusting the contrast factor 'c' and threshold value it is possible to tailor the amount of lightening and darkening to control the overall contrast enhancement

Geometric Transformations

The earlier methods in this article deal with the colour and brightness/contrast. With geometric transformation, positions of pixels in an image are modified but the colours are unchanged.

Geometric transforms permit the elimination of geometric distortion that occurs when an image is captured. The normal Geometric transformation operations are rotation, scaling and distortion (or undistortion!) of images.

There are two basic steps in geometric transformations:

1. Spatial transformation of the physical rearrangement of pixels in the image

2. Grey level interpolation, which assigns grey levels to the transformed image

change the perspective of a given image or video for getting better insights about the required information. Here the points needs to be provided on the image from which want to gather information by changing the perspective.

Interpolation Methods :

After the transformation methods, the new point co-ordinates (x',y') were obtained. Lets suppose these new points do not in general fit the discrete raster of the output image. So Each pixel value in the output image raster can be obtained by interpolation methods.

The brightness interpolation problem is usually expressed in a dual way. The brightness value of the pixel (x',y') in the output image where x' and y' lie on the discrete raster and it is

Different types of Interpolation methods are

1. Nearest neighbor interpolation is the simplest technique that re samples the pixel values present in the input vector or a matrix

2. Linear interpolation explores four points neighboring the point (x,y), and assumes that the brightness function is linear in this neighborhood.

3. Bicubic interpolation improves the model of the brightness function by approximating it locally by a bicubic polynomial surface.sixteen neighboring points are used for interpolation.

Image Filtering and Segmentation

The goal of using filters is to modify or enhance image properties and/or to extract valuable information from the pictures such as edges, corners, and blobs. A filter is defined by a kernel, which is a small array applied to each pixel and its neighbors within an image

Some of the basic filtering techniques are

1. Low Pass Filtering (Smoothing) : A low pass filter is the basis for most smoothing methods. An image is smoothed by decreasing the disparity between pixel values by averaging nearby pixels

2. High pass filters (Edge Detection, Sharpening) : High-pass filter can be used to make an image appear sharper. These filters emphasize fine details in the image – the opposite of the low-pass filter. High-pass filtering works in the same way as low-pass filtering; it just uses a different convolution kernel.

3. Directional Filtering : Directional filter is an edge detector that can be used to compute the first derivatives of an image. The first derivatives (or slopes) are most evident when a large change occurs between adjacent pixel values.Directional filters can be designed for any direction within a given space

4. Laplacian Filtering : Laplacian filter is an edge detector used to compute the second derivatives of an image, measuring the rate at which the first derivatives change. This determines if a change in adjacent pixel values is from an edge or continuous progression. Laplacian filter kernels usually contain negative values in a cross pattern, centered within the array. The corners are either zero or positive values. The center value can be either negative or positive.

Computer Vision: Low-level Vision

Image Segmentation

Image segmentation is a commonly used technique in digital image processing and analysis to partition an image into multiple parts or regions, often based on the characteristics of the pixels in the image. Image segmentation could involve separating foreground from background, or clustering regions of pixels based on similarities in colour or shape.

Image Segmentation mainly used in

- Face detection
- Medical imaging
- Machine vision
- Autonomous Driving

There are two types of image segmentation techniques.

1. Non-contextual thresholding : Thresholding is the simplest non-contextual segmentation technique. With a single threshold, it transforms a greyscale or colour image into a binary image considered as a binary region map. The binary map contains two possibly disjoint regions, one of them containing pixels with input data values smaller than a threshold and another relating to the input values that are at or above the threshold. The below are the types of thresholding techniques

1. Simple thresholding
2. Adaptive thresholding
3. Colour thresholding

1. Contextual segmentation : Non-contextual thresholding groups pixels with no account of their relative locations in the image plane. Contextual segmentation can be more successful in separating individual objects because it accounts for closeness of pixels that belong to an individual object. Two basic approaches to contextual segmentation are based on signal discontinuity or similarity. Discontinuity-based techniques attempt to find complete boundaries enclosing relatively uniform regions assuming abrupt signal changes across each boundary. Similarity-based techniques attempt to directly create these uniform regions by grouping together connected pixels that satisfy certain similarity criteria. Both the

approaches mirror each other, in the sense that a complete boundary splits one region into two. The below ate the types of Contextual segmentation.

1. Pixel connectivity
2. Region similarity
3. Region growing
4. Split-and-merge segmentation

1. Texture Segmentation : Texture is most important attribute in many image analysis or computer vision applications. The procedures developed for texture problem can be subdivided into four categories.
    1. structural approach
    2. statistical approach
    3. model based approach
    4. filter based approach

Fourier transform

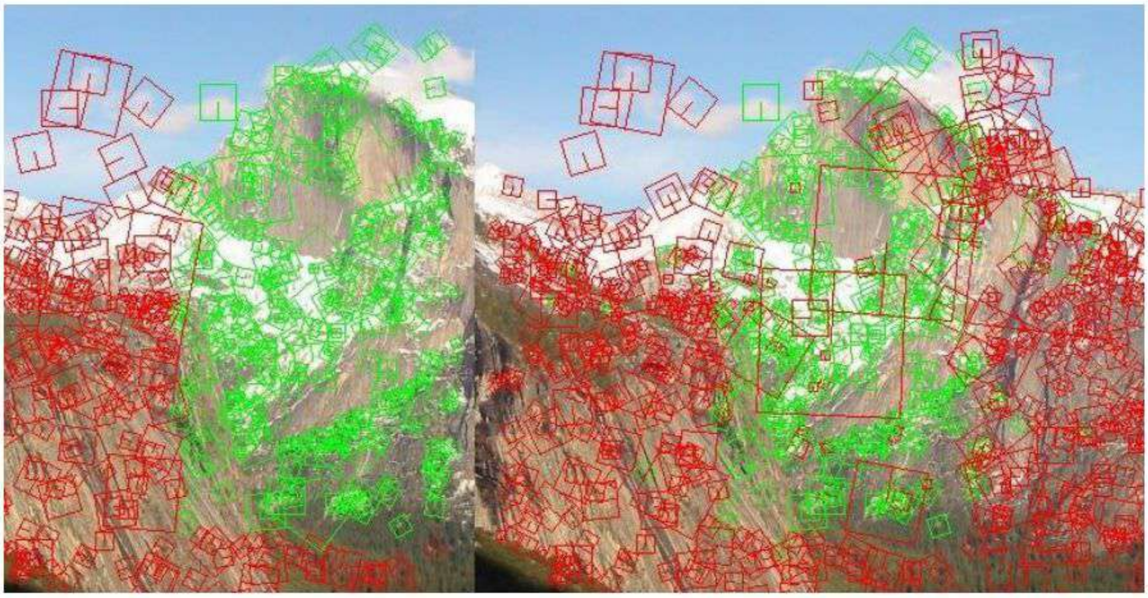
The Fourier Transform is an important image processing tool which is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image.

The Fourier Transform is used in a wide range of applications, such as image analysis, image filtering, image reconstruction and image compression.

The DFT(Discrete Fourier Transform) is the sampled Fourier Transform and therefore does not contain all frequencies forming an image, but only a set of samples which is large enough to fully describe the spatial domain image. The number of frequencies corresponds to the number of pixels in the spatial domain image, i.e. the image in the spatial and Fourier domain are of the same size.

Feature detection and matching

Feature detection and matching is an important task in many computer vision applications, such as structure-from-motion, image retrieval, object detection, and more. In this series, we will be talking about local feature detection and matching.



Application Of Feature Detection And Matching

- Automate object tracking
- Point matching for computing disparity
- Stereo calibration(Estimation of the fundamental matrix)
- Motion-based segmentation
- Recognition
- 3D object reconstruction
- Robot navigation
- Image retrieval and indexing

Feature

A feature is a piece of information which is relevant for solving the computational task related to a certain application. Features may be specific structures in the image such as points, edges or objects. Features may also be the result of a general neighborhood operation or feature detection applied to the image. The features can be classified into two main categories:

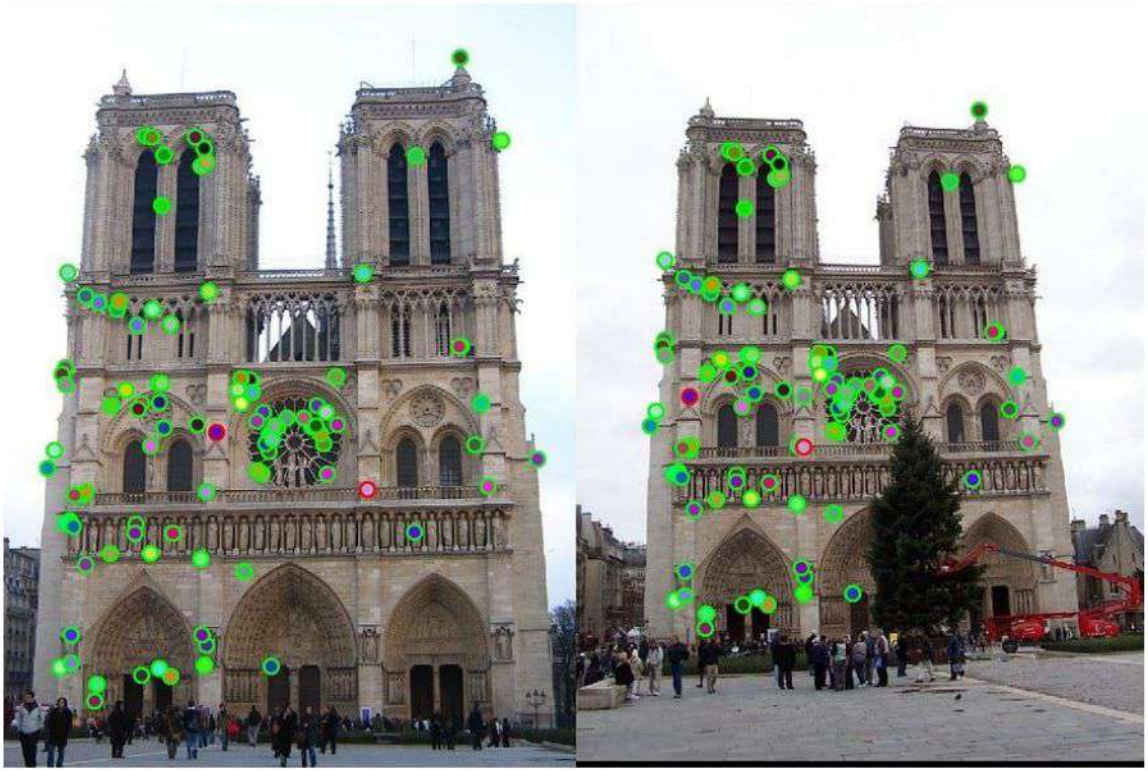
- The features that are in specific locations of the images, such as mountain peaks, building corners, doorways, or interestingly shaped patches of snow. These kinds of localized features are often called **keypoint features** (or even corners) and are often described by the appearance of patches of pixels surrounding the point location.
- The features that can be matched based on their orientation and local appearance (edge profiles) are called **edges** and they can also be good indicators of object boundaries and occlusion events in the image sequence.

Main Component Of Feature Detection And Matching

- **Detection:** Identify the **Interest Point**
- **Description:** The local appearance around each feature point is described in some way that is (ideally) invariant under changes in illumination, translation, scale, and in-plane rotation. We typically end up with a descriptor vector for each feature point.
- **Matching:** Descriptors are compared across the images, to identify similar features. For two images we may get a set of pairs (***Xi, Yi***) ↔ (***Xi`, Yi`***), where (***Xi, Yi***) is a feature in one image and (***Xi`, Yi`***) its matching feature in the other image.

Interest Point

Interest point or Feature Point is the point which is expressive in texture. Interest point is the point at which the direction of the boundary of the object changes abruptly or intersection point between two or more edge segments.



Properties Of Interest Point

- It has a well-defined *position* in image space or well localized.
- It is *stable* under local and global perturbations in the image domain as illumination/brightness variations, such that the interest points can be reliably computed with a high degree of *repeatability*.
- Should provide efficient detection.

Possible Approaches

- Based on the brightness of an image(Usually by image derivative).
- Based on Boundary extraction(Usually by Edge detection and Curvature analysis).

Image Classification Techniques

Image classification refers to a process in computer vision that can classify an image according to its visual content.

Structure for performing Image Classification

1. ***Image Pre-processing:*** The aim of this process is to improve the image data (features) by suppressing unwanted distortions and enhancement of some important image features so that the computer vision models can benefit from this improved data to work on. Steps for image pre-processing includes Reading image, Resizing image, and Data Augmentation (Gray scaling of image, Reflection, Gaussian Blurring, Histogram, Equalization, Rotation, and Translation).

2. ***Detection of an object:*** Detection refers to the localization of an object which means the segmentation of the image and identifying the position of the object of interest.

3. ***Feature extraction and training:*** This is a crucial step wherein statistical or deep learning methods are used to identify the most interesting patterns of the image, features that might be unique to a particular class and that will, later on, help the model to differentiate between different classes. This process where the model learns the features from the dataset is called model training.

4. ***Classification of the object:*** This step categorizes detected objects into predefined classes by using a suitable classification technique that compares the image patterns with the target patterns.

Supervised Classification

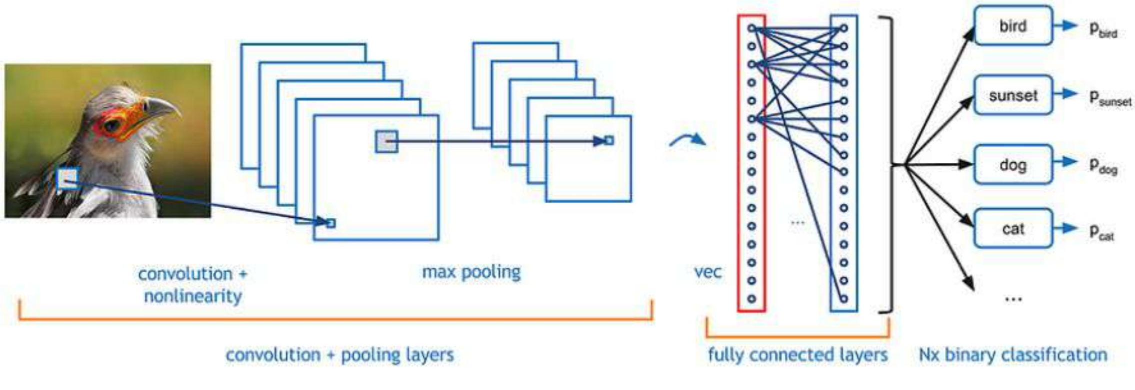
Supervised classification is based on the idea that a user can select sample pixels in an image that are representative of specific classes and then direct the image processing software to use these training sites as references for the classification of all other pixels in the image. Training sites (also known as testing sets or input classes) are selected based on the knowledge of the user. The user also sets the bounds for how similar other pixels must be to group them together. These bounds are often set based on the spectral characteristics of the training area. The user also designates the number of classes that the image is classified into. Once a statistical characterization has been achieved for each information class, the image is then classified by examining the reflectance for each pixel and making a decision about which of the signatures it resembles most. Supervised classification uses classification algorithms and regression techniques to develop predictive models. The algorithms include linear regression, logistic regression, neural networks, decision tree, support vector machine, random forest, naive Bayes, and k-nearest neighbor.

Unsupervised Classification

Unsupervised classification is where the outcomes (groupings of pixels with common characteristics) are based on the software analysis of an image without the user providing sample classes. The computer uses techniques to determine which pixels are related and groups them into classes. The user can specify which algorithm the software will use and the desired number of output classes but otherwise does not aid in the classification process. However, the user must have knowledge of the area being classified when the groupings of pixels with common characteristics produced by the computer have to be related to actual features on the ground. Some of the most common algorithms used in unsupervised learning include cluster analysis, anomaly detection, neural networks, and approaches for learning latent variable models.

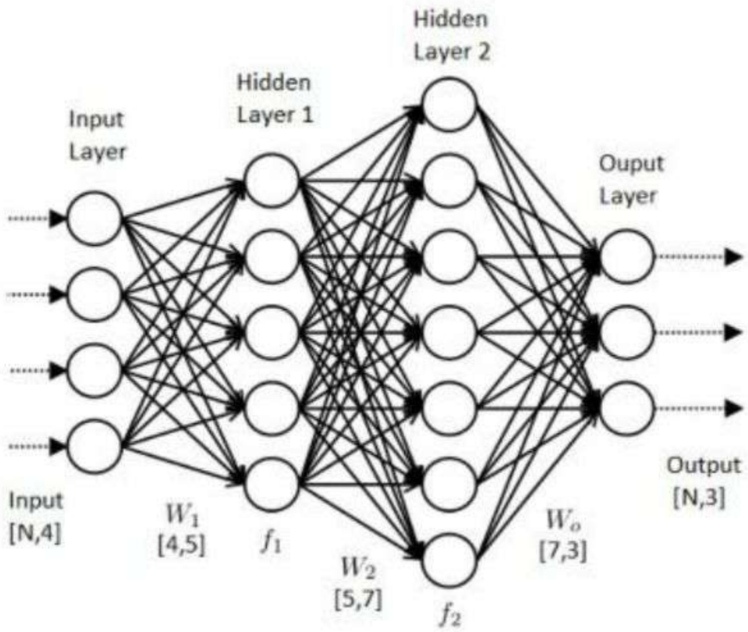
Convolutional Neural Network

Convolutional Neural Network (CNN, or ConvNet) are a special kind of multi-layer neural networks, designed to recognize visual patterns directly from pixel images with minimal pre-processing. It is a special architecture of artificial neural networks. Convolutional neural network uses some of its features of visual cortex and have therefore achieved state of the art results in computer vision tasks. Convolutional neural networks are comprised of two very simple elements, namely convolutional layers and pooling layers. Although simple, there are near-infinite ways to arrange these layers for a given computer vision problem. The elements of a convolutional neural network, such as convolutional and pooling layers, are relatively straightforward to understand. The challenging part of using convolutional neural networks in practice is how to design model architectures that best use these simple elements. The reason why convolutional neural network is hugely popular is because of their architecture, the best thing is there is no need of feature extraction. The system learns to do feature extraction and the core concept is, it uses convolution of image and filters to generate invariant features which are passed on to the next layer. The features in next layer are convoluted with different filters to generate more invariant and abstract features and the process continues till it gets final feature/output which is invariant to occlusions. The most commonly used architectures of convolutional neural network are LeNet, AlexNet, ZFNet, GoogLeNet, VGGNet, and ResNet.



Artificial Neural Network

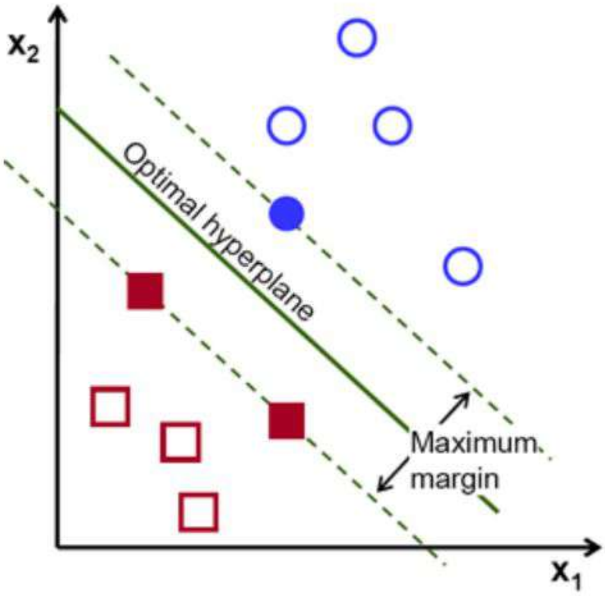
Inspired by the properties of biological neural networks, Artificial Neural Networks are statistical learning algorithms and are used for a variety of tasks, from relatively simple classification tasks to computer vision and speech recognition. Artificial neural networks are implemented as a system of interconnected processing elements, called nodes, which are functionally analogous to biological neurons. The connections between different nodes have numerical values, called weights, and by altering these values in a systematic way, the network is eventually able to approximate the desired function. The hidden layers can be thought of as individual feature detectors, recognizing more and more complex patterns in the data as it is propagated throughout the network. For example, if the network is given a task to recognize a face, the first hidden layer might act as a line detector, the second hidden takes these lines as input and puts them together to form a nose, the third hidden layer takes the nose and matches it with an eye and so on, until finally the whole face is constructed. This hierarchy enables the network to eventually recognize very complex objects. The different types of artificial neural network are convolutional neural network, feedforward neural network, probabilistic neural network, time delay neural network, deep stacking network, radial basis function network, and recurrent neural network.



Support Vector Machine

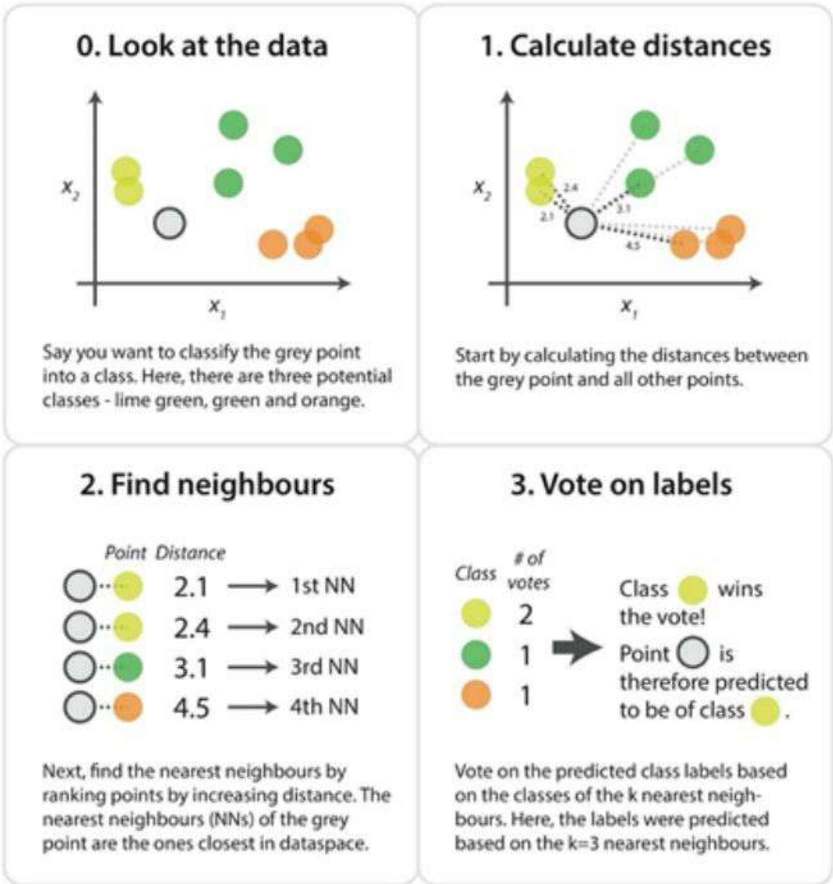
Support vector machines (SVM) are powerful yet flexible supervised machine learning algorithms which are used both for classification and regression. Support vector machines have their unique way of implementation as compared to other machine learning algorithms. They are extremely popular because of their ability to handle multiple continuous and categorical variables. Support Vector Machine model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by support vector machine so that the error can be minimized. The goal is to divide the datasets into classes to find a maximum marginal hyperplane. It builds a hyper-plane or a set of hyper-planes in a high dimensional space and good separation between the two classes is achieved by the hyperplane that has the largest distance to the nearest training data point of any class.

The real power of this algorithm depends on the kernel function being used. The most commonly used kernels are linear kernel, gaussian kernel, and polynomial kernel.



K-Nearest Neighbor

K-Nearest Neighbor is a non-parametric method used for classification and regression. In both cases, the input consists of the *k* closest training examples in the feature space. It is by far the simplest algorithm. It is a non-parametric, lazy learning algorithm, where the function is only approximated locally and all computation is deferred until function evaluation. This algorithm simply relies on the distance between feature vectors and classifies unknown data points by finding the most common class among the k-closest examples. In order to apply the k-nearest Neighbor classification, we need to define a distance metric or similarity function, where the common choices include the Euclidean distance and Manhattan distance. The output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor. Condensed nearest neighbor (CNN, the Hart algorithm) is an algorithm designed to reduce the data set for K-Nearest Neighbor classification.



Naïve Bayes Algorithm

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. All naive bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. Naive Bayes algorithm is a fast, highly scalable algorithm, which can be used for binary and multi-class classification. It depends on doing a bunch of counts. It is a popular choice for text classification, spam email classification, etc. It can be easily trained on small dataset. It has limitation as it considers all the features to be unrelated, so it cannot learn the relationship between features. Naive Bayes can learn individual features importance but can't determine the relationship among features. Different types of naïve bayes algorithms are gaussian naïve bayes, multinomial naïve bayes, and bernoulli naïve bayes.

Random Forest Algorithm

Random forest is a supervised learning algorithm which is used for both classification as well as regression. As we know that a forest is made up of trees and more trees means more robust forest, similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result. The random forest is a classification algorithm consisting of many decision trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.