الجامعة التكنولوجية

قسم علوم الحاسوب

Digital video processing

4th class –Multimedia branch

Lect. Teaba W. khairi

First course 2023-2024

University of Technology- Iraq

Computer science Dept.

DIGITAL VIDEO PROCESSONG

Fourth Class,
M M branch, First semester, 2022-2023.

SIS Id : UG_S1_004

Lecturer : Teaba W. khairi

computer sciences Dept. \ University of technology - Iraq

To contact me :

https://www.youtube.com/channel/UC8oLRPZj2gL0wRIO1UWZZMw	
@teaba83	
teaba.w.khairi@uotechnology.edu.iq	
https://www.linkedin.com/in/teaba-wala-aldeen-b54261160/	
https://www.researchgate.net/profile/Teaba-Khairi?ev=hdr_xprf&_sg=G77Gv0gFRDKb3j4awR2tyvXkMnTW1QRXiezatIzmOEbtwYJAuGpAc2gTYknUHHy7KqmSuk7bFfpS8iJBi1yayaxw	

4th Class first course/ Digital video processing

- Basics of video (types of video signals, component video, composite video, s-video).
- Analog video (NTSC video, PAL video, SECAM video).
- Digital video.
- Color video representations.
- Characteristics of video streams.
- Video formats.
- Video compression.
- Video equipment and applications.
- Motion Estimation and Transform Coding.
- Video Modeling and Retrieval.
- Video Transcoding.
- Video quality evaluation methods and metrics (Monitoring and QoS Measurement, Video Quality Measurements).

References:

1 . MULTIMEDIA IMAGE and VIDEO PROCESSING, Ling Guan, 2000.

2. Video and Multimedia Transmissions over Cellular Networks Analysis, Modelling and Optimization in Live 3G Mobile Communications, **Editor Markus Rupp, 2009**

Lecture one

1.1 The History of Digital Video

Digital video is an electronic representation of moving visual images (video) in the form of encoded digital data. This is in contrast to analog video, which represents moving visual images in the form of analog signals. Digital video comprises a series of digital images displayed in rapid succession.

Digital Video is audio and visual mixed together to make a production. the data gathered used to create a video, rather than a series of photos put together. Digital video have many advantages such as easy copying, multicasting, easy sharing and storage. Digital video is made of images displayed rapidly frequencies of 15, 24,30, and 60 frames per second. There is a saying "A picture is worth a thousand words." Pertaining to Digital Video the saying is "a video represents a million of those words strung together"

Digital video was first introduced commercially in 1986 with the Sony D1 format, which recorded an uncompressed standard-definition component video signal in digital form. In addition to uncompressed formats, popular compressed digital video formats today include H.264 and MPEG-4. Modern interconnect standards used for playback of digital video include HDM as an example.

Digital video can be copied and reproduced with no degradation in quality. In contrast, when analog sources are copied, they experience generation loss. Digital video can be stored on digital media such as Blu-ray Disc, on computer data storage, or streamed over the Internet to end users who watch content on a desktop computer screen or a digital smart TV. Today, digital video content such as TV shows and movies also include a digital audio soundtrack.

1.2 Basics of Video

Definition: *It is the technology of electronically capturing, recording, processing, storing, transmitting, and display of moving visual sequence of still images representing scenes in motion.* Video systems vary greatly in the resolution of the display and refresh rate. Video can be carried on a variety of media, including broadcast, tapes, DVDs, computer files etc.

1.3 Types of color video signals

Video signals can be organized in three different ways: Component video, Composite video, and S-video.

i- Component Video (3 wires)

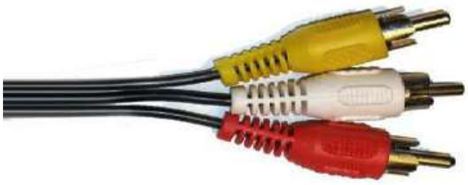
One way of maintaining signal clarity is by separating the components of a video signal so that they do not interfere with each other.

Component video: Higher-end video systems make use of three separate video signals for the red, green, and blue image planes. Each color channel is sent as a separate video signal.

(a) Most computer systems use Component Video, with separate signals for R, G, and B signals.

(b) For any color separation scheme, Component Video gives the best color reproduction since there is no "crosstalk" between the three channels.

(c) Component video requires more bandwidth and good synchronization of the three components.




ii- Composite Video- (1 wires)

In contrast to component video, all video information, chrominance and luminance are mixed together into a single carrier wave.

a) Chrominance is a composition of two color components.

b) The chrominance and luminance components can be separated at the receiver end and then the two color components can be further recovered.

c) When connecting to TVs Composite Video uses only one wire and video color signals are mixed, not sent separately. The audio and sync signals are additions to this one signal.

Dot-Crawl is a defect that results from crosstalk due to the intermodulation of the chrominance and luminance components of the signal, where Dot-crawl affects the edges of color and manifests itself as moving dots of color along these edges. Dot-Crawl can be eliminated by using an S-Video, or component video connection. Composite-video cables do not carry audio and are often paired with audio cables.This is the type of signal used by broadcast color TVs.

Enlarged detail from a video source exhibiting dot crawl.
Note the distinctive checkerboard pattern on the vertical edges between yellow and blue areas.

iii- Separate Video (S-Video)

S-Video: e.g., in S-VHS uses two wires, one for luminance and another for a composite chrominance signal.

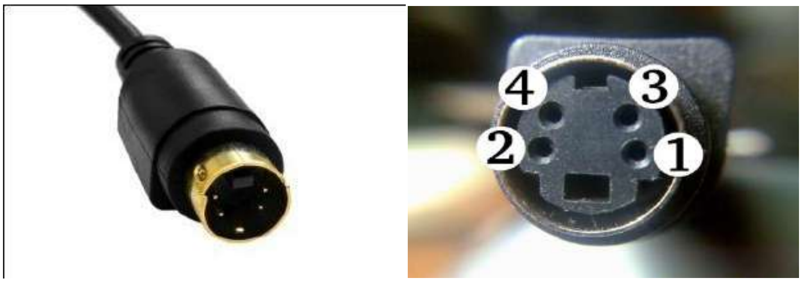
a) As a result, there is less crosstalk between the color information and the crucial gray-scale information.

b) The reason for placing luminance into its own part of the signal is that black-and-white information is most crucial for visual perception.

– In fact, humans are able to differentiate spatial resolution in grayscale images with a much higher acuity than for the color part of color images.

– As a result, we can send less accurate color information than must be sent for intensity information — we can only see fairly large blobs of color, so it makes sense to send less color detail.

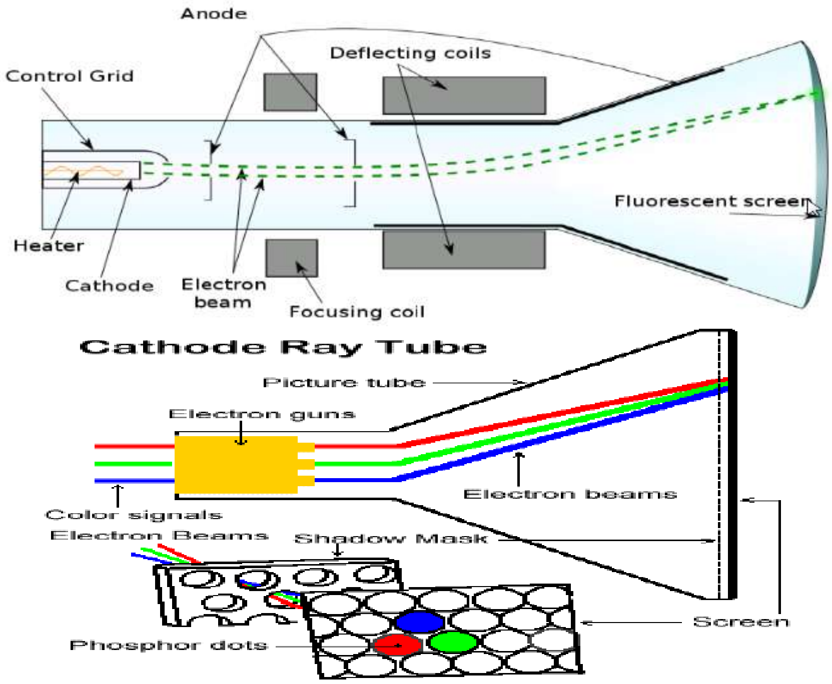
It does not carry audio on the same cable. The infamous dot crawl is eliminated.



A standard 4-pin S-Video cable connector, with each signal pin (3, 4) paired with its own ground pin (1,2)

1.4 Video Display

In conventional TV sets or monitors, the video signal is displayed using a CRT (Cathode Ray Tube). An electron beam sweeps the screen from top to bottom beam carrying the corresponding pattern information, such as intensity in a viewed scene.



1.5 Analog video

The analog signals were used in many systems to produce signals to carry information. These signals are continuous in both values and time. The use of analog signals has been declined with the arrival of digital signals. In short, to understand analog signals ,all signals that are natural or come naturally are analog signals.

- NTSC is an abbreviation for National Television Standards Committee, named for the group that originally developed the black & white and subsequently color television system that is used in the United States, Japan and many other countries. An NTSC picture is made up of 525 interlaced lines and is displayed at a rate of 29.97 frames per second.

- PAL is an abbreviation for Phase Alternate Line. This is the video format standard used in many European countries. A PAL picture is made up of 625 interlaced lines and is displayed at a rate of 25 frames per second.

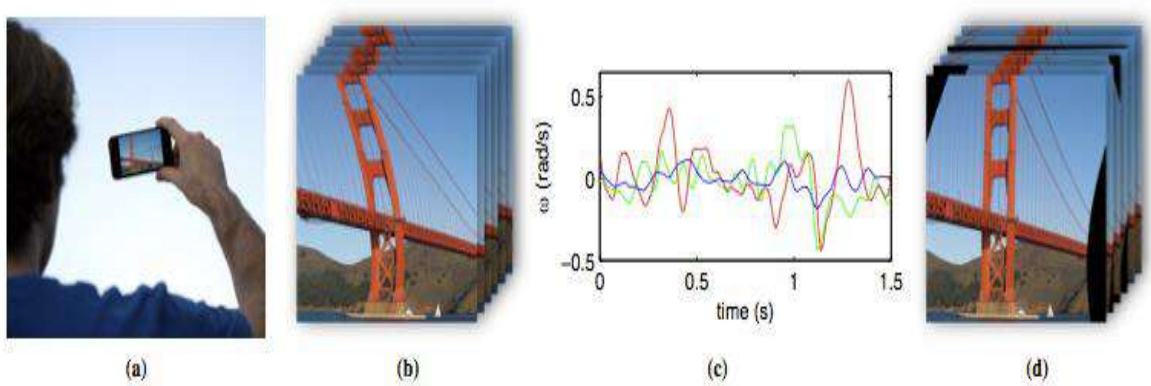
- SECAM is an abbreviation for Sequential Color and Memory. This video format is used in many Eastern countries such as the USSR, China, Pakistan, France, and a few others. Like PAL, a SECAM picture is also made up of 625 interlaced lines and is displayed at a rate of 25 frames per second. However, the way SECAM processes the color information, it is not compatible with the PAL video format standard.

1.6 Digital Video

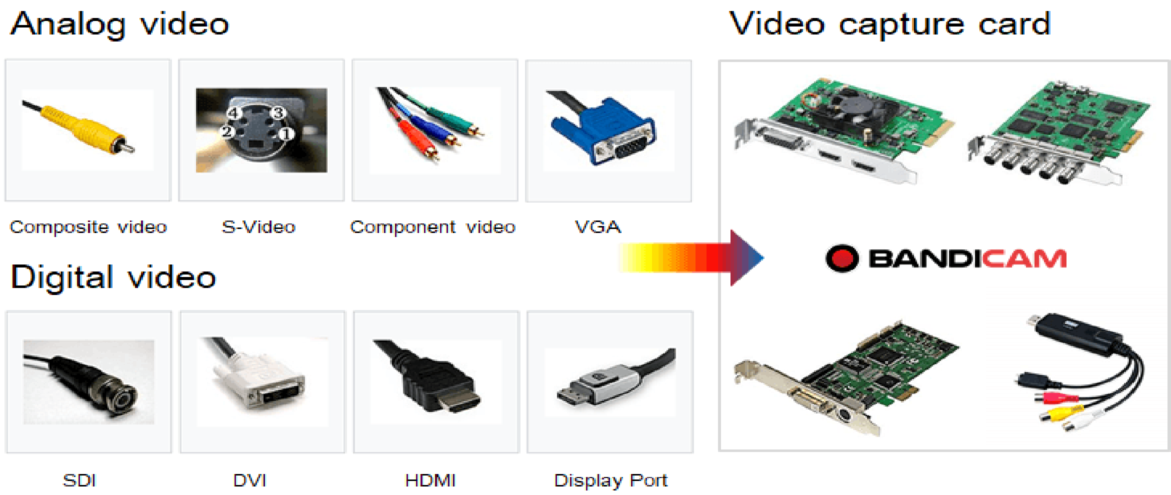
Digital video is an electronic representation of moving visual images (video) in the form of encoded digital data. This is in contrast to analog video, which represents moving visual images in the form of analog signals. Digital video comprises a series of digital images displayed in rapid succession.

Digital Video is audio and visual mixed together to make a production. the data gathered used to create a video, rather than a series of photos put together. Digital video have many ***advantages*** such as easy copying, multicasting, easy sharing and storage. Video recorded on tape is used on a computer on media player. Digital video is made of images displayed rapidly frequencies of 15, 24,30, and 60 frames per second. There is a saying "A picture is worth a thousand words." Pertaining to Digital Video the saying is "a video represents a million of those words strung together".

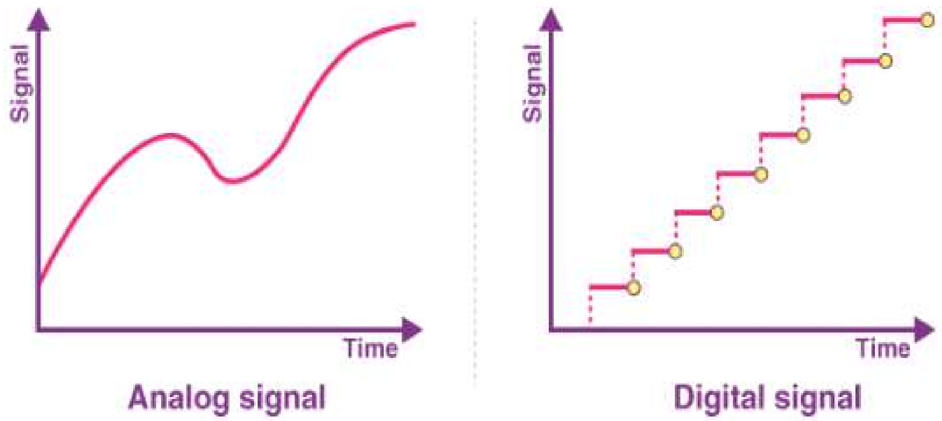
Digital video can be copied and reproduced with no degradation in quality. In contrast, when analog sources are copied, they experience generation loss. Digital video can be stored on digital media such as Blu-ray Disc, on computer data storage, or streamed over the Internet to end users who watch content on a desktop computer screen or a digital smart TV. Today, digital video content such as TV shows and movies also include a digital audio soundtrack.



Digital video representation.



Digital video and analog video capture.



Digital signal and analog signal.

Difference Between Analog And Digital Signal	
Analog Signals	**Digital Signals**
Continuous signals	Discrete signals
Represented by sine waves	Represented by square waves
Human voice, natural sound, analog electronic devices are a few examples	Computers, optical drives, and other electronic devices
Continuous range of values	Discontinuous values
Records sound waves as they are	Converts into a binary waveform.
Only used in analog devices.	Suited for digital electronics like computers, mobiles and more.

Some available analog to digital video converters for PC is not all effective, while others might just be expensive. One of the best analog to digital video converter available in the open market is Wondershare UniConverter (originally Wondershare Video Converter Ultimate). This software is a complete video conversion tool with over a thousand conversion formats and an attractive, user-friendly interface.

Lecture 2

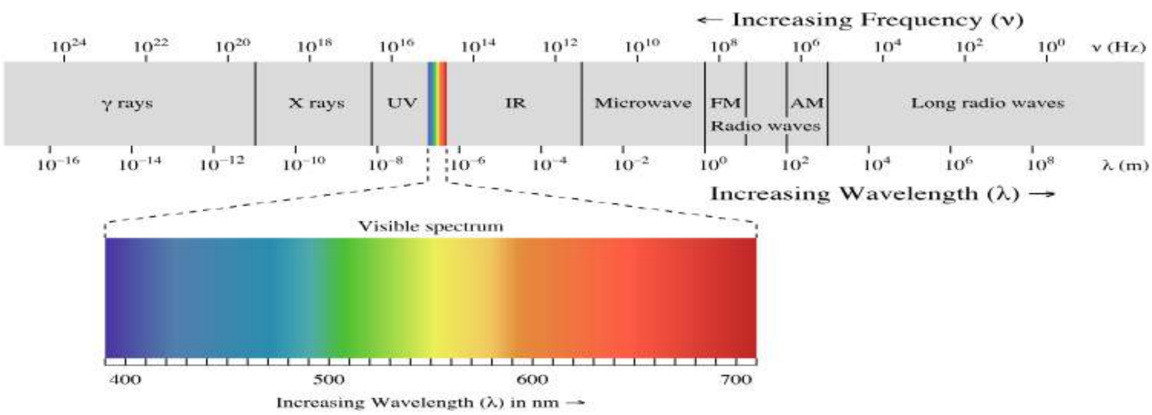
2.1 Color in Image and Video

a-Basics of Color

Here is some basic color representations :

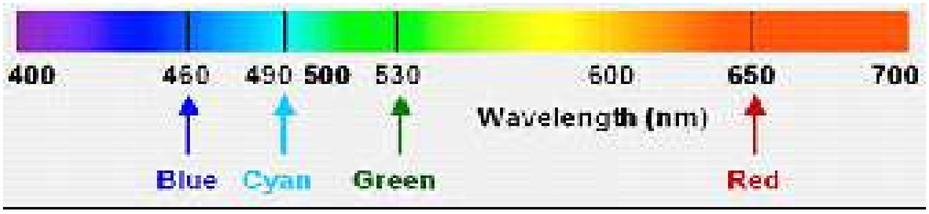
i-Light and Spectra

most light sources produce contributions over many wavelengths. Humans cannot detect all light , just contributions that fall in the visible wavelength.



Electromagnetic spectrum (visible-light range highlighted).

The human eye is sensitive to electromagnetic radiation with wavelengths between about 380 and 700 nanometers. This radiation is known as *light*. The eye has three classes of color-sensitive light receptors , which respond roughly to red, green and blue light (around 650, 530 and 460 nm, respectively). Short wavelengths produce a blue sensation, and long wavelengths produce a red one.



Some of invisible Electromagnetic spectrum are:

a) **X-rays** is a form of electromagnetic radiation. X-ray wavelengths are shorter than those of UV rays. X-rays can identify bone structures and have been used for medical imaging.

b) **Ultraviolet (UV)** is an electromagnetic radiation with a wavelength from 10 nm to 400 nm, shorter than that of visible light but longer than X-rays. UV radiation is present in sunlight. It is also produced by specialized lights.

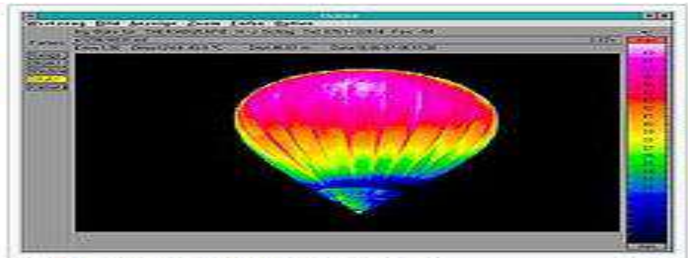
Suntan, freckling and sunburn are familiar effects of over-exposure, along with higher risk of skin cancer. Ultraviolet is also responsible for the formation of bone-strengthening vitamin D in most land vertebrates, including humans. The UV spectrum thus has effects both beneficial and harmful to human health.

UV light is used to kill or inactivate microorganisms by destroying nucleic acids and disrupting their DNA, leaving them unable to perform vital cellular functions. UV is used in a variety of applications, such as food, air, and water purification.

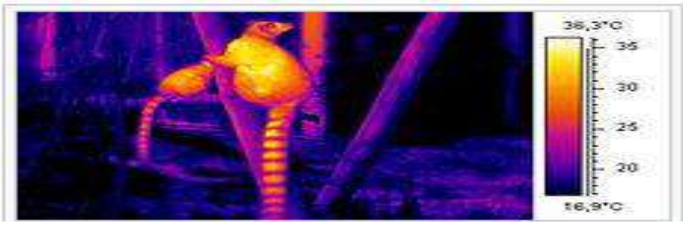
c) **Infrared** (**IR**) is an invisible electromagnetic radiation with longer wavelengths than those of visible light, extending from the nominal red edge of the visible spectrum at 700 nanometers.

IR application: Infrared radiation is used in industrial, scientific, and medical applications. Night-vision devices using active near-infrared illumination allow people or animals to be observed without the observer being detected. Infrared astronomy uses sensor-equipped telescopes to penetrate dusty regions of space such as molecular clouds, detect objects such as planets. Infrared thermal-imaging cameras are used to detect heat loss in insulated systems, to observe changing blood flow in the skin, and to detect overheating of electrical equipment.

Thermal imaging camera & screen, in an airport terminal in Greece.
Thermal imaging can detect fever, one of the signs of infection.



A thermal image showing temperature variation in a hot air balloon.

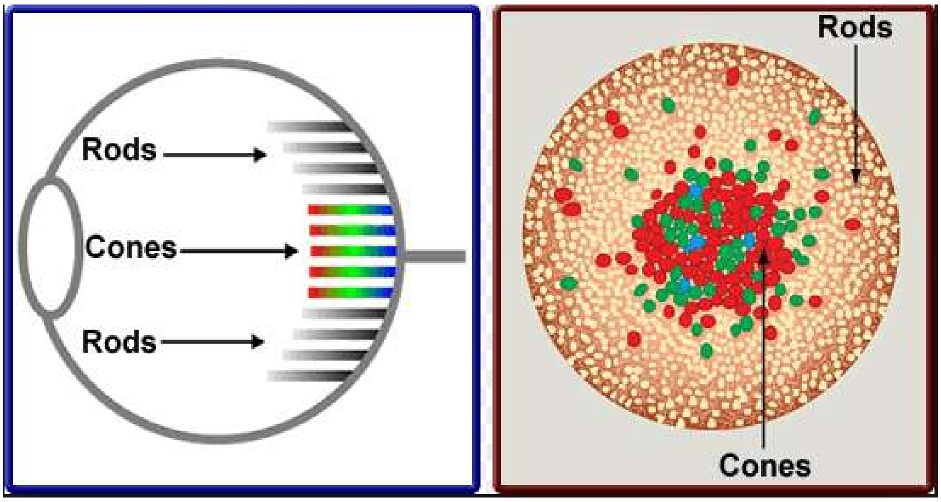


d) **Microwave** Are a form of electromagnetic radiation with wavelength between radio waves and infrared.

Microwave technology is extensively used for telecommunications. Microwaves are used in spacecraft communication, and much of the world's data, TV, and telephone communications are transmitted long distances by microwaves between ground stations and communications satellites. Microwaves are also employed in microwave ovens and in radar technology.

ii- Human Vision

The eye works like a camera, with the lens focusing an image onto the retina. The retina consists of an array of *rods* and three kinds of *cones*, so named because of their shape. The rods come into play when light levels are low and produce an image in shades of gray. For higher light levels, the cones each produce a signal. Because of their differing pigments, the three kinds of cones are most sensitive to red *(R)*, green (*G*), and blue *(B)* light.

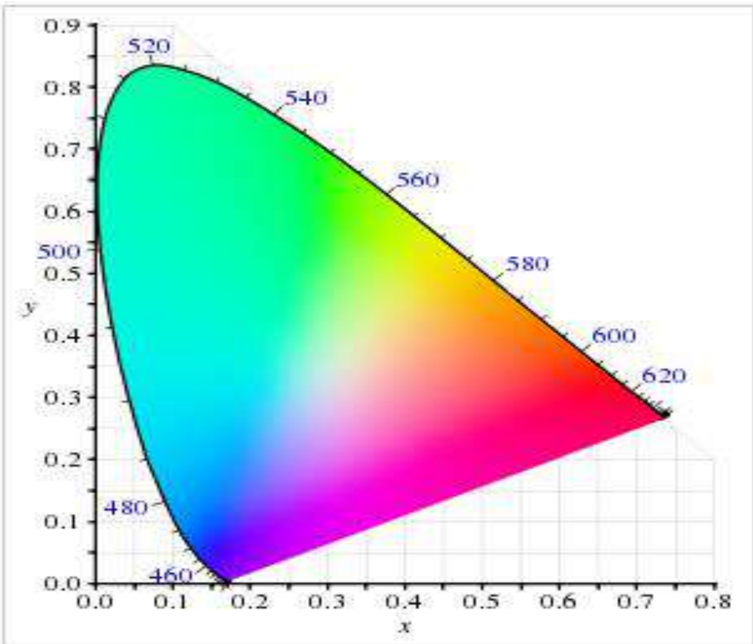


iii-Cones and perception

The response in each color channel in the eye is proportional to the number of neurons firing. For the red channel, any light falling anywhere in the nonzero part of the red cone will generate some response. So the total response of the red channel is the sum over all the light falling on the retina to which the red cone is sensitive, weighted by the sensitivity at that wavelength.

b-CIE Chromaticity (اللونية) Diagram

Since the human eye has three types of color sensors that respond to different ranges of wavelengths, a full plot of all visible colors is a three-dimensional figure. However, the concept of color can be divided into two parts: brightness and chromaticity. For example, the color white is a bright color, while the color grey is considered to be a less bright version of that same white. In other words, the chromaticity of white and grey are the same while their brightness differs.



The CIE 1931 color space chromaticity diagram.

i-Camera Systems

Humans develop camera systems in a similar fashion. A good camera has three signals produced at each pixel location (corresponding to a retinal position). Analog signals are converted to digital, truncated to integers, and stored. If the precision used is 8·bit, the maximum value for any of *R*, G, *B* is 255, and the minimum is 0.

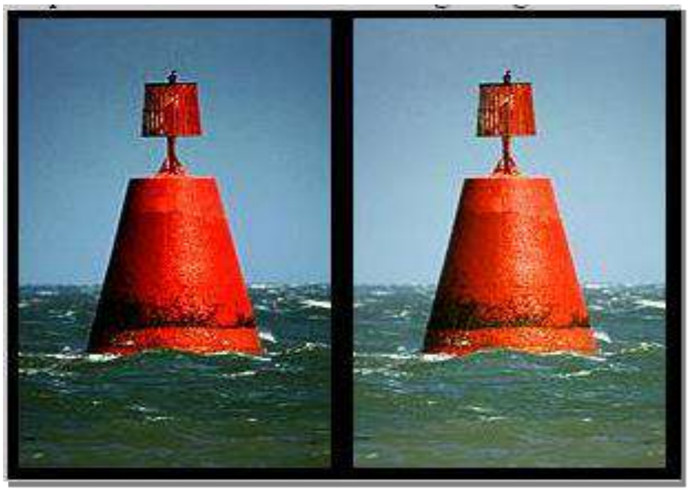
ii-CRT display and Gamma Correction

The cathode ray tube (CRT) is a vacuum tube that contains one or more electron guns and a phosphorescent screen, and is used to display images.

Our eyes do not perceive light the way cameras do. With a digital camera, when twice the number of photons hit the sensor, it receives twice the signal (a "linear" relationship). That's not how our eyes work. Instead, we perceive twice the light as being only a fraction brighter and increasingly so for higher light intensities (a "nonlinear" relationship).

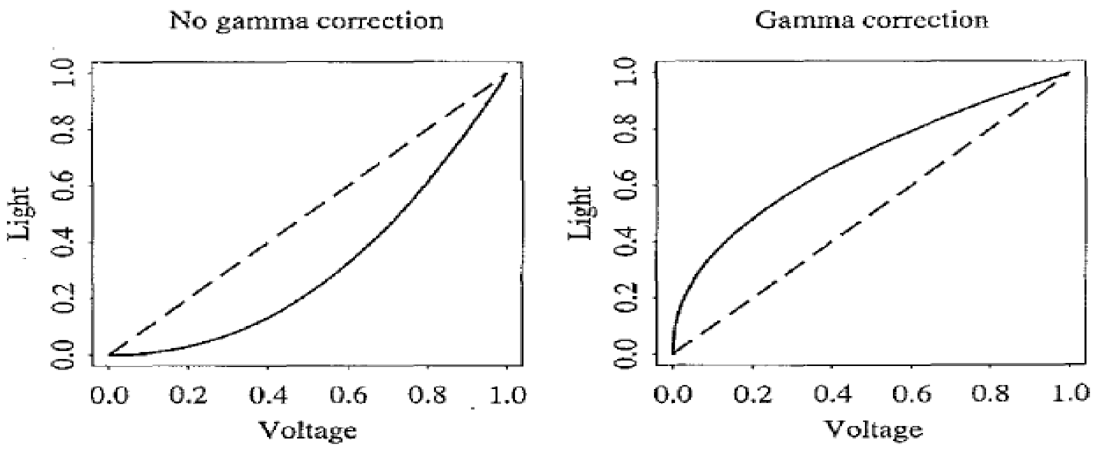
Compared to a camera, we are much more sensitive to changes in dark tones than we are to similar changes in bright tones. There's a biological reason for this peculiarity: it enables our vision to operate over a broader range of luminance. Otherwise the typical range in brightness we encounter outdoors would be too overwhelming.

But how does all of this relate to gamma? In this case, gamma is what translates between our eye's light sensitivity and that of the camera. When a digital image is saved, it's therefore "gamma encoded" so that twice the value in a file more closely corresponds to what we would perceive as being twice as bright.



On the left is the image as it might appear on an un-corrected monitor. The right image should look right on a monitor with a gamma of around 1.8

The RGB numbers in an image file are converted back to analog and drive the electron guns in the cathode ray tube (CRT). Electrons are emitted proportional to the driving voltage, and we would like to have the CRT system produce light linearly related to the voltage. Thus, if the file value in the red channel is R, the screen emits light proportional to R^{Y}, with the red phosphor paint on the screen that is the target of the red-channel electron gun.



(a) (b)

FIGURE 4: Effect of gamma correction: (a) no gamma correction (b) gamma correction of signal.

Figure 4(a) shows the light output with no gamma correction applied. We see that darker values are displayed too dark. This is also shown in Figure 5, which display a linear ramp from left to right. Figure 4(b) shows the effect of precorrecting signals by applying the power law $R^{1/Y}$,

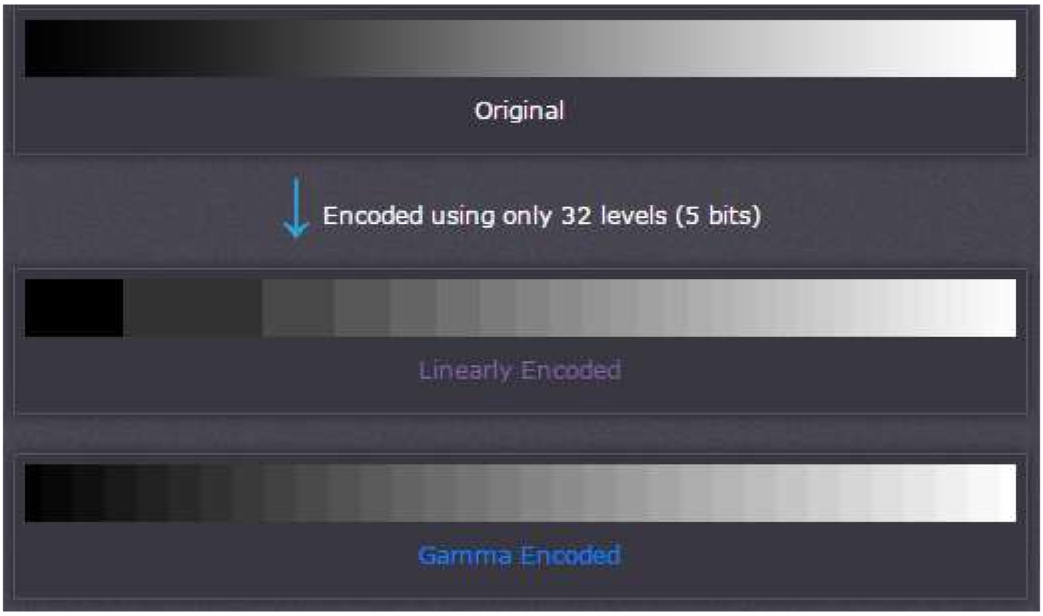


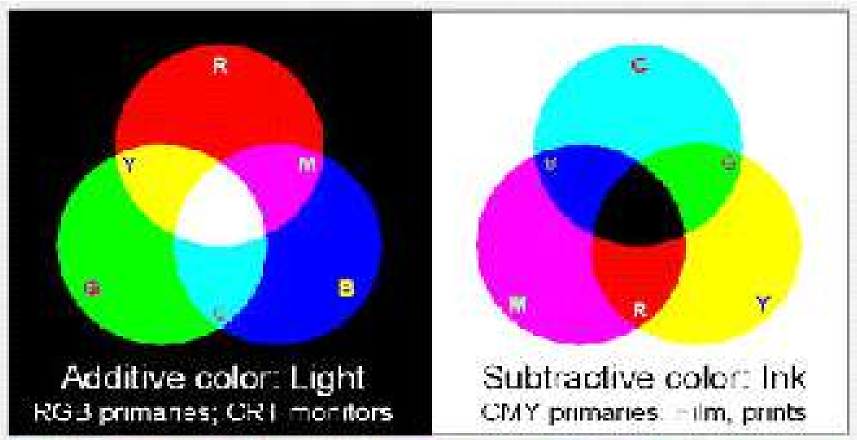
Figure 5

1.5 Color image and video representations

A color model is a mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values. We will focus on some models:

The RGB (CMY) Color Model

RGB and its subset CMY form the most basic and well-known color model. This model bears closest resemblance to how we perceive color. It also corresponds to the principles of additive and subtractive colors.

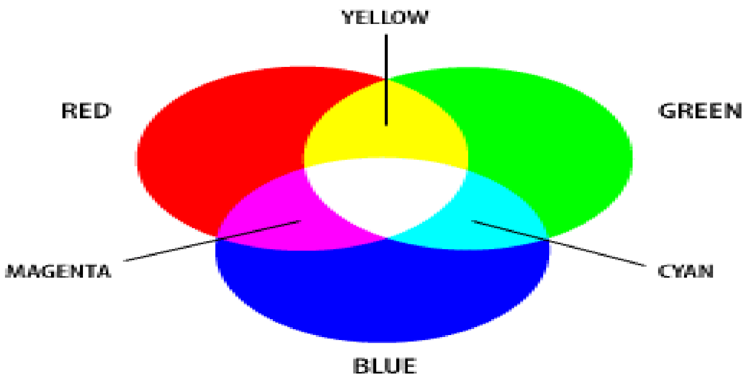


The objects may either transmit light (transparencies) or reflect light . A range of colors can be reproduced by one of two complimentary approaches:

1)The RGB (Additive Colors)

Additive colors are created by mixing spectral light in varying combinations. The most common examples of this are television screens and computer monitors, which produce colored pixels by firing red, green, and blue electron guns at phosphors on the television or monitor screen. It is used for Web graphics, but it cannot be used for print production.

The *additive primary* colors are red (R), green (G), and blue (B). Adding R and G light makes yellow (Y). Similarly, G + B = cyan (C) and R + B = magenta (M). Combining all three additive primaries makes white.

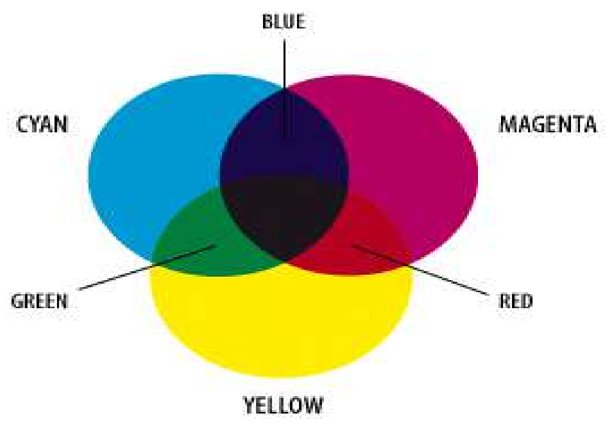


The importance of RGB as a color model is that it relates very closely to the way we perceive color with the r g b receptors in our retinas. It is the basic color model on computers.

2)CMY(K) ***(Subtractive color)***

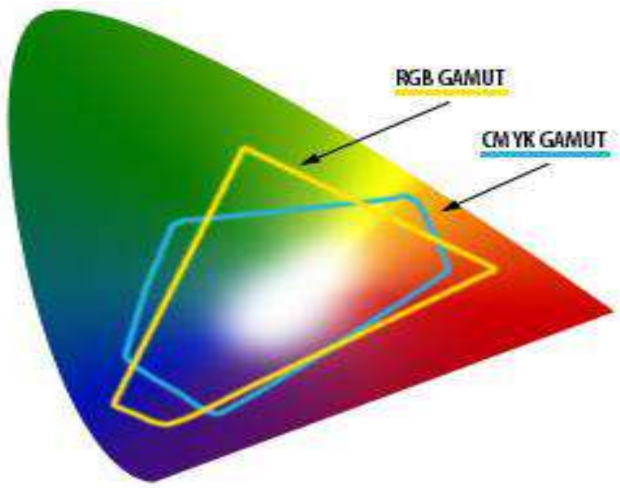
The *subtractive primaries* are C, M and Y. Cyan absorbs red; hence C is sometimes called "minus red" (-R). Similarly, M is -G and Y is -B.

Cyan, magenta, and yellow correspond roughly to the primary colors in art production: red, blue, and yellow. In the illustration below, you can see the CMY counterpart to the RGB model shown above:



Just as the primary colors of CMY are the secondary colors of RGB, the primary colors of RGB are the secondary colors of CMY. But as the illustrations show, the colors created by the subtractive model of CMY don't look exactly like the colors created in the additive model of RGB. Particularly, CMY cannot reproduce the brightness of RGB colors.

The illustration below shows the representative RGB and CMY gamut's over the 1931 CIE Chromaticity Diagram (representing the whole gamut of human color perception):



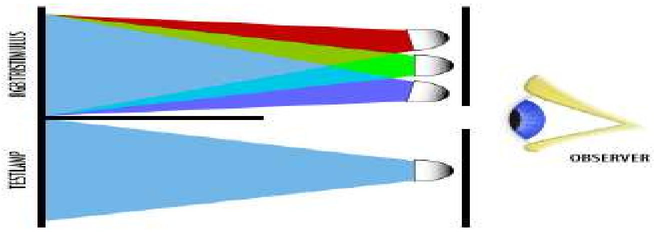
Both models fall short of reproducing all the colors we can see. Furthermore, they differ to such an extent that there are many RGB colors that cannot be produced using CMY(K), and similarly, there are some CMY colors that cannot be produced using RGB.

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix}$$

3)The CIE Color Models

The CIE color model was developed to be completely independent of any device and is based as closely as possible on how humans perceive color.



The significant difference between the 1931 and 1964 standard observers was the field of vision used to view the screens. The 1931 observer had a 2° field of vision. This was later considered inadequate in many cases since it did not take in enough of the observer's peripheral vision. The 1964 specification widened the observer's field of vision to 10° in order to get tristimulus values that reflect a wider retinal sensitivity.

Designation	*Name*	*Notes*
RGB	Red, Green, Blue	The native format
CYMK	Cyan, Yellow, Magenta, Black	For color printing
HSB	Hue, Saturation, Brightness	Related to human perception
L-a-b	Luminance, a (green to red) and b (blue to yellow)	The CIE model

Video color representation :

Beside the RGB representation, YIQ and YUV are the two commonly used in vide :

1- YIQ Color Space

- YIQ is used in color TV broadcasting.
- Y (luminance) is the CIE Y primary.

$$Y = 0.299R + 0.587G + 0.114B$$

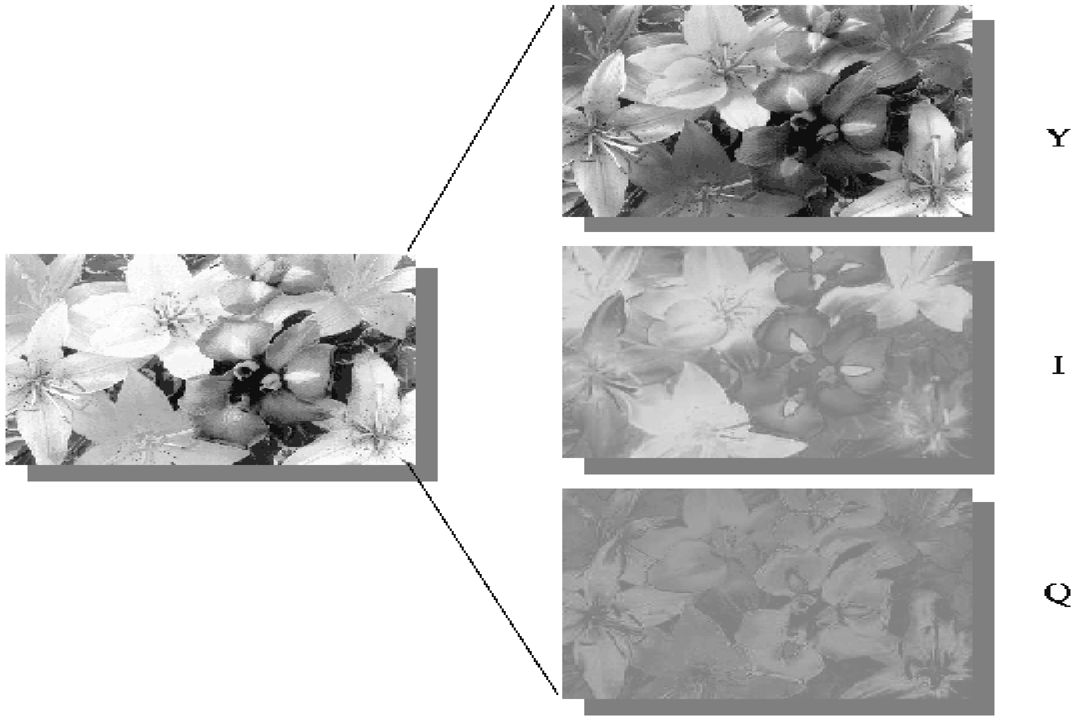
- the other two vectors:

$$I = 0.596R - 0.275G - 0.321B \quad Q = 0.212R - 0.528G + 0.311B$$

- The YIQ transform:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

An Example YIQ Decomposition is shown in Fig.



2- YUV color space

The YUV colorspace encodes the brightness in the first components and uses two difference components, U and V, that somewhat correspond to the yellow-blue and the red-cyan differences:

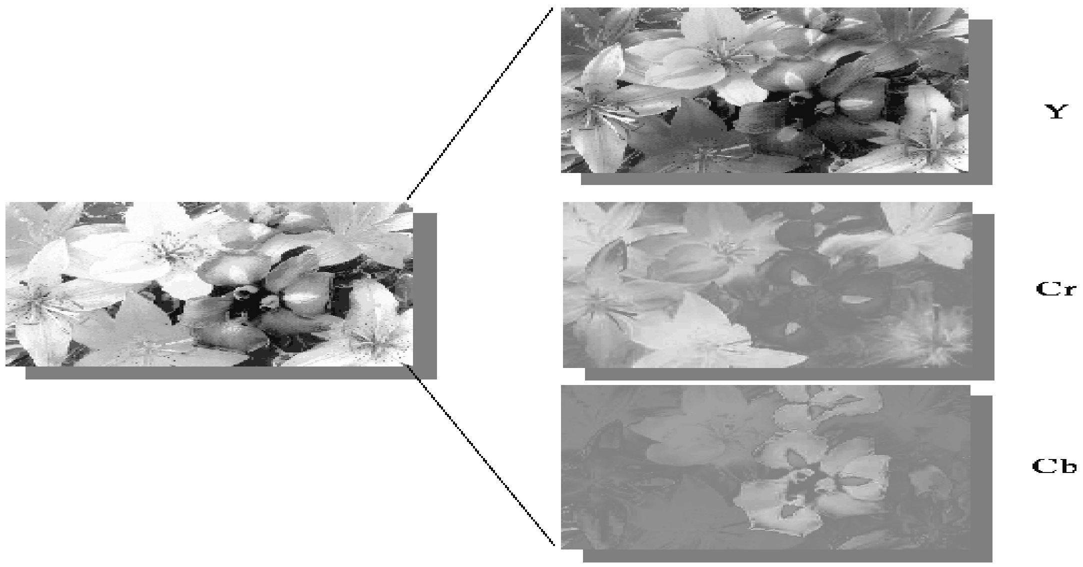
$$\begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.1 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} Y \\ U \\ V \end{bmatrix}$$

with inverse:

$$\begin{bmatrix} 1 & 0 & 1.340 \\ 1 & -0.395 & -0.581 \\ 1 & 2.032 & 0 \end{bmatrix} \begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} r \\ g \\ b \end{bmatrix}$$

3- **YCrCb of a color image** :

Is a family of color spaces used as a part of the color image pipeline in video and digital photography systems. Y′ is the luma component and CB and CR are the blue-difference and red-difference chroma components. Y′ (with prime) is distinguished from Y, which is luminance, meaning that light intensity is nonlinearly encoded based on gamma corrected RGB primaries. YCbCr prioritizes what to store and display based on how we see. This means the most brightness and green are saved in every image. Then some red, and a little bit of blue. We use chroma subsampling to process and remove some of the red and blue color data and gamma, contrast, and brightness processing for the luminance data. An example YCrCb Decomposition is shown in Fig.



Color	Y Value	Cr Value	Cb Value
White	235 (EB)	128 (80)	128 (80)
Black	16 (10)	128 (80)	128 (80)
Red	81 (51)	240 (F0)	90 (5A)
Green	145 (91)	34 (22)	54 (36)
Blue	41 (29)	110 (6E)	240 (F0)

Yellow 210 (D2) 146 (92) 16 (10)

Cyan 170 (AA) 16 (10) 166 (A6)

Magenta 106 (6A) 222 (DE) 202 (CA)

Note that:

1. The Y value ranges from 16 to 235 (220 levels), with 16 being black.
2. The Cr and Cb values range from 16 to 240 with 128 as the mid-point.
3. The above are for 8-bit values. 10-bit values can be achieved by multiplying the decimal values by 4.
4. The values in brackets are in hexadecimal.

See also RGB to YUV calculator to create your own values.

$$Y = w_R \cdot R + (1 - w_B - w_R) \cdot G + w_B \cdot B,$$

$$C_b = \frac{0.5}{1 - w_B} \cdot (B - Y),$$

$$C_r = \frac{0.5}{1 - w_R} \cdot (R - Y),$$

$$R = Y + \frac{1 - w_R}{0.5} \cdot C_r,$$

$$G = Y - \frac{w_B \cdot (1 - w_B) \cdot C_b \; - \; w_R \cdot (1 - w_R) \cdot C_r}{0.5 \cdot (1 - w_B - w_R)},$$

$$B = Y + \frac{1 - w_B}{0.5} \cdot C_b.$$

Lecture 3

Video Visual Effect of Motion

The visual effect of motion is due to biological phenomenon of Persistence of vision where an object seen by the human eye remains mapped on the eye's retina for a brief time after viewing (approximately 25 ms).

Due to this phenomena of our vision system, a discrete sequence of individual pictures can be perceived as a continuous sequence

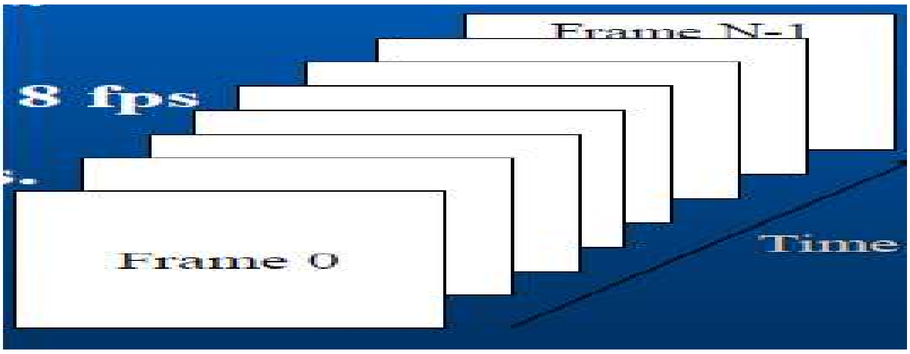
Lecture 3 Characteristics of video streams.

Characteristics of video streams

It is possible to describe the characteristics of the video by many parameters, some of them are: Number of frames per second, Aspect ratio, Color space and bits per pixel, video quality, formats (analog or digital), and video compression method (for digital only).

1- Number of frames per second (Frame rate)

Frame rate, the number of still pictures per unit of time of video, ranges from 6 or 8 frames per second (fps) for old mechanical cameras to 120 or more frames per second for new professional cameras.



- PAL standards (Europe, Asia, Australia, etc.) and SECAM (France, Russia, parts of Africa etc.) specify 25 frame/s,

- while NTSC standards (USA, Canada, Japan, etc.) specify 29.97 frames.

- The minimum frame rate to achieve a comfortable illusion of a moving image is about 16 frames per second.

- The higher a camera's FPS, the smoother motion will appear to the viewer.
- FPS stands for frames per second and is a metric used to describe a camera's frame rate. This represents the number of photos, or frames, that your camera can capture per second.
- A lower frame rate will result in broken or choppy motion.
- A higher frame rate necessitates a larger file size. It is for this reason that many security cams and smartphone cameras shoot video in 15 FPS, as it helps keep the files small. Some webcams also operate at 15 FPS so they can send a steady stream of content to the Internet without

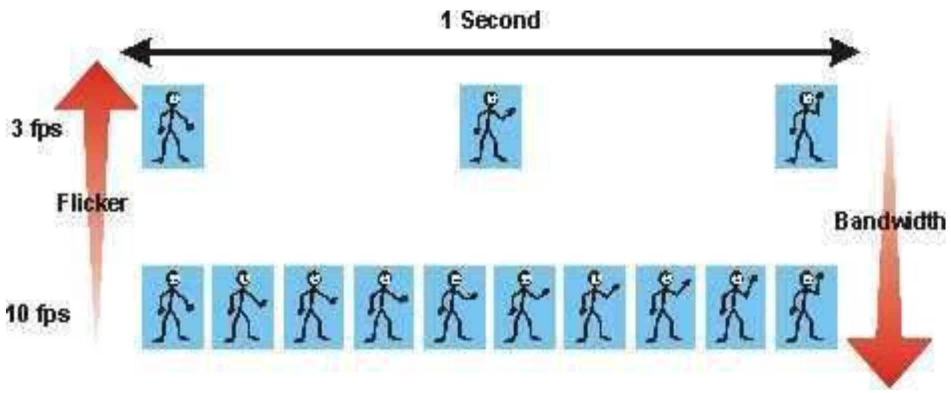
clogging up the bandwidth.

2- Aspect ratio

The **aspect ratio** describes the dimensions of video screens or the proportional relationship between its width and its height. It is commonly expressed as two numbers separated by a colon (W:H), for example; the screen aspect ratio of a traditional television screen is 4:3 and high definition televisions use an aspect ratio of 16:9.

3- Interlaced vs progressive

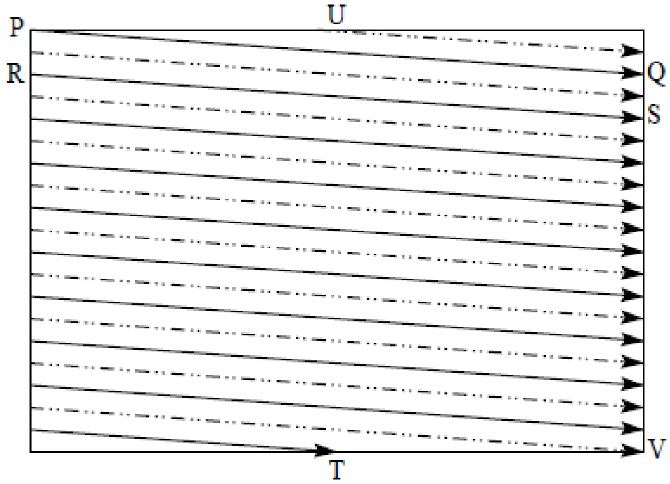
Flicker: is a visible fading between cycles displayed on video displays, especially the refresh interval on cathode ray tube (CRT) based computer screens.



There are two common methods for "painting" a video image on an electronic display:

a- Interlaced video is a technique for doubling the perceived frame rate of a video display without consuming extra bandwidth. The interlaced signal contains two fields to create a frame. One field contains all odd-numbered lines in the image; the other contains all even-numbered lines. This enhances motion perception to the viewer, and reduces flicker.

Interlacing was invented because, when standards were being defined, it was difficult to transmit the amount of information in a full frame quickly enough to avoid flicker. The double number of fields presented to the eye reduces perceived flicker.



Interlaced raster scan

Because of interlacing, the odd and even lines are displaced in time from each other - generally not noticeable except when very fast action is taking place on screen, when blurring may occur.

Initially the odd-numbered lines are scanned and then the process is repeated for even-numbered lines -this time starting at the second row.

- For example, in the video in Fig. 1, the moving helicopter is blurred more than is the still background.

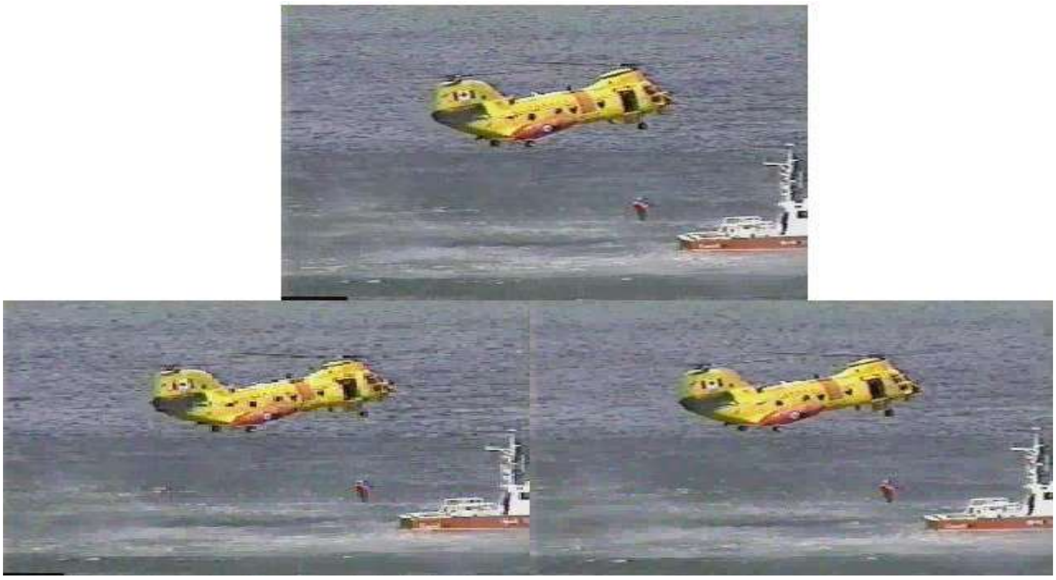
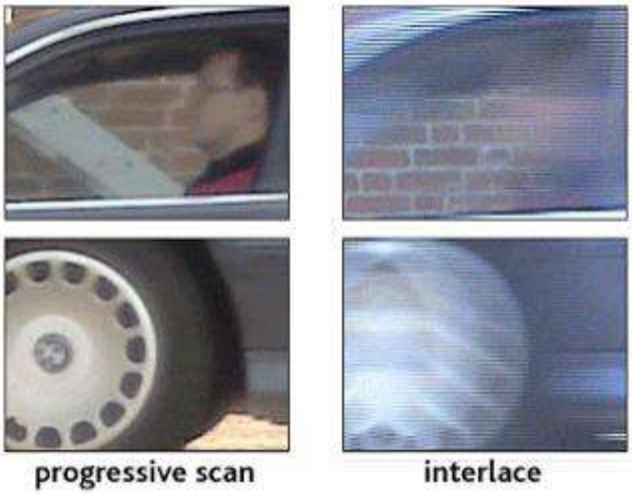


Fig. 1: Interlaced scan produces two fields for each frame.
At top: The video frame, at left: odd field, at right: even field,

A Phase Alternating Line (PAL)-based television set display, for example, scans 50 fields every second (25 odd and 25 even). The two sets of 25 fields work together to create a full frame

NTSC, PAL and SECAM are interlaced formats. Abbreviated video resolution specifications often include an i to indicate interlacing. For example, PAL video format is often specified as 576i50, where 576 indicates the total number of horizontal scan lines, i indicates interlacing, and 50 indicates 50 fields (half-frames) per second.

b- Progressive scanning (alternatively referred to as noninterlaced scanning, is a way of displaying, storing, or transmitting moving images in which all the lines of each frame are drawn in sequence.



So interlacing is in fact a clever way to compress a movie when one cannot use digital compression methods. Interlacing reduces the bandwidth (= storage space) by half, without losing vertical resolution in quiet areas.

THE DIFFERENCE BETWEEN INTERLACED AND NON-INTERLACED IMAGES

These images look and function almost the same way as a non-interlaced image with one exception–how it appears to load to your visitor. If you have a large image on your site and someone with a slower Internet connection comes to view that image, a non-interlaced image will simply be blank until the data transfers and then slowly it will appear from top to bottom.

An interlaced image will appear completely, but it will be highly pixelated. As the data transfers, the picture will begin to get clearer and clearer until the full resolution becomes apparent. A perfect example of the difference between the two can be seen below:

(a) non-interlaced (b) interlaced

Half of figure downloaded

Deinterlacing is the process of converting interlaced video, such as common analog television signals into a non-interlaced form.

To display interlaced video on a progressive scan display requires a process called deinterlacing. This is an imperfect technique, and causes various artifacts—particularly in areas with objects in motion. Providing the best picture quality for interlaced video signals requires expensive and complex devices and algorithms.

Method 1: Capturing one field and combining it with the next field, Problem: "combing" effect.

Method 2: Line doubler

The most basic and literal way to double lines is to repeat each scanline. Most line doublers use digital interpolation to recreate the missing lines in an interlaced signal, and the resulting quality depends on the technique used. Generally a line doubler will only interpolate within a single field, rather than merging information from adjacent fields, to preserve the smoothness of motion.

When interlaced video is watched on a progressive monitor with very poor deinterlacing, it exhibits combing when there is movement between two fields of one frame.

4- Video formats:

a- Analog Video: Analog video is represented as a continuous (time-varying) signal.

i- NTSC Video

- NTSC (National Television System Committee) TV standard is mostly used in North America and Japan. It uses the familiar 4:3 aspect ratio (i.e., the ratio of picture width to its height) and uses 525 scan lines per frame at 30 frames per second (fps). NTSC follows the interlaced scanning system, and each frame is divided into two fields, with 262.5 lines/field**.**

ii- PAL Video

- PAL (Phase Alternating Line) is a TV standard widely used in Western Europe, China, India, and many other parts of the world.
- PAL uses 625 scan lines per frame, at 25 frames/second, with a 4:3 aspect ratio and interlaced fields.

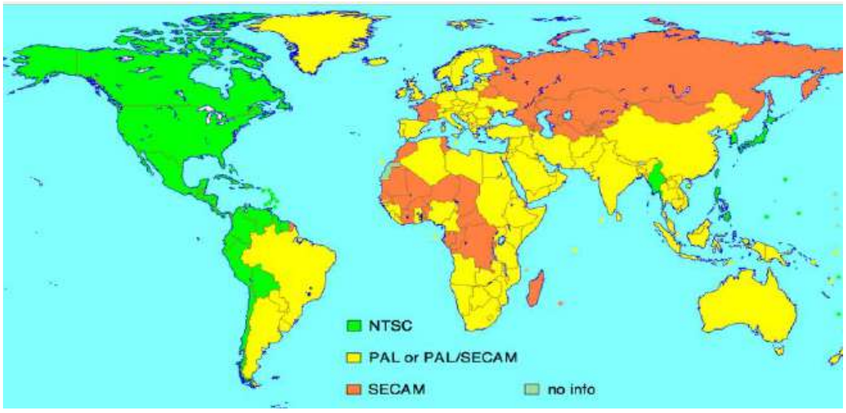
iii- SECAM Video

SECAM stands for Systeme Electronique Couleur Avec Memoire,the third major broadcast TV standard.

- SECAM also uses 625 scan lines per frame, at 25 frames per second, with a 4:3 aspect ratio and interlaced fields.

- SECAM and PAL are very similar. They differ slightly in their color coding scheme.

	Line duration (µS)	*Picture height (line)*	*Line rate (Hz)*	*Frame rate (Hz)*	*Active picture area*	
					Width (µS)	*Height (lines)*
RS-170		525	15,750	30		486
NTSC	63.49	525	15,734.26	29.96	720	486
PAL, SECAM	64	625	15,625	25	52	576



b- Digital Video

Digital video is represented as a sequence of digital images.

The advantages of digital representation for video are many. For example:

(a) Video can be stored on digital devices or in memory, ready to be processed (noise removal, cut and paste, etc.), and integrated to various multimedia applications;

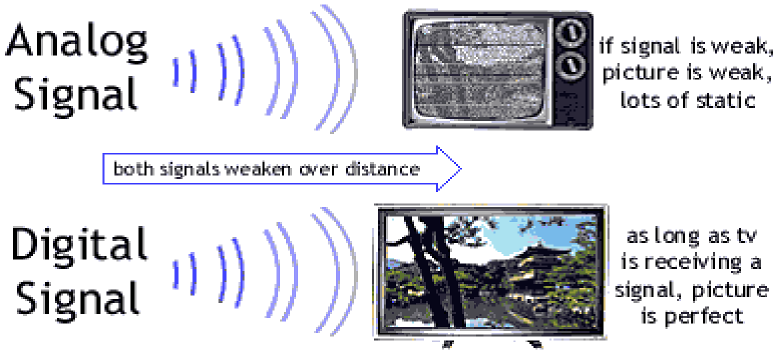
(b) Direct random access is possible, which makes nonlinear video editing achievable as a simple, rather than a complex, task;

(c) Repeated recording does not degrade image quality;

(d) Ease of encryption and better tolerance to channel noise.

(e) An advantage digital has over analog is that analog signals can't be compressed as well as a digital signal can.

Almost all digital video uses component video.



Homework: why can't analog video be compressed like digital?

Chroma Subsampling

Digital signals are often compressed to save transmission time and reduce file size. Since the human visual system is much more sensitive to variations in brightness than color, a video system can be optimized by devoting more bandwidth to the luma component (usually denoted Y'), than to the color difference components Cb and Cr.

This became a central motivation behind early forms of analog and digital compression. Video signals would be separated into a lightness or “luma” component and two color or “chroma” components, similar to how images can be separated into three red, green and blue (RGB) components. The luma and chroma components would then be referred to as YUV (with analog) or YCbCr (with digital).

Once separated, the chroma resolution would then be reduced by half or more through a process called "chroma subsampling." The end result is a video signal that appears more detailed at the same broadcast bandwidth, since the luma component occupies a greater fraction of the video signal:

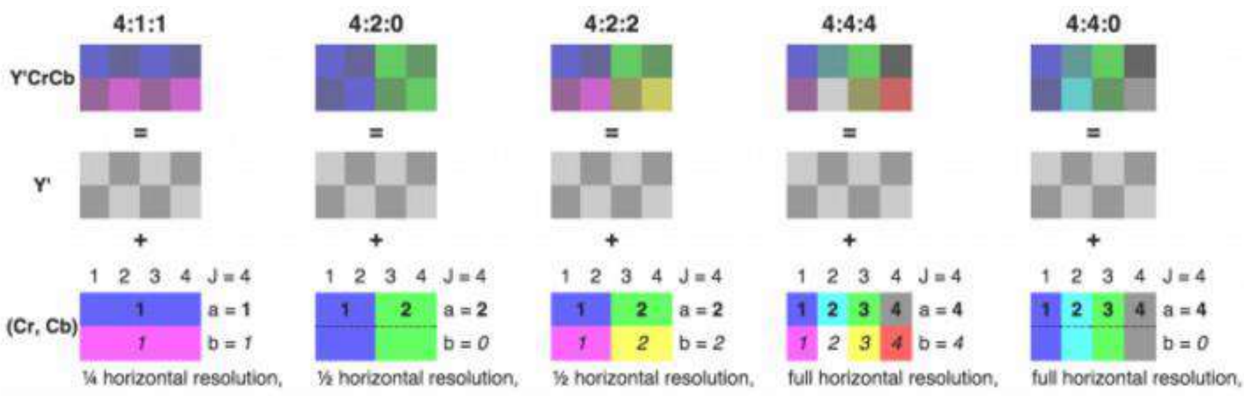
In compressed images, for example, the 4:2:2 Y'CbCr scheme requires two-thirds the bandwidth of (4:4:4) R'G'B'. This reduction results in almost no visual difference as perceived by the viewer.

Sampling systems and ratios

The subsampling scheme is commonly expressed as a three part ratio J:a:b (e.g. 4:2:2) that describe the number of luminance and chrominance samples in a conceptual region that is J pixels wide, and 2 pixels high.

The scheme"4:2:2" indicates horizontal subsampling of the Cb and C*r* signals by a factor of 2. That is, of four pixels horizontally labeled 0 to 3, all four *Y* s are sent, and every two Cbs and two *Crs* are sent.

The chroma subsampling scheme "4:4:4" indicates that no chroma subsampling is used: each pixel's Y, Cb and Cr values are transmitted, 4 for each of Y, Cb, Cr.



mapping examples

Scheme 4:2:0, along with others, is commonly used in JPEG and MPEG.

High Definition TV (HDTV)

The introduction of wide-screen movies brought the discovery that viewers seated near the screen enjoyed a level of participation not experienced with conventional movies. Apparently the exposure to a greater field of view, especially the involvement of peripheral vision, contributes to the sense of "being there". The main thrust of High Definition TV (HDTV) is not to increase the "definition" in each unit area, but rather to increase the visual field, especially its width.

The salient difference between conventional TV and HDTV:

(a) HDTV has a much wider aspect ratio of 16:9 instead of 4:3 where 16:9 is closer to aspect ratio of the human eye sight.

(b) HDTV moves toward progressive (non-interlaced) scan. The rationale is that interlacing introduces serrated edges to moving objects and flickers along horizontal edges.

(c) HDTV has higher resolution 1280×720 or 1920×1080.

Lecture 4

What is Video Resolution ? (تكملة)

Resolution is defined as the number of pixels organized or arranged by width and height on a screen or an image. In simple words, resolution measures the number of pixels an image has. The more pixels, the higher the quality. The most commonly used video resolutions are known as 'Formats', which are technically termed as Ultra HD or Full HD.

Pixel: It is the smallest unit of a graphic or an image, which can be represented or displayed on any digital device.

Resolution = Pixel width x Pixel height

A high number of pixels signifies a better resolution whereas a low number of pixels indicates a poor-resolution video. Resolution helps determine a video's quality and how clear or realistic it can appear. Generally, a higher resolution means a clearer video. It is measured by the number of pixels in the standard 16:9 aspect ratio, which is common for computer monitors and television sets.

Resolution measures the image size as width x height, in pixels. For example, a resolution of 640x480, indicates the width is 640 pixels and the height is 480 pixels. Full HD resolution starts at 1920x1080, and increases upwards to 3840x2160 or 4K, then 7680x4320, also known as 8K. 10K resolution is the top end at 10240x4320. When it comes to common resolutions like 1080 and 720, the naming conventions are based on the number of pixels displaying in the vertical line of the screen. For 8K, 4K, and 2K videos, the naming convention depends on the number of pixels running in the horizontal line throughout the frame.
Technically, resolution can be classified into two types:

- Spatial resolution: This is the height and width of the picture, which is measured in pixels. Simply put, it's the total number of pixels in every individual frame.
- Temporal resolution: This is the frame rate (the number of frames displayed per second), or ‘resolution over time.'

Note: Spatial is the default resolution type.

Resolution also impacts a video's file size and due to this, HD (High Definition) videos tend to have a larger file size than SD (Standard Definition), even though it's of the same duration.

From shooting, to editing, and uploading the videos online, the resolution is something that affects the overall visibility of the content.

The 16:9 aspect ratio is widely used in HDTV, non-HD TV, Full HD, and analog television screens. The ratio is obtained when you divide vertical pixels by horizontal pixels (in numbers), containing 1920 vertical lines and 1080 horizontal lines of resolution .

For reference, a 16:9 aspect ratio on television simply means that the horizontal viewing area is wider (16) than the vertical viewing area. Moreover, the 16:9 format is natively supported in DVD format and is further used on Vimeo, YouTube, and other popular video websites attracting a very high view rate . In the research area, standardization of TVs and computer displays with this aspect ratio saved producers time and money.

What Is The Resolution To The 16:9 Aspect Ratio؟

The sharpness and clarity of your screen's image are also determined by its resolution . It's calculated by multiplying the number of horizontal pixels by the number of vertical pixels.

Width	Height	Resolution
640	360	nHD
854	480	FWVGA
960	540	qHD
1024	576	WSVGA
1280	720	HD/WXGA
1366	768	FWXGA
1600	900	HD+
1920	1080	FHD
2048	1152	QWXGA
2560	1440	QHD
3200	1800	WQXGA+
3840	2160	UHD
5120	2880	UHD+
7680	4320	FUHD
15360	8640	QUHD
30720	17280	HHD
61440	34560	FHHD
122880	69120	QHHD

Video quality

Video quality is a characteristic of a video passed through a video transmission or processing system, representing a measure of perceived degradation with respect to the original source video. Video processing systems may introduce some amount of distortion or artifacts in the video signal, which negatively impacts the user's perception of a system. For many stakeholders such as content providers, service providers and network operators, the assurance of video quality is an important task.

What is (UHD vs HD vs SD)

UHD, HD, and SD all refer to different types of video formats. Video formats are often described in terms of their horizontal pixels x vertical pixels. For example, you'll often see 1,280 x 720p, or just 720p for short.

Here's a breakdown of these different formats, and how we handle them on the SproutVideo platform.

- **SD Video**
  A standard definition video loosely refers to any video that is not UHD or HD. SD videos are low quality, with lower bit rates and smaller file sizes. However, they are a good option for viewers on very slow internet connection speeds because the lower bit rate means they will stream more smoothly with less chance of buffering compared to higher quality videos. A broadly accepted cut-off point for SD video is 480p, meaning any video with 480 vertical pixels or less is standard definition.
- **HD Video**
  A high definition video is a high quality video. It will look sharp and clear on the majority of screens. Typically, HD refers to any

video that is at least 720p, meaning any video with 720 vertical pixels or greater.

- **UHD Video**

  UHD stands for ultra-high definition. This video format is extremely high quality, and is suitable for display on very large screens. UHD refers to 4K and 8K video, as well as any video format larger than that. For online video, 4K refers to any video that has at least 2,160 vertical pixels. Meanwhile, 8K video is any video that has at least 4,320 vertical pixels.

One complicating factor is that UHD formats are defined differently for different screens. On a UHD TV in your home, 4K content is typically streamed at 3,840 x 2,160p. This format will fill the screen, and look extremely sharp.

However, at the cinema, 4K is defined by the Digital Cinema Initiatives as having a horizontal resolution of 4,096 pixels, with no vertical resolution defined.

Q \ How This Works on the SproutVideo Platform ?

The higher quality you upload, the better quality your viewers will be able to enjoy. We cannot upgrade or improve the quality of your video files . Depending on the resolution of the video you uploaded, we will encode your video in the following formats:

240p, 360p, 480p, 720p, 1080p, 2K, 4K and 8K.

For example, if you uploaded a 1080p video, we will encode it in every format listed from 240p up to 1080p. However, we won’t be able to create a 2K, 4K, or 8K version of that video.

What is 4K?

4K is made up of 3860 x 2160 pixels , the 4K comes from the horizontal pixels almost reaching 4,000. 4K has been around for a few years Netflix started offering 4K streaming way back in 2014. But only recently has it become more affordable and, therefore, popular . 4K is sometimes referred to as UHD (ultra high definition).

With all the big streaming services now offering 4K, as well as 4K Blu-ray players providing exceptional viewing quality, 4K is setting the visual standard for mainstream entertainment.

What is 8K?

Not content with the incredible clarity of 4K, the biggest tech developers in the world have doubled their efforts. It has a 7,680 x 4,320 screen , with more than four times the level of detail of current 4K screens.

Video file format and codec basics.

Because video files can be large, programs called codecs were developed to make them easier to store and share. Codecs encode data to compress it for storing and sharing. Then they decode that data to decompress it for viewing and editing like H.264 or AVC. Which format you choose depends on the balance you want to strike between quality and ease of use.

Understand the top video file extensions.

These are the most common digital video formats and their most frequent uses.

- **MP4**

MP4 (MPEG-4 Part 14) is the most common type of video file format. Apple's preferred format, MP4 can play on most other devices as well. It uses the MPEG-4 encoding algorithm to store video and audio files and text, but it offers lower definition than some others. MP4 works well for videos posted on YouTube, Facebook, Twitter, and Instagram .

- **MOV**

  MOV (QuickTime Movie) stores high-quality video, audio, and effects, but these files tend to be quite large. Developed for QuickTime Player by Apple, MOV files use MPEG-4 encoding to play in QuickTime for Windows. MOV is supported by Facebook and YouTube, and it works well for TV viewing.

- **WMV**

  WMV (Windows Media Viewer) files offer good video quality and large file size like MOV. Microsoft developed WMV for Windows Media Player. YouTube supports WMV, and Apple users can view these videos, but they must download Windows Media Player for Apple. Keep in mind you can't select your own aspect ratio in WMV.

- **AVI**

  AVI (Audio Video Interleave) works with nearly every web browser on Windows, Mac, and Linux machines. Developed by Microsoft, AVI offers the highest quality but also large file sizes. It is supported by YouTube and works well for TV viewing.

- **AVCHD**

  Advanced Video Coding High Definition is specifically for high-definition video. Built for Panasonic and Sony digital camcorders, these files compress for easy storage without losing definition.

- **FLV, F4V, and SWF**

  Flash video formats FLV, F4V, and SWF (Shockwave Flash) are designed for Flash Player, but they're commonly used to stream video on YouTube. Flash is not supported by iOS devices.

- **MKV**

  Developed in Russia, Matroska Multimedia Container format is free and open source. It supports nearly every codec, but it is not itself supported by many programs. MKV is a smart choice if you expect your video to be viewed on a TV or computer using an open-source media player like VLC or Miro.

- **WEBM or HTML5**

  These formats are best for videos embedded on your personal or business website. They are small files, so they load quickly and stream easily.

- **MPEG-2**

  If you want to burn your video to a DVD, MPEG-2 with an H.262 codec is the way to go.

Editing and exporting video files.

Whether you shoot your footage with a video camera, you can work with your format in Adobe Premiere Pro to create the video you want, even on mobile workstations, you can export to the latest broadcast formats.

What's the difference between an MP3 and an MP4?

While similar in name, these two file types are very different. MP3 stands for MPEG-1 Audio Layer III, and it is an audio-only container. This format compresses audio files for an easy and versatile way to store sound bites and music. An MP4 (MPEG-4 Part 14), on the other hand, is a multimedia container that can store audio, video, still images, subtitles, and text.

What's the difference between an MOV and an MP4?

MOV is a proprietary video file format made for Apple's QuickTime player. This large format is great for editing videos or viewing on a larger screen. MP4s are a universal file format compatible with most streaming platforms,

Lecture 5

Digital Video Compression

Video compression is the art of throwing as much data away as possible without it showing. Video compression methods tend to be lossy that is, what comes out after decoding isn't identical to what was originally encoded. By cutting video's resolution, color depth and frame rate, PCs managed postage stamp-size windows at first, but then ways were devised to represent images more efficiently and reduce data without affecting physical dimensions. The technology by which video compression is achieved is known as a codec, an abbreviation of compression/decompression. Various types of codec have been developed implementable in either software and hardware.

Lossy techniques reduce data both through complex mathematical encryption and through selective intentional shedding of visual information that our eyes and brain usually ignore and can lead to perceptible loss of picture quality.

Lossless compression, by contrast, discards only redundant information. Codecs can be implemented in hardware or software, or a combination of both.

Compression Loss, Artifacts, and Visual Quality:

Compression artifacts are noticeable distortions in compressed video, when it is subsequently decompressed and presented to a viewer. One of the goals of compression algorithms is to minimize the distortion while maximizing the amount of compression. However, depending on the algorithm and the amount of compression, the output has varying levels of diminishing quality or introduction of artifacts.

Compression loss is manifested in many different ways and results in some sort of visual impairment. Quantization is the process of mapping a large set of input values to a smaller set for example, rounding the input values to some unit of precision. The round-off error introduced by the process is referred to as *quantization error* or the *quantization noise*. In other words, the difference between the input signal and the quantized signal is the quantization error. مهم

There are two major sources of quantization noise in video applications:

1) when an analog signal is converted to digital format;
2) when high-frequency components are discarded during a lossy compression of the digital signal.

1- Quantization of Samples:

The digitization process of an image converts the continuous-valued brightness information of each sample at the sensor to a discrete set of integers representing distinct gray levels that is, the sampled image is quantized to these levels.

2-Frequency Quantization

The human eye is fairly good at seeing small differences in brightness over a relatively large area, but not so good at distinguishing the exact strength of a high frequency (rapidly varying) brightness variation. This fact allows one to reduce the amount of information required by ignoring the high frequency components. This is done by simply dividing each component in the frequency domain by a constant for that component, and then rounding to the nearest integer. This is the main lossy operation in the whole process. As a result of this, it is typically the case that many of the higher frequency components are rounded to zero, and many of the rest become small positive or negative numbers.

Color Quantization

Color quantization is a method to reduce the number of colors in an image. As the HVS is less sensitive to loss in color information, this is an efficient compression technique.

Common Artifacts

some common artifacts that are typically found in various image and video compression applications.

a- Blurring Artifact

Blurring of an image refers to a smoothing of its details and edges, and it results from direct or indirect low-pass filter effects of various processing. Blurring of an object appears as though the object is out of focus. Also blur can appear in:

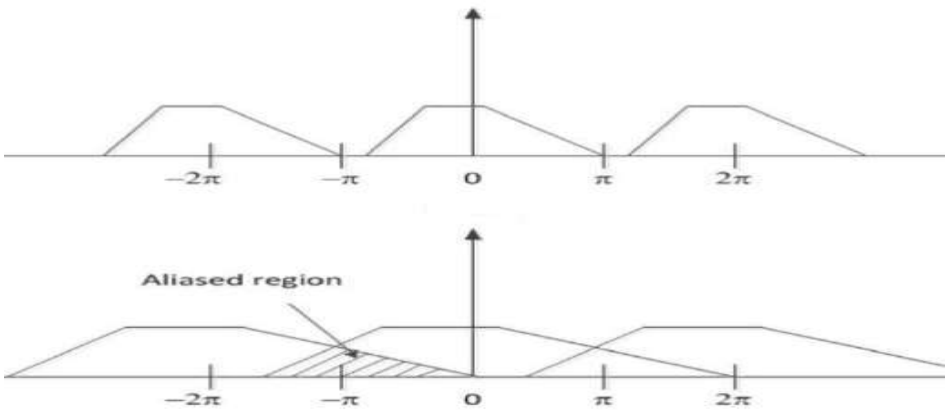
1) Graphics, image, or video editing tools may also generate the motion blur effect .
2) Motion blur appears in the direction of motion corresponding to rapidly moving objects in a still image or a video.
3) Deinterlacing by the display
4) Compression artifacts present in digital video streams can contribute additional blur during fast motion.
5) Motion blur has been a more severe problem for LCD displays, owing to their sample-and-hold nature, where a continuous signal is sampled and the sample values are held for a certain time to eliminate input signal variations.

b-**Block Boundary Artifact**

The *block boundary artifact* is the result of independently quantizing the blocks of transform coefficients, leading to discontinuities in the reconstructed block boundaries. These block-boundary discontinuities are usually visible, especially in the flat color regions such as the sky, faces, and so on, where there are little details to mask the discontinuity.

c-Aliasing Artifacts

If the transformed signal overlap with its shifted replicas. In case of such overlap, the original signal cannot be unambiguously recovered from its downsampled version, as the overlapped region represents two copies of the transformed signal at the same time. One of these copies is an *alias*, or replica of the other. This overlapping effect is called *aliasing*.



Transform domain effect of downsampling, causing aliasing

aliasing is an effect that causes different signals to become indistinguishable (or aliases of one another).

Aliasing is generally avoided by applying low pass filters anti-aliasing filters to the analog signal before sampling.

Aliasing can **occur** in :

1- signals sampled in time, for instance digital audio, and is referred to as temporal aliasing.

2- in spatially sampled signals, for instance moiré patterns in digital images. Aliasing in spatially sampled signals is called spatial aliasing.

To prevent this an anti-aliasing filter is used to remove components above the Nyquist frequency prior to sampling.

Factors Affecting Visual Quality

Visual artifacts resulting from loss of information due to processing of digital video signals usually degrade the perceived visual quality. The following are important contributing factors that affect visual quality:

a- Sensor noise: Sensor noise, is an undesirable by-product of image capture that affects visual quality.

b- Characteristics of video: Visual quality is affected by digital video characteristics including:

1- bit depth: Typical video frames use 8 bits for each pixel component, while premium quality videos allocate 10 to 16 bits.

2- Resolution: high-definition video frames are four to six times as large as standard.

3- frame rate: Frame rate is another important factor; although the HVS can perceive slow motion at 10 frames per second (fps) and smooth motion at 24 fps, higher frame rates imply smoother motion, especially for fast-moving objects. For example, a moving ball may be blurry at 30 fps, but would be clearer at 120 fps. Very fast motion is more demanding.

4- frame complexity: One measure of the complexity of a frame is the amount of details or *spatial business* of the frame.

c- Amount of compression: Highly compressed video has lower visual quality than lightly compressed video.

d- Methods of compression: Lossless compression retains all the information present in the video signal, so it does not introduce quality degradation.

e- Multiple generations of compression: Some video applications may employ multiple generations of compression, where a compressed video signal is decompressed before compressing again with possibly different parameters.

Video Quality Evaluation Methods and Metrics

Video quality evaluation is performed to describe the quality of a set of video sequences under study. Video quality can be evaluated with two approaches:

1- Subjectively: the actual visual quality of the image or video content is determined based on subjective evaluation done by humans.

2- Objectively: Objective video models are mathematical models that can automatically evaluate the quality of the multimedia content, predicting the subjective judgment. In this context, the term model may refer to a simple statistical model which can be implemented in software or hardware.

Classification of objective video quality metrics:

Objective metrics can be classified by the amount of information available about the original signal, the received signal, or whether there is a signal present at all:

1- Full Reference Methods (FR): FR metrics compute the quality difference by comparing the original video signal against the received video signal. Typically, every pixel from the source is compared against the corresponding pixel at the received video, with no knowledge about the encoding or transmission process in between. FR metrics are usually the most accurate at the expense of higher computational effort.

2- Reduced Reference Methods (RR): RR metrics extract some features of both videos and compare them to give a quality score. They are used when all the original video is not available, or when it would be practically impossible to do so, e.g. in a transmission with a limited bandwidth. This makes them more efficient than FR metrics.

3- No-Reference Methods (NR): NR metrics try to assess the quality of a distorted video without any reference to the original signal. Due to the absence of an original signal, they may be less accurate than FR or RR approaches, but are more efficient to compute. NR can be divided into:

   a- Pixel-Based Methods (NR-P): Pixel-based metrics use a decoded representation of the signal and analyze the quality based on the pixel information. Some of these evaluate specific degradation types only, such as blurring or other coding artifacts.

   b- Parametric/ Bitstream Methods (NR-B): These metrics make use of features extracted from the transmission container and/or video bitstream, e.g. MPEG-TS packet headers, motion vectors and quantization parameters. They do not have access to the original signal.

c- Hybrid Methods (Hybrid NR-P-B): Hybrid metrics combine parameters extracted from the bitstream with a decoded video signal. They are therefore a mix between NR-P and NR-B models.

Simple full-reference metrics

The most traditional ways of evaluating quality of digital video processing system (e.g. a video codec) are FR-based. Among the oldest FR metrics are signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR), which are calculated between every frame of the original video signal and the video passed through a system (e.g., an encoder or a transmission channel). PSNR is the most widely used objective image quality metric, and the average PSNR over all frames can be considered a video quality metric. However, PSNR values do not correlate well with perceived picture quality due to the complex, highly non-linear behavior of the human visual system HVS.

Lecture 6

Video equipment and applications.

When it comes to content creation these days, video production can be as simple or as complex as you need it to be. There's a wide choice of video making softwares on the market with tools for every type of creator, from hobbyist to professional , you need the right gear to film it. This can include:

- Filming equipment, like a camera, camcorder, or mobile phone with HD video capabilities
- Microphones and video lighting kits
- Video editing software for post-production

What kind of video equipment do you need ?

It's easy to assume that the latest gadgets on the market are your best options. However, this isn't always the case. Before spending half your budget on new equipment, it's crucial to prepare first by identifying the purpose of your video and your marketing goals. This process can be separated into two distinct parts.

- For this first part, you should answer the following questions:

1- What is the objective of your video? To drive brand awareness, increase sales, teach a skill…

2- What types and formats of videos will I need? Explainer videos, how-to videos, advertisements…

3- Where will these videos live and be seen? Social media, YouTube, company website...
4- What are my resources, team skill level, and budget ?

- The second part to consider before investing in video equipment expands on this last question: resources, skill level, and budget. Mapping out your project budget in advance gives you a more accurate idea of what you can afford and will realistically be able to do without running out of money in the middle of production.

Determining the setup of your video making team depends not only on the budget, but also on the experience level of each role. The number of people involved in the filming process can vary widely, from a one-man show to a cast and crew of 100+. Here's a breakdown of different types of video production teams:

- **Beginner** - Smaller budgets usually means smaller teams. Resources are limited, so starting off with basic or beginner equipment is the best bet here. Usually a beginner level team would be composed of one or two individuals, each covering multiple responsibilities, including script writing, shooting/directing, editing, and publishing.
- **Intermediate** - A step up from beginners, this type of video production team typically includes video producers, a camera team, and a copywriter.
- **Advanced** - These are your video production agencies and consultant video makers, with specialized professionals in every role. An advanced team would include video producer, copywriter, video writer, storyboard artist, director, camera team, production

assistants, sound team, makeup team, editors, graphic/motion designers, and special/visual effects artists.

Video equipment recommendations for every team

Always read reviews online for video equipment to understand their context and make sure it’s the best gear for your unique production. High quality video equipment can get costly very quickly, so you don’t want to waste resources on something that won’t work for your needs.

Video equipment for beginners and small budgets

If you’re working with a beginner level team or going the DIY route, sticking to a smaller budget is usually safer. This range of equipment is easy for any type of creator to learn to use, and is great for creating video for social media, blog, or email content.

Best beginner video editing software

Video editing can be one of the most daunting aspects of video creation for beginners. There are many web and smartphone apps designed to make the post-production process intuitive and easy to learn .Five great video editing apps for beginners

- PlayPlay - Paid (Web-based platform)
- Adobe Spark - Free (Web, Android, iOS)
- Quik - Free (Android, iOS)

- Horizon - Free (Android, iOS)
- Magisto - Free (Android, iOS)

If you want to make money with your videos online, you need to step up your production game.

Video Production Equipment

Here's a checklist of 12 pieces of video production equipment that you can use to get professional-quality video:

1. Camera

If you're making a video, you need a good camera , there are a lot of opinions out there. You can definitely use your smartphone camera for filming. It won't get you the same level of quality as a professional-level full-frame mirrorless video camera. But it's more accessible, and the video quality improves all the time.

Many creators and filmmakers who monetize videos online use DSLRs and mirrorless cameras, and the quality is great. Most of these cameras are capable of shooting in 4K. Some can record videos in 6K and even 8K. But what it really comes down to is the frame rates of the camera.

2. Highly portable camera

Why do you need another camera? Because you never know when inspiration or opportunity will strike. Video from your smartphone or a compact camera can be great for marketing, documentation, and lots of other things. If you're all-in on video, you should be able to capture it at any time. And having a phone or small camera with solid video-recording capabilities lets you capture, edit, and upload your video in a matter of minutes. Editing is where a phone will serve you better than a compact camera; there are plenty of apps that let you perform basic edits before you upload your videos to improve the viewing experience. But if you want to be able to capture quality-consistent videos at any time and shoot vlogs and videos on the go, a portable camera can come in handy.

3. Tripod

No matter how good your camera's stabilization is, you need a tripod. Even small shakes can be visible in a video especially if you're using a DSLR with a zoom lens. A tripod turns even a basic camera into a much better video production tool. A tripod will seriously improve the video quality from any camera.



4. External microphone

While some video cameras, camcorders, phones have great audio pickups, they still leave much to be desired. And if you're using something smaller, the audio will likely be terrible. These devices aren't designed to capture high-quality sound.

External microphones significantly improve the quality of your audio. This is especially important when you're filming video lessons, courses, live streams, interviews, or anything else that contains speech. Fortunately, you have some solid external mic options without breaking the bank.

5. Lighting

Good lighting makes a huge difference in the final quality of your video. If you've ever seen a video where an interviewee is poorly lit, you know that it's distracting. Getting your lighting right is not easy. You need to consider different types of light, foreground and background lighting, shadows, and equipment. And that equipment can get expensive. If you're just getting started, you can use a reflector to take advantage of ambient light. It's just a matter of getting it set up to properly reflect the light. (You’ll also need someone to hold it unless you get a reflector holder as well.)

6. Gimbal

A gimbal is like a stabilizer for your camera. There are many different types and sizes; you can find gimbals for cameras as small as your iPhone and as big as a professional-level video camera. By using pivots and weights, a gimbal stabilizes your shot, which is especially important when you’re moving. When your camera is still, a tripod will work. If you’re moving around, you’ll need a gimbal.

7. Editing/Production Software

With some basic editing skills, your video quality will go way up. You don’t even need to get into advanced techniques like color correcting or complicated cuts. Just editing out your vocal pauses and inserting some text is enough to take your video from amateurish to pro-quality. Mastering editing and post-production software is a long process. But you can get started with just a few basic tutorials and you’ll see an improvement in your video quality right away.

Here are some of the best editing software for creators at all levels:

- iMovie
  If you have a Mac computer to edit on, this software comes free with it. iMovie is simple to learn and easy to use. This is a great place to start for beginners.
- Final Cut Pro
  Final Cut Pro is a fantastic production software. It can be as simple or as complex as you want. You can use hotkeys to quickly fly around your editing timeline to make necessary adjustments.

- DaVinci Resolve
  DaVinci Resolve is one of the best editing tools available on both Mac, Windows, and Linux. The software offers pretty much all you need to create, organize, edit, and render your video project.

8. Video Editing Hardware

You can edit videos on almost any device. Most smartphones can download capable video-editing apps, and just about every computer can run one as well. But if you’re going to be doing a lot of video editing, you may want to consider upgrading your hardware. Many top videographers use Apple hardware for their video editing, but Windows PCs have caught up in graphics processing power, too.

The main thing to pay attention to is that your computer has enough graphics power to let you edit at full speed. Lower-end graphics cards and wifi 6 laptops may have difficulty making changes to very large files. And that slows down your computer. You'll also need a big enough hard drive to manage your growing file library. In general, the more power the better when it comes to editing. The last thing you want is to start editing your video and discover you're constantly waiting for your computer to render your editing timeline or stuttering during video playback because that makes editing nearly impossible to do effectively.

For most people, editing on your current device will be fine. If you notice that the process is slow and you think you might benefit from something faster, it's time to consider making a change.

9. Transcription Service

Adding text to your video is a great way to increase engagement, no matter where you're going to be posting it. Videos on social media and OTT platforms are often watched without sound, so you'll lose a lot of viewers if you don't have subtitles. Even if you're not posting on social media, you'll still want to have captions available for viewers with hearing difficulties (and those who like to read along as they watch). The problem is that if you're going to be creating a lot of videos, transcription takes a lot of time. That's why using a transcription service is a good idea.

10. High-Quality Memory Cards

High-quality SD cards can make a big difference when shooting videos. You want high capacity (64GB or more) with high read and write speeds.

A great option is the 64GB SanDisk Extreme Pro card with speeds of 300MB/s. This allows you to shoot long videos and never have to worry about running out of space. If you’re planning to shoot longer videos, it’s best to go for 128GB or larger though.

11. Capture Cards

If you’re planning on live streaming or hosting a virtual event, a capture card is an essential gear that you need in your video production arsenal .This tool allows your computer to process the data from your camera .

12. Dummy Battery

Dummy battery is yet another useful tool you need for live streaming and video production. Having a dummy battery is a game-changer. It allows you to easily plug your camera into the wall so you wouldn’t have to worry about your battery level when shooting a long video or live streaming. Dummy batteries are very affordable and can drastically improve and speed up your video production process.

Lecture 7

Motion Estimation and Transform Coding

Motion estimation is the process of determining *motion vectors* that describe the transformation from one 2D image to another; usually from adjacent frames in a video sequence. It is an ill-posed problem as the motion is in three dimensions but the images are a projection of the 3D scene onto a 2D plane. The motion vectors may relate to the whole image (global motion estimation) or specific parts, such as rectangular blocks, arbitrary shaped patches or even per pixel. The motion vectors may be represented by a translational model or many other models that can approximate the motion of a real video camera, such as rotation and translation in all three dimensions and zoom.

The methods for finding motion vectors can be categorised into pixel based methods ("direct") and feature based methods ("indirect").

❖ Direct methods

- Block-matching algorithm
- Phase correlation and frequency domain methods
- Pixel recursive algorithms
- Optical flow

❖ Indirect methods

Indirect methods use features, such as corner detection, and match corresponding features between frames, usually with a statistical function applied over a local or global area. The purpose of the statistical function is to remove matches that do not correspond to the actual motion. Statistical functions that have been successfully used include RANSAC.

Additional note on the categorization

All methods require some kind of definition of the matching criteria. The difference is only whether you :

1- summarise over a local image region first and then compare the summarisation (such as feature based methods), or you compare each pixel first (such as squaring the difference) and then summarise over a local image region (block base motion and filter based motion).

2- An emerging type of matching criteria summarises a local image region first for every pixel location (through some feature transform such as Laplacian transform), compares each summarised pixel and summarises over a local image region again.

Some matching criteria have the ability to exclude points that do not actually correspond to each other albeit producing a good matching score, others do not have this ability, but they are still matching criteria.

Video coding

Applying the motion vectors to an image to synthesize the transformation to the next image is called motion compensation. It is most easily applied to discrete cosine transform (DCT) based video coding standards, because the coding is performed in blocks.

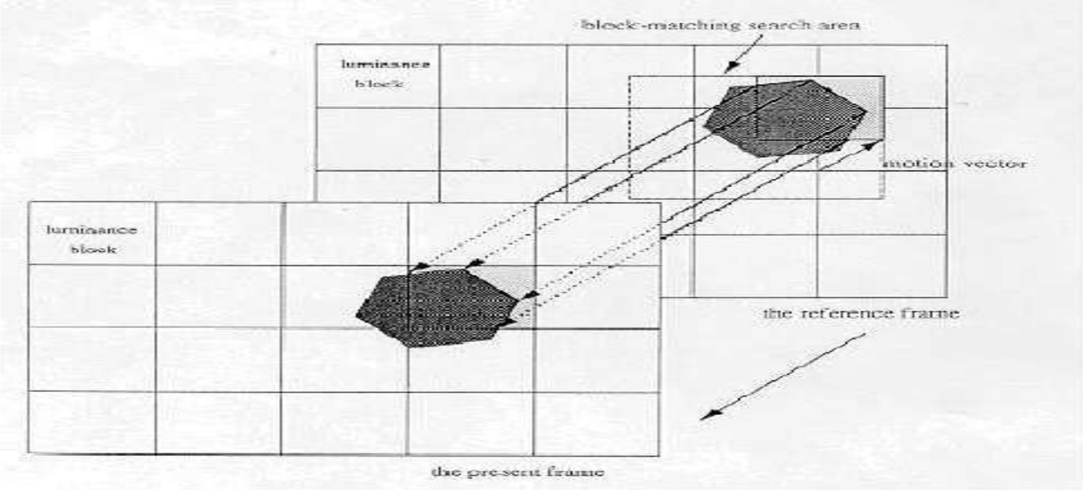
Motion compensation *in computing, is an algorithmic technique used to predict a frame in a video, given the previous and/or future frames by accounting for motion of the camera and/or objects in the video.*

It is employed in the encoding of video data for video compression, for example in the generation of MPEG-2 files. Motion compensation describes a picture in terms of the transformation of a reference picture to the current picture. The reference picture may be previous in time or even from the future. When images can be accurately synthesized from previously transmitted/stored images, the compression efficiency can be improved.

Almost all video coding standards use block-based motion estimation and compensation such as the MPEG series including the most recent HEVC.

A Block Matching Algorithm *is a way of locating matching macroblocks in a sequence of digital video frames for the purposes of motion estimation.*

The underlying supposition behind motion estimation is that the patterns corresponding to objects and background in a frame of video sequence move within the frame to form corresponding objects on the subsequent frame. This can be used to discover temporal redundancy in the video sequence, increasing the effectiveness of inter-frame video compression by defining the contents of a macroblock by reference to the contents of a known macroblock which is minimally different. A metric for matching a macroblock with another block is based on a cost function.



Advantages of BMA (Block Matching Algorithm)

- Low overhead: one vector per block
- Straightforward, regular parallel procedure good for VLSI implementation, Low cost VLSI implementation.
- Robust (immunity to noise)

Disadvantages:

- A block may contain several moving objects.
- Minimizing a numerical criterion may not give us the true movement
- Fail to zoom rotational motion, local deformation
- Blocking artifact

1. Mean squared error (MSE): *(min.)*

$$MSE(d_1,d_2)=\frac{1}{N_1N_2}\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}[f(n_1,n_2,t)-f(n_1-d_1,n_2-d_2,t-1)]^2$$

2. Mean absolute difference (MAD): *(min.)*

$$MAD(d_1,d_2)=\frac{1}{N_1N_2}\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}|f(n_1,n_2,t)-f(n_1-d_1,n_2-d_2,t-1)|$$

3. Normalized cross-correlation function (NCF): *(max., most complex)*

$$NCF(d_1,d_2)=\frac{\sum\sum f(n_1,n_2,t)f(n_1-d_1,n_2-d_2,t-1)}{[\sum\sum f^2(n_1,n_2,t)]^{1/2}[\sum\sum f^2(n_1-d_1,n_2-d_2,t-1)]^{1/2}}$$

Moving object detection *is a technique used in computer vision and image processing. Multiple consecutive frames from a video are compared by various methods to determine if any moving object is detected.*

Moving objects detection has been used for wide range of applications like video surveillance, activity recognition, road condition monitoring, airport safety, monitoring of protection along marine border, etc..

Moving object detection is to recognize the physical movement of an object in a given place or region. By acting segmentation among moving objects and stationary area or region, the moving objects' motion can be tracked and thus analyzed later. To achieve this, consider a video is a structure built upon single frames, *moving object detection is to find the foreground moving target(s),* either in each video frame or only when the moving target shows the first appearance in the video.

Traditional methods

Among all the traditional moving object detection methods, we could categorize them into four major approaches: Background subtraction, Frame differencing, Temporal Differencing, and Optical Flow.

1-Frame differencing

Instead of using traditional approach, to use image subtraction operator by subtracting second and images afterwards, the frame differencing method makes comparisons between two successive frames to detect moving targets.

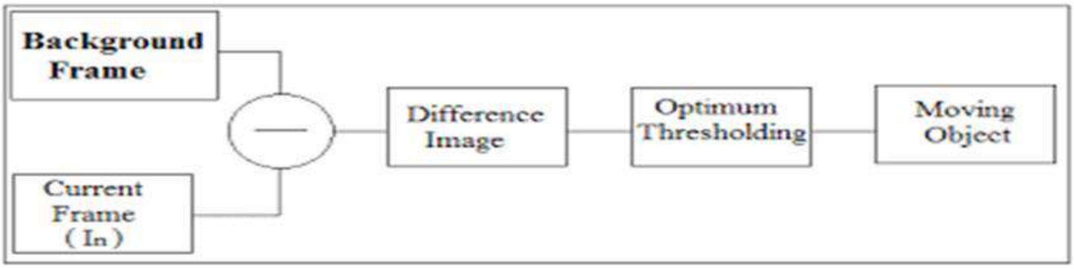
2- Temporal differencing

The temporal differencing method identifies the moving object by applying pixel-wise difference method with two or three consecutive frames.

3- Background subtraction

Background subtraction is any technique which allows an image's foreground to be extracted for further processing .

Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called "background image", or "background model". Background subtraction is mostly done if the image in question is a part of a video stream.

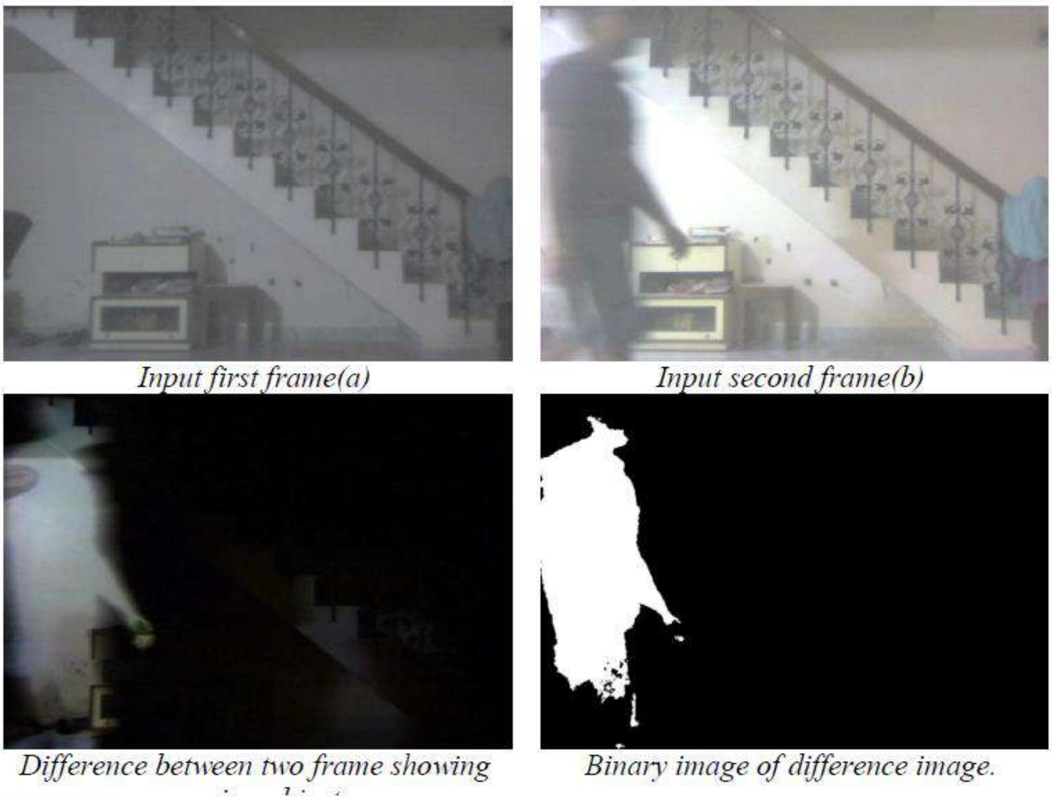


Motion Estimation

The basic premise of motion estimation is that in most cases, consecutive video frames will be similar except for changes induced by objects moving within the frames. In the trivial case of zero motion between frames (and no other differences caused by noise, etc.), it is easy for the encoder to efficiently predict the current frame as a duplicate of the prediction frame. When this is done, the only information necessary to transmit to the decoder becomes the syntactic overhead necessary to reconstruct the picture from the original reference frame. When there is motion in the images, the situation is not as simple.

When there is movement in the frame:

When there is movement in the scenes then the binary image of the difference between the two frames shows motion having white color and where there is no change shows black color.



Input first frame(a) *Input second frame(b)*

Difference between two frame showing *Binary image of difference image.*

Limitations

As the air moves, the camera not remains in the position of static so when there is no movement of object then also it results motion and shows holes in the binary output image.

Lecture 8

Video Modeling and Retrieval.

An *image retrieval* system is a computer system used for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding *metadata* such as *captioning, keywords, title or descriptions* to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.

Content-Based Image Retrieval is a well studied problem in computer vision, with retrieval problems generally divided into two groups:

- Category-level retrieval
- Instance-level retrieval.

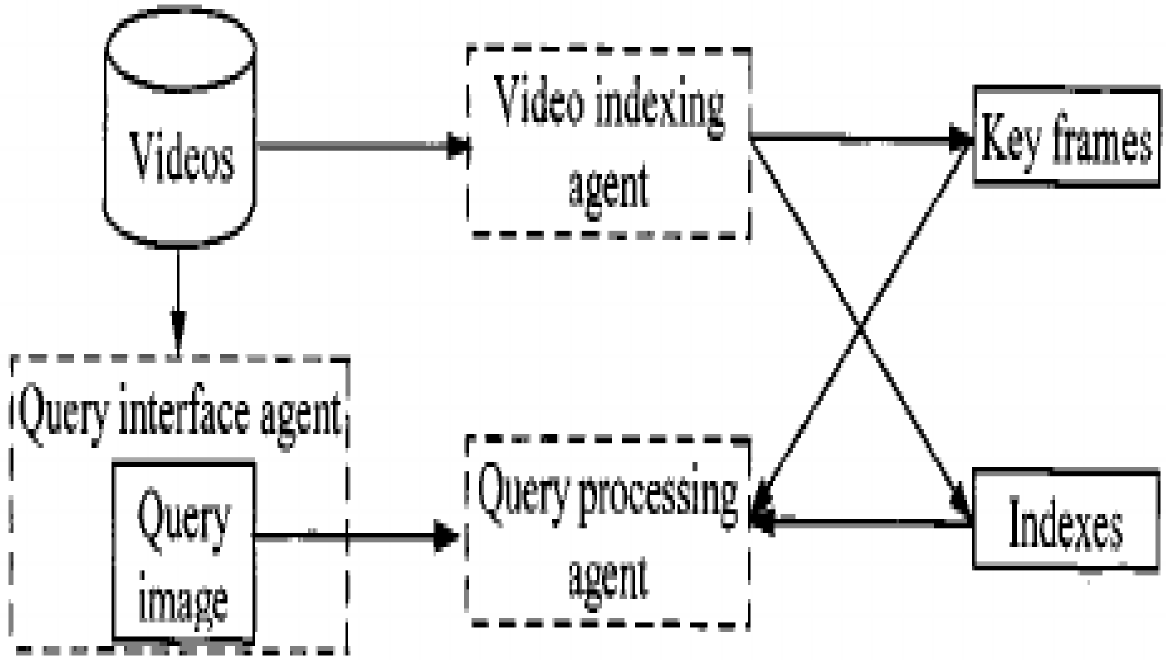
Given a query image of the Sydney Harbour bridge, for instance, category-level retrieval aims to find any bridge in a given dataset of images, whilst instance-level retrieval must find the Sydney Harbour bridge to be considered a match.

A complete video consist of these things :

•Subtitle .

•Sound track .

•Images recorded or played out continuously at a fixed rate



Video IR Method

The ultimate goal of video retrieval is to find relevant video content for a specific information need of the user.

There are so many types of retrieval in digital video such as :

•**Metadata-based method** :Structured metadata using traditional DBMS .

•**Text-based method** : Associated subtitles using IR technique .

•**Audio-based method** : Associated soundtracks using Audio IR .

•**Content-based method** :Video as collection of independent images , Video sequences divides into groups of similar frames .

•**Integrated approach** : Combination of two or more above methods.

What is the video made of ?

Video is made of video ***shot.*** A ***shot*** is a sequence of contiguous frames have one or more following features :

•The frames depict the same scene .

•The frames signify a single camera operation .

•The frames contain a distinct event .

•The frames are chosen as a single indexable entity by the user.

Shot-based video IR consist of few steps:
•Segment the video into shots This is called video temporal segmentation •Index each shot The steps: identify keyframes for each shot, and use image IR •Apply a similarity measurement between queries and video shots, then retrieve shots with high similarities. Use image IR based on feature vectors in 2nd step

Video Shot Detection or Segmentation

Segmentation is a process for dividing a video sequence into shots .

Segmentation require suitable *quantitative measure* that captures the differences between a pair of frames . If the differences exceeds a threshold, it may be interpreted as indicating a segment boundary . Establishing suitable differences metrics and technique for applying them are the key issues in automatic partitioning.

Basic Video Segment Techniques
The key issue is how to measure the frame-to-frame differences , there are two basic techniques : •*Sum pixel-to-pixel differences between neighboring frames* If the sum is larger than a press threshold, a shot boundary exist between these two frames •*Measures color histogram distance between neighboring frames* Object motion causes histogram differences. If large differences is found, a camera break occurred

Video Indexing and Retrieval

Video Indexing and Retrieval Shots are needs to be represented and indexed, in order to locate and retrieved the shots quickly. We represents each shot by one or more keyframes or r frames (representative frames). Video retrieval is based on similarity between the query and r frames.

IR Based on r Frames of Video Shots How to represent and index each shot in IR based on r frames?
1. Use r frames (representative frames) to represent each shot. 2. During retrieval, features on the r frames are compared with queries. 3. If the frames is similar/relevant, it is presented to the user.

How many representative frame(s) should be used in a shot, and how we select the representative frames?

How Many R Frames Should be Used in a Shot? There are few methods:

• Uses one r frames per shot.

• Assigns the number of r frames to shots according to their length.

• Divides a shot into subshots, and assign one r frame into each subshot. Subshot are detected based on changes on the content.

How to Select the R Frames? There are few methods:

• Use first frame for each segment/shot as the r frame.

• Use a frame which is similar with average frame. Each pixel in this frame is the average of pixel values at the same grid point in all frames of the segment.

• Use a frame whose its histogram is closest to the average histogram. The histograms of all the frames in the segment are averaged

1-IR Based on Motion Information

Motion information is derived from optical flow or motion vectors. Parameters used are:

- Motion content Total amount of motion within a given video
- Motion uniformity Smoothness of the motion within a video as a function of time
- Motion panning Left-to-right or Right-to-left motion of the camera
- Motion tilting Vertical motion component of the motion within a video sequence

2-IR Based on Objects Object

IR Based on Objects Object is a group of pixels that move together. To index the video, use the segmented objects. Object motion can help to construct a description of that motion for use in subsequent retrieval of the video shot. Object-based video IR is easy when the video is compressed using the MPEG-4 object-based coding standard.

3-IR Based on Metadata Video

IR Based on Metadata Video IR based on metadata is using conventional DBMS. Metadata such as: title, video type, directors, genre, etc.

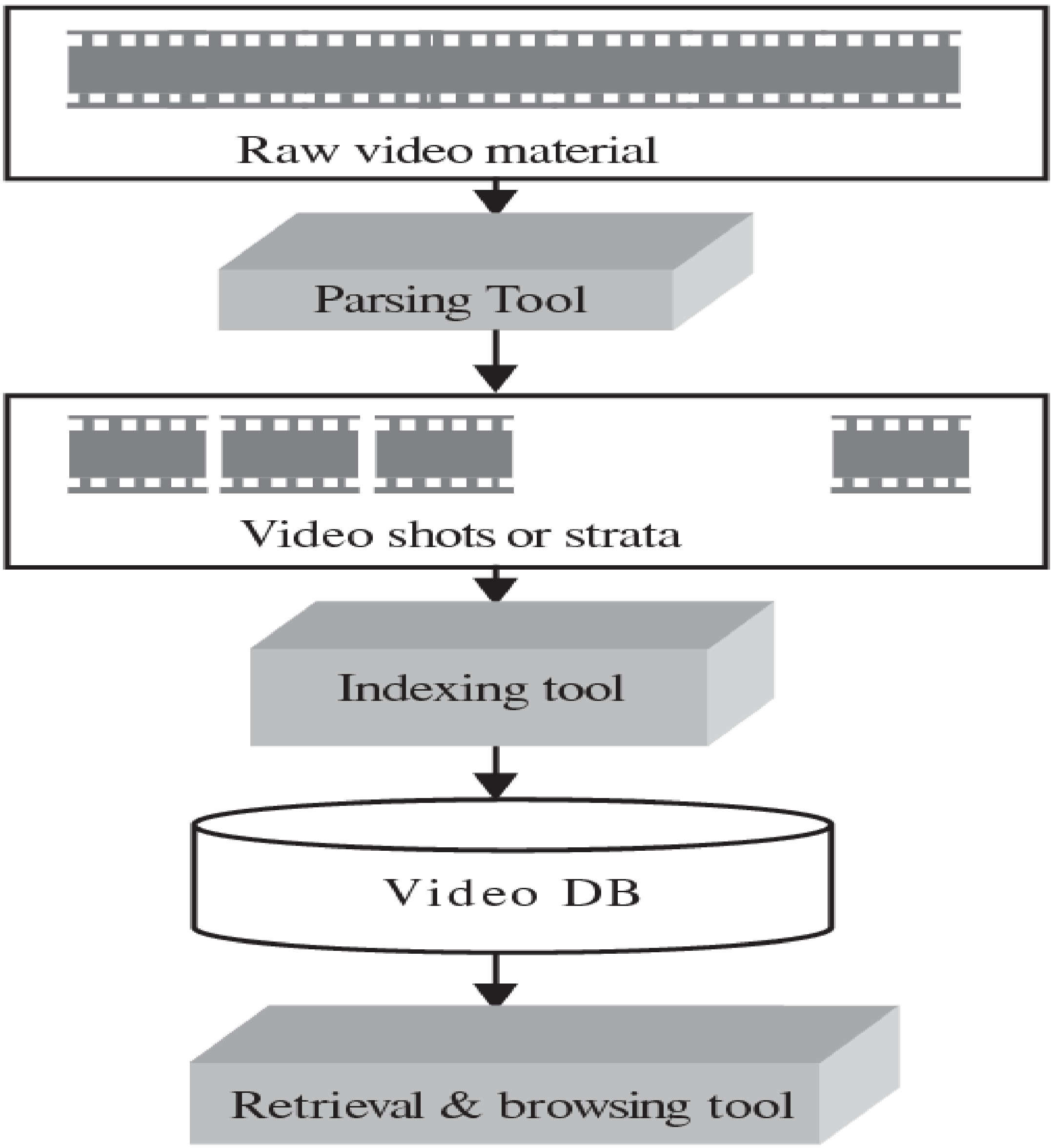
4-IR Based on Annotation Video

IR Based on Annotation Video IR based on annotation is using IR technique. Annotation is obtained in ways:

• Video is manually interpreted and annotated. A time consuming task, automatic high-level video content understanding is currently not possible for general video.

• Videos have associated transcripts and subtitles.

• Speech recognition within video. Extract spoken words for indexing and retrieval.



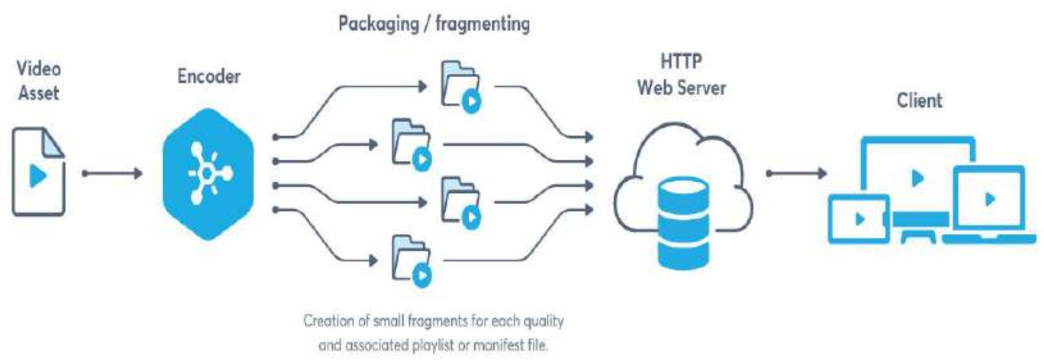
Video metadata data attached to images, audio and video files, etc. video metadata is data that describes the characteristics of a video file. It doesn't reveal what is in the video but rather what the video is. Metadata also makes video content easier to find on your website , types of metadata:

1- Administrative Video Metadata
2- Descriptive Metadata
3- Structural Metadata

Lecture 10

Video Transcoding

In the context of video Transcoding refers to *the process of compressing video files as much as possible at minimal quality loss to represent (and transfer) information by using less data.* Essentially, video transcoding online *is the conversion of a video file from one format to a better-compressed version to ensure consumers can stream content without buffering and at the highest possible qualities.* It's easy to get Transcoding mixed up with Encoding. A good encoding *definition is 'the process of converting a raw file (codec) in to a compatible, compressed and efficient digital format.'* The transcoding process creates a copy of a video file in a format that's suitable for playback on your platform of choice at the quality and in the file size you desire. In a nutshell, transcoding decodes encoded data into an intermediate format and then encodes it into the target format.



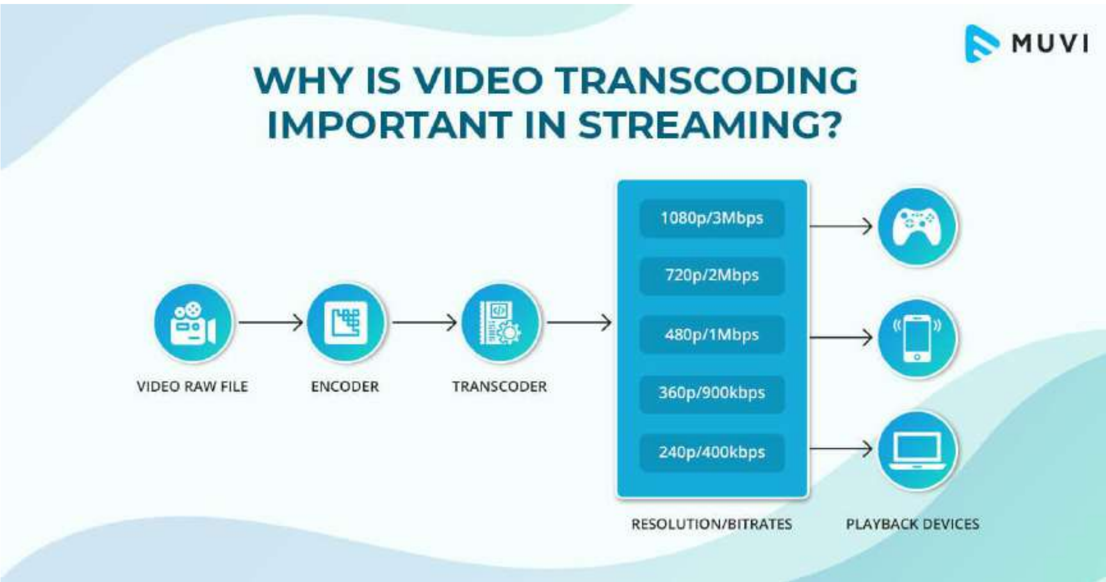
What is transcoding and why is it important?

Video file sizes are large. The better the quality of the video, or the more footage you need to work with, the more those video files will tax your computer system. Transcoding helps improve your workflow by creating a copy of your video files in a new format that will deliver better playback in your editing platform.

When Should You Transcode Video?

By transcoding video, you accomplish the following:

- Simultaneously create a set of time-aligned video streams, each with a different frame size and bitrate.
- Convert codecs and protocols for a larger audience.
- Package those internet-friendly streams into adaptive streaming formats, such as **HLS**, enabling playback on almost any screen.



Why is Transcoding Critical?

Video transcoding is critical when you want your content to reach more end users. *For example*, you're preparing to do a live stream from your office. It's a possibility that you're capturing webcam audio and video with a browser-based desktop application like Adobe Flash, which generates 1080p H.264 video and Speex audio.

With video transcoding, you can create a set of time-aligned video streams, each with a different bitrate and frame size, while converting the Speex audio to AAC audio. This group of streams is internet-friendly, allowing you reach virtually any screen on the planet.

How is Video Transcoding Done?

Video begins life in a format unique to the camera or program used to capture your images and sounds. This is typically not the format needed to deliver the video online or play it from another device. Video transcoding is used to change the original format into one allowing the video and audio to play properly on a user's devices be they computer, tablet, smartphone, or tv. *Video encoding*, or *video transcoding*, involves a two-step process.

- First, the original file is decoded to an uncompressed format.
- Second, this uncompressed format is then encoded into the target format.

What Are the Types of Video Transcoding?

There are three main types:

- **Standard transcoding,** which involves changing the video or audio to transcode a video or stream. For example, streaming a digital conference usually requires working with IP cameras set within a conference space. If those cameras use the Real Time Streaming Protocol (RTSP) and cannot create a video stream suitable for online playback, you can convert the content into an adaptive bitrate stream by transcoding the video.

- **Transsizing (image scaling),** which involves resizing a video frame. For example, you can lower a 4K resolution to 1,080 pixels.
- **Transrating,** which changes the bitrate without modifying the video content, format, or codec. For example, to ensure that the video can fit into less storage space or can be broadcast over a lower bandwidth connection, reduce an 8-Mbps bitrate to 3-Mbps.

Transrating comprises three types:

- **Lossless to lossless,** which maintains the quality of the video across formats, enabling you to take advantage of more effective hardware or compression algorithms.
- **Lossless to lossy,** which reduces the quality of the video but yields a smaller and faster file or one that’s compatible with the requirements of a certain platform, browser, or player.
- **Lossy to lossless,** which ensures that the video quality does not deteriorate further during the conversion process. Note that, by adopting this process, you **cannot** regain the data and quality previously lost through compression.

What are Formats and File Types in Video Transcoding?

When talking about video transcoding, we refer to file formats, file types, and something called codecs.

- Video format specifies how video and audio have been combined and tells playback devices how to play the files.
- Video and audio codec formats refer to the technologies used to both create and play back the digital video and audio.

Examples of file types are MPEG-4 (MP4), Quicktime (MOV), and Flash Video (FLV). The key variable in most cases is the streaming platform you use. Always check what formats and codecs are supported by the platform you wish to use, and whether your video file must be converted to a supported format.

MP4 video provides better quality and compression over WMV, allowing you to store better quality video for the same file size. A good video transcoding program will be able to convert multiple file formats and codecs.

What is the Future of Video Transcoding?

The demand for streaming video has exponentially increased the demand for video transcoding, and clearly, streaming video is here to stay. Why?

- **Convenience**

No one likes being tied to a schedule, and streaming video allows you to watch your programs when and where you have the time to watch them. You can watch them on a bus, you can watch them on a train. You can watch them…well, you get the picture.

- **Demographics**

Millennials and Generation Xers are leading the trend toward streaming video. 63% of Millennials watch live video, and streaming services put that capability in the palm of their hands.

- **Personalization**

Video streaming apps and services make recommendations based on user viewing profiles. In other words, they customize your buffet of recommended content based on your personal interests.

- **Scalability**

Many streaming video services allow you to pick and choose the features and functions you want, and only pay for those selections. If you don't use it, you don't pay for it.

While these are the necessary video transcoding problems :

- Increasing Need for Storage
- Increasing Online Traffic
- Increasing Need for Video Compression

The video transcoding process

When an editing app works with multiple file types, it can drop frames as it struggles to decode media in real time, meaning you lose footage.

Transcoding (which is a process of decoding, reformatting, and re-encoding files) takes source footage of various types and recodes it into a single video codec or file format. This improves performance of the editing program as well as the filmmaker’s user experience especially if a creator or editor is collaborating on work. Transcoding gives projects a common file type that all collaborators can work with.

While transcodes are the same frame rates and frame sizes of your original video files, and an hour of footage will take an hour to transcode, it’s faster to edit transcoded footage than to juggle multiple video formats while you work. Transcoding will help your editing program and computer work more quickly so you can work more efficiently.

For example, many editors work with smartphone footage. These clips are often in the H264 or HEVC file format, a format that can be very processor intensive, which means it will slow down your machine. If you transcode those files into video formats like ProRes or DNx, you can more quickly edit your work.

Lecture 10

Video quality evaluation methods and metrics

Video quality and performance directly impact a viewer's quality of experience, which directly impacts media organizations' business outcomes. In this white paper, we will provide an understanding of the video quality metrics, methodology, and measurement tools, as well as some of the best practices in designing a video quality and performance measurement framework that will help your organization improve viewer experience, retain viewership, and ensure that your services can stand up to increasing competition.

Video quality is a characteristic of a video passed through a video transmission or processing system that describes perceived video degradation (typically, compared to the original video). Video processing systems may introduce some amount of distortion or artifacts in the video signal that negatively impacts the user's perception of a system. For many stakeholders in video production and distribution, assurance of video quality is an important task. Video quality evaluation is performed to describe the quality of a set of video sequences under study. Video quality can be evaluated objectively (by mathematical models) or subjectively (by asking users for their rating). Also, the quality of a system can be determined offline (i.e., in a laboratory setting for developing new codecs or services), or in-service (to monitor and ensure a certain level of quality).

Objective video quality models are mathematical models that approximate results from subjective quality assessment, in which human observers are asked to rate the quality of a video.[1] In this context, the term model may refer to a simple statistical model in which several independent variables (e.g. the packet loss rate on a network and the video coding parameters) are fit against results obtained in a subjective quality evaluation test using regression techniques. A model may also be a more complicated algorithm implemented in software or hardware.

Objective models can be classified by the amount of information available about the original signal, the received signal, or whether there is a signal present at all:[5] Full Reference Methods (FR): FR models compute the quality difference by comparing the original video signal against the received video signal. Typically, every pixel from the source is compared against the corresponding pixel at the received video, with no knowledge about the encoding or transmission process in between. More elaborate algorithms may choose to combine the pixelbased estimation with other approaches such as described below. FR models are usually the most accurate at the expense of higher computational effort. As they require availability of the original video before transmission or coding, they cannot be used in all situations (e.g., where the quality is measured from a client device).

Reduced Reference Methods (RR): RR models extract some features of both videos and compare them to give a quality score. They are used when all the original video is not available, or when it would be practically impossible to do so, e.g. in a transmission with a limited bandwidth. This makes them more efficient than FR models at the expense of lower accuracy. No-Reference Methods (NR): NR models try to assess the quality of a distorted video without any reference to the original signal. Due to the absence of an original signal, they may be less accurate than FR or RR approaches, but are more efficient to compute. Pixel-Based Methods (NR-P): Pixel-based models use a decoded representation of the signal and analyze the quality based on the pixel information. Some of these evaluate specific degradation types only, such as blurring or other coding artifacts. Parametric/Bitstream Methods (NR-B): These models make use of features extracted from the transmission container and/or video bitstream, e.g. MPEG-TS packet headers, motion vectors and quantization parameters. They do not have access to the original signal and require no decoding of the video, which makes them more efficient. In contrast to NRP models, they have no access to the final decoded signal. However, the picture quality predictions they deliver are not very accurate. Hybrid Methods (Hybrid NR-P-B): Hybrid models combine parameters extracted from the bitstream with a decoded video signal.[6] They are therefore a mix between NR-P and NRB models.

All the visual artifacts are still valuable for video quality. Unique not mentioned attributes include Spatial Blurring — a result of loss of high spatial frequency image detail, usually at sharp edges. Blocking — is caused by multiple algorithms because of the internal representation of an image with blocks size 8, 16, or 32. With specific parameters, they can average pixels inside a block making blocks distinct Ringing, echoing or ghosting - takes the form of a “halo,” band, or “ghost” near sharp edges. Color bleeding — occurs when the edges of one colour in the image unintentionally bleeds or overlaps into another colour Staircase noise — is a special case of blocking along a diagonal or curved edge. Rather than rendering as smooth, it takes on the appearance of stair steps Temporal Flickering — is usually frequent brightness or colour changes along the time dimension. It is often broken out as fine-grain flickering and coarse-grain flickering. Mosquito noise — a variant of flickering, it’s typified as haziness and/or shimmering around high-frequency content (sharp transitions between foreground entities and the background or hard edges). Floating — refers to illusory motion in certain regions while the surrounding areas remain static. Visually, these regions appear as if they were floating on top of the surrounding background Jerkiness or judder — is the perceived uneven or wobbly motion due to frame sampling. It’s often caused by the conversion of 24 fps movies to a 30 or 60 fps video format. The majority of them can be grouped into compression artifacts

The main goal of many-objective video quality metrics is to automatically estimate the average user's (viewer's) opinion on the quality of a video processed by a system. Procedures for subjective video quality measurements are described in ITU-R recommendation B