MMR COMP5425

MULTIMEDIA RETRIEVAL

THE UNIVERSITY OF SYDNEY

Week04

Semester 1, 2025

Multimedia Basics

- Digital Image
- Digital Video
- Digital Audio

Visual Perception

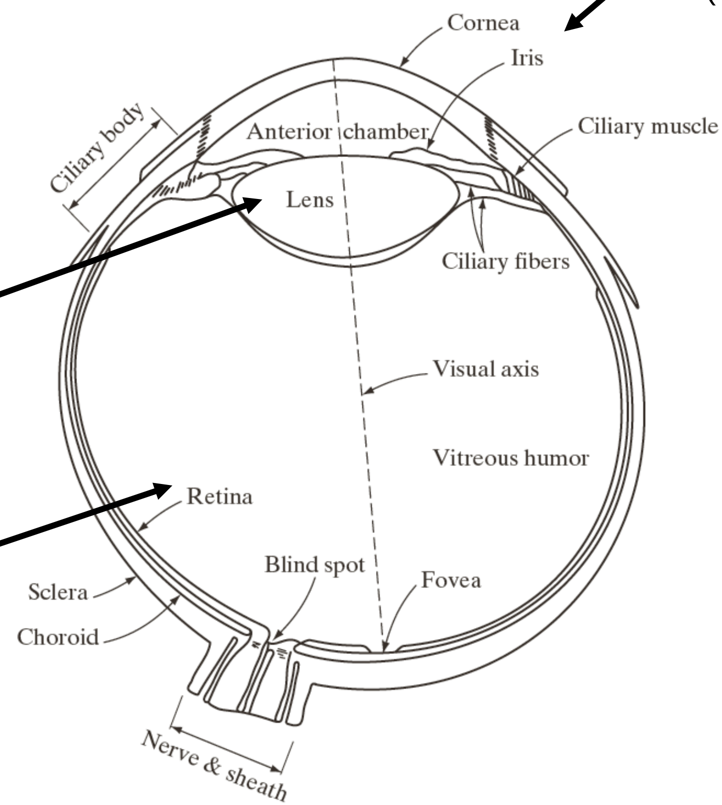
- Structure of the Human Eyes

Contracts & expands to control the amount of light entering the eye

Central opening of the Iris: Pupil (diameter: ~2 to 8 mm)

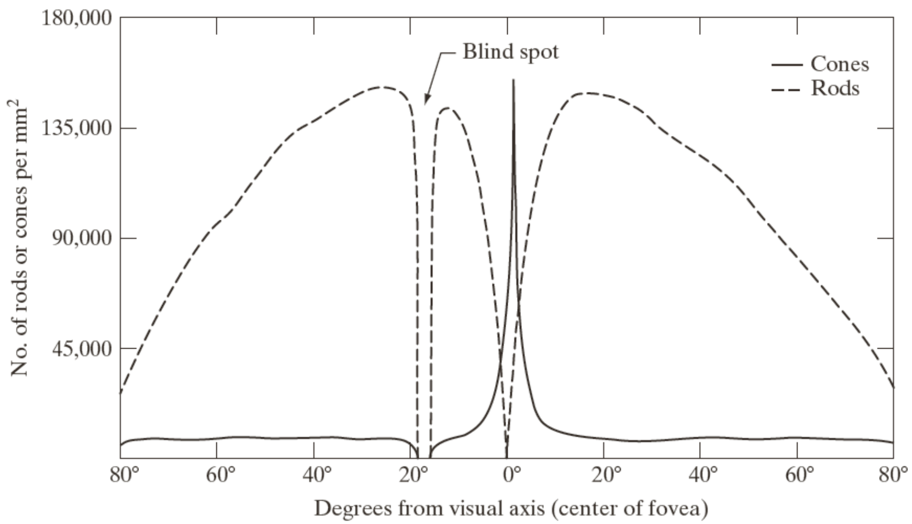
Concentric layers of fibrous cells
Absorb ~8% of the visible light spectrum

Innermost membrane of the eye
Light from object imaged on retina



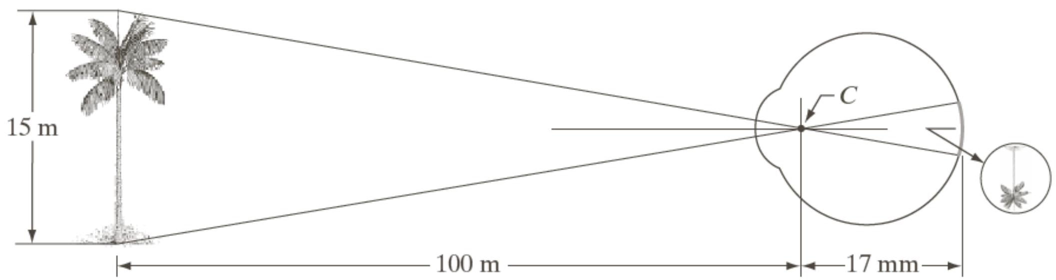
Visual Perception

- Structure of the Human Eyes
  - Distribution of discrete light receptors over the surface of the retina
  - 2 classes of receptors: cones and rods
    - Cones: 6 – 7 million in each eye, mainly located in the fovea. Highly sensitive to color, fine details. (Photopic or bright-light vision)
    - Rods: 75 – 150 million, distributed. Sensitive to low level illumination, not involved in color vision. (Scotopic or dim-light vision)



Visual Perception

- Image Formation in the Eye



- Photo camera: lens has fixed local length. Focusing at various distances by varying distance between the lens and an imaging plane (location of film or sensor chip)
- Human eye: converse. Distance lens-imaging region (retina) is fixed. Focal length for proper focus obtained by varying the shape of the lens.

Visual Perception

- Structure of the Human Eyes
  - Approximation: fovea ≈square sensor array of size 1.5mm x 1.5mm
  - Density of cones in this area: 150,000 elements/mm^2
    - Number of cones in the region of highest acuity in the eye: ~337,000 elements
  - Just in terms of raw resolving power, a CCD can have this number of elements in a receptor array no larger than 5mm x 5mm.
    - Basic ability of the eye to resolve details is comparable to current electronic imaging sensors

Visual Perception

- Perceived intensity is not a simple function of actual intensity

Although the intensity of the stripes is constant, we actually perceive a brightness pattern that is strongly scalloped, especially near the boundaries.
These seemingly scalloped bands are called *Mach bands* after Ernst Mach, who first described the phenomenon in 1865.

This phenomenon clearly demonstrate that perceived brightness is not a simple function of intensity. The human visual system tends to undershoot or overshoot around the boundary of regions of different intensities.

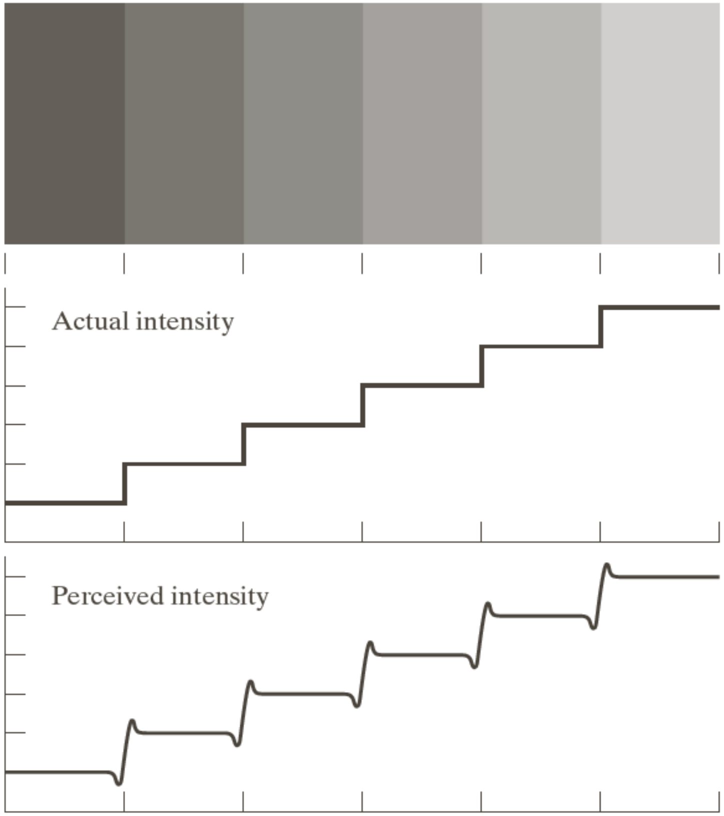


Illustration of the Mach band effect.

Visual Perception

- Perceived intensity is not a simple function of actual intensity

a b c

FIGURE 2.8 Examples of simultaneous contrast. All the inner squares have the same intensity, but they appear progressively darker as the background becomes lighter.

All the center squares have exactly the same intensity. However, they appear to the eye to become progressively darker as the background becomes lighter.

A more familiar example is a piece of paper that seems white when lying on a desk, but can appear totally black when used to shield the eyes while looking directly at a bright sky.

Light

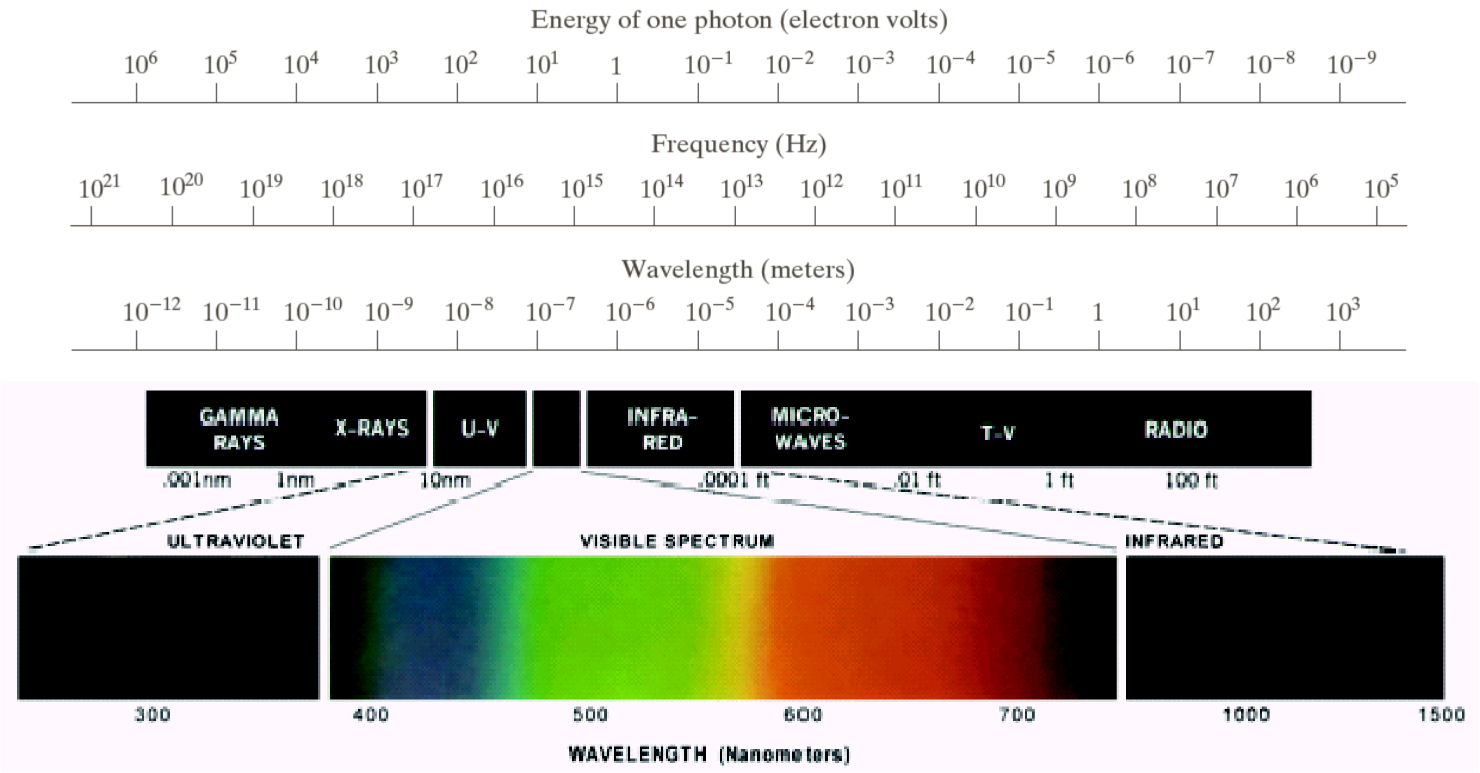
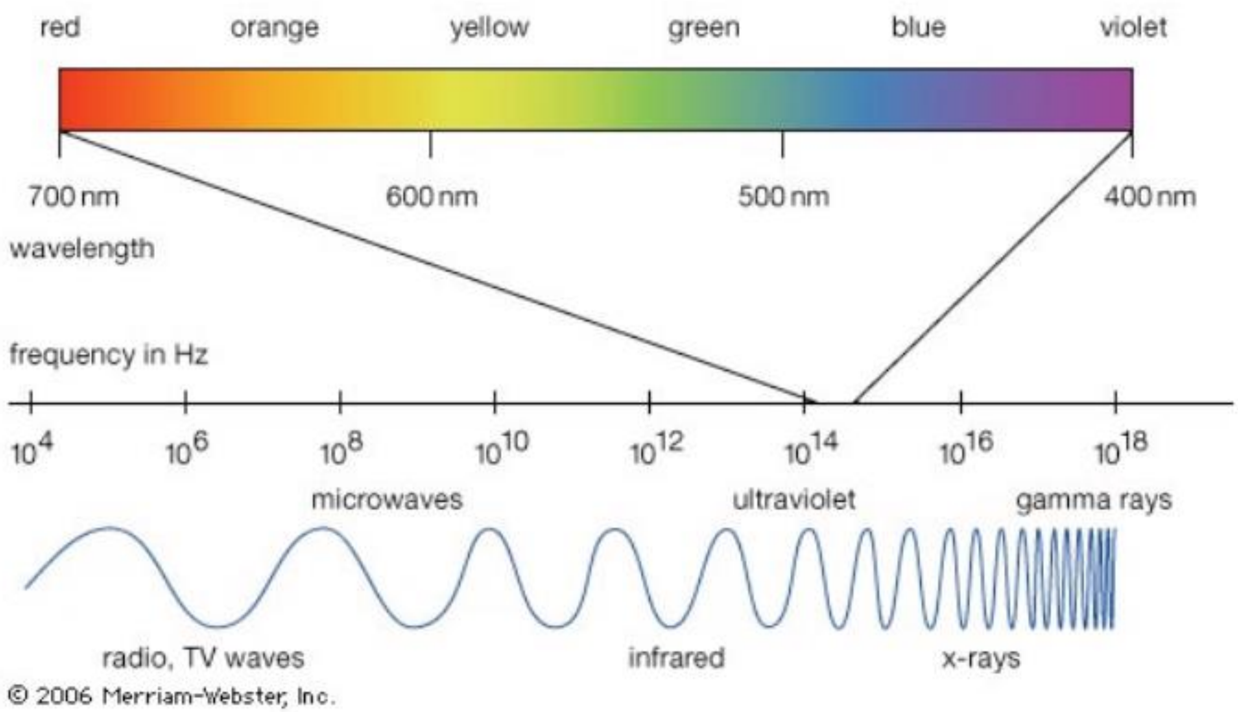


FIGURE 2.10 The electromagnetic spectrum. The visible spectrum is shown zoomed to facilitate explanation, but note that the visible spectrum is a rather narrow portion of the EM spectrum.

Light



Wavelength (λ) and frequency (ν) related by: $\lambda = \frac{c}{\nu}$

$c \approx 2.998 \times 10^8$ m/s
λ in *microns* (μm=10^{-6} m) or *nanometers* (nm=10^{-9} m)

Energy (eV): $E = h\nu$ (h: Planck's constant)

Light

- Light void of color = monochromatic (or achromatic) light
  - Only attribute: intensity or gray level
- Range of measure values = gray scale
  - Monochromatic images = gray-scale images
- Chromatic light source: frequency + radiance, luminance, brightness
  - Radiance = total amount of energy that flows from the light source (W)
  - Luminance (in lumens, lm) = measure the amount of energy an observer perceives from a light source
  - Brightness = subjective descriptor of light perception practically impossible to measure

Digital Image Acquisition

- There are numerous ways to acquire images, but our objective in all is the same: to generate digital images from sensed data. The output of most sensors is a continuous voltage phenomenon being sensed.

- To create a digital image, we need to convert the continuous sensed data into digital form. This involves two processes: *sampling* and *quantization*.

Digital Image Acquisition

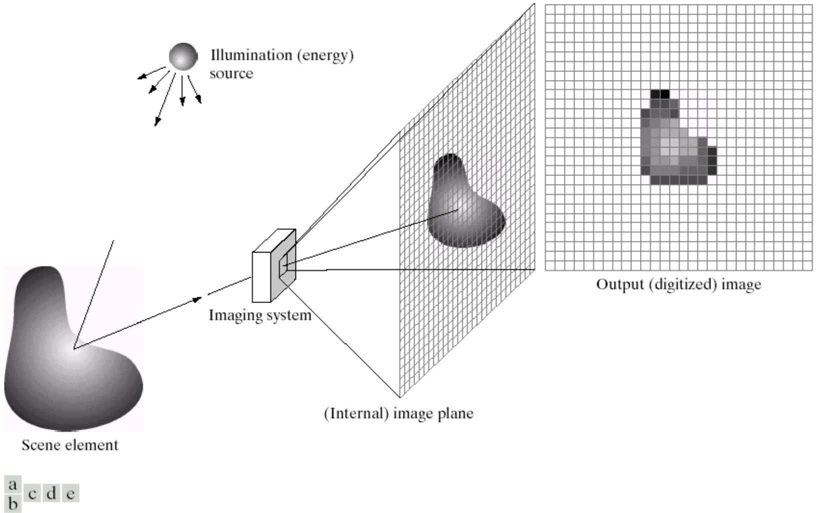
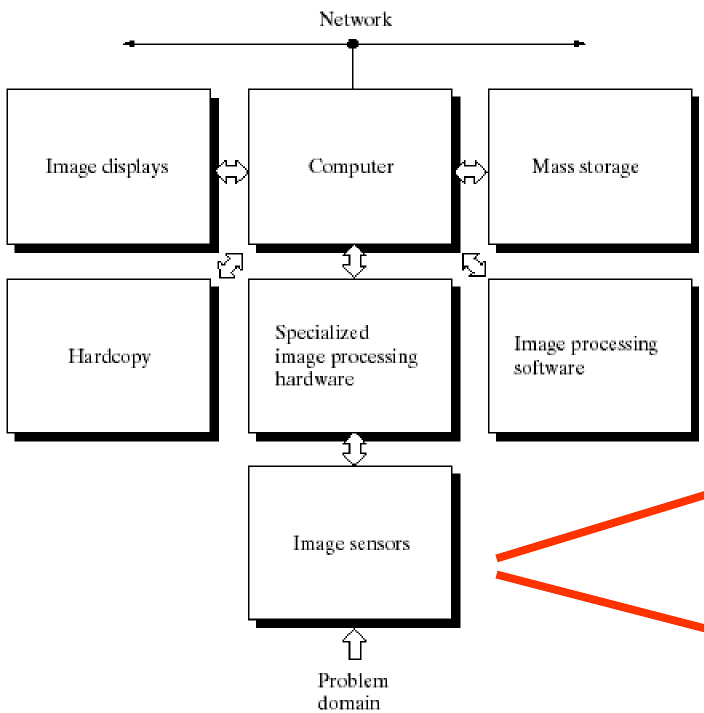


FIGURE 2.15 An example of the digital image acquisition process. (a) Energy ("illumination") source. (b) An element of a scene. (c) Imaging system. (d) Projection of the scene onto the image plane. (e) Digitized image.

- Illumination source reflected from a scene element
- Imaging system collects the incoming energy and focus it onto an image plane (sensory array)
- Response of each sensor proportional to the integral of the light energy projected
- Sensor output: analogue signal → digitized

Digital Image Acquisition

- A general-purpose image processing system

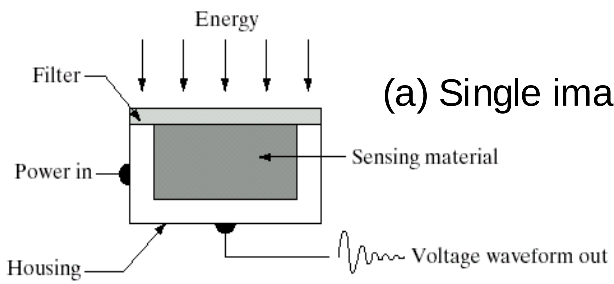


Example: In a digital video camera, the sensors produce an electrical output proportional to light intensity. The digitizer converts these outputs to digital data.

<u>A physical device</u> that is sensitive to the energy radiated by the object we wish to image

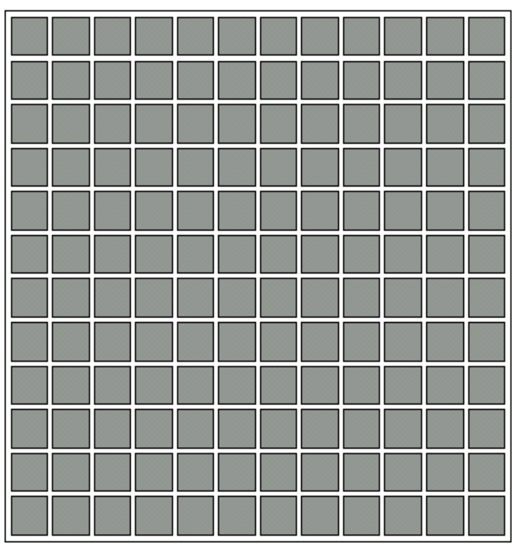
<u>A digitizer</u> is a device for converting the output of the physical sensing device into digital form.

Digital Image Acquisition



(a) Single imaging sensor

(b) Line sensor

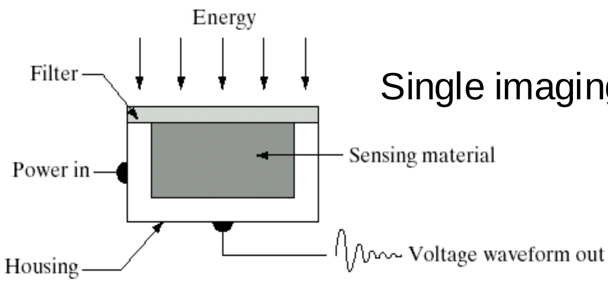


(c) Array sensor

The three principal sensor arrangements used to transform illumination energy into digital images.

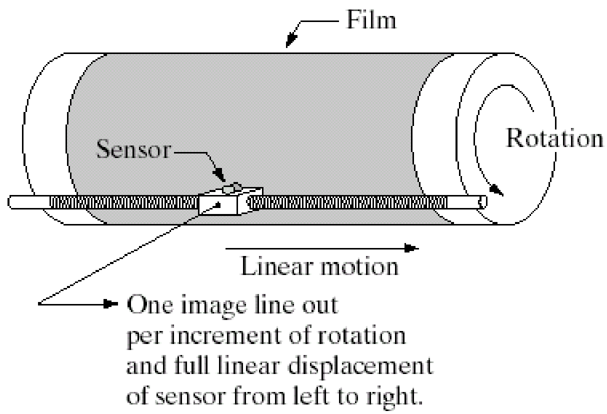
Incoming energy is transformed into a voltage by the combination of input electrical power and sensor material that is responsive to the particular type of energy being detected. The output voltage waveform is the response of the sensor(s), and a digital quantity is obtained from each sensor by digitizing its response.

Digital Image Acquisition



Single imaging sensor

In order to generate a 2D image using a single sensor, there has to be relative displacements in both the *x*- and *y*-directions between the sensor and the area to be imaged.



Combining a single sensor with motion to generate a 2D image

Example: an arrangement used in high-precision scanning, where a film negative is mounted onto a drum whose mechanical rotation provides displacement in one dimension. The single sensor is mounted on a lead screw that provides motion in the perpendicular direction. Since mechanical motion can be controlled with high precision, this method is an inexpensive (but slow) way to obtain high-resolution images.

Digital Image Acquisition

Line sensor

A geometry that is used much more frequently than single sensors consists of an in-line arrangement of sensors in the form of a sensor strip.

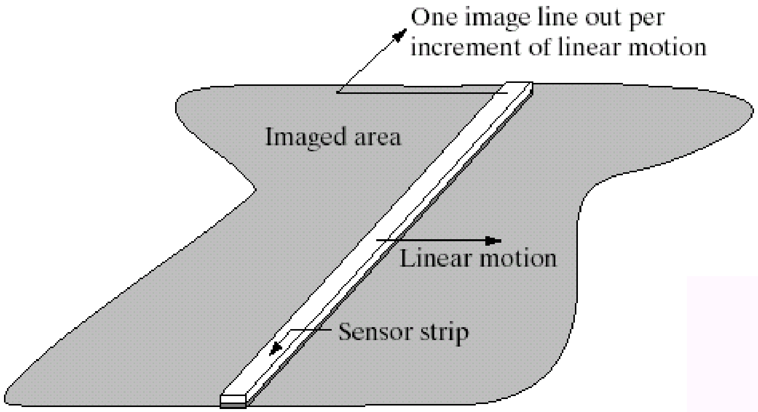


Image acquisition using a linear sensor strip

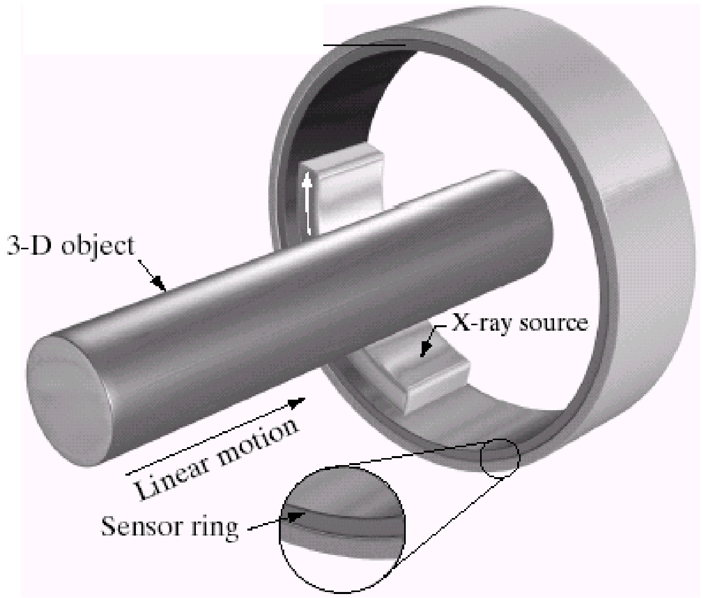
The strip provides imaging elements in one direction. Motion perpendicular to the strip provides imaging in the other direction.

The imaging strip gives one line of an image at a time, and the motion of the strip completes the other dimension of a 2D image.

This is the type of arrangement used in most flat bed scanners.

Digital Image Acquisition

Image acquisition using a circular sensor strip



Sensor strips mounted in a ring configuration are used in medical and industrial imaging to obtain cross-sectional ("slice") images of 3D objects.

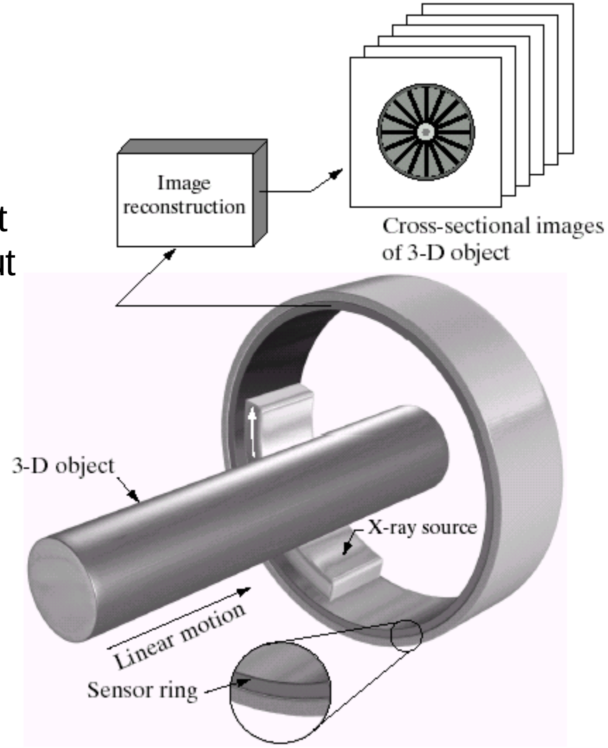
A rotating X-ray source provides illumination and the portion of the sensors opposite the source collect the X-ray energy that pass through the object (the sensors obviously have to be sensitive to X-ray energy).

This is the basis for medical and industrial computerized tomography (CT) imaging.

Digital Image Acquisition

- Image acquisition using a circular sensor strip

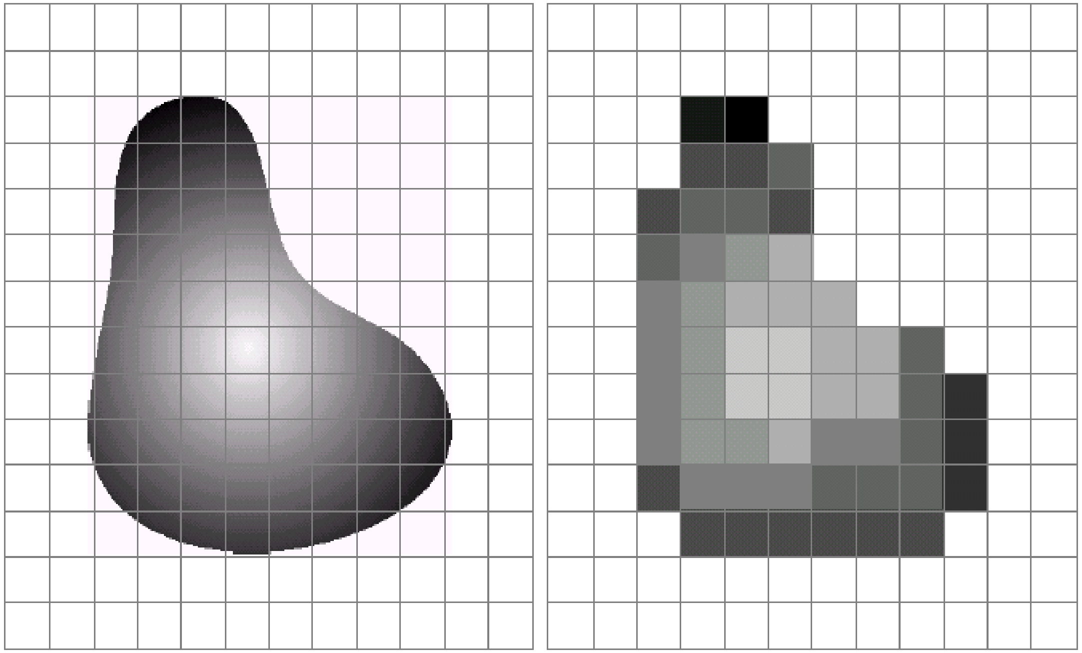
The output of the sensors must be processed by image reconstruction algorithms which can transform the sensed data into meaningful cross-sectional images, i.e., images are not obtained directly from the sensors but require extensive processing.



A 3D digital volume consisting of stacked images is generated since the object is moved in a direction perpendicular to the sensor ring.

Digital Image Acquisition

- Image sampling and quantization

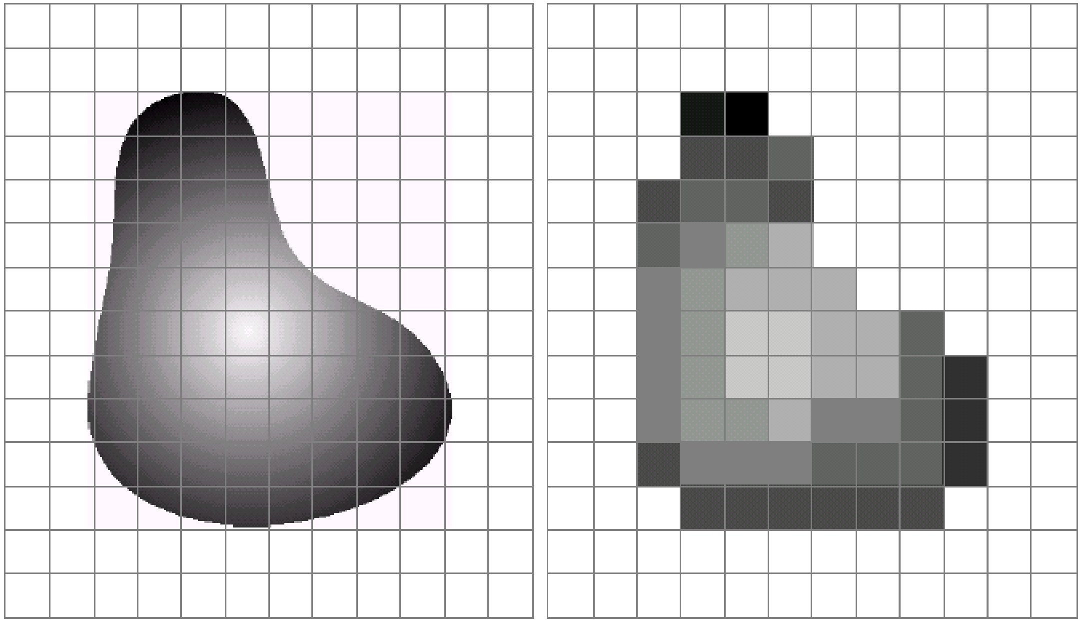


(a) Continuous image projected onto a sensor array

(b) Result of image sampling and quantization

Digital Image Representation

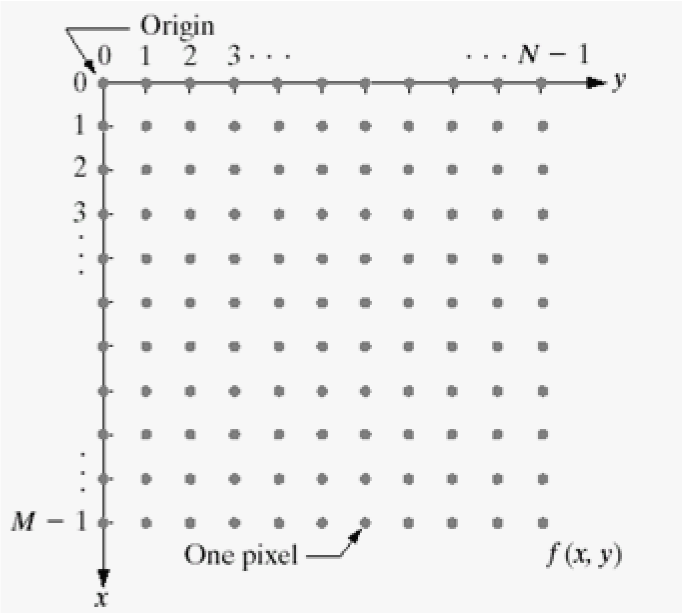
Image sampling and quantization



(a) Continuous image projected onto a sensor array

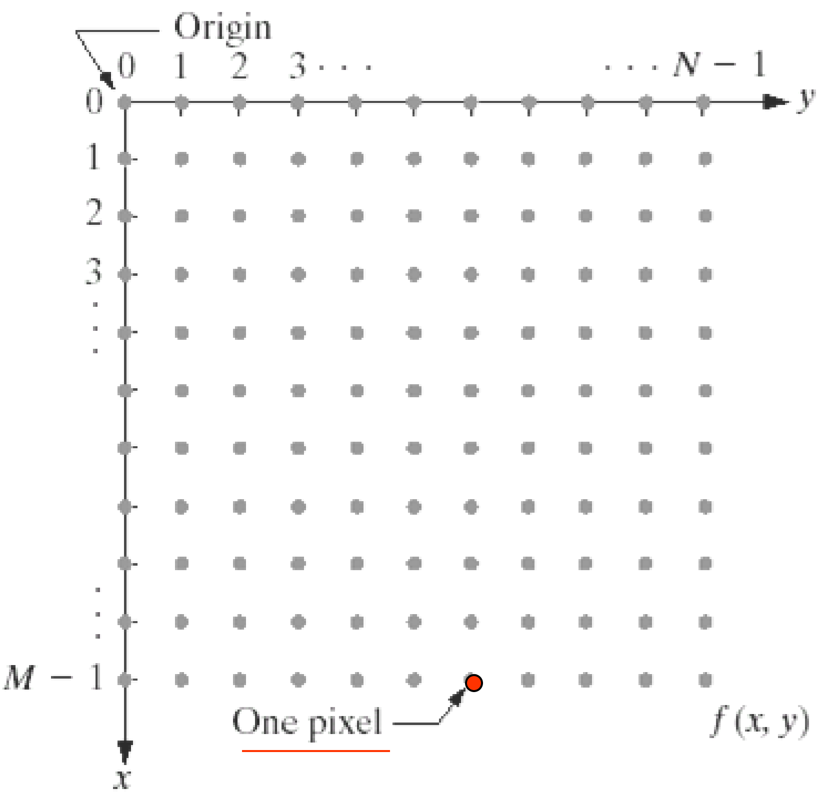
(b) Result of image sampling and quantization

Representing digital images



The result of sampling and quantization is a matrix of real numbers.

Digital Image Representation



$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \cdots & f(M-1,N-1) \end{bmatrix}$$

$$\mathbf{A}=\begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \cdots & a_{M-1,N-1} \end{bmatrix}$$

Each element of the matrix array is called an *image element*, *picture element*, *pixel*, or *pel*.

Digital Image Representation

- An image is referred to as a 2D light intensity function f(x,y) where
  - (x,y) denotes the spatial coordinate, and
  - f is a function of (x,y) and is proportional to the brightness or grey level of the image at that point
- Geometrically

Digital Image Representation

- The image digitization process requires decisions about values for *M* (rows), *N* (columns), and for the number , *L*, of discrete gray levels allowed for each pixel.

$$L = 2k$$

  - *Due to processing, storage, and sampling hardware considerations, the number of gray levels typically is an integer power of 2.*

- The number, b, of bits required to store a digitized image is

$$b = M \times N \times k$$

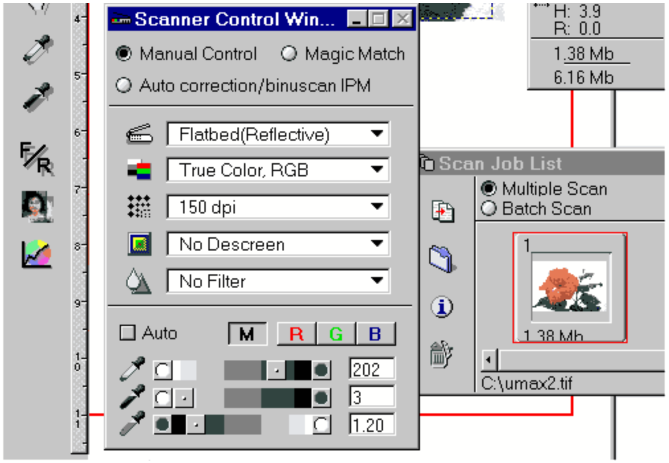
- When an image can have 2^k gray levels, it is common practice to refer to the image as a “k-bit image”.

Digital Image Representation

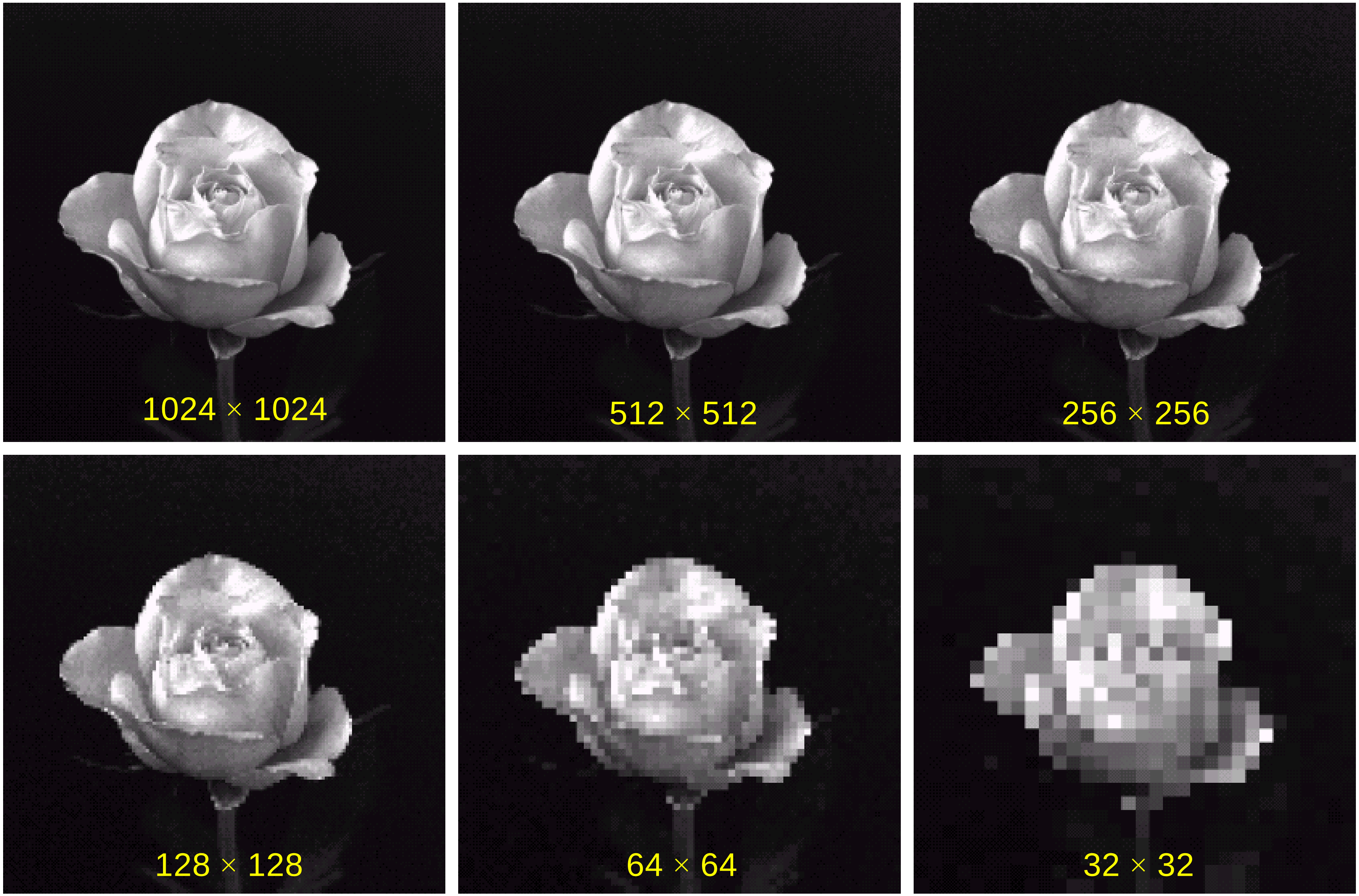
- Sampling is the principal factor determining the <u>spatial resolution </u>of an image.

- Basically, spatial resolution is the smallest discernible detail in an image.

- In terms of acquisition, spatial resolution can be obtained by multiplying the physical size of sensors and sampling resolution (e.g. dpi or ppi)
- Print size can be determined vice versa



Spatial Resolution of Image

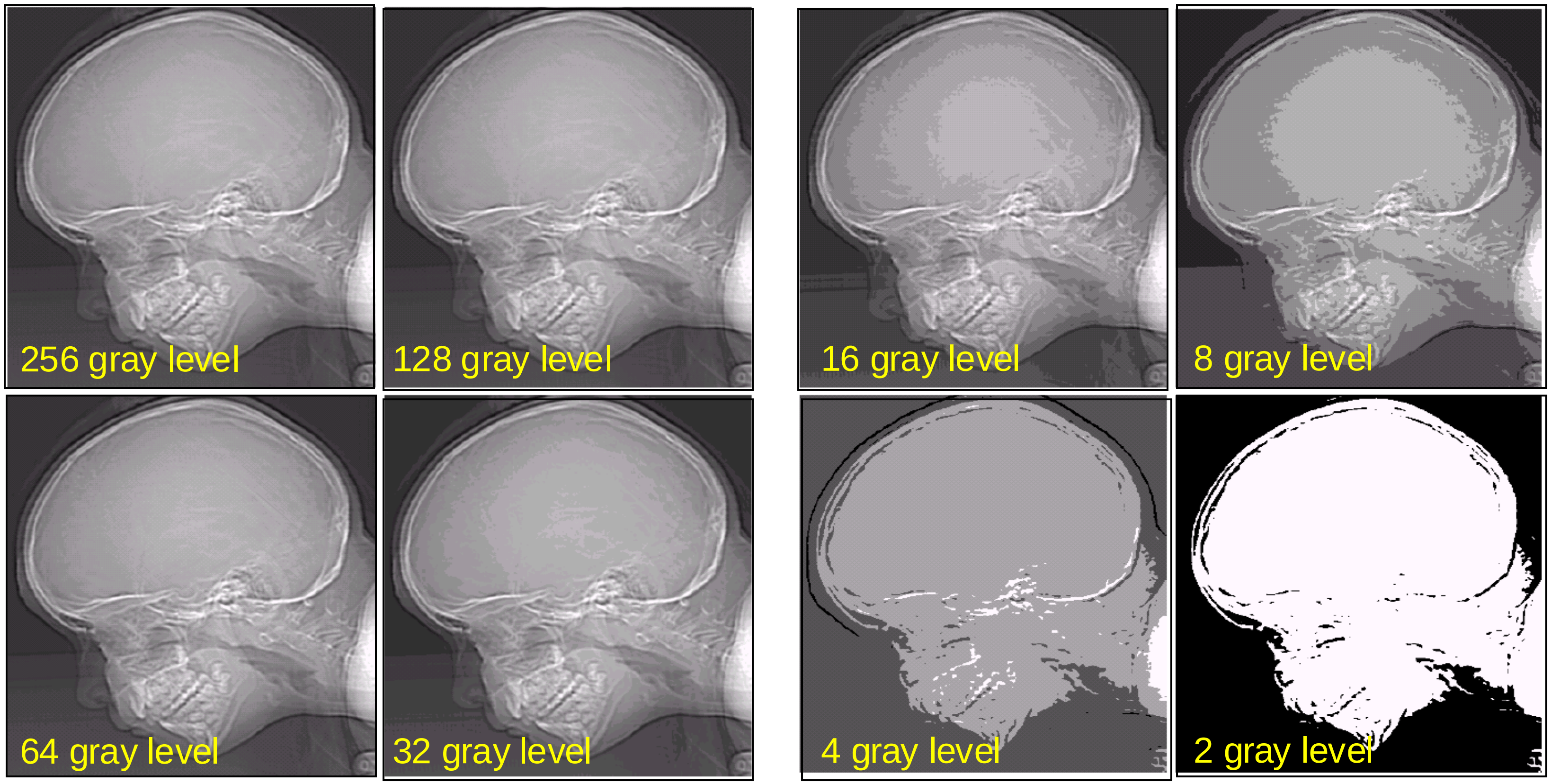


Gray-Level Resolution of Image

- Gray-level resolution refers to the smallest discernible change in gray level.

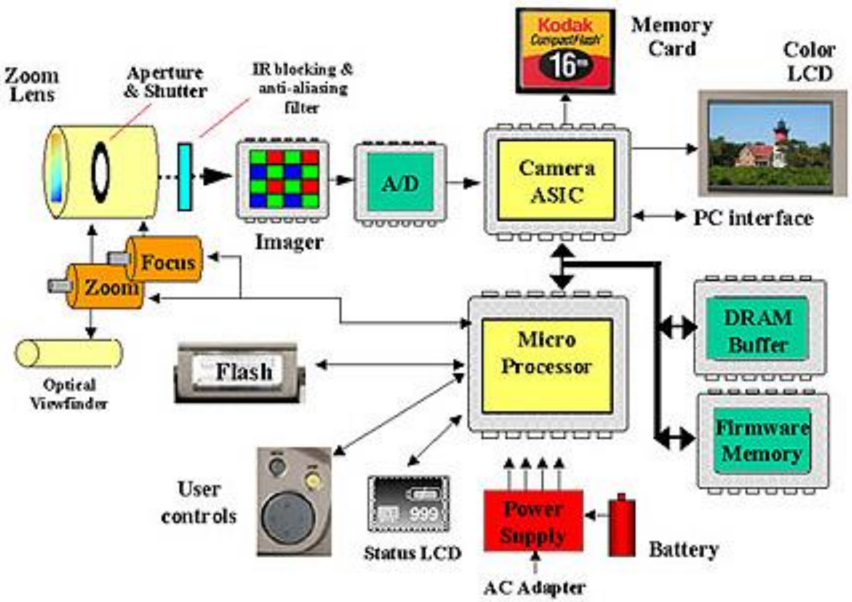
- The most common number is 8 bits, with 16 bits being used in some applications where enhancement of specific gray-level ranges is necessary.

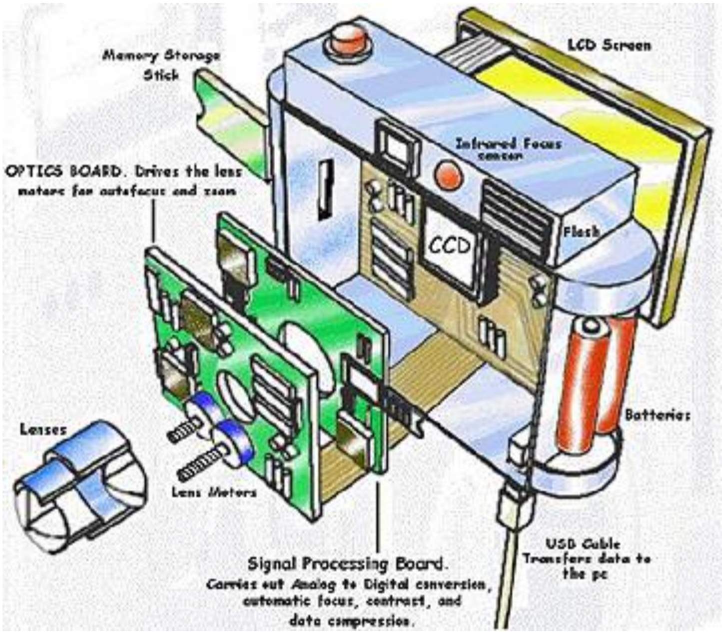
Gray-Level Resolution of Image



Color Imaging

- Digital Camera



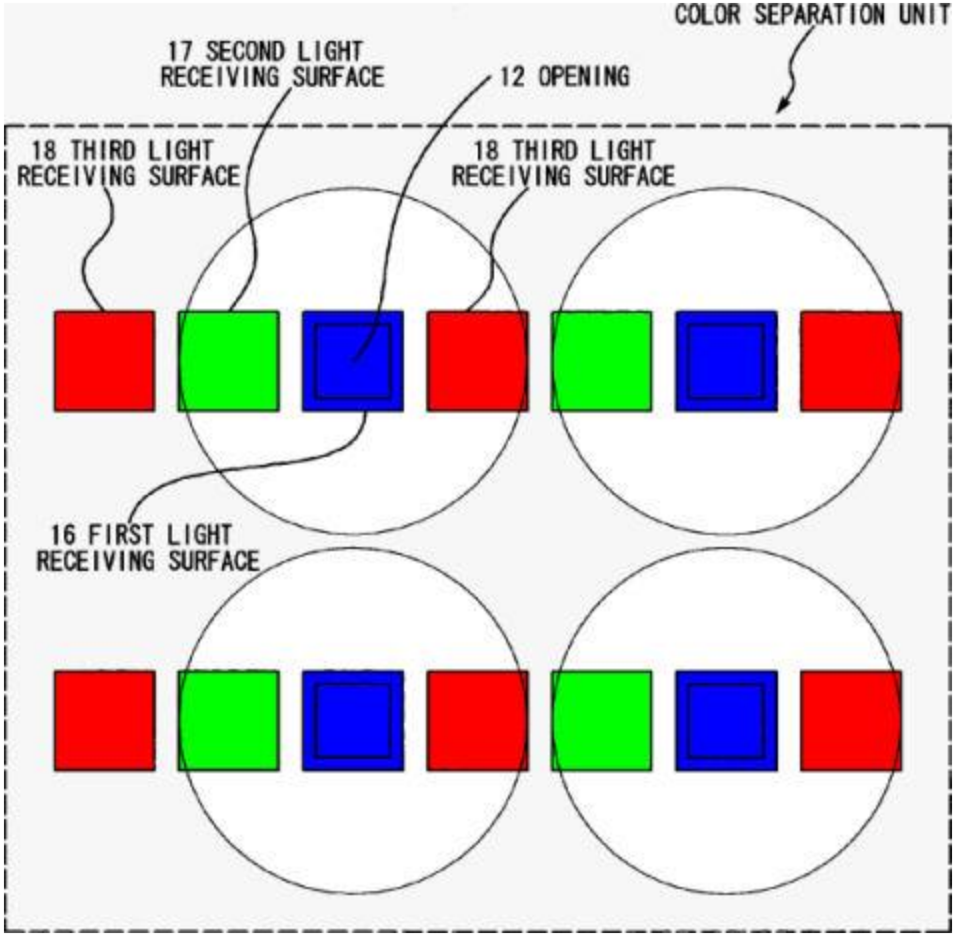


- Sensors: CCD (Charge Coupled Device) vs CMOS (Complementary Metal Oxide Semiconductor)
- http://www.dalsa.com/markets/ccd_vs_cmos.asp

Digital Camera

http://www.fas.org/irp/imint/docs/rst/Intro/Part2_5a.html

Color Imaging

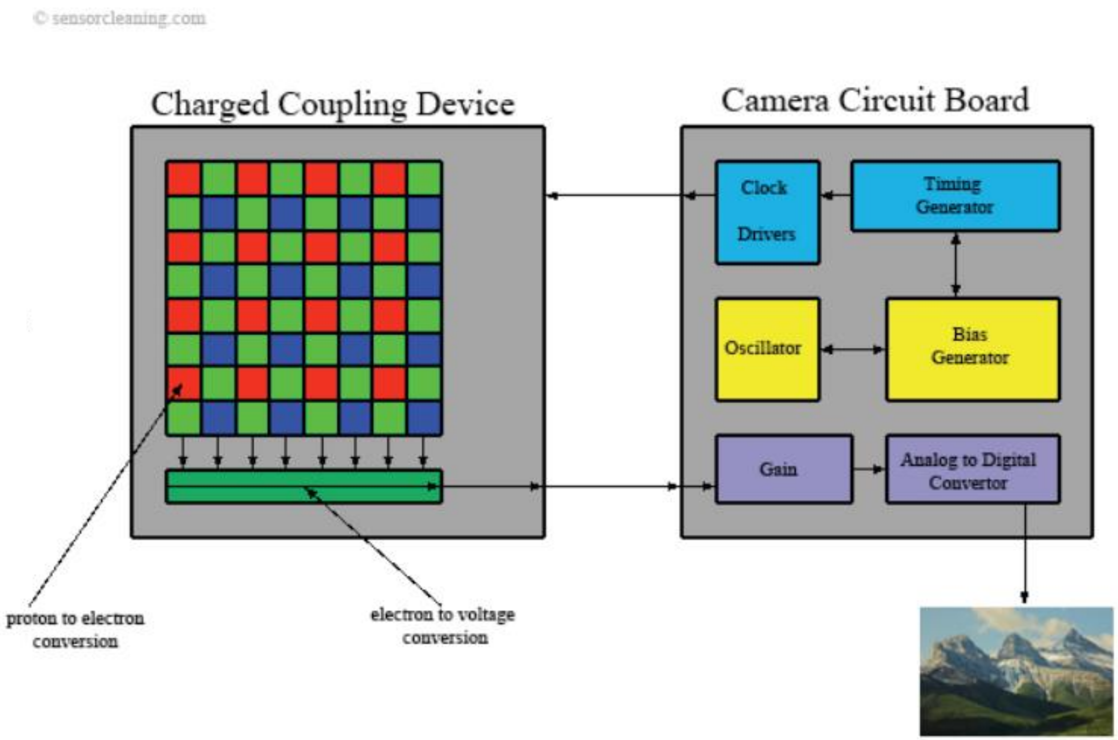


Nikon patents full-color RGB sensor

http://dslr-cameras.blogspot.com/2007/08/nikon-patents-full-color-rgb-sensor.html

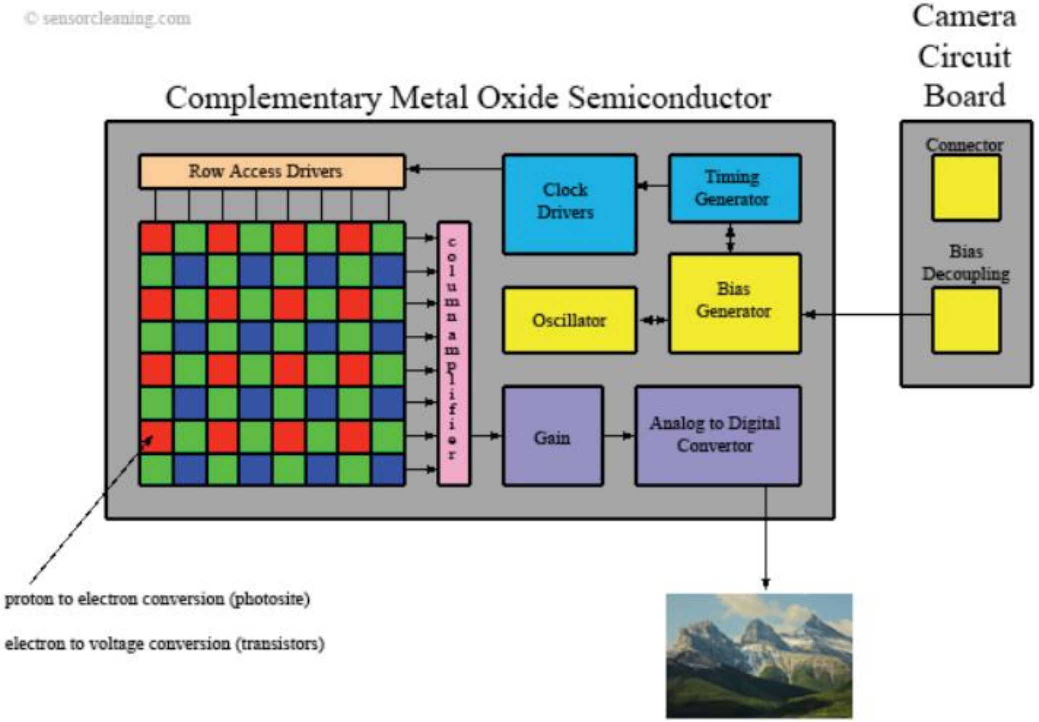
Color Imaging

- CCD Sensor
  - Rectangular grid of electro-collection sites laid over a thin silicon wafer
  - Image readout of the CCD one row at a time, each row transferred in parallel to a serial output register



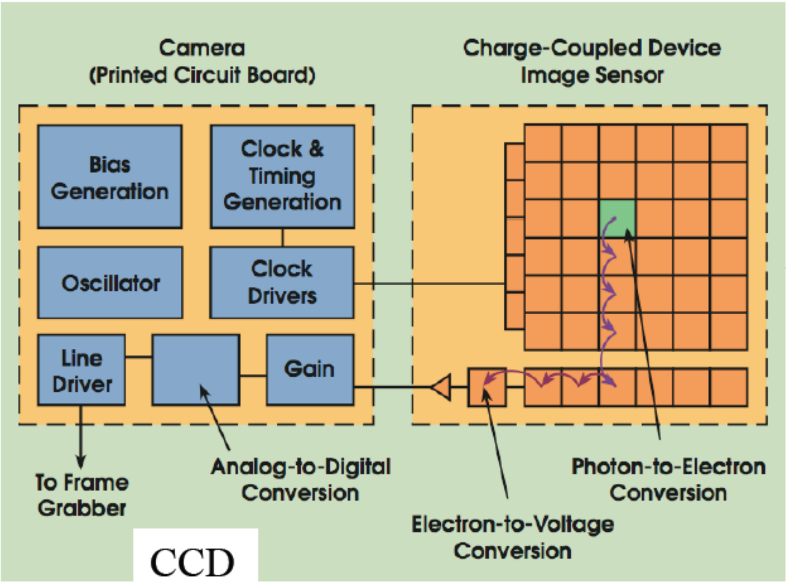
Color Imaging

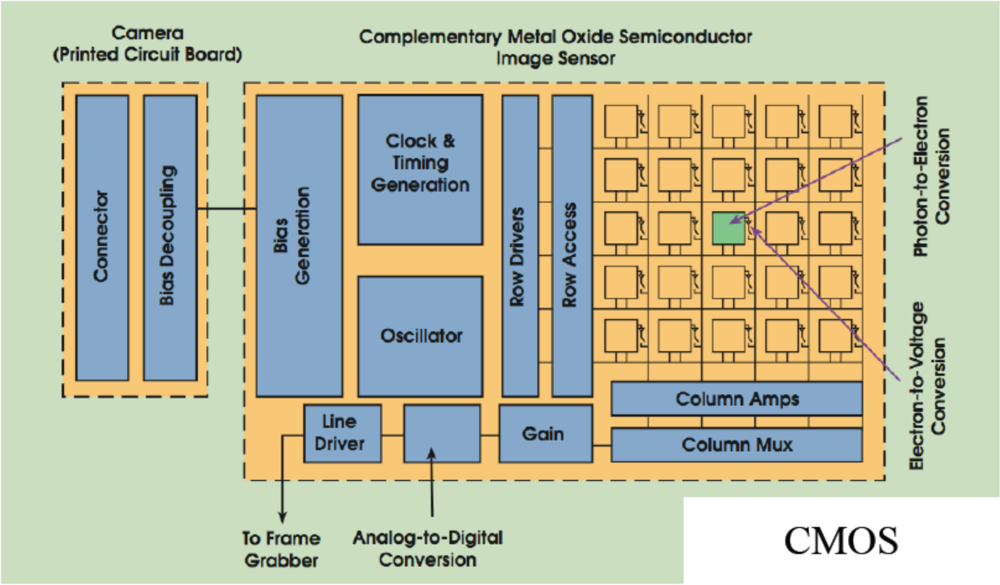
- CMOS Sensor
  - CMOS can potentially be implemented with fewer components, use less power and provide data faster than CCDs
  - Image readout of the CCD one row at a time, each row transferred in parallel to a serial output register



Color Imaging

- CCD vs CMOS





CCD: when exposure complete, transfers each pixel's charge packet sequentially to a common output structure, which converts the charge to voltage, buffer it and send it off-chip.

CMOS imager: the charge-to-voltage conversion takes place in each pixel

D. Litwiller, CCD vs CMOS: Facts and Fiction, Photonics Spectra, Jan 2001, Laurin Publishing Co. Inc.

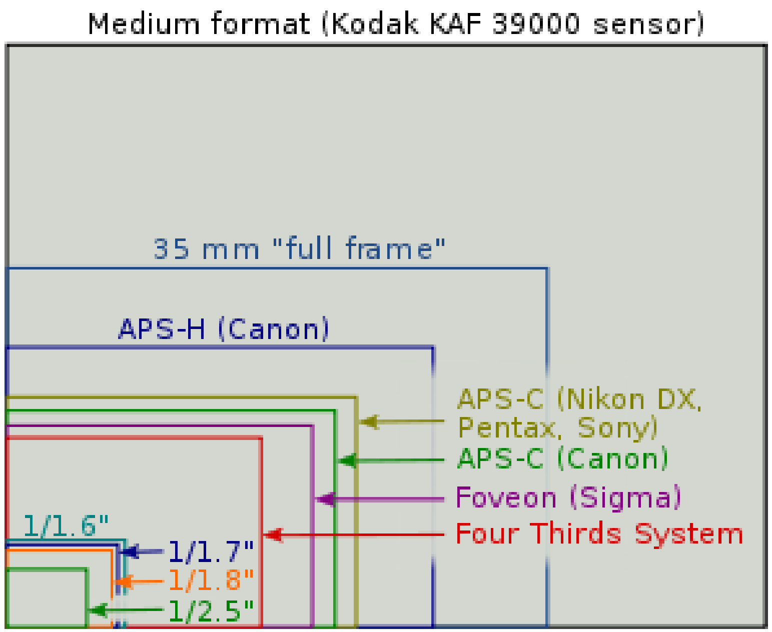
Color Imaging

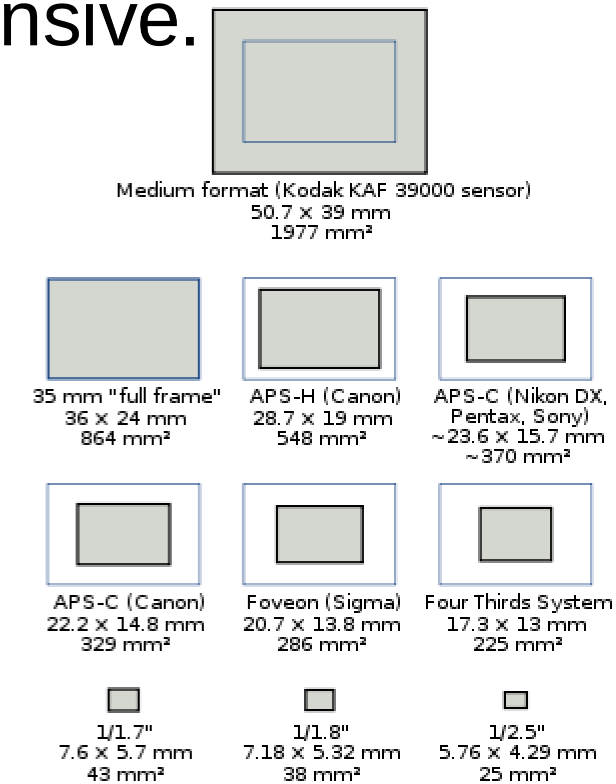
- CCD vs CMOS
  - Responsivity (amount of signal the sensor delivers per unit of input optical energy): CMOS imagers marginally superior to CCDs
  - Dynamic range (ratio of a pixel's saturation level to its signal threshold): CCDs have advantage by factor of 2 in comparable circumstances
  - Uniformity (consistency of response for different pixels under identical illumination conditions): CMOS imagers "Traditionally worse"
  - Shuttering (ability to start and stop exposure arbitrarily): standard feature of virtually all consumer and industrial CCDs
  - Speed: CMOS argubly has the advantage over CCDs (all camera functions can be placed on the image sensor)
  - Windowing: CMOS has ability to read out a portion of the image sensor
  - Anti-blooming (ability to gracefully drain localized overexposure without compromising the rest of the image in the sensor): CMOS generally has natural blooming immunity, CCDs require specific engineering
  - Reliability: CMOS have advantage (all circuit functions can be placed on a single integrated circuit chip)

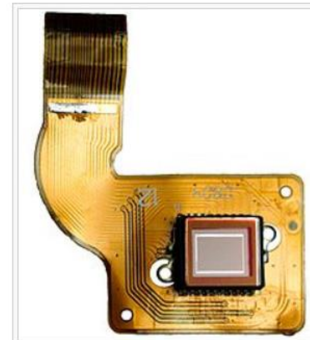
D. Litwiller, CCD vs CMOS: Facts and Fiction, Photonics Spectra, Jan 2001, Laurin Publishing Co. Inc.

Sensor Types

- In general, the bigger the sensor, the better quality, the more expensive.



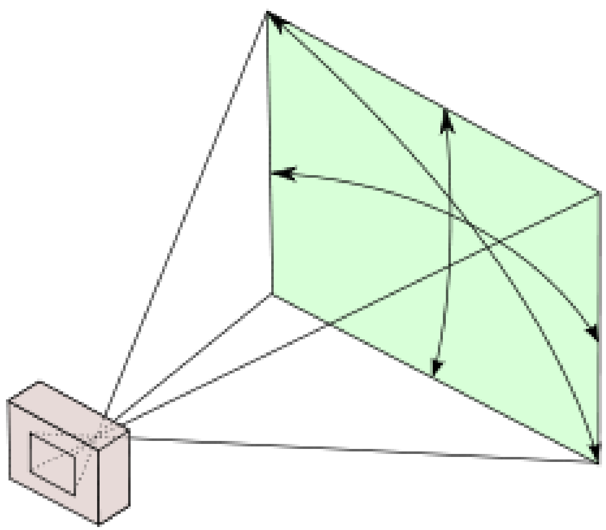


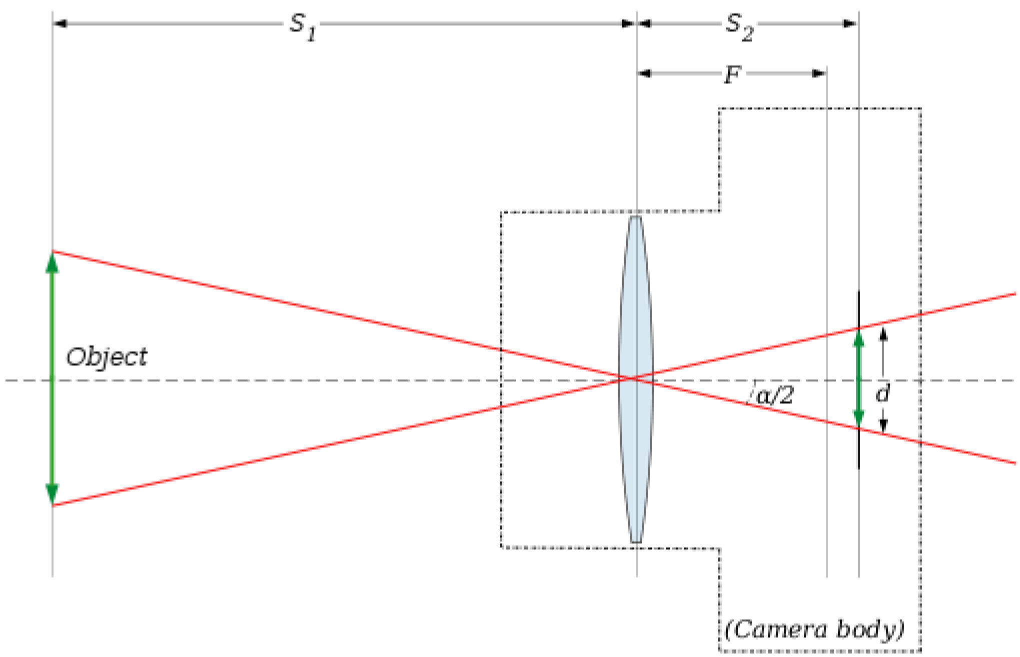

A CCD image sensor on a flexible circuit board

http://en.wikipedia.org/wiki/Image_sensor_format

Photography – Angle of View

- In photography, angle of view describes the angular extent of a given scene that is imaged by a camera. It is used interchangeably with the more general term field of view (FOV).



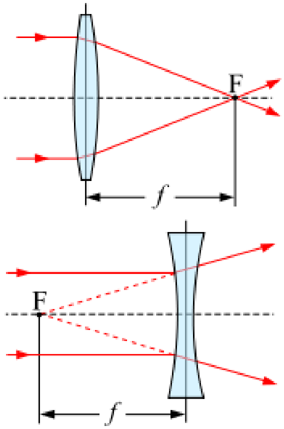


http://en.wikipedia.org/wiki/Angle_of_view

Focal Length

- The focal length of an optical system is a measure of how strongly the system converges (focuses) or diverges (defocuses) light.
  - The focal length (f), the distance from the front nodal point to the object to photograph (S_1), and the distance from the rear nodal point to the image plane (S_2)

$$\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f}.$$



28 mm lens | 50 mm lens | 70 mm lens | 210 mm lens

An example of how lens choice affects angle of view. The photos above were taken by a 35 mm camera at a constant distance from the subject.

Focal Length and Angle of View

28 mm lens, 65.5° × 46.4°

50 mm lens, 39.6° × 27.0°

70 mm lens, 28.9° × 19.5°

210 mm lens, 9.8° × 6.5°

http://en.wikipedia.org/wiki/Angle_of_view

Focal Length and Angle of View

- Common lens angles of view
  - 36 mm × 24 mm format (that is, 135 film or full-frame 35mm digital using width 36 mm, height 24 mm, and diagonal 43.3 mm).

Focal Length (mm)	13	15	18	21	24	28	35	43.3	50	70	85	105	135	180	200	300	400	500	600	800	1200
Diagonal (°)	118	111	100	91.7	84.1	75.4	63.4	53.1	46.8	34.4	28.6	23.3	18.2	13.7	12.4	8.25	6.19	4.96	4.13	3.10	2.07
Vertical (°)	85.4	77.3	67.4	59.5	53.1	46.4	37.8	31.0	27.0	19.5	16.1	13.0	10.2	7.63	6.87	4.58	3.44	2.75	2.29	1.72	1.15
Horizontal (°)	108	100.4	90.0	81.2	73.7	65.5	54.4	45.1	39.6	28.8	23.9	19.5	15.2	11.4	10.3	6.87	5.15	4.12	3.44	2.58	1.72

http://en.wikipedia.org/wiki/Angle_of_view

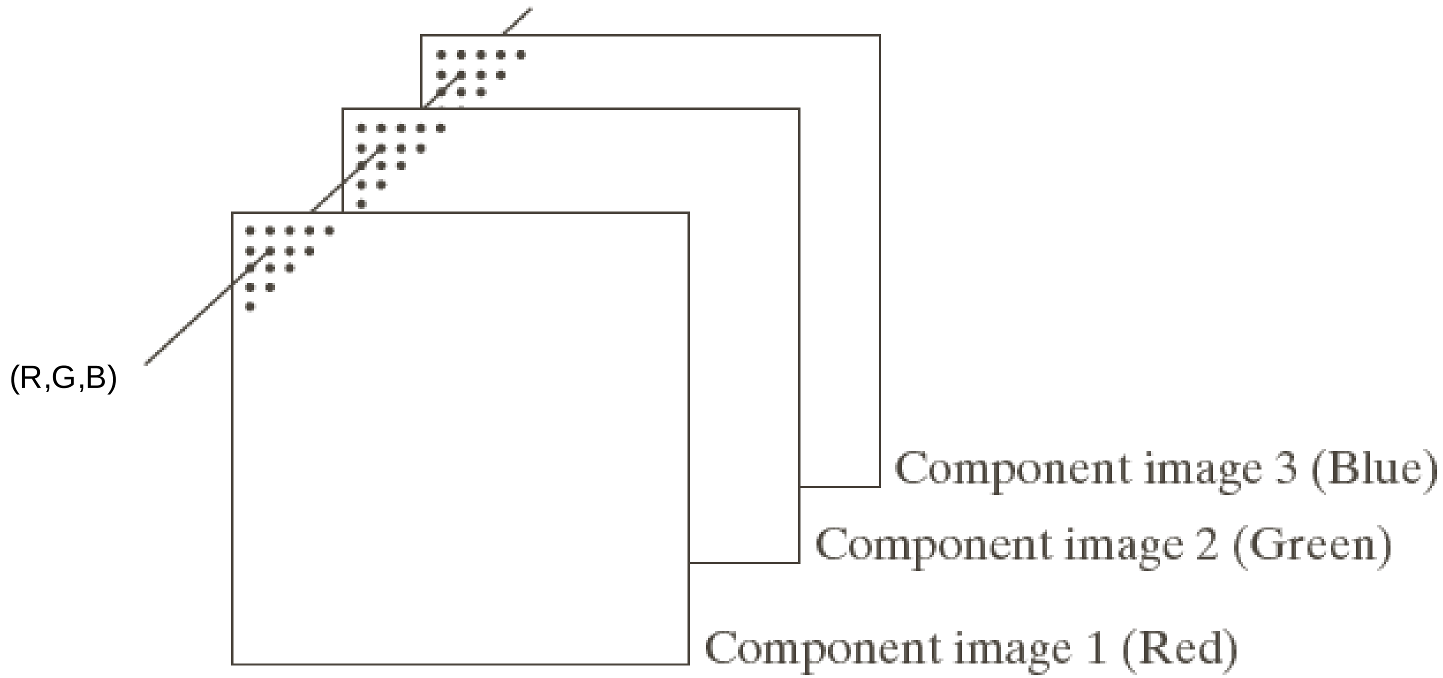
Focal Length and Angle of View

Five images using 24, 28, 35, 50 and 72mm equivalent zoom lengths, portrait format, to illustrate angles of view [17]

Five images using 24, 28, 35, 50 and 72mm equivalent step zoom function, to illustrate angles of view

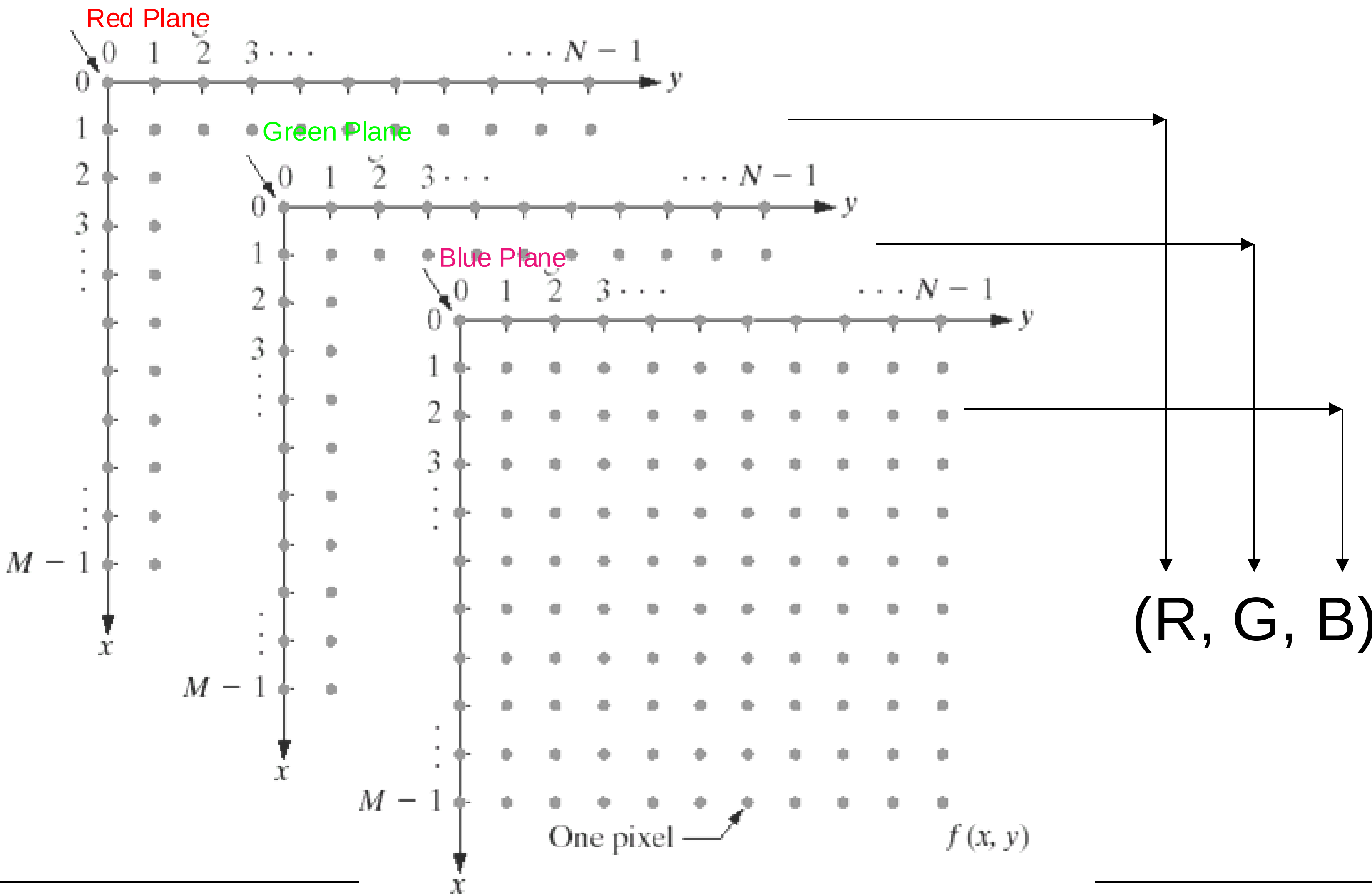
http://en.wikipedia.org/wiki/Angle_of_view

Color Image Representation



Formation of a vector from corresponding pixel values in three RGB component images.

Color Image Representation

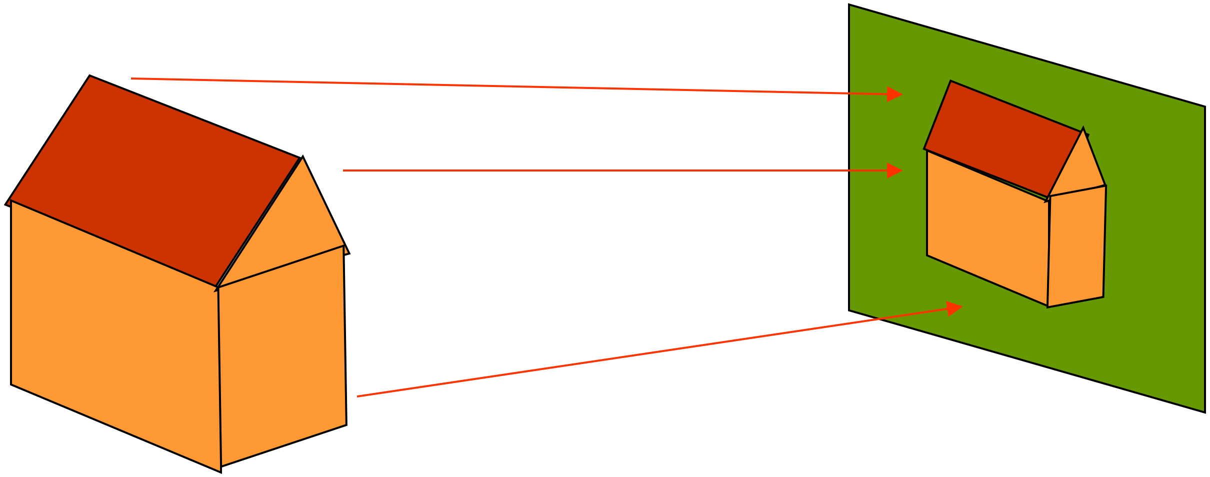


Video

- There are two different ways of generating moving pictures in a digital form for inclusion in a multimedia production.
- Video – we can use a video camera to capture a sequence of frames recording actual motion as it is occurring in the real world;
- Animation – we can create each frame individually, either within the computer or by capturing single images one at a time.

The Video Image

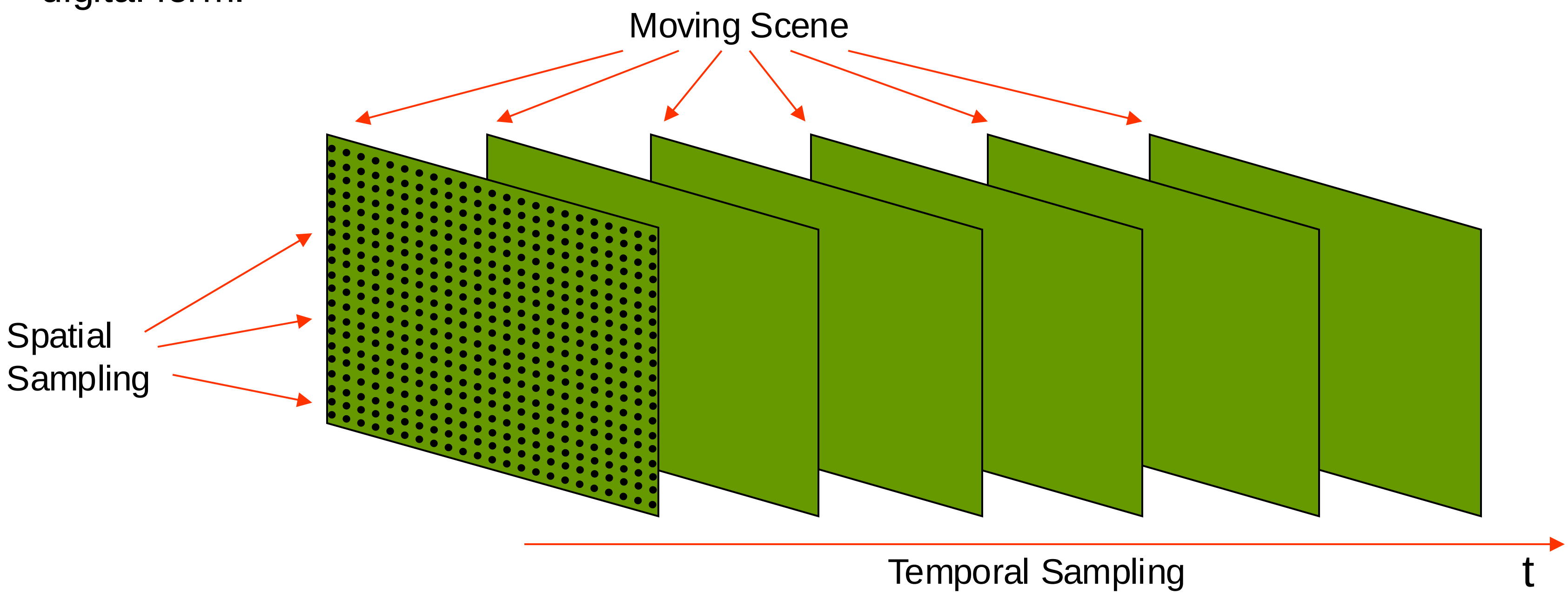
- A video image is a projection of a 3D scene onto a 2D plane.
  - A 3D scene consisting of a number of objects each with depth, texture and illumination is projected onto a plane to form a 2D representation of the scene.
  - The 2D representation contains varying texture and illumination but no depth information.
- A still image is a "snapshot" of the 2D representation at a particular instant in time whereas a video sequence represents the scene over a period of time.



Digital Video Sequence

- A “real” visual scene is continuous both spatially and temporally.
  - In order to represent and process a visual scene digitally, it is necessary to sample the real scene spatially (typically on a rectangular grid in the video image plane) and temporally (typically as a series of “still” images or frames sampled at regular intervals in time).

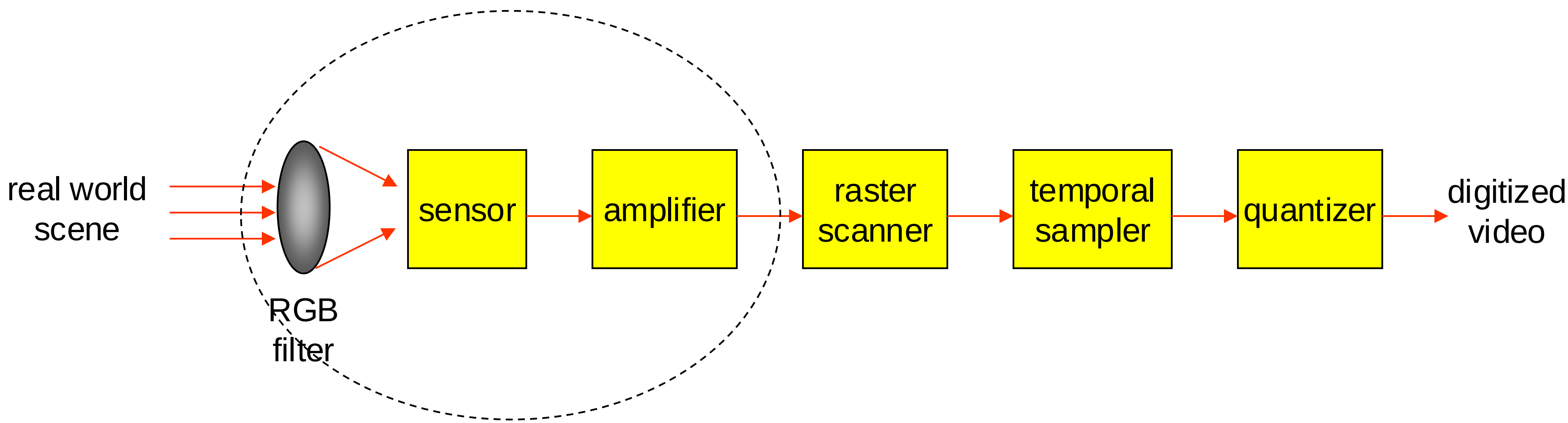
- Digital video is the representation of a spatio-temporally sampled video scene in digital form.



Digitization

- The typical processing steps involved in the digitization of video.

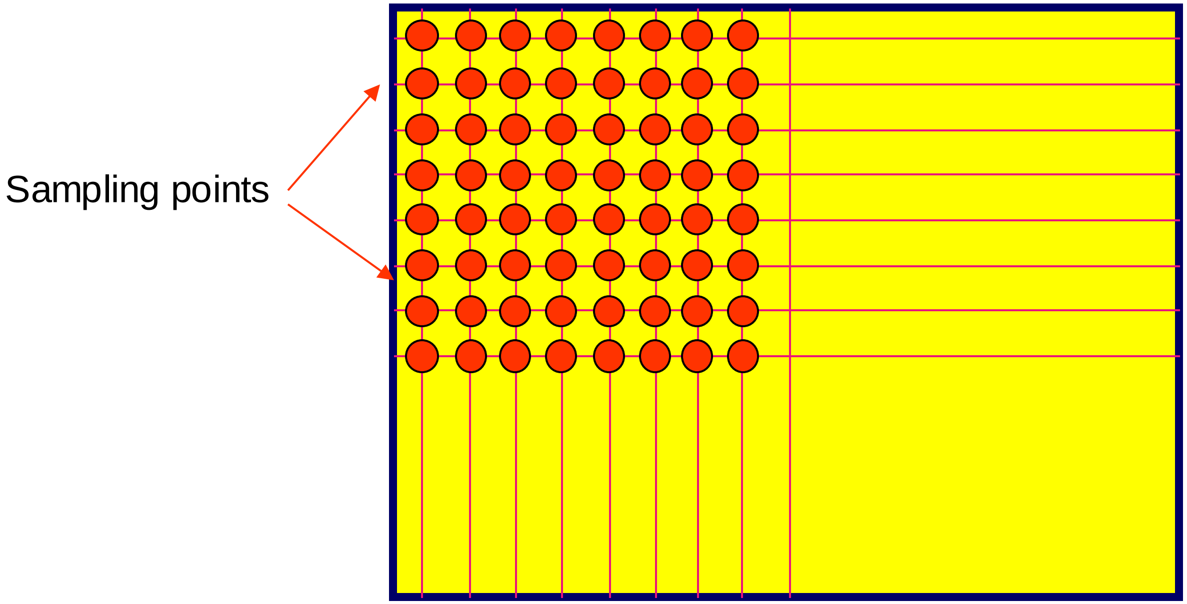
  After signal acquisition and amplification, the key processing steps are spatial sampling, temporal sampling, and quantization.



Spatial Sampling

- Spatial sampling consists of taking measurements of the underlying analog signal at a finite set of sampling points in a finite viewing area (or frame).

- To simplify the process, the sampling points are restricted to lie on a lattice, usually a rectangular grid.



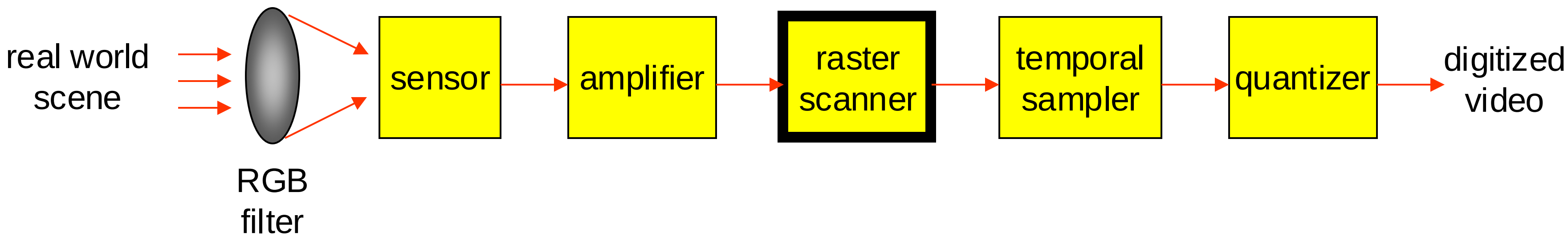
Video Image Resolutions

- The visual quality of the video image is influenced by the number of sampling points. More sampling points (a higher sampling resolution) give a "finer" representation of the image: however, more sampling points require higher storage capacity.

Typical video image resolutions

Video image resolution	Number of sampling points	Analogue video "equivalent"
352 x 288	101,376	VHS video
704 x 576	405,504	Broadcast television
1920 x 1152	2,211,840	High-definition television

Raster Scanning

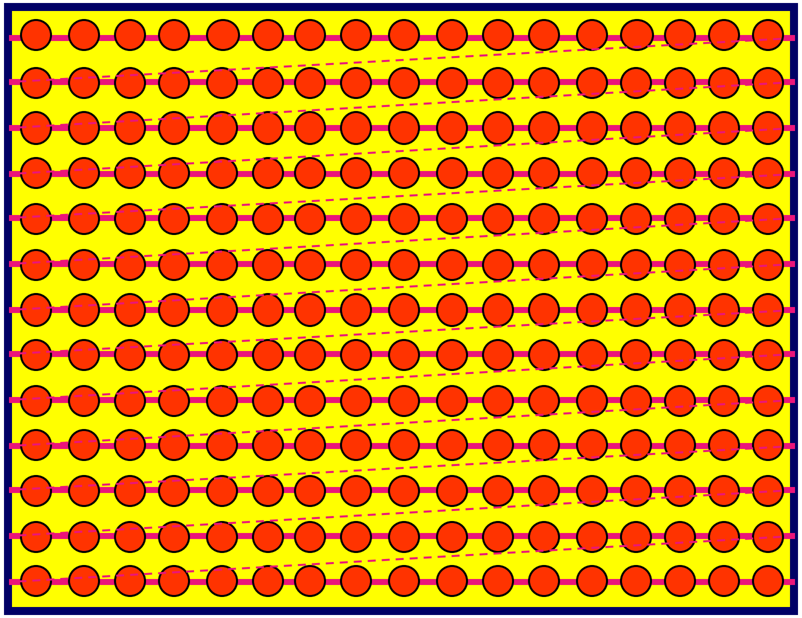


- The two-dimensional set of sampling points are transformed into a one-dimensional set through a process called raster scanning.
- The two main ways to perform raster scanning are progressive scanning and interlaced scanning.

Progressive Scanning

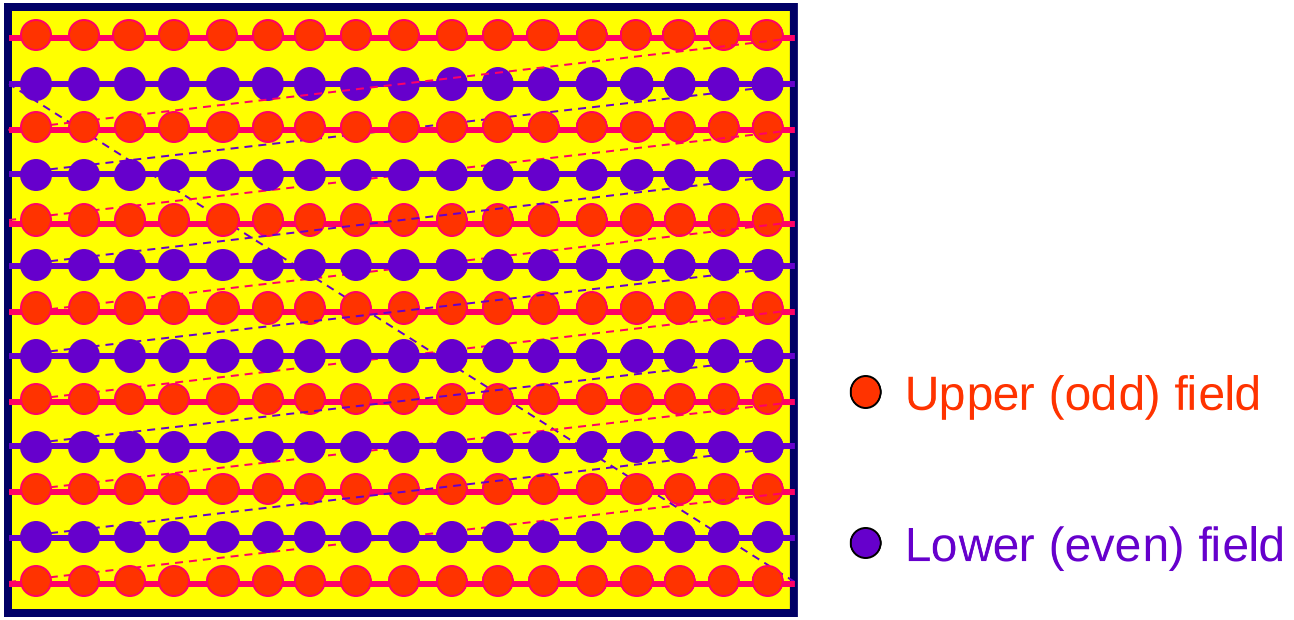
- In a progressive (or non-interlaced) scan, the sampling points are scanned from left to right and top to bottom.

- Progressive scanning is typically used for film and computer displays.



Interlaced Scanning

- In an interlaced scan, the points are divided into odd and even scan lines. The odd lines are scanned first from left to right and top to bottom. Then the even lines are scanned.
- The odd (respectively, even) scan lines make up a field. In an interlaced scan, two fields make up a video frame.
- Interlaced scanning is commonly used for television signals.



De-interlace Processing

- Interlaced video sometimes may produce unpleasant visual artifacts when displaying certain textures or types of motion.
- If you are capturing images from a video signal, you can filter them through a de-interlacing filter provided by image-editing applications

Original video image with interlaced scanning

Improved video image with de-interlacing processing

Temporal Sampling

- A moving video image is formed by sampling the video signal temporally, taking a rectangular “snapshot” of the signal at periodic time intervals.
- The human visual system is relatively slow in responding to temporal changes. By showing at least 16 frames of video per second, an illusion of motion is created. This observation is the basis for motion picture technology, which typically performs temporal sampling at a rate of 24 frames / sec.

Video frame rate	Appearance
Below 10 frames/sec	“Jerky”, unnatural appearance to movement
10 ~ 20 frames/sec	Slow movements appear OK; rapid movement is clearly “jerky”
20 ~ 30 frames/sec	Movement is reasonably smooth
50 ~ 60 frames/sec	Movement is very smooth

DEMO – "Bad Day"

Video frame rate – 30 fps

Video frame rate – 10 fps

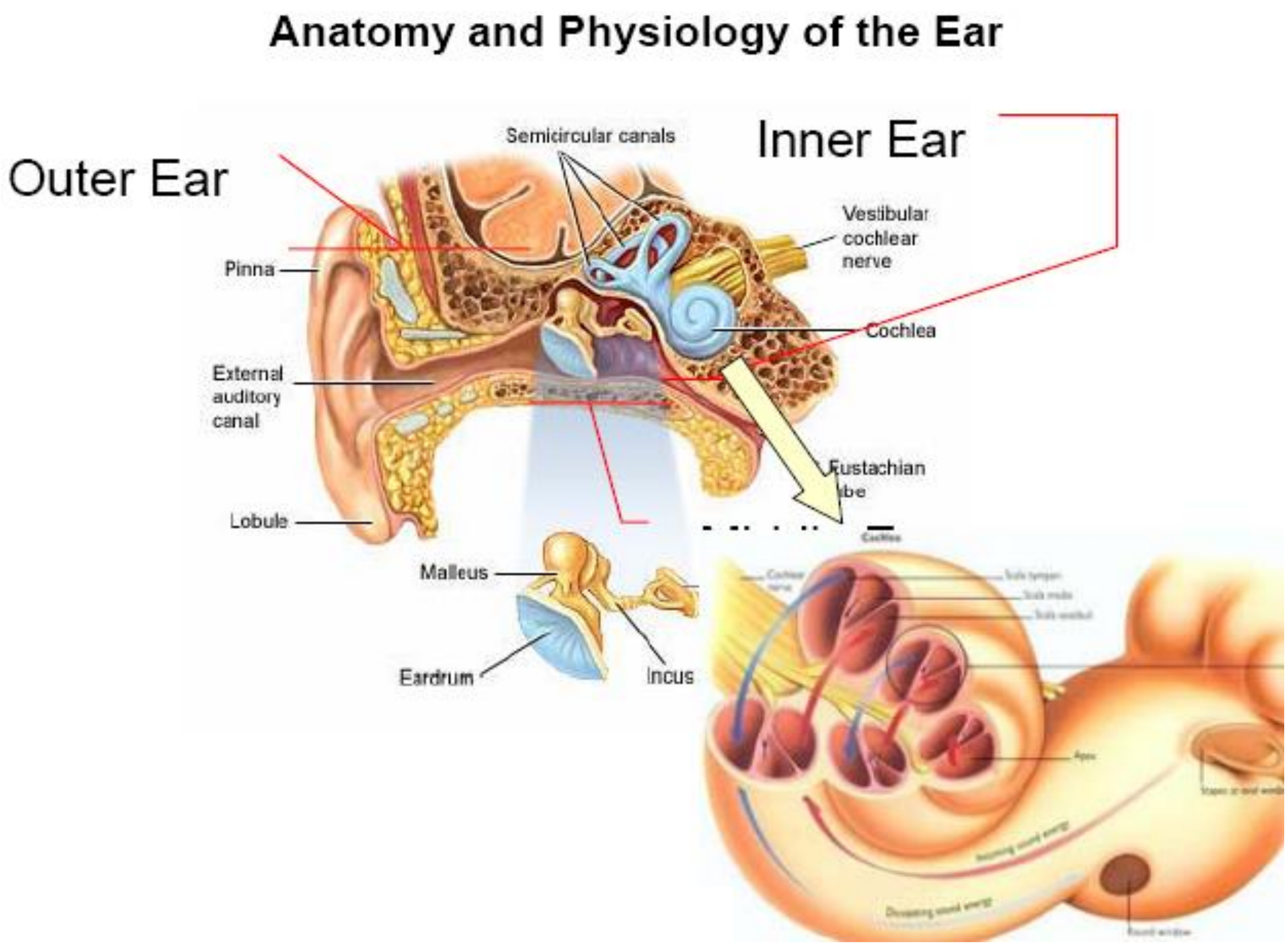
Computer Graphics & Animation

- Software (not an exhaustive list)
  - GIMP (Open Source)
  - Illustrator
  - Paint Shop Pro
  - Photoshop

  - 3DS MAX
  - Blender 3d (Open Source)
  - Bryce 3d
  - Cinema 4D
  - Lightwave 3D
  - Maya

  - AVID
  - Final Cut Studio
  - Premiere

Sound


Anatomy and Physiology of the Ear
Outer Ear
Inner Ear
Semicircular canals
Vestibular cochlear nerve
Pinna
Cochlea
External auditory canal
Eustachian
Lobule
Malleus
Eardrum
Incus

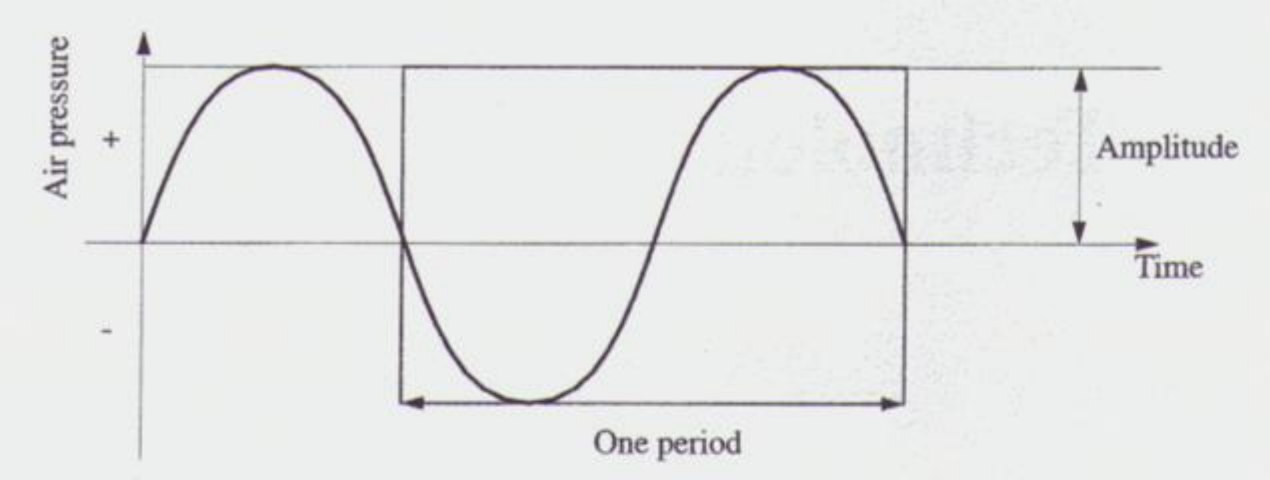
Sound

- Sound is different in kind from any of the other digital media types. All other media are primarily visual, being perceived through our eyes, while sound is perceived through the different sense of hearing.

- Although sound is, for most of us, a familiar everyday phenomenon, it is a complex mixture of physical and psychological factors, which is difficult to model accurately.

Sound

- Sound is a physical phenomenon caused by vibration of material, such as a violin string or a wood log.
- This type of vibration triggers pressure wave fluctuations in the air around the material. The pressure waves propagate in the air. The pattern of this oscillation is called wave form. We hear a sound when such a wave reaches our ears.
- We can display the wave form of any sound by plotting its amplitude against time.



The Amplitude of Sound

- A sound has a property called amplitude, which humans perceive subjectively as loudness or volume.
- The amplitude of a sound is a measuring unit used to deviate the pressure wave from its mean value (idle state).
- The size of a sound pressure level (SPL) is measured in decibels (dB).
- An audibility threshold value of 20 microPascal is the limit value above which a sound can just about be perceived. This value functions as a basis of the sound pressure, measured in decibels. It means that a decibel is the smallest variation in amplitude that can be detected by the human ear.

The Amplitude of Sound

Sound example	Sound pressure size
Rustling of paper	20 dB
Spoken language	60 dB
Heavy road traffic	80 dB
Rock band	120 dB
Pain sensitivity threshold	130 dB

The dynamic range of the human ear's sound recognition is in the range of up to 130 dB.

The Frequency of Sound

- A sound's frequency is the reciprocal value of its period.
- Similarly, the frequency represents the number of periods per second and is measured in hertz (Hz) or cycles per second (cps).
- Sound processes that occur in liquids, gases, and solids are classified by frequency range:
  - Infrasonic – 0 to 20 Hz
  - Audiosonic – 20 Hz to 20kHz
  - Ultrasonic – 20 kHz to 1 GHz
  - Hypersonic – 1 GHz to 10 THz
- The human ear is generally considered to be able to detect frequencies in the range between 20 Hz and 20 kHz, although individuals frequency responses vary greatly.
  - In particular, the upper limit decreases fairly rapidly with increasing age: few adults can hear sounds as high as 20 kHz, although children can.

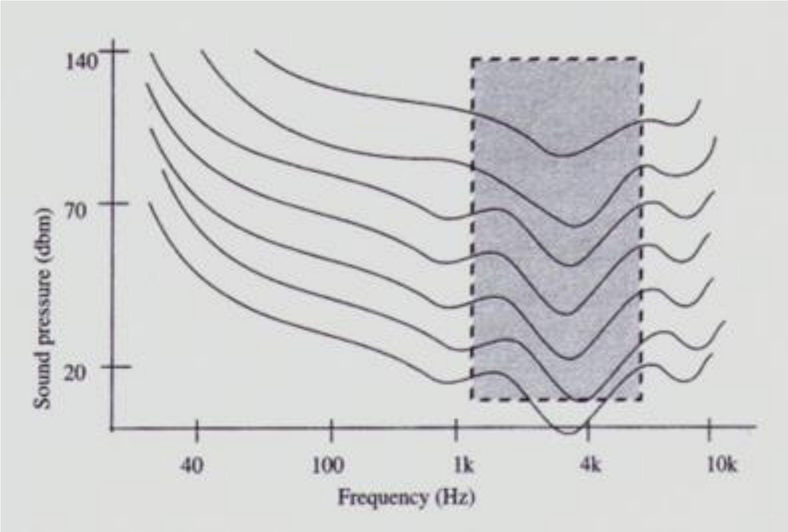
The Frequency of Sound

Test Signals from 25Hz to 20kHz

Low Freq.	25Hz	31Hz	40Hz	56Hz	63Hz	100Hz	125Hz	160Hz	200Hz	250Hz	315Hz	400Hz	500Hz	630Hz
Mean Freq.	800Hz	1kHz	1.25kHz	1.6kHz	2kHz	2.5kHz	3.15kHz	4kHz	5kHz	6.3kHz				
High Freq.	8kHz	10kHz	12.5kHz	16kHz	20kHz									

- The 29 band signals from 25Hz to 20kHz. Each band lasts 10 seconds.
- They can be used to test the equipment's frequency response and room acoustics.
  - When the extremely low frequencies like 25, 30 and 40Hz are played on small loud-speakers, you probably can only "See the woofer shaking, but not hear the sound".
  - Similarly, the extremely high frequencies like 16 and 20 kHz will also vanish in some loud-speakers.

The Frequency of Sound



- The perception sensitivity we call loudness is not linear across all frequencies.
- Human sound perception is most sensitive in the mean frequency ranges between 700Hz and approximately 6.6kHz.
- The human hearing system responds much better to the mean frequency range than it does to low and very high frequencies.
- The above phenomena offer the potential for audio compression

Spectrum Analysis

- The spectrum analysis allows you to examine the fundamental frequency and overtones present in a recording.

- Unlike the wave form, which represents audio in the time domain (amplitude vs. time), spectrum analysis allows you to examine sound by representing the sound in the frequency domain (amplitude vs. frequency).

- Data displayed in the frequency domain shows the amplitudes and frequencies of sine waves that, if mixed together, would sound very much like the original sound. Since it's relatively easy to remember how a sine wave sounds at different frequencies, it's possible to visualize how simple wave forms sound by looking at the spectrum of the sound.

Spectrum Analysis

- Fast Fourier Transform (FFT) is usually used to perform spectrum analysis.
- A Fourier Transform is the mathematical method used to convert a waveform from the Time Domain to the Frequency Domain.
- Since the Fourier Transform is computationally intensive, it is common to use a technique called a Fast Fourier Transform (FFT) to perform spectral analysis. The FFT uses mathematical shortcuts to lower the processing time at the expense of putting limitations on the analysis size.
- The analysis size, also referred to as the FFT size, indicates the number of samples from the sound signal used in the analysis and also determines the number of discrete frequency bands. When a high number of frequency bands are used, the bands have a smaller bandwidth, which allows for more accurate frequency readings.

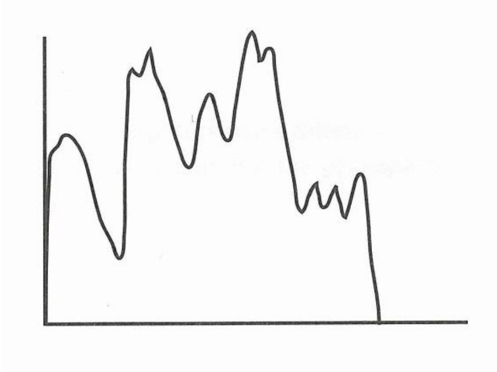
Digitizing Sound

- Digitization consists of two steps:

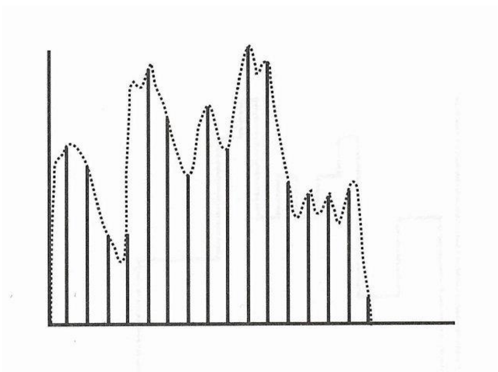
  Sampling, when we measure the signal's value at discrete intervals;

  Quantization, when we restrict the value to a fixed set of levels.

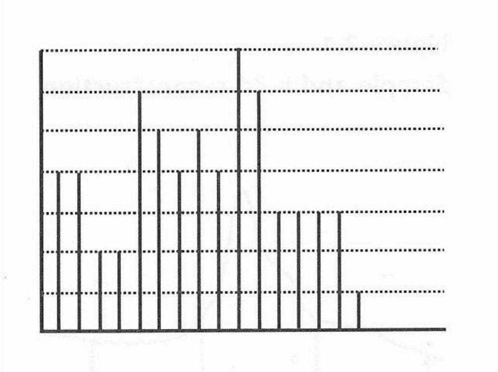
- Sampling and quantization can be carried out in either order, by special hardware devices – analogue to digital converters (ADCs).



An analogue signal

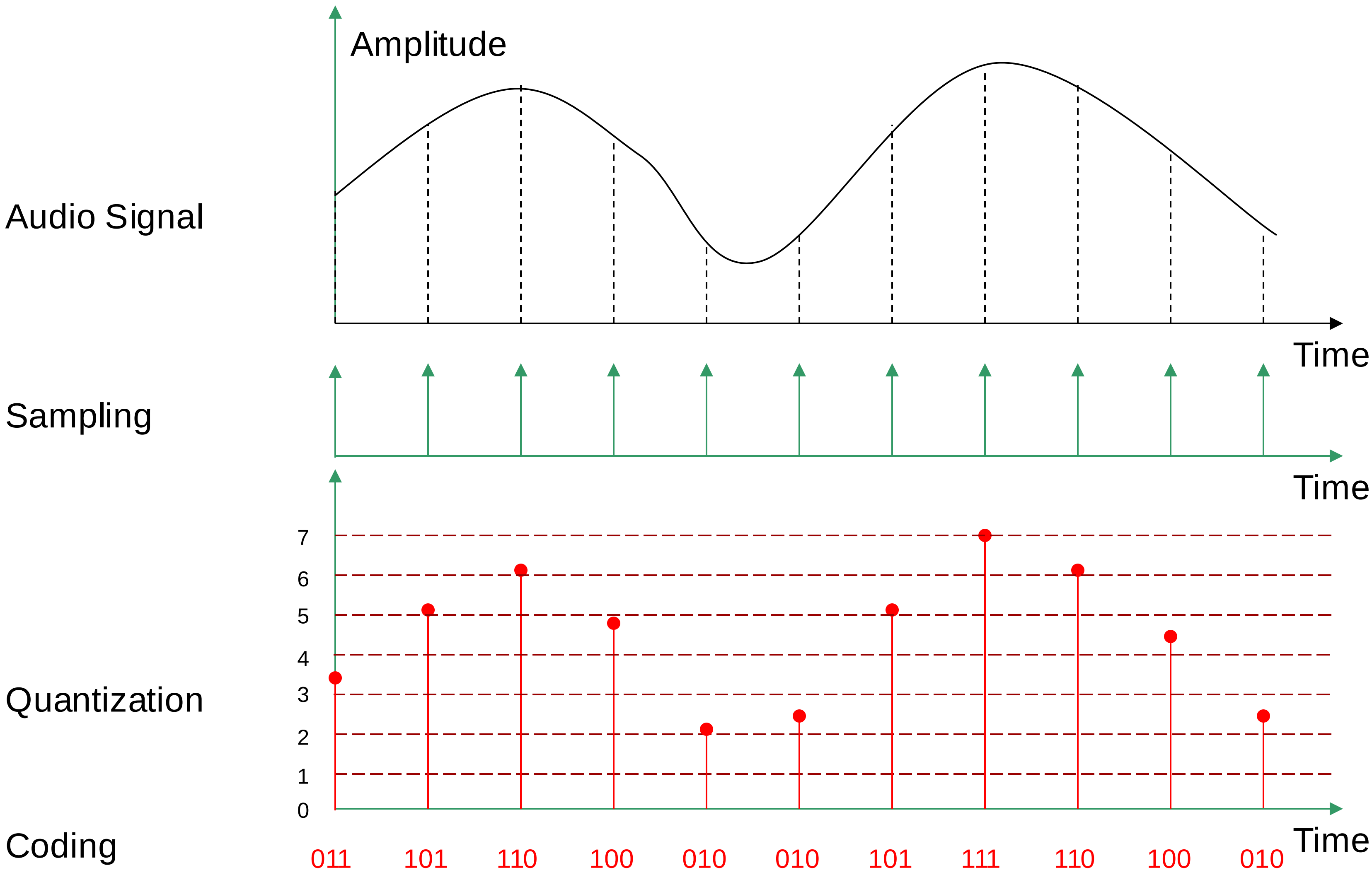


Sampling



Quantization

Digitizing Sound



Sampling Rate

- A sampling rate must be chosen that will preserve at least the full range of audible frequencies, if high-fidelity reproduction is desired.

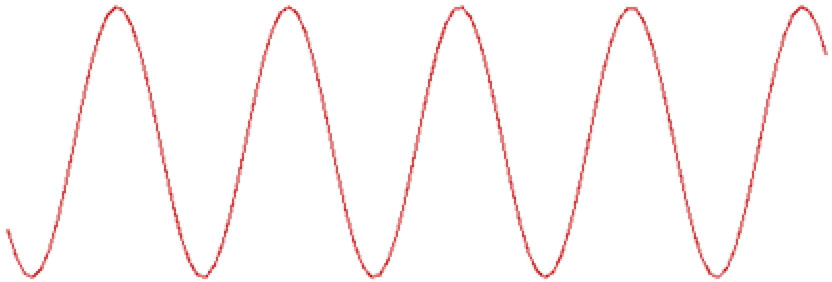
- The Sampling Theorem – if the highest frequency component of a signal is at f_h , the signal can be properly reconstructed if it has been sampled at a frequency greater than $2f_h$.
  - This limiting value is known as the *Nyquist rate*.

- If the limit of hearing is taken to be 20kHz, a minimum sampling rate of 40 kHz is required by the Sampling Theorem.

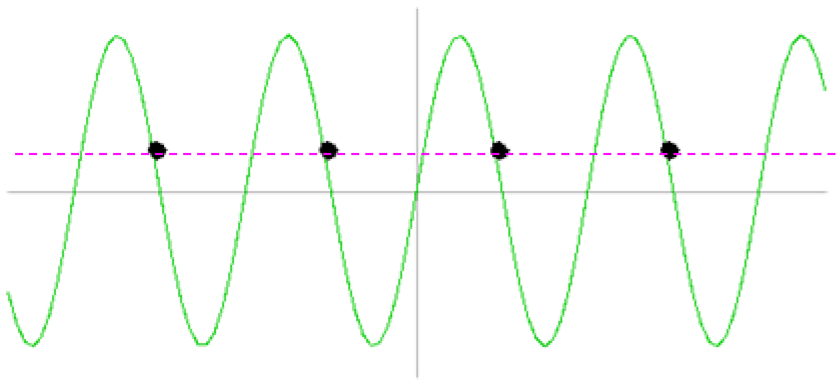
Sampling Rate

- If the limit of hearing is taken to be 20kHz, a minimum sampling rate of 40 kHz is required by the Sampling Theorem.
- The sampling rate used for audio CDs is 44.1kHz – the precise figure being chosen by manufacturers to produce a desired playing time given the size of the medium.
- Where a lower sound quality is acceptable, or is demanded by limited bandwidth, sub-multiples of 44.1kHz are used: 22.05kHz is commonly used for audio destined for delivery over the Internet, while 11.025kHz is sometimes used for speech.
- 48kHz is used when the best quality is desired, e.g., DAT (digital audio tape) recorders and the better sound cards.

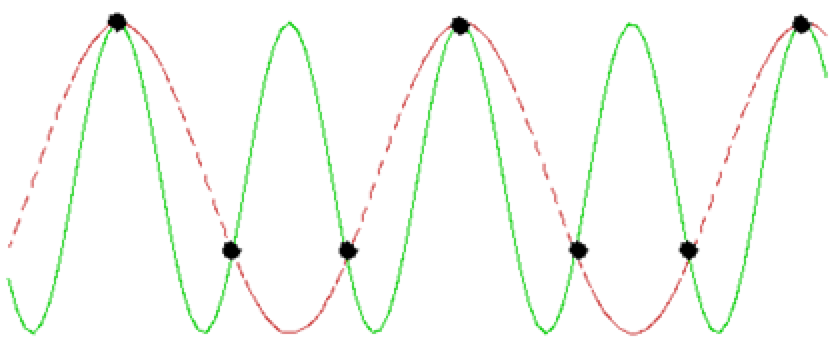
Example_



Original signal
with a single frequency

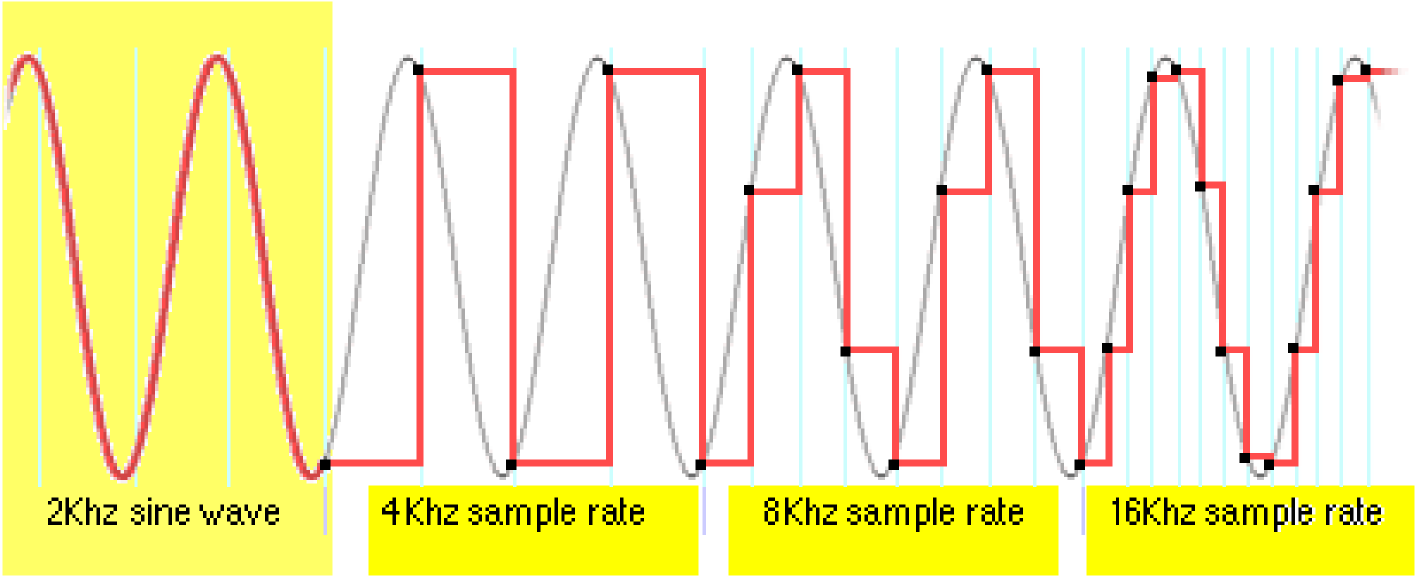


Sampling at exactly the frequency produces a constant



Sampling at 1.5 times per cycle produces an alias frequency that is perceived

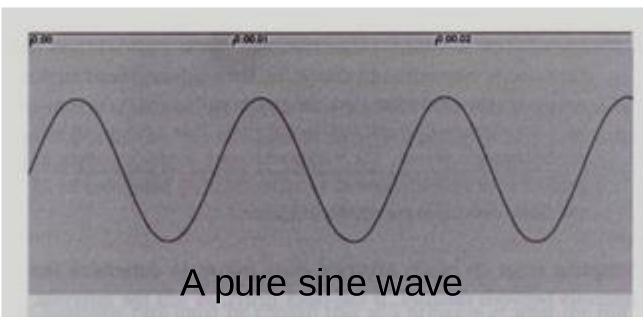
Example_



2Khz sine wave
4Khz sample rate
8Khz sample rate
16Khz sample rate

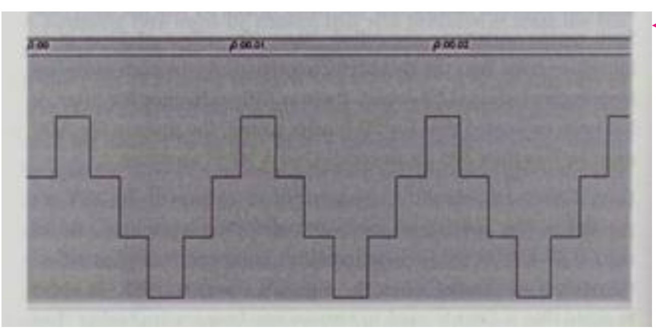
Quantization

- For sound, the most common choice of sample size is 16 bits, as used for CD audio, giving 65536 quantization levels.
- Smaller samples sizes (lower bit-depths) are sometimes needed to maintain small file sizes and bit rates. The minimum acceptable is 8-bit sound, and even this has audible quantization noise, so it can only be used for applications such as voice communication, where the distortion can be tolerated.



A pure sine wave

- ADC – sampling with only four quantization levels (2 bits)



Evidently, the sampled wave form is a poor approximation of the original, with high quantization error.

Recording Sound

- Before recording, it is necessary to select a sampling rate and sample size.
- As a general rule, the highest possible sampling rate and sample size should be used, to minimize deterioration of the signal when it is processed.
- If a compromise must be made, the effect on quality of reducing the sampling size is more drastic than that of reducing the sampling rate – the same reduction in size can be produced by halving the sampling rate or halving the sample size; the former is better.

Audio File Size

- A simple calculation suffices to show the size of digitized audio.

- The sampling rate is the number of samples generated each second, so if the rate is r Hz and the sample size is s bits, each second of digitized sound will occupy $r \times s / 8$ bytes.

  - For example, for CD-quality, $r = 44.1 \times 10^3$ and $s = 16$, so each second occupies just over 86 kbytes, each minute roughly 5 Mbytes.

  - These calculations are based on a single channel, but audio is almost always recorded in stereo, so the estimates should be double.

Sampling Rate and File Size

- One-Minute Digital Audio Recordings at Common Sampling Rates and Resolutions

Sampling Rate	Resolution	Stereo or Mono	Bytes / min	Comments
44.1kHz	16-bit	Stereo	10.5MB	CD-quality recording
44.1kHz	16-bit	Mono	5.25MB	A good trade-off for HQ recordings of voice-overs
44.1kHz	8-bit	Stereo	5.25MB	Achieves highest playback quality on low-end devices
44.1kHz	8-bit	Mono	2.6MB	An appropriate trade-off for recording a mono source
22.05kHz	16-bit	Stereo	5.25MB	Darker sounding than CD-quality; For CD-ROM projects
22.05kHz	16-bit	Mono	2.5MB	Not a bad choice for speech
22.05kHz	8-bit	Stereo	2.6MB	A popular choice for reasonable stereo recording
22.05kHz	8-bit	Mono	1.3MB	Sound from TV set, or from good AM radio
11kHz	8-bit	Stereo	1.3MB	Few advantages to using stereo
11kHz	8-bit	Mono	650KB	Very dark and muffled
5.5kHz	8-bit	Stereo	650KB	Stereo not effective
5.5kHz	8-bit	Mono	325KB	About as good as a bad telephone connection

Need To Know

- Imaging process
- Video acquisition
- Audio acquisition
- Digitization process
  - Sampling
  - Quantization

References

- Rafael Gonzalez and Richard Woods, Digital Image Processing (3rd Edition), Prentice Hall, 2008.
- The Science of Digital Media, Jennifer Burg, Prentice Hall, 2009.
- Linda G. Shapiro and George C. Stockman, Computer Vision, Prentice Hall, 2001.
- D. Litwiller, CCD vs. CMOS: Facts and Fiction, Photonics Spectra, January 2001, LaurinPublishing Co. Inc.