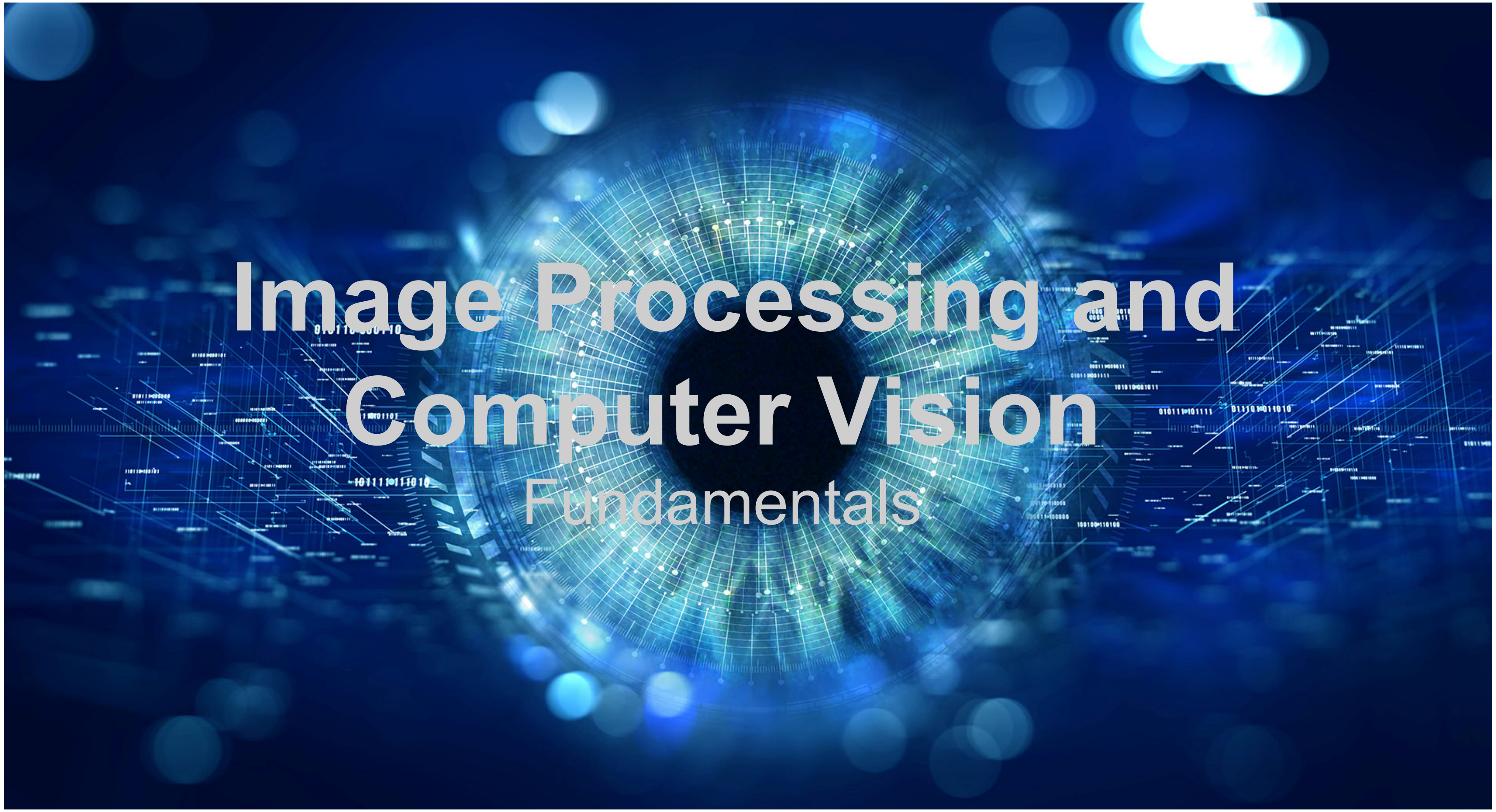


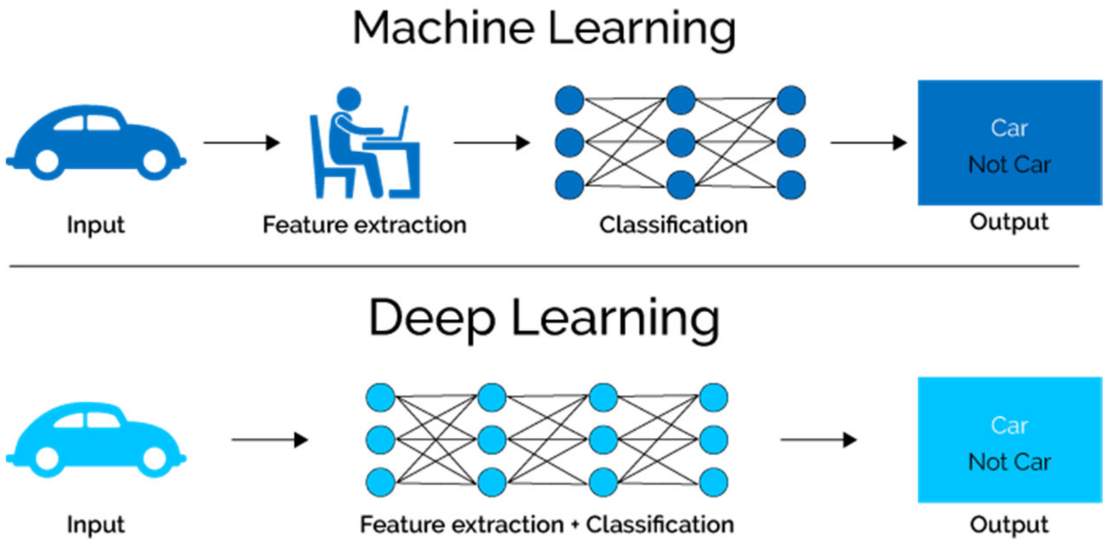
Image Processing and Computer Vision
Fundamentals

Some Areas of Interest

- Image representation, coding, and transmission
  - GIF, JPEG, Multimedia applications, HDTV
- Image enhancement
  - Making images easier for humans to interpret
  - sharpening and false colouring of medical images
- Image alteration
  - often for aesthetic reasons, but sometimes to meet scientific needs
  - image warping, morphing, correcting for geometrical distortions
- Image analysis
  - computer-based analysis of images
  - computer vision, image segmentation
  - deep learning

What is Deep Learning (DL) ?

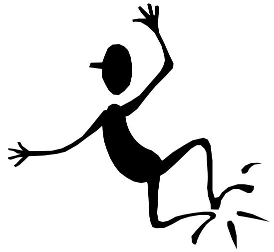
A machine learning subfield of learning representations of data. Exceptionally effective at learning patterns.
Deep learning algorithms attempt to learn (multiple levels of) representation by using a hierarchy of multiple layers of artificial neurons
If you provide the system lots of information, it begins to understand it and respond in useful ways.
Why is it called deep? Because it has many layers.



https://www.xenonstack.com/blog/static/public/uploads/media/machine-learning-vs-deep-learning.png

Example Problem: Number Plate Recognition

- Since humans can easily perform these tasks, sometimes the problems look deceptively simple

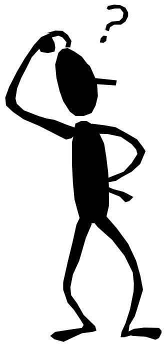


Still a bit noisy

Example: Number Plate Recognition

- Robust real world computer vision has to be able to cope with much more complicated examples

How about this for geometrical distortion?



Much Harder Problem

- How do we recognize cats?
- Deep Learning comes to the rescue

Why is Computer Vision Hard?

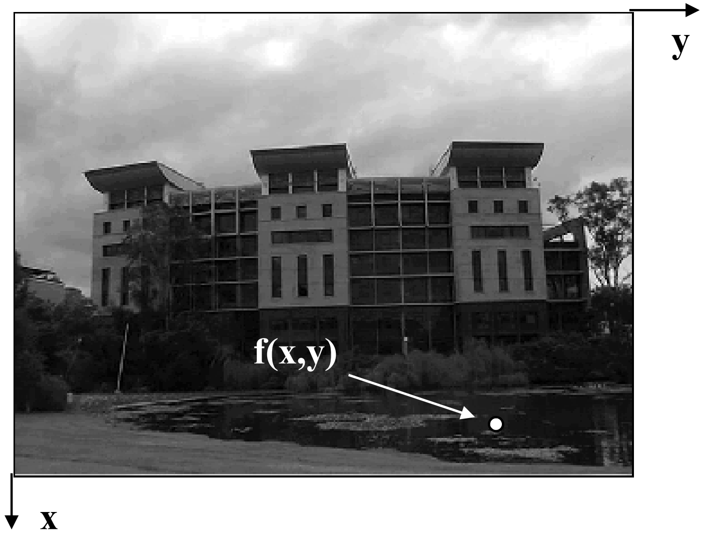
The camera sees :

194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

Image courtesy Andrew Ng CS229 Notes

Image Representation

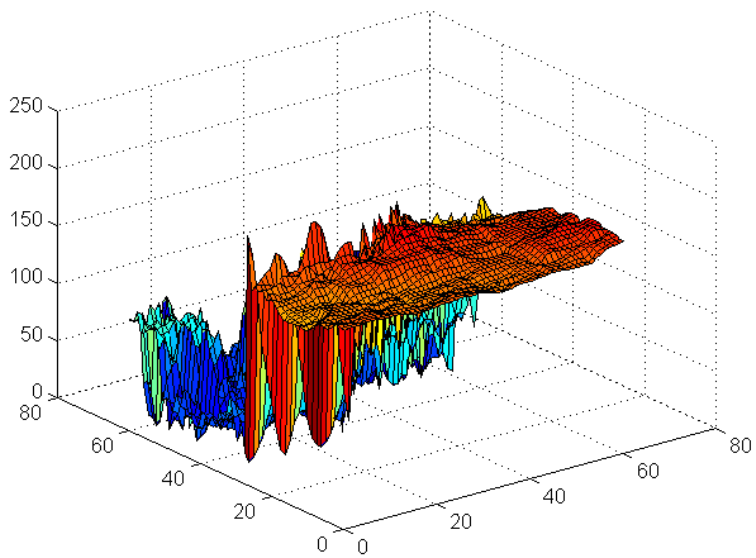
- A monochrome image or simply image, refers to a 2D light intensity function f(x,y) where x and y denote spatial coordinates and f is proportional to the brightness or grayscale (gray level) at that point.



Note the direction of the axes!

Image as a Surface

- Sometimes it is more useful to think of an image as a surface where gray level is represented by height

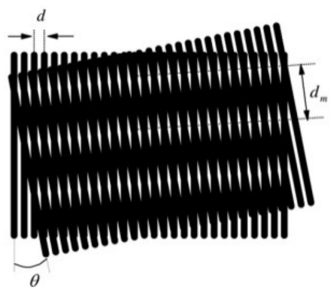


Digital Images

- A digital image is an image f(x,y) that has been discretized in both spatial coordinates and brightness
- The most common way to discretize spatial coordinates is regular sampling on a rectangular grid as is usually obtained from CCD and CMOS cameras.
- A digital image can be considered as a matrix where matrix element values represent grayscale
- Each element at a given spatial location of the digital image is called a pixel
- A typical size comparable in quality to a monochrome Analog TV image is 512 x 512 with 128 gray levels (7-bits). This is sometimes referred to as standard definition (SD) and is about 0.3 MP.
- Modern Digital TVs are 2MP (HD) or 4MP (4K).

Digital Image Formats

- Binary Format, $f \in \{0,1\}$
  - Intensity is indicated by just a 0 or 1
  - Good for representing text and line diagrams
  - Can be used to represent grayscale image data with half toning for printing and photocopying. Hard to scan due to Moire fringe effects!
  - Not easy to scale and resize due to aliasing effects

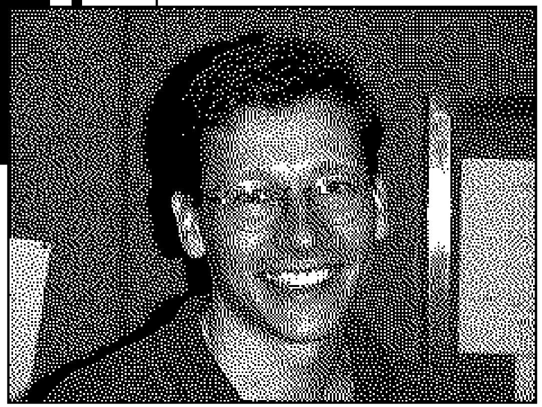


- Grayscale Format, $f \in Z^+$
  - Intensity is typically represented in 8 bits, 0..255
  - Good for representing monochrome pictures
  - may be used to represent colour images if each grayscale is mapped to a colour through a palette (pseudo-colour format)
  - relatively easy to scale images and text
- Colour Format, $f \in (Z^+)^3$
  - Intensity is typically represented by 3 8-bit numbers representing, say, red, green, and blue intensities.

Examples

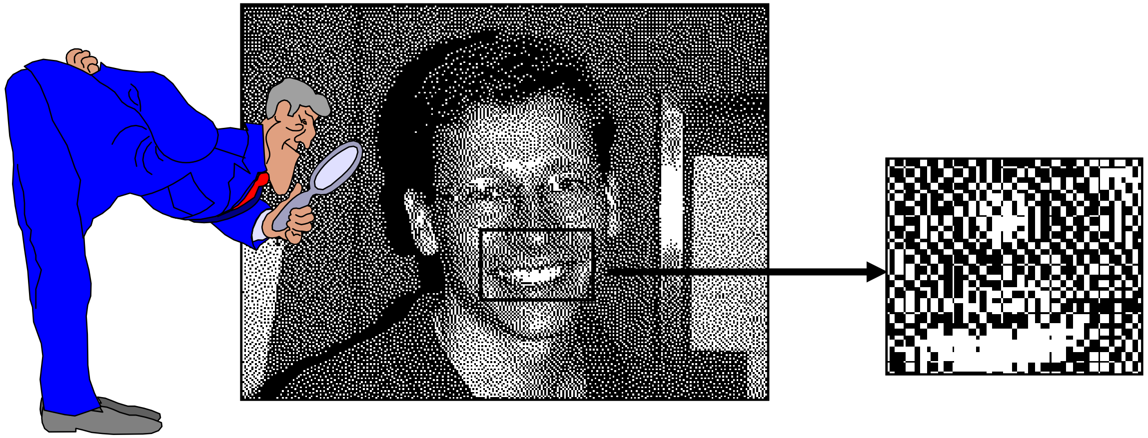
Original

Binary
Thresholded

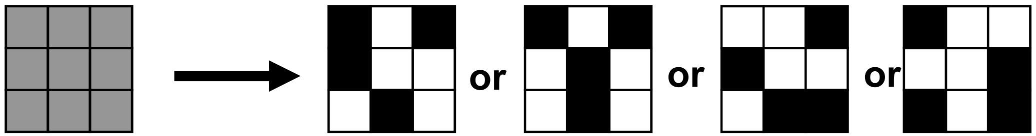


Binary
Halftoned

Half toning



A block of grayscale pixels replaced by a block of binary pixels with the same average intensity. The binary pixels are randomized to reduce visible artifacts.



Binary File Formats

0101001
1100101
1010011

- Simple formats
  - ascii text: inefficient
  - raw: 1 bit per pixel written out a byte at a time
- Some standard formats
  - Portable bitmap format
    - pbm and pbmraw as above, with headers
    - part of netpbm utilities
    - uncompacted
  - TIFF
    - Tagged Image File Format, compacted, run length encoding
  - G3 Fax standard
    - Used in fax machines, compacted, run length encoding
  - JPEG
    - For photography, not so good for text

Simple Image Compaction

- Ordinary images often contain runs of the same gray level over large areas. It suffices to precede the gray level by the run length or number of repeats.
- This is called run length encoding
  - Example: a grayscale image may contain the following sequence of gray levels
    - 003 004 004 004 008 008 006 007
  - Preceding each gray level with the number of repeats yields
    - 000 003 002 004 001 008 000 006 000 007
  - This is actually a little longer in this example, but would usually be much shorter when there is a great deal of repetition

TIFF Format

- Another format for RLE is the Tagged Image File Format (TIFF) where the run length is indicated by a preceding tag equal to 256 minus the number of repeats (as far as 127).
- Then to save too frequent repetition of 000, where a byte is not repeated, a tag in the range 1-127 indicated the number of following bytes which are to be read literally.
- Thus

  003 004 004 004 008 008 006 007

  becomes

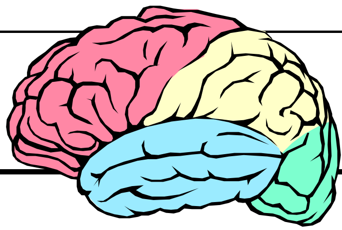
  001 003 254 004 255 008 002 006 007

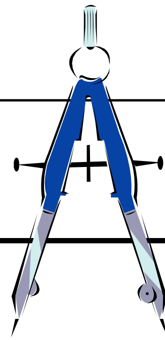
We still use TIFF for medical images but the files are very large.

Perception

- We generally process visual information with our eyes and brain, but we can also use instruments. Often the two do not agree due to the different processing mechanisms.

Perceived Quantity	**Physical Quantity**
Brightness	Light Intensity
Colour	Light Frequency





Colour Image Processing

- Why use Colour?
  - Colour is a powerful descriptor that often simplifies object identification and extraction from a scene
  - Humans can discern thousands of colour shades and intensity compared to about two dozen shades of gray.
- Two major areas
  - True colour
    - needs a full colour sensor such as a colour TV camera
    - may be converted to a reduced colour palette (e.g., 256 colours for GIF images); sometimes called pseudo-colour
  - False colour
    - assigning a shade of colour to a particular grayscale or range of grayscale; for example colour CAT or MRI scan or X-ray baggage inspection.

Examples

True Colour

Grayscale

False Colour

Colour Spectrum

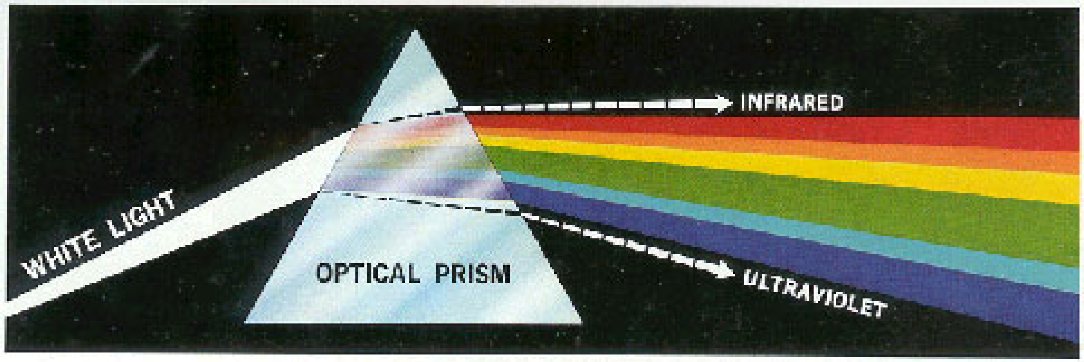
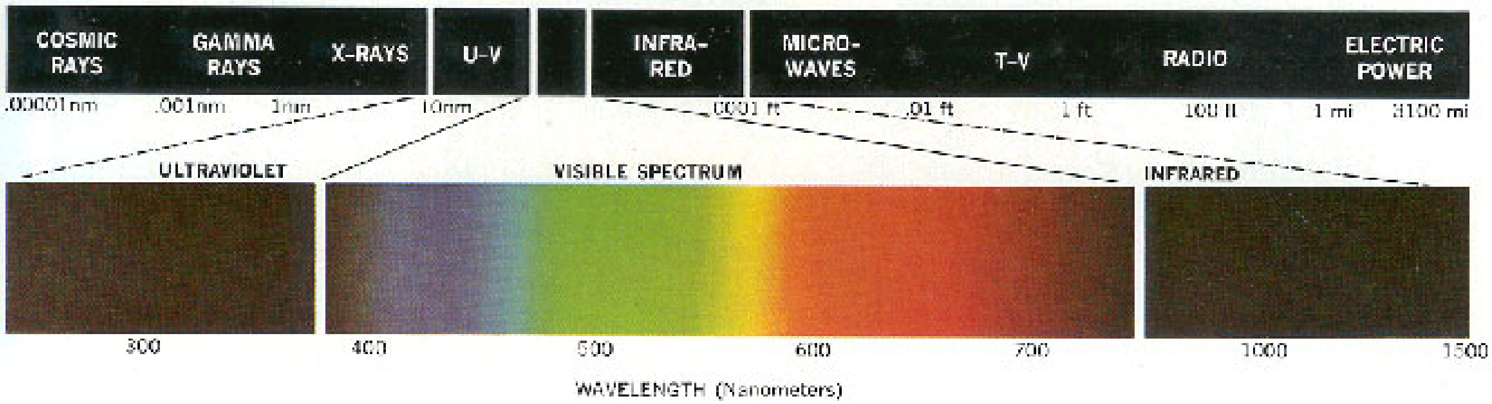
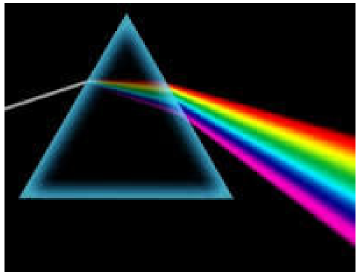


Plate I. Color spectrum seen by passing white light through a prism. (Courtesy of General Electric Co., Lamp Business Division.)



Fundamentals



- When a beam of light is passed through a prism, the emerging light is not white; instead it is a continuous spectrum of colours ranging from red at one end to violet at the other.
- Light consists of electromagnetic radiation in the wavelength range of about 400 (violet) to 700 (red) nm.
- Colours that humans perceive are determined by the nature of the light reflected from the object
- A body that reflects light that is relatively balanced in all visible wavelengths appears *white* to the observer

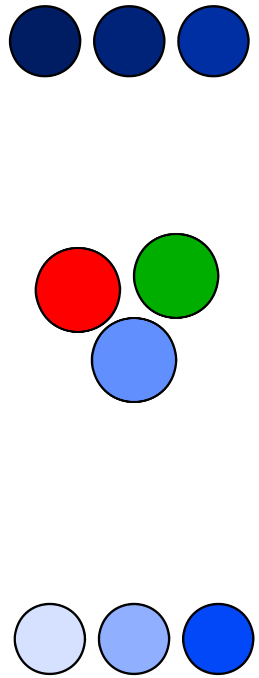
Fundamentals

- For example, green objects reflect light with wavelengths primarily in the 500 to 570 nm range.
- If light is achromatic (void of colour), its only attribute is its intensity or value. This is sort of light that is seen emanating from a B&W TV and is implicit is grayscale processing.
- Chromatic (coloured) light can be described in terms of radiance, luminance, and brightness

Radiance, Luminance, and Brightness

- Radiance is is the total amount of energy that flows from the light source (usually expressed in Watts)
- Luminance (measured in lumens (lm)) gives a measure of the energy an observer perceives from a light source. For example, an infrared lamp could have significant radiance but it would be hardly visible; its luminance would be almost zero.
- Brightness is a subjective descriptor that is practically impossible to measure. It embodies the achromatic notion of intensity and is one of the key factors in describing colour sensation.

Brightness, Hue, and Saturation



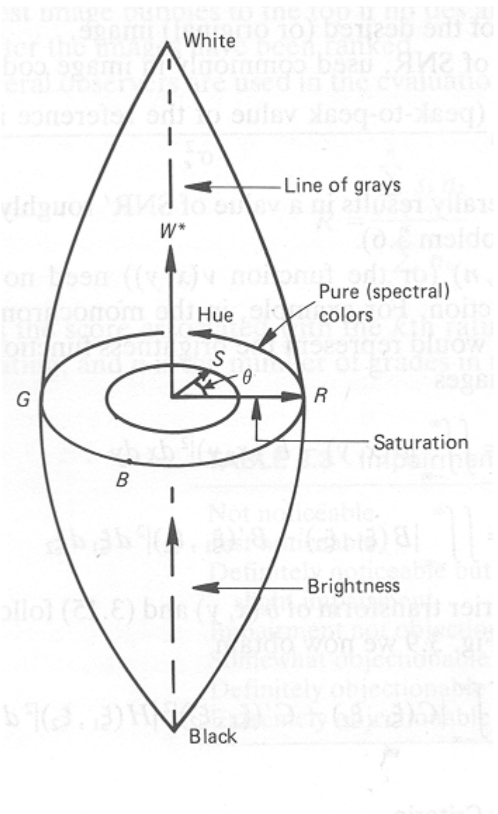
- The perceptual attributes of colour are brightness, hue, and saturation.
- Brightness represents perceived luminance
- Hue refers to the redness, greenness etc of the colour. For monochromatic light sources, differences in hue are manifested by differences in wavelength.
- Saturation is the aspect of color that varies as white light is added to monochromatic light. Fully saturated colours are vivid, less saturated colours are described as washed-out or pastel.

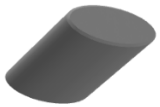
Perceptual Representation of Colour Space

The centre axis represents transitions from black to white through all the shades of gray

As we go away from this axis, we have increasingly saturated colours

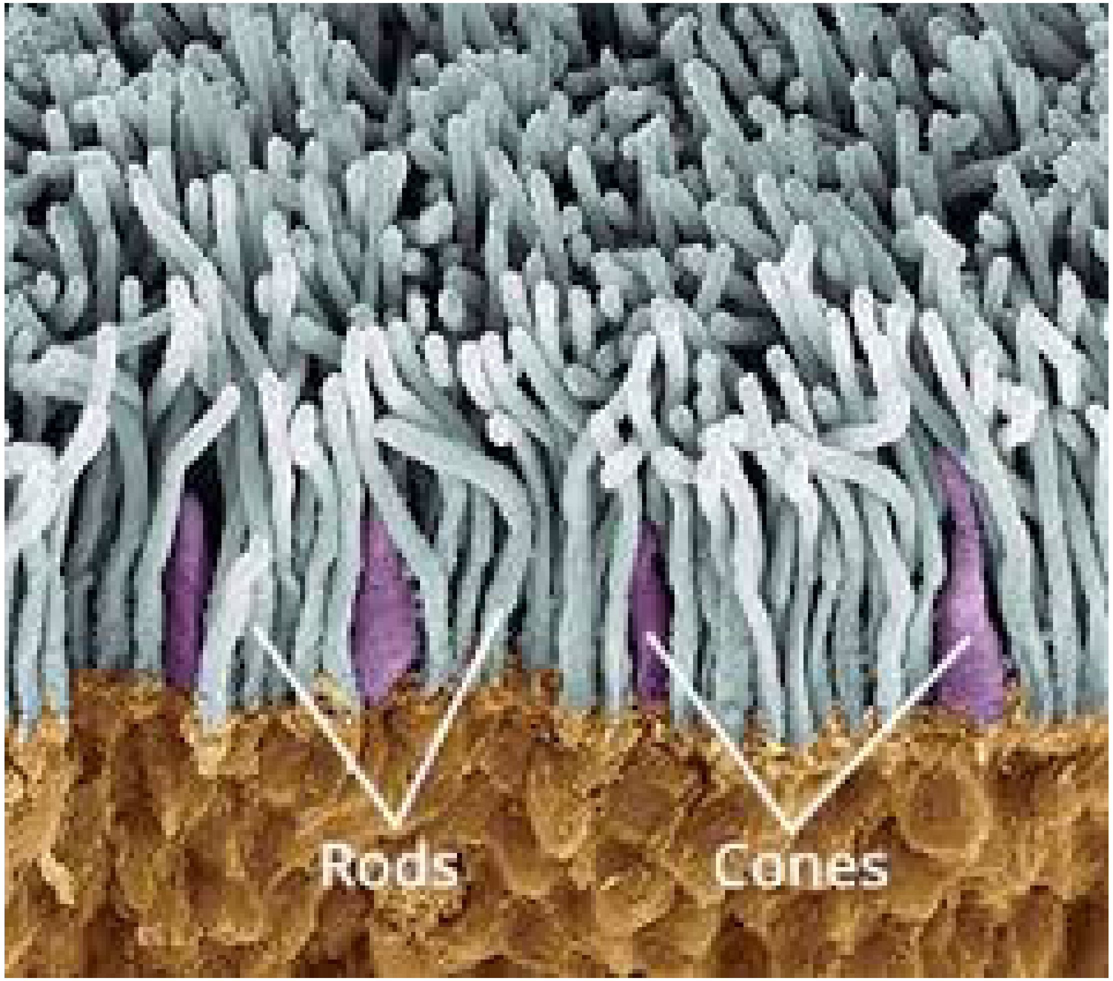
The angle of departure represents the hue





Human Eye

- The eye has two types of light receptors: rods and cones
- The rods measure light intensity only and are used mainly for night and peripheral vision
- There are three types of cone receptors;each one sensitive to a different wavelength of light: red, green, and blue
- Thus the three cones form a type of light spectrum analyzer that is used to interpret incoming visible light in terms of perceived colour.


Rods
Cones

Sensitivities of Cone Receptors

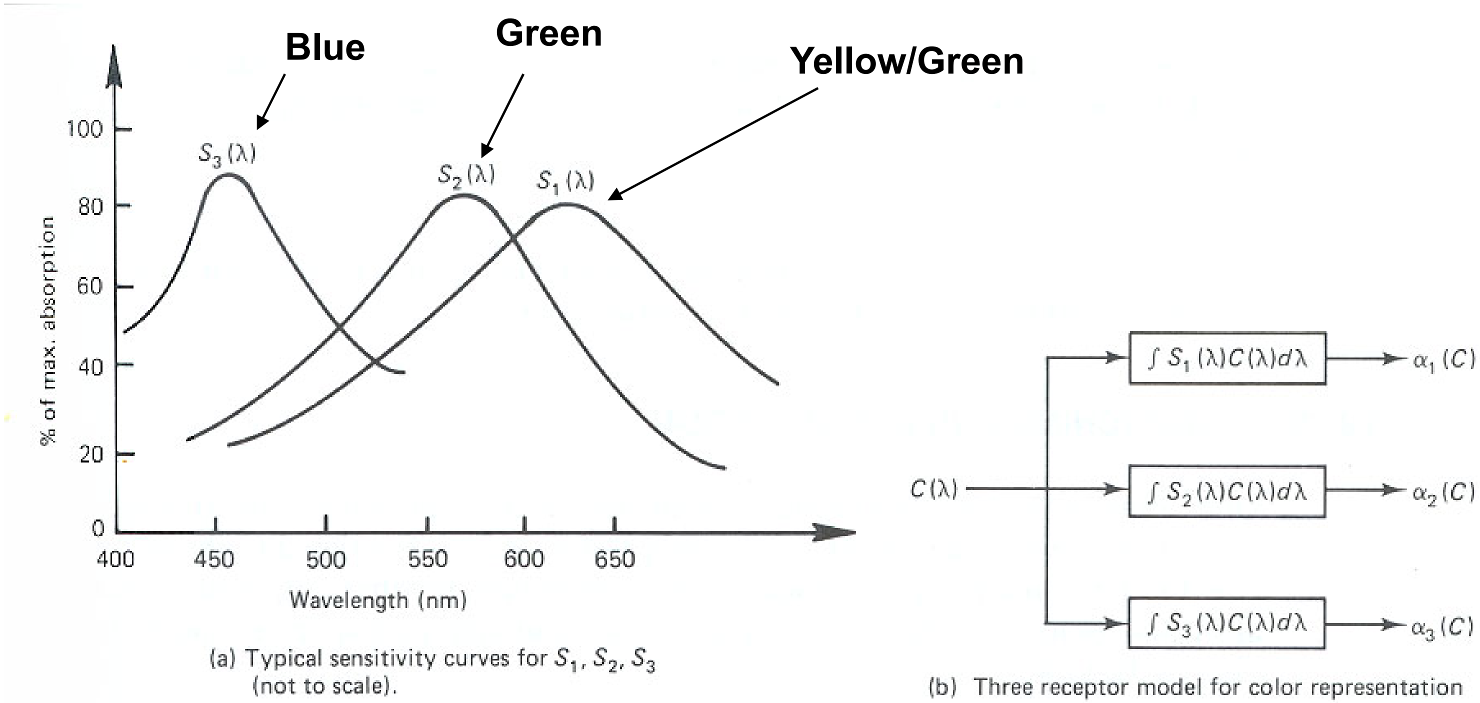
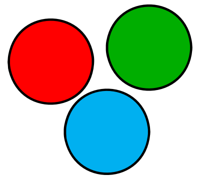


Figure 3.11 (a) Typical absorption spectra of the three types of cones in the human retina; (b) three-receptor model for color representation.

Primary Colours

- For the purpose of standardization, the CIE (Commission Internationale de l'Eclairage – International Commission on Illumination) assigned the following wavelengths to the three primary colours
  - red=700nm, green=546.nm, blue=435.nm
- Note that no single colour can be called red, green, or blue. These attributes apply to a range of wavelengths
- By mixing the light from three primary colour light sources, we can create a very wide range of colours – but not all possible colours. This is an example of the *additive* primary colour model as used in colour TV sets and computer monitors.

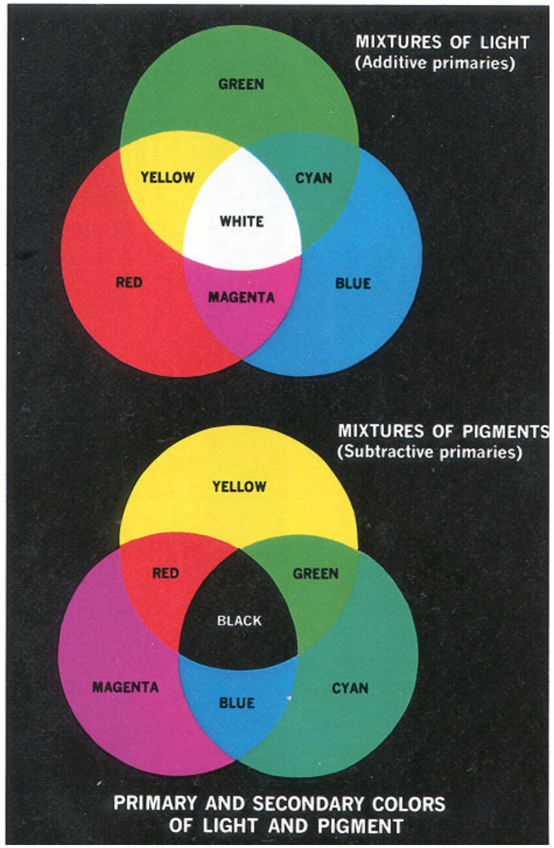


Additive and Subtractive Models

If we add the primary colours pair wise, we get the colours cyan, yellow, and magenta.

These are called the subtractive primary colours which is the model required for mixing paint and pigment.

Thus many printers use the CYMK model where CYM stands for the subtractive primaries and K stands for black ink.



CIE Chromaticity Diagram

Only colours within the triangle can be represented by mixing the three primary colours red, green, and blue. Some computer monitors have 4 phosphors to improve the colour *gamut*.

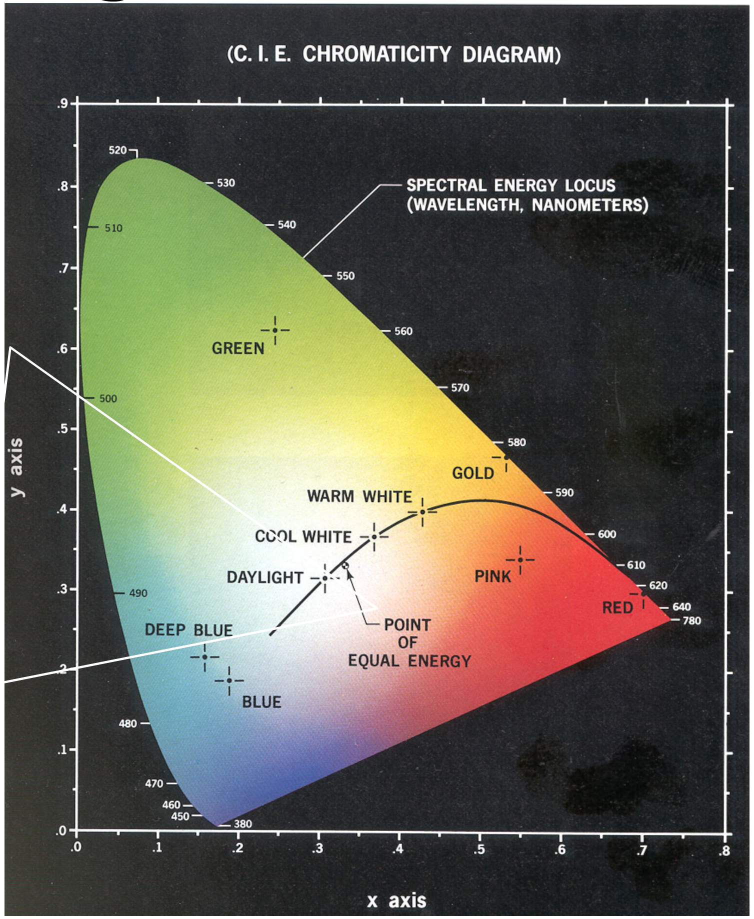
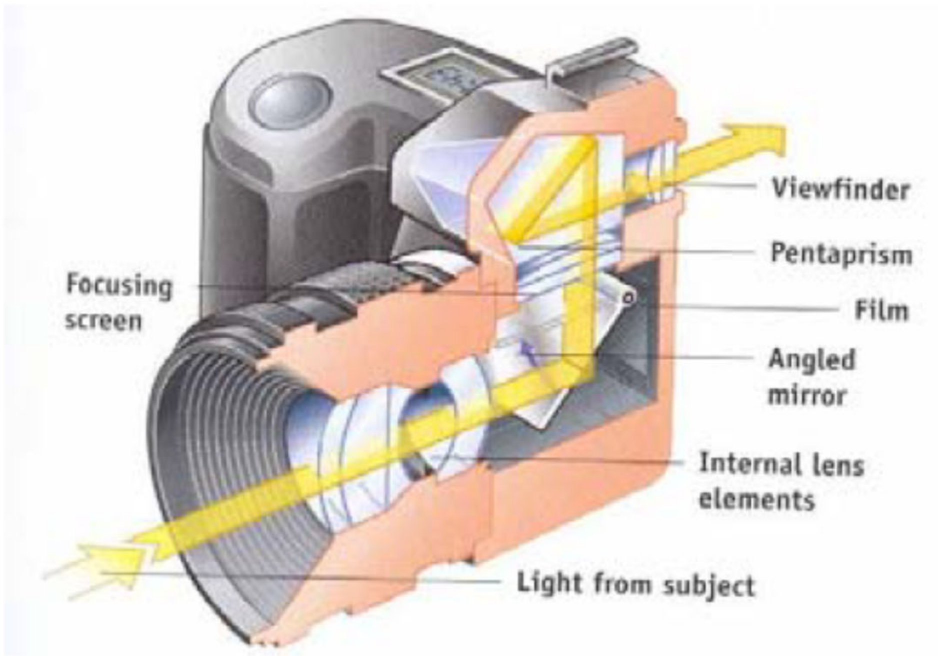
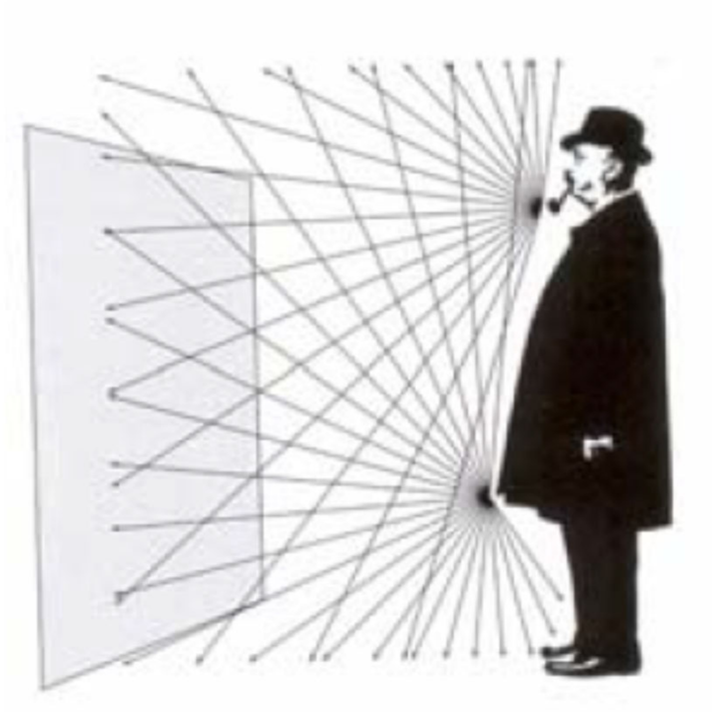


Image Formation

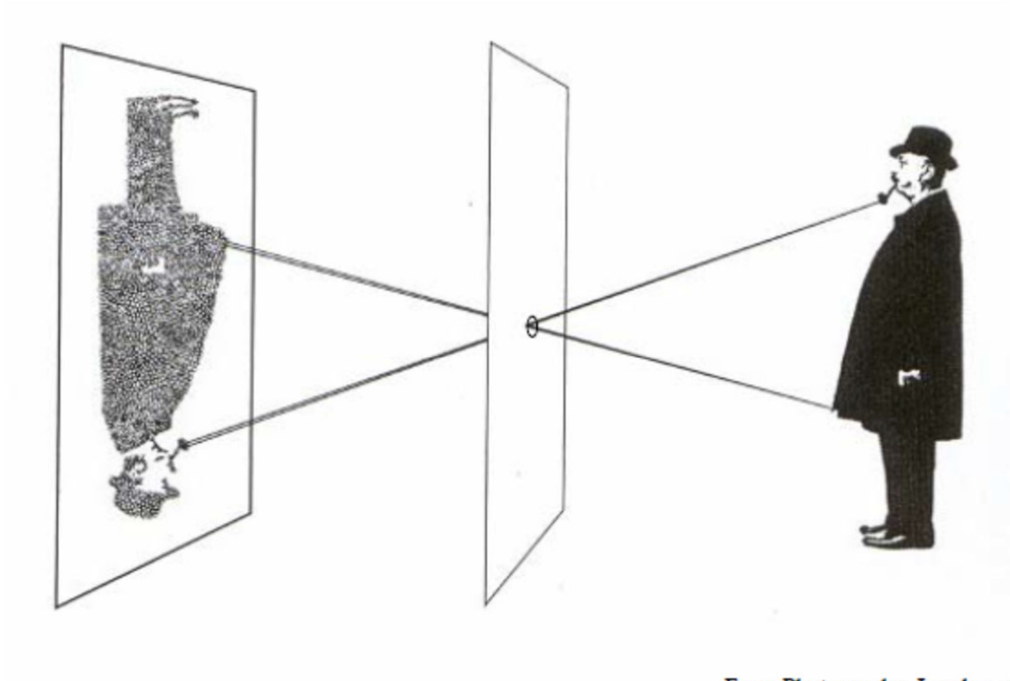


Why is there no image on a White Piece of Paper?

It receives Light from all directions.

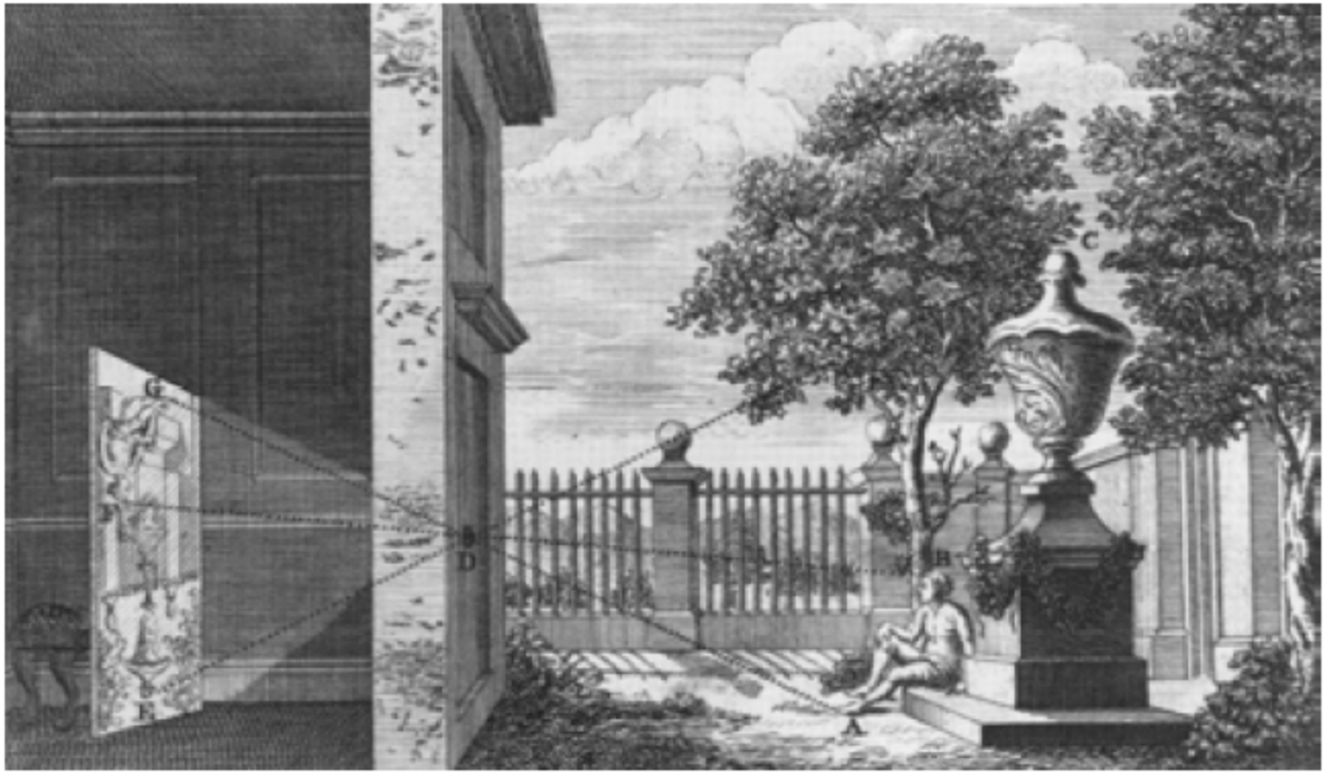


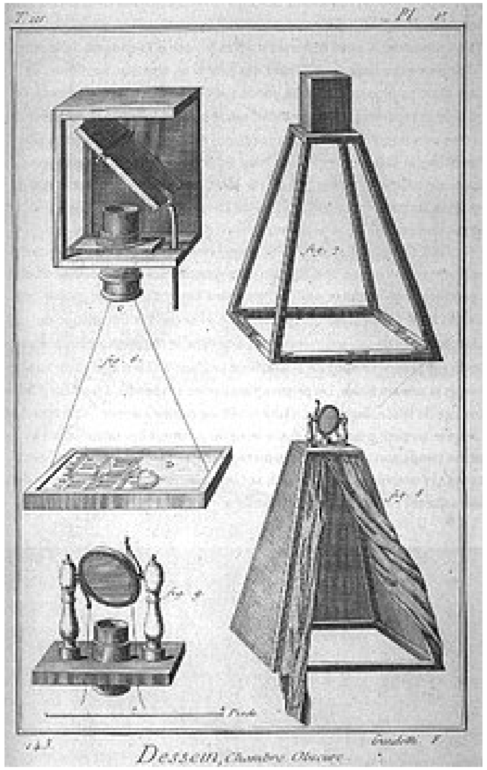
Pinhole Image Formation



From Photography, London et al.

Camera Obscura





Can Also use Concave Mirrors

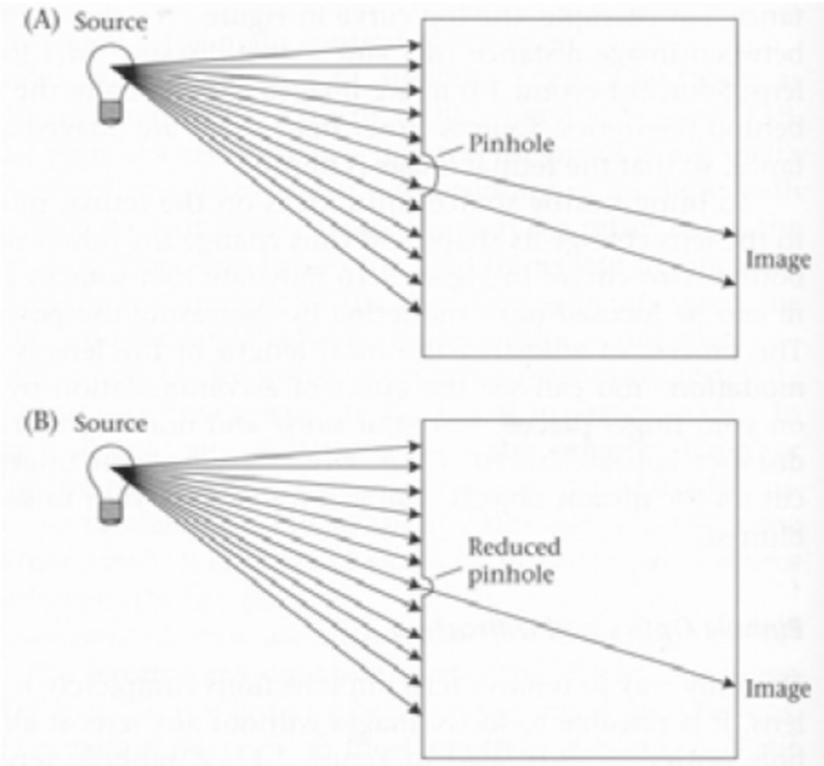
This could explain sudden increase in realism paintings from about 1420. Photographic techniques may have guided these artists.

Examples of out-of-focus paintings
Also left-right reversed paintings
Many portraits of left-handed people

Detail of the chandelier and mirror from Jan van Eyck's 1434 *Arnolfini Portrait*, one of Hockney's key examples

https://en.wikipedia.org/wiki/Hockney%E2%80%93Falco_thesis

Pinhole Size

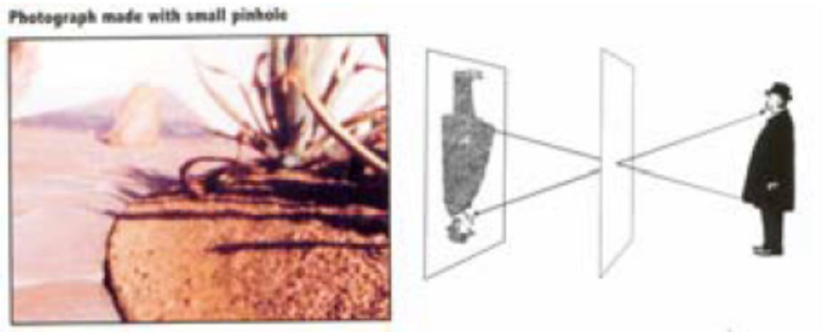


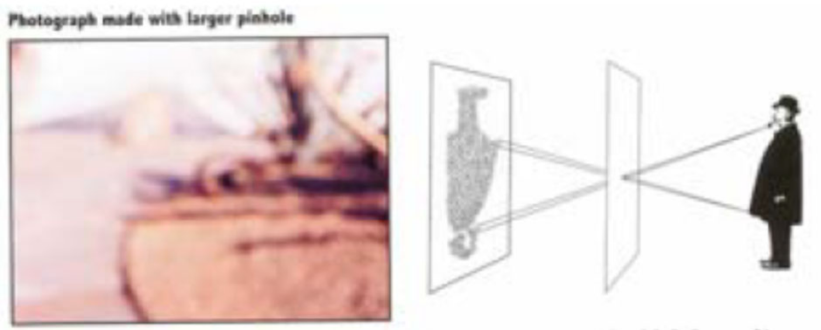
Wandell, Foundations of Vision, Sinauer, 1995

Blurred but more light

Clear but very dark

Pinhole Size?

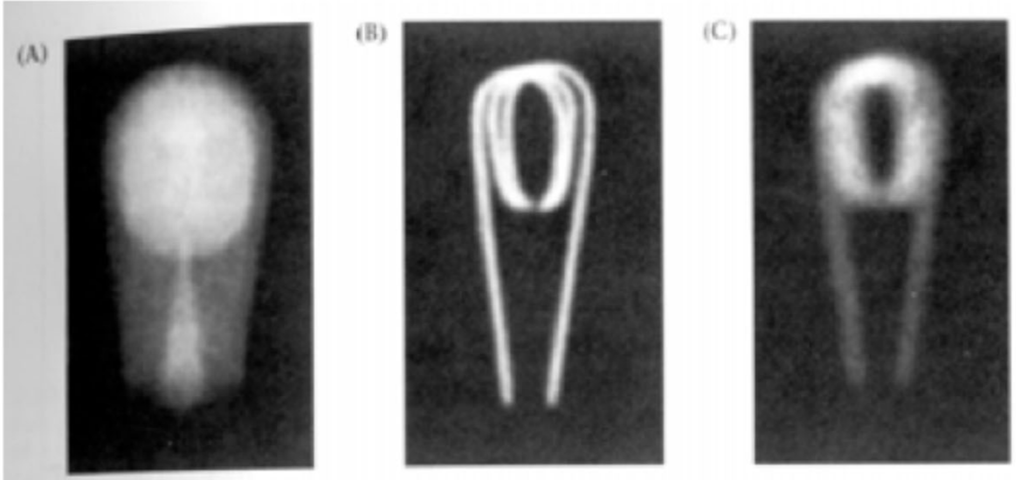




From Photography, London et al.

Diffraction Limit

- Optimal size for visible light:
  - sqrt(f)/28 (in millimiters) where f is focal length

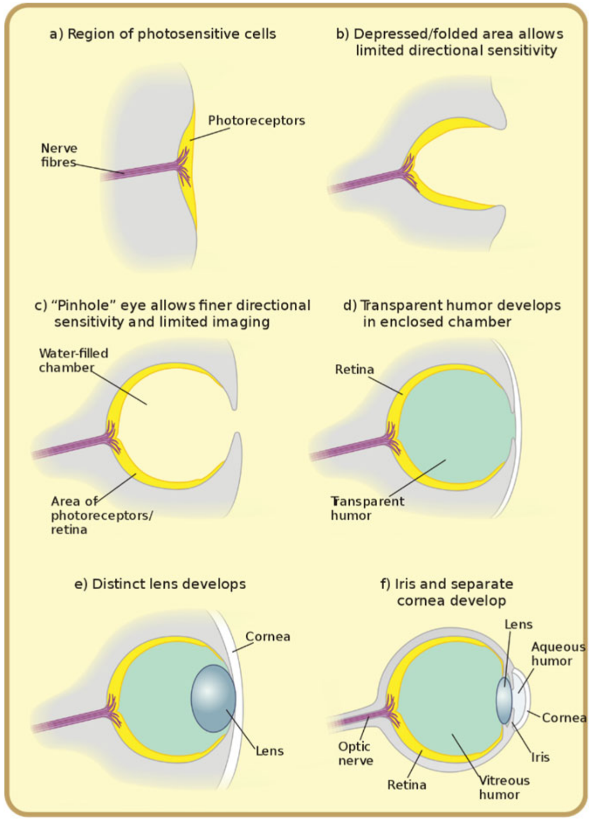


2.18 DIFFRACTION LIMITS THE QUALITY OF PINHOLE OPTICS. These three images of a bulb filament were made using pinholes with decreasing size. (A) When the pinhole is relatively large, the image rays are not properly converged, and the image is blurred. (B) Reducing the size of the pinhole improves the focus. (C) Reducing the size of the pinhole further worsens the focus, due to diffraction. From Ruechardt, 1958.

From Wandell

Evolution of the Eye

Some people claim they the eye is so complex that it could not have arisen be chance. This image suggests the various stages in the evolution of the eye. The development of vision caused an explosion in Life on Earth in 537 million years BC.



Problems with Pinhole

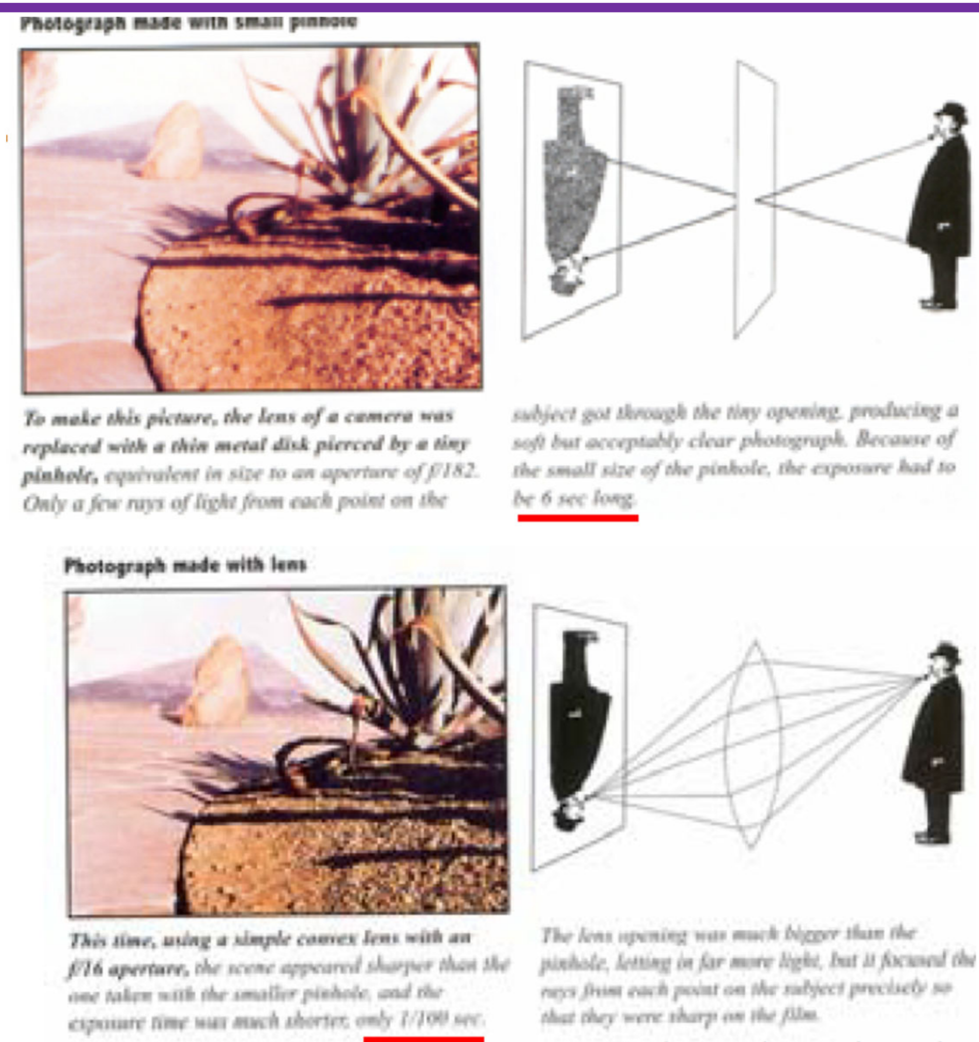
- Not enough light
  - Requires long exposure and leads to motion blur
- Diffraction limits sharpness of images

Solution

- Refraction and Reflection
- Refraction is responsible for image formation by lenses and the eye
- Reflection by mirrors is used in large telescopes

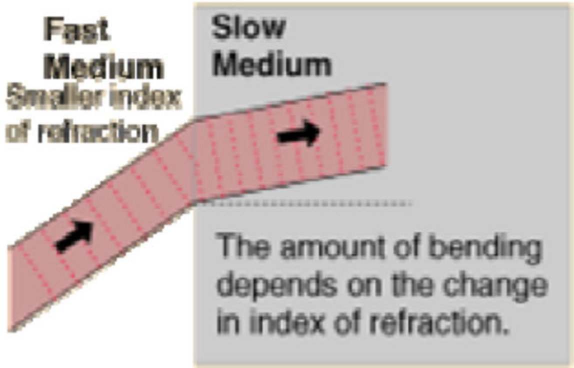
Lenses

- Gathers more light
- But needs to be focussed



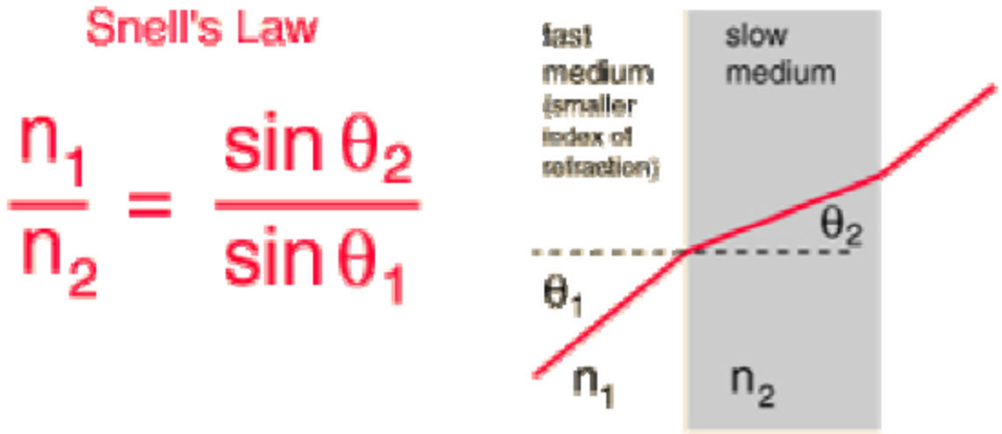
From Photography, London et al.

Refraction



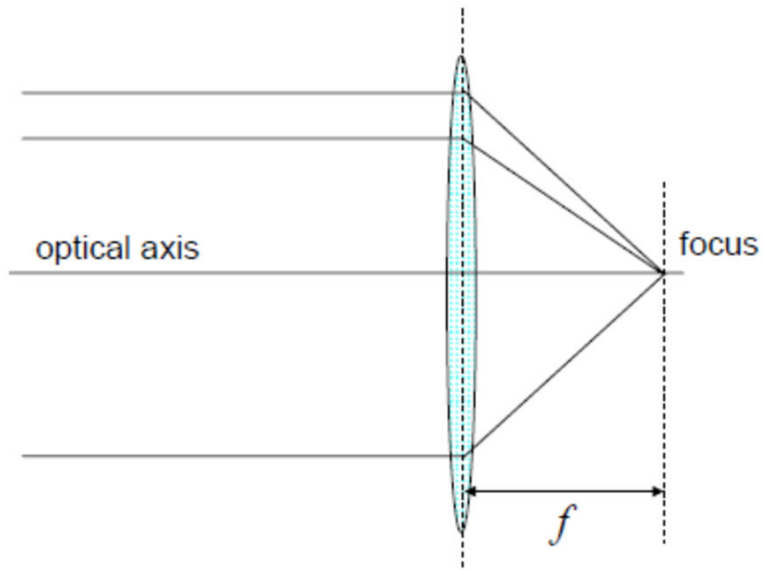
- Refraction is the bending of a wave when it enters a medium where it's speed is different.
- The refraction of light when it passes from a fast medium to a slow medium bends the light ray toward the normal to the boundary between the two media.
- The amount of bending depends on the indices of refraction of the two media and is described quantitatively by Snell's Law.

Snell’s Law



n_i -- is the index of refraction which is defined as the speed of light in vacuum divided by the speed of light in the medium.

Thin Lens



Spherical lense surface: Parallel rays are refracted to single point

Gathers more light
Image plane needs to
Placed at focal plane
For optimal sharpness

Cameras and Computer Vision

- Computer vision works best with clear images
- We encounter issues due to
  - Low light grain
  - Motion blur due to long exposure
  - Deformation due to rolling shutter in CMOS devices
  - Codec issues JPEG and H265 artefacts
  - Depth of field and focussing issues
  - Low resolution of small objects
- Some of these problems are best solved by getting a better camera

Some Image Capture Issues

Motion Blur

Low Light Grain

Rolling Shutter Distortion

H264 CCTV Artefacts

Fence in front of Subject!!

7/7/2005 London Tube Bombers