EEE116 Multimedia Systems

Topic 6 – Visual content - 1

So far in EEE116

- Communication systems, Network requirements, latency, reliability and connectivity.
- Signal digitisation, sampling rate, fixed length codes, variable length codes, information & entropy, Huffman coding, necessity for data compression, data modelling for compression, lossless coding model.
- Digital Audio – sampling, quantisation, audio quality, PCM, DPCM, speech coding, LPC, perceptual coding, masking, standards, mp3 codec.

In this topic

- Visual content
  - Elements of visual perception.
  - Image acquisition and digital representation.
- Mid Term Test on 27th March (Friday) 2009 at **9 am** in the first year lab

Dr. Charith Abhayaratne
c.abhayaratne@sheffield.ac.uk

Visual Content (image/video)

- Elements of visual perception. (Topic 6)
  - Functionalities of the human visual system (HVS).
  - Image formation in the eye.
  - Brightness adaptation and discrimination mechanisms.

- Image acquisition and digital representation. (Topic 6)
  - Image sensing and acquisition
  - Image formation model
  - Digital representation (sampling and quantisation revisited)
  - Display systems

- Preparation for image/video communication (Topic 7)
  - How to reduce the data rate for digital image/video content

Capture/Display/Storage/Transmission of Visual Content

Human Visual System (HVS)

Why should engineers know about the HVS?

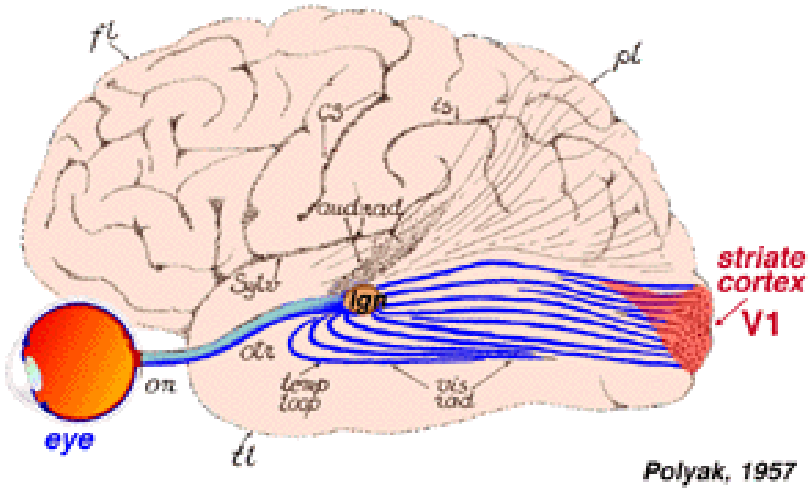
- Provides an existence proof for vision
- Understand the limits of the HVS – hence design systems to match (economy of representation) or to extend
- Provides clues as to how to design vision systems

What are the main components of the HVS?

-
-

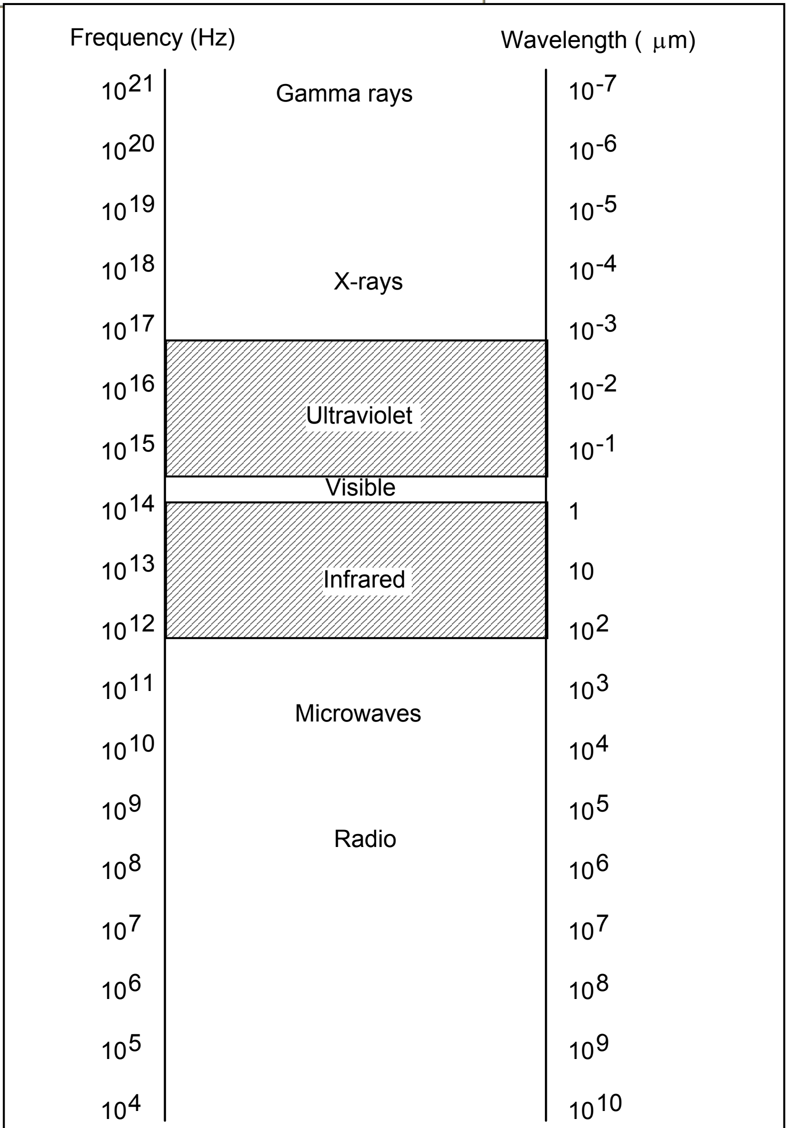
Human Visual System (HVS)

- Half of our brain is devoted to understanding what we see
- 50,000 times more information enters our eyes than our ears



Demos: 1. Card experiment
2. Basketball experiment

Colour Vision



Electromagnetic Spectrum

Speed of light, $c = \nu\lambda$

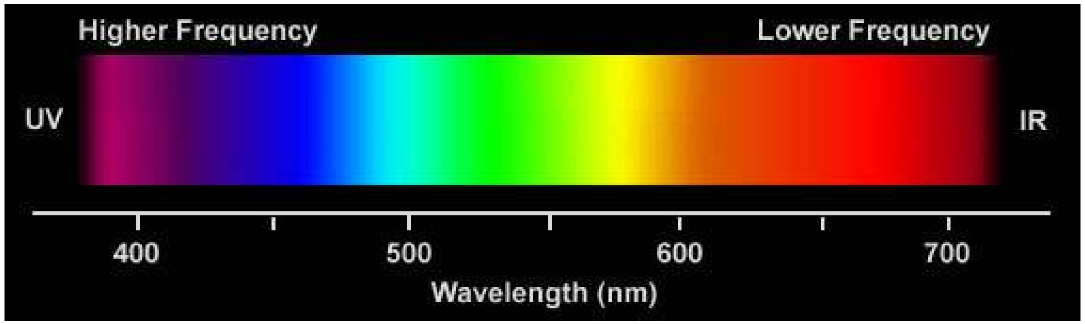
The electromagnetic spectrum is expressed in terms of wavelength or frequency.

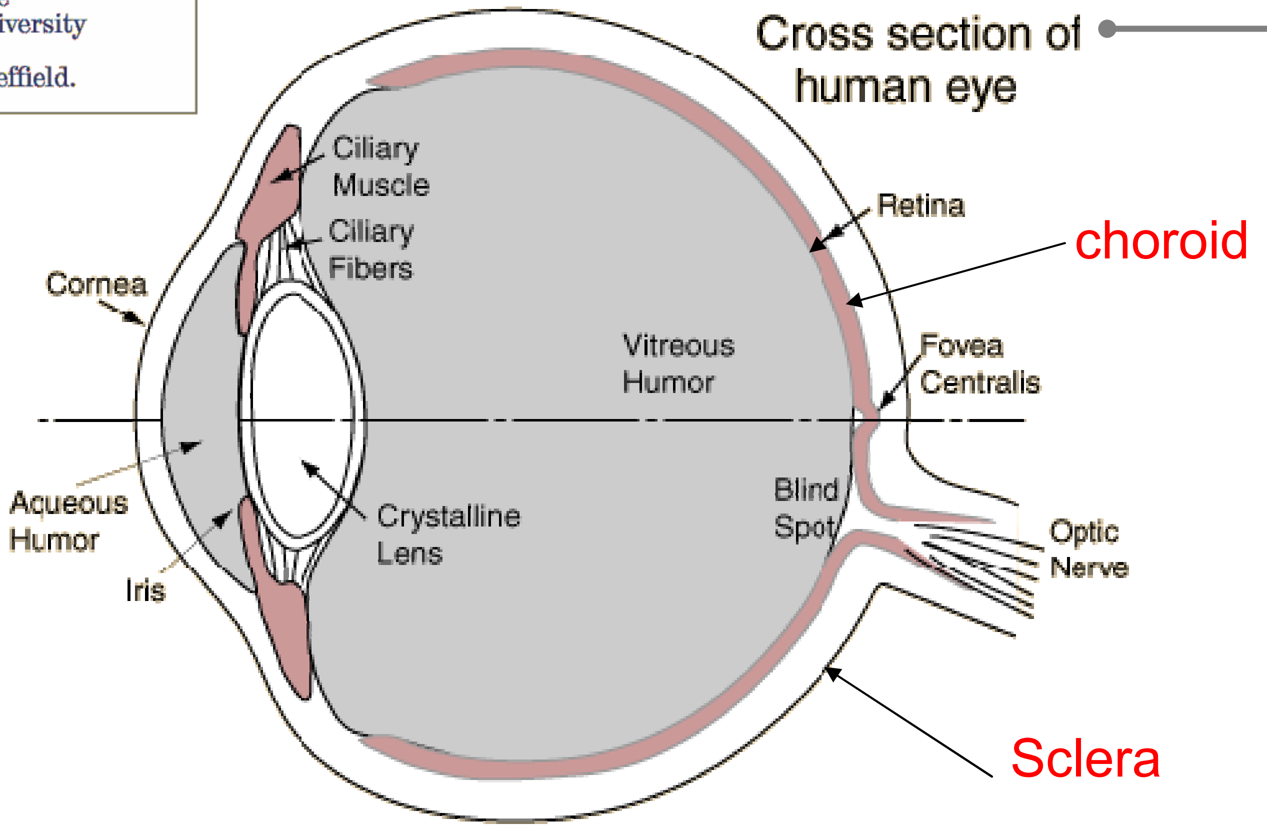
wavelength x frequency = speed of light

Visible light represents only a small portion of the Electromagnetic spectrum.

The HVS can detect the range of light spectrum from about 400 nm (violet) to about 700 nm. (red)

HVS perceives this range of light wave frequencies as a smoothly varying rainbow of colours.



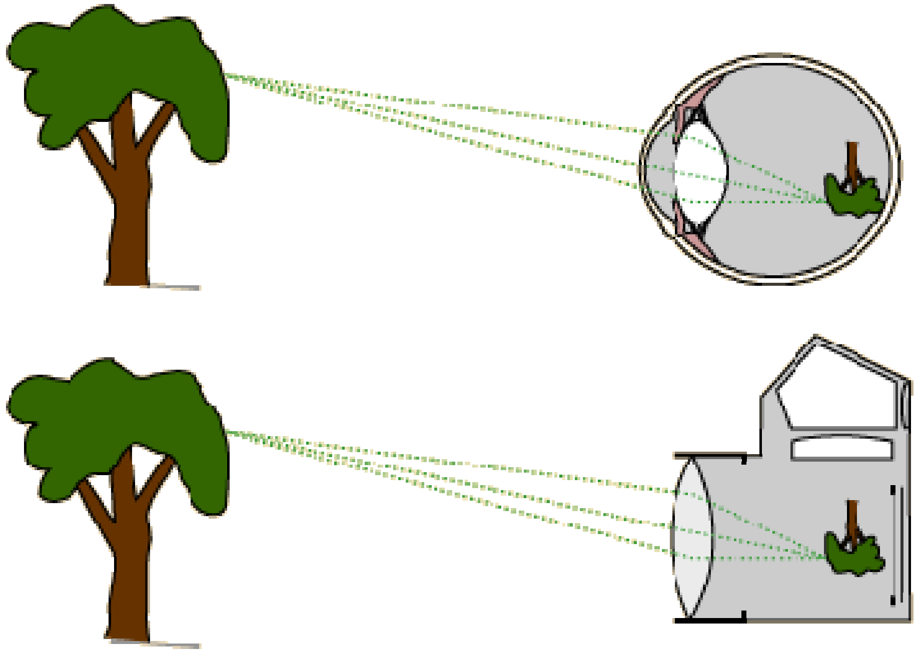


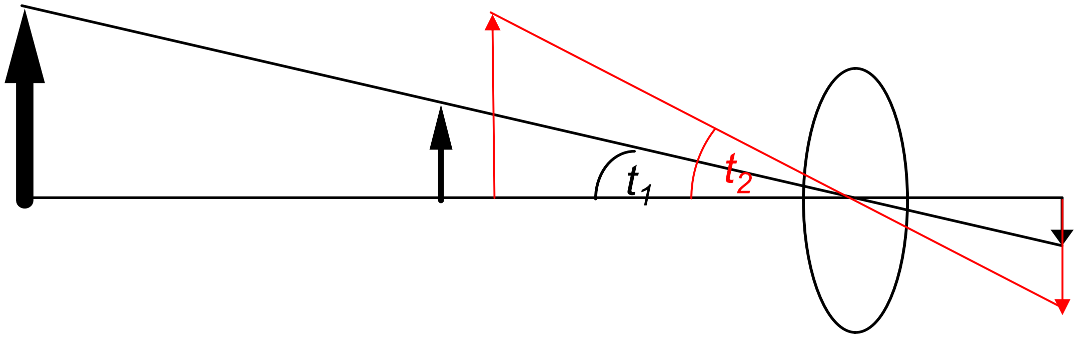
- The eyeball is about 20 mm in diameter three main membranes - the cornea and sclera outer cover, the choroid and the retina.
- The cornea is a tough, transparent tissue that covers the anterior surface. The sclera is the opaque membrane that covers the remainder of the eyeball.
- The choroid contains a network of blood vessels that provide nutrition to the eye. The choroid coat is heavily pigmented to limit extraneous light entering eye and reduce backscatter.
- Iris (pupil) can expand/dilate from 2 mm to 8 mm - front of iris contains the visible pigment of the eye while the back contains a black pigment.

Image formation on retina

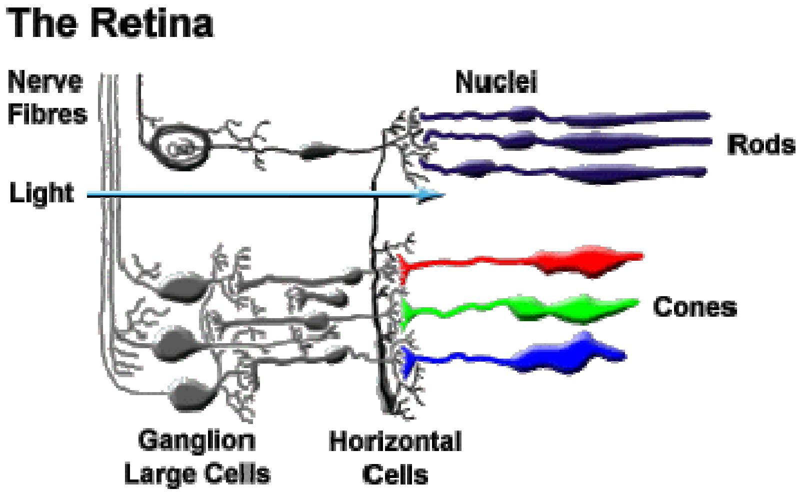
In a camera, the exposure is controlled by varying the aperture and the shutter speed

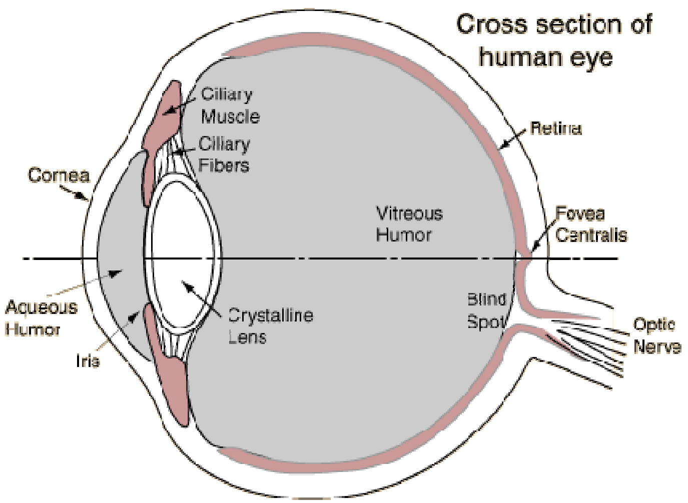
what similar mechanisms exist in the human eye?





Measure size of objects in terms of angle subtended at the eye, θ

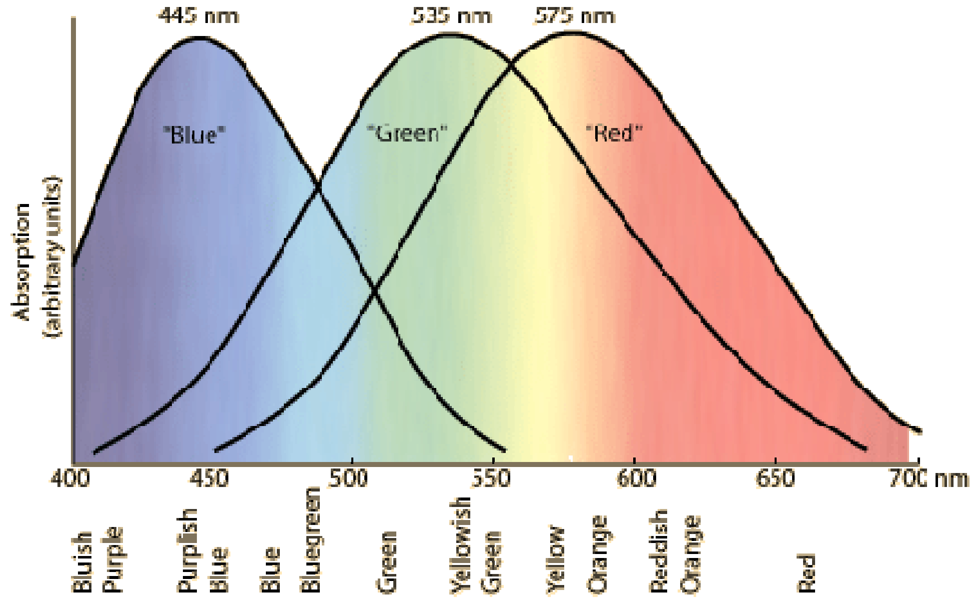


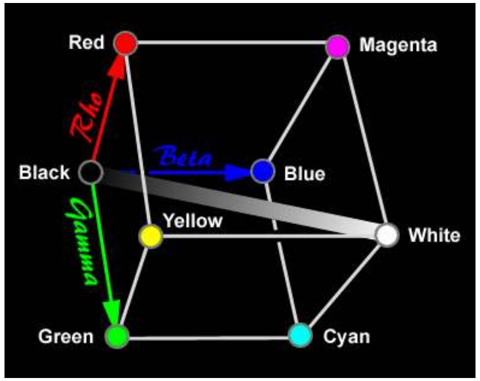


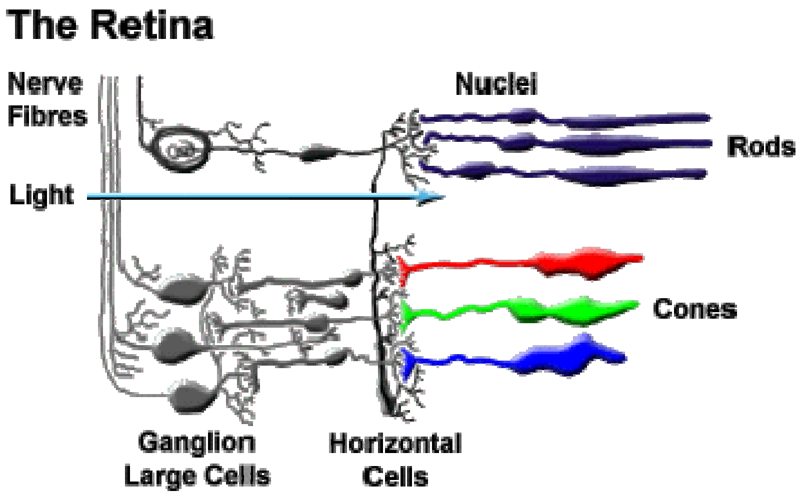
- Retina contains two types of photoreceptors, **rods** and **cones**.
- The 6 to 7 million cones provide the eye's colour sensitivity and they are much more concentrated in the central spot known as is the "**fovea centralis**", a 1.5 mm diameter rod-free area with very thin, densely packed cones.
- Humans can resolve fine details with cones, mainly because each cone is connected to its own nerve-end.
- Cone vision is called **photopic vision** or bright-light vision.

Colour sensitivity of cones

- The human eye has three sets of sensors with peak sensitivities at light frequencies *that we call*
  - o red (575 nm), 65% of cones
  - o green (535 nm) 33% of cones
  - o blue (445 nm). 2% of cones
- Light at any wavelength in the visual spectrum range from 400 to 700 nanometres will excite one or more of these three types of sensors.
- Our perception of which colour we are seeing is determined by which combination of sensors are excited and by how much.
- Using this, visible colour can be mapped in terms of three numbers called tri-stimulus values.

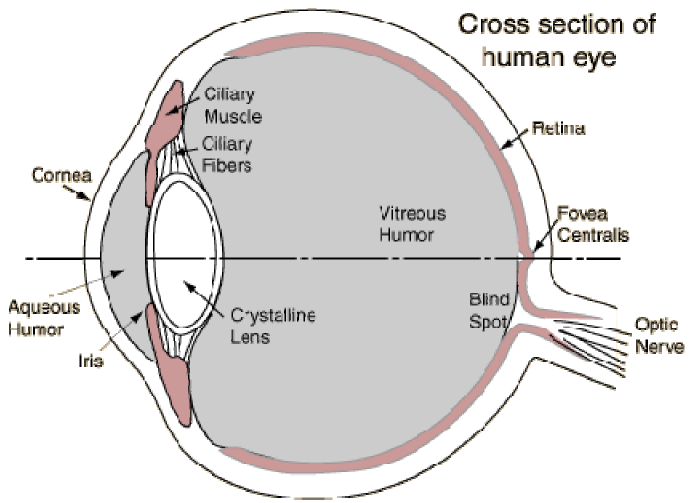






- The number of **rods** is much larger
- 75-150 million distributed over the retinal surface.

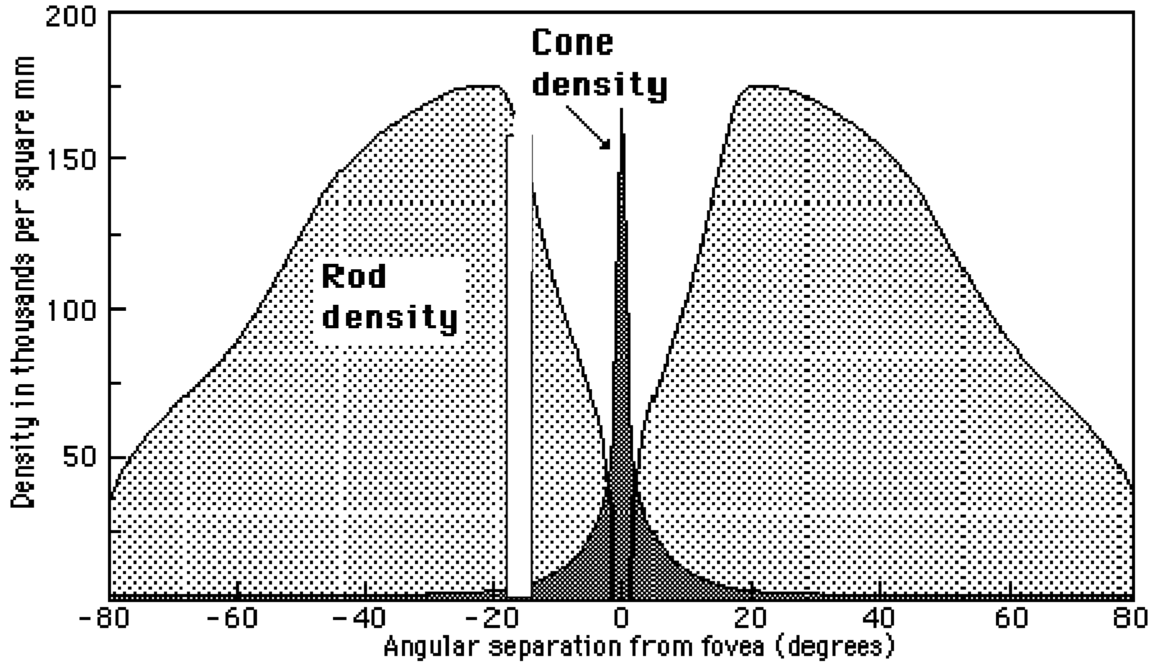
- Several rods are connected to one nerve-end.
- Therefore, limits the amount of details perceived by rod vision.

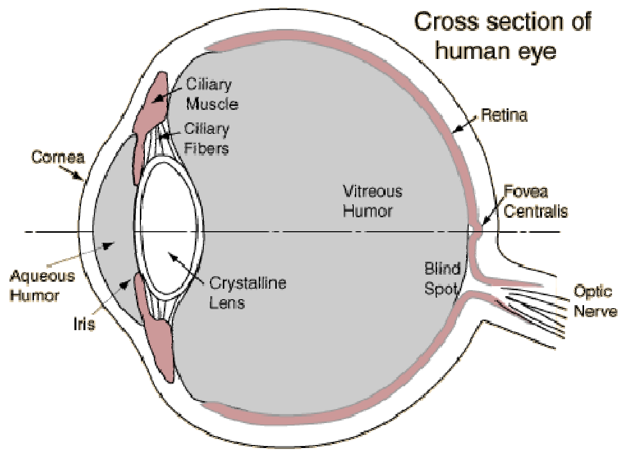


- They are not involved in colour vision and are sensitive to low levels of illumination.
- This is called **scotopic vision** or dim-light vision.

The University Of Sheffield.

Rods and cones density

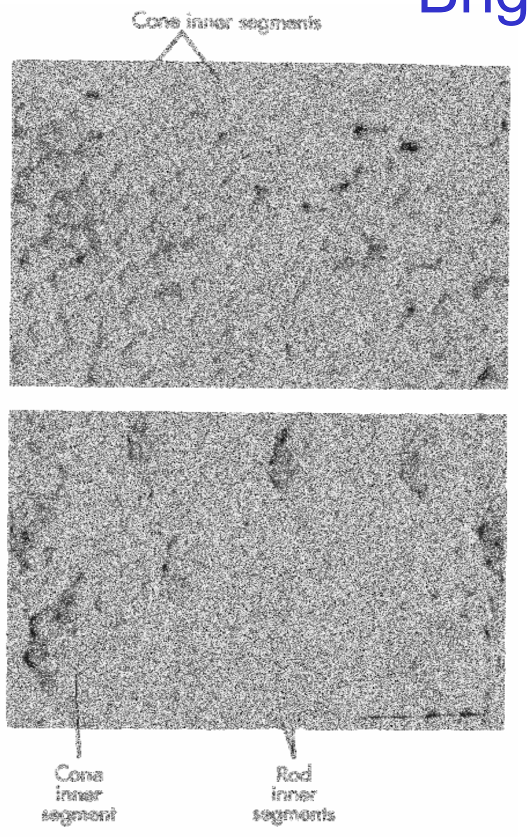




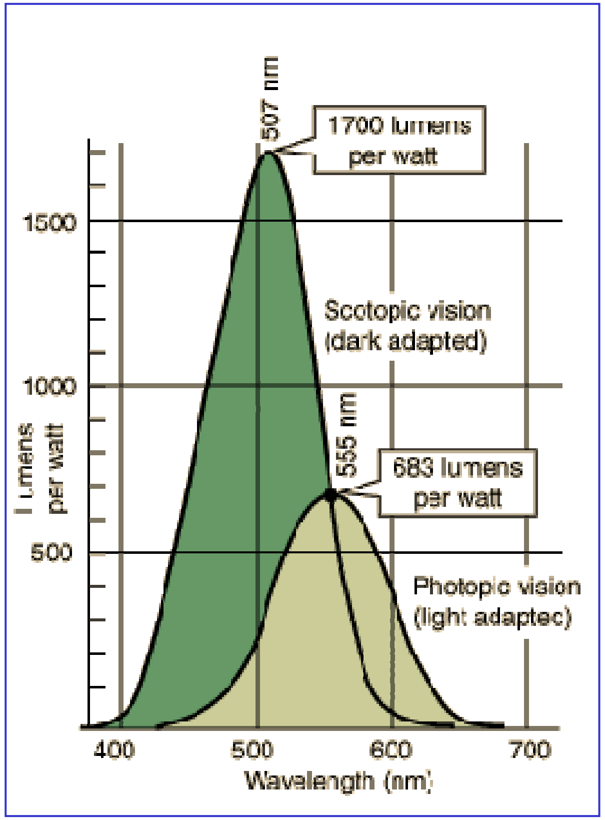
- Density of rods and cones for a cross section of right eye shown in the figure.
- The nerves to the rods and cones exit the eye through the optic nerve - gives rise to the **Blind Spot.**

- The green and red cones are concentrated in the fovea centralis. The "blue" cones have the highest sensitivity and are mostly found outside the fovea, leading to some distinctions in the eye's blue perception.

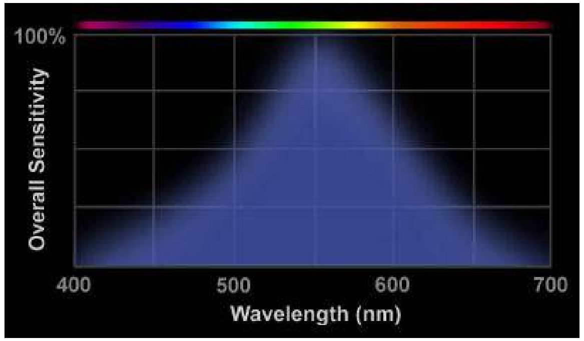
Rods and Cones: Brightness adaptation



Spatial arrangement



- The HVS has much greater sensitivity in low ambient illumination.
- The cones contribute little or no sensitivity under this condition.
- Imaging is primarily accomplished by the rods when illumination levels are very low.



Spectral sensitivity of the rods

The scotopic vision is primarily rod vision, and the photopic vision includes the cones.

Brightness Adaptation

The human eye's ability to discriminate between different intensity levels is an important consideration for HVS models.
Brightness is the intensity as perceived by the HVS.
Although, the range of light intensity levels to which the HVS can adapt is huge, it cannot operate on such a large range simultaneously.

It accomplishes this large variation by changes in its overall sensitivity. This is called brightness adaptation.

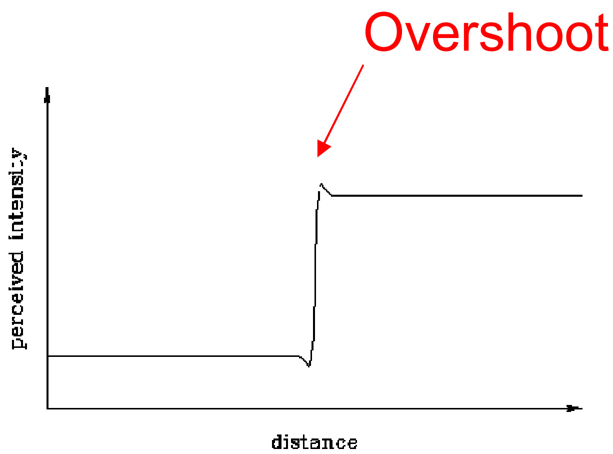
The daylight vision (cone vision) adapts much more rapidly to changing light levels.
E.g., adjusting to a change like coming indoors out of sunlight in a few seconds.
E.g., Can look into a flash light in daytime, but not in the night

The perceived brightness is not a simple function of intensity. This can be explained using 2 phenomena.

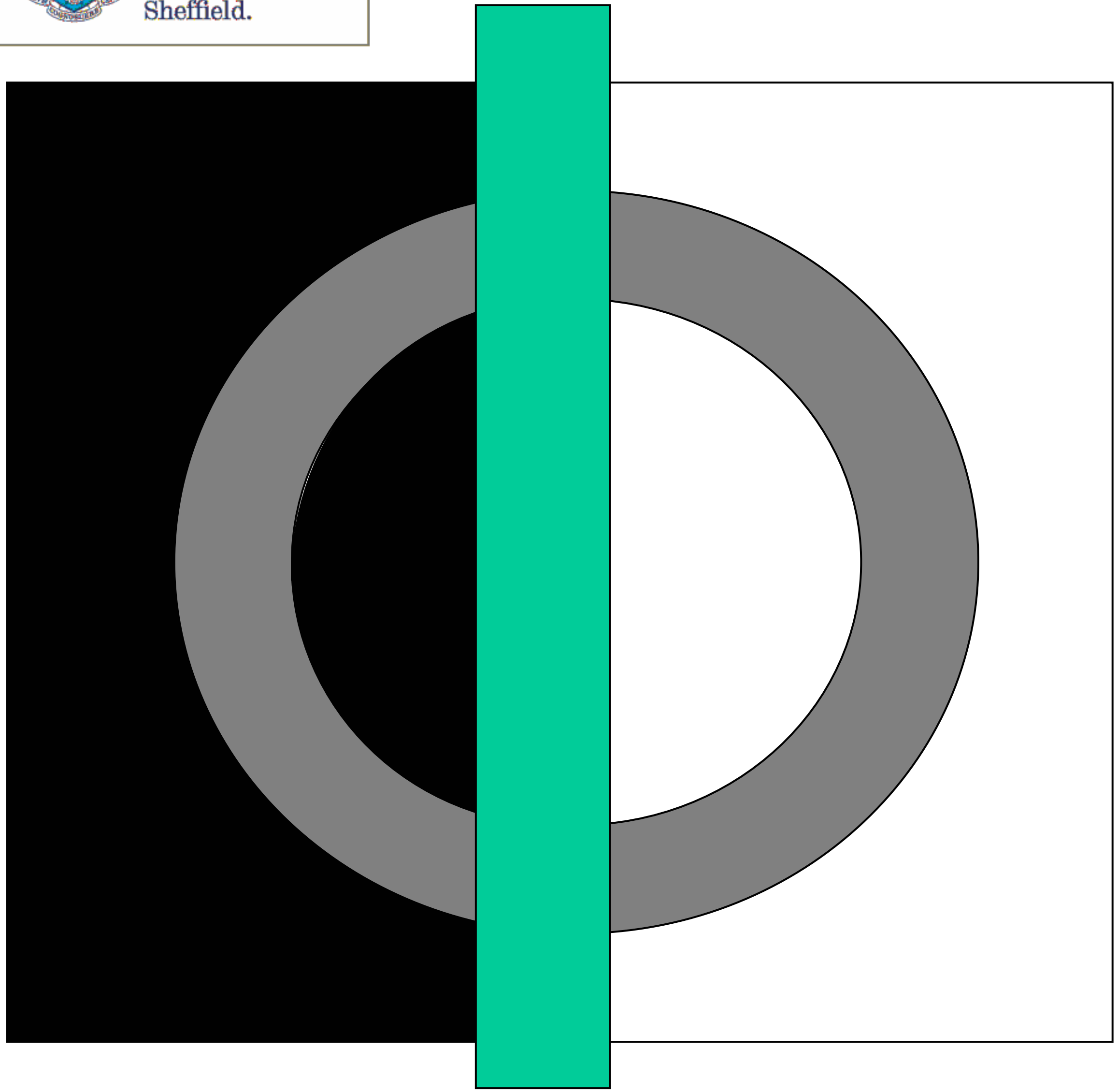
1. Mach bands
2. Contrast sensitivity

Mach band effect

Although the intensity of the blocks is constant, the HVS tends to undershoot or overshoot around the boundaries of regions of different intensities.



Contrast Sensitivity



The perceived brightness depends on the contrast with respect to the background.

Compare the shade of the ring before and after placing the solid bar

Other examples of HVS perception phenomena are optical illusions.

-Eye fills in non-existing information
-Wrongly perceives geometrical properties of objects

-A few examples to follow.

Optical illusions is one characteristic of the HVS that is not fully understood yet

☹

http://www.illusionworks.com/html/hall_of_illusions.html

Factors Affecting Visibility

Luminance - proportion of incident light reflected into the eye. A measure of the density of the luminance intensity.

Contrast - relationship between the luminance of an object and the luminance of the background.

Size - The larger an object, the easier it is to see.

Time - There is a time lag in the photochemical processes of the retina, therefore the time available for viewing is important. When objects are briefly viewed we need bright light, when lots of time is available even small details can be seen.

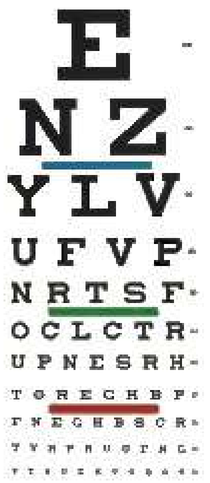
Colour - not really a factor by itself, but related both to contrast and luminance factors.

Acuity

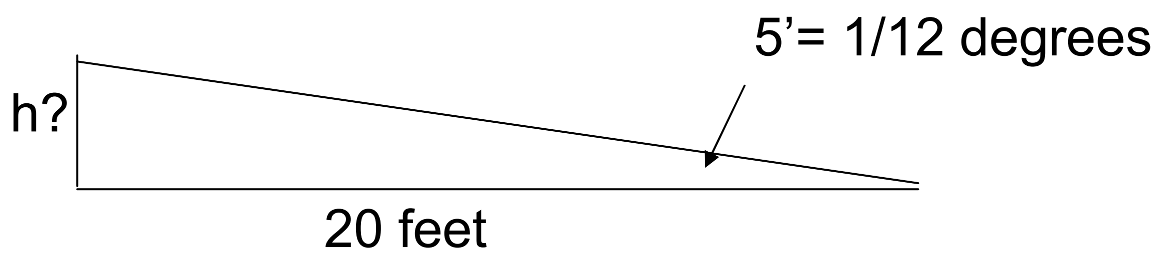
<u>Recognition acuity</u>
Ability to identify objects. Not a suitable engineering measure

<u>Detection acuity</u>
Ability to perceive that an object is present (Limit is about 10 photons)

<u>Resolution (or visual) acuity</u>
Ability to resolve fine details (e.g., Snellen chart) –clarity or clearness of one's vision



20/20 visual acuity



HVS Model

The HVS model addresses 3 main sensitivity variations as a function of

- light level
- spatial frequency
- signal content

Sensitivity $s = \frac{1}{C}$

C = Contrast

Contrast $C = \frac{L_{\max} - L_{\min}}{L_{mean}}$

L = Luminance

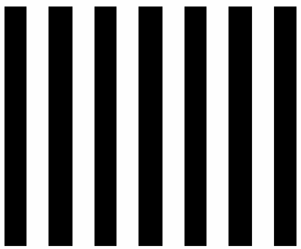
Variable	**Function**	**The modelled component in the eye**
Light level	Amplitude nonlinearity	Retina
Spatial frequency	Contrast Sensitivity Function (CSF)	Optics of the eye
Signal content	Masking	Neural Circuitry

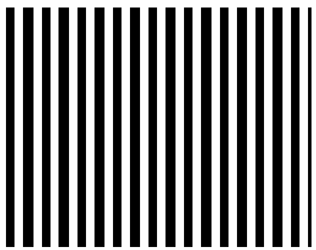
Modulation Transfer Function (mtf)

- Most optical devices (such as lenses), including the HVS are not perfect.
- Consequently, when light (or visual stimulus) is passed through, it can be degraded.
- Modulation transfer function is used as a measure for such degradation of a system.
- Just as we can characterise electronic systems by considering their amplitude frequency response, a similar approach can be employed for imaging systems (including the optics, detector, electronics and display).

- We consider a grating with dark and light bars. We can measure the amount of light coming from each of the bars. The maximum amount of light will come from the light bars and the minimum from the dark bars.

Grating with low spatial frequency

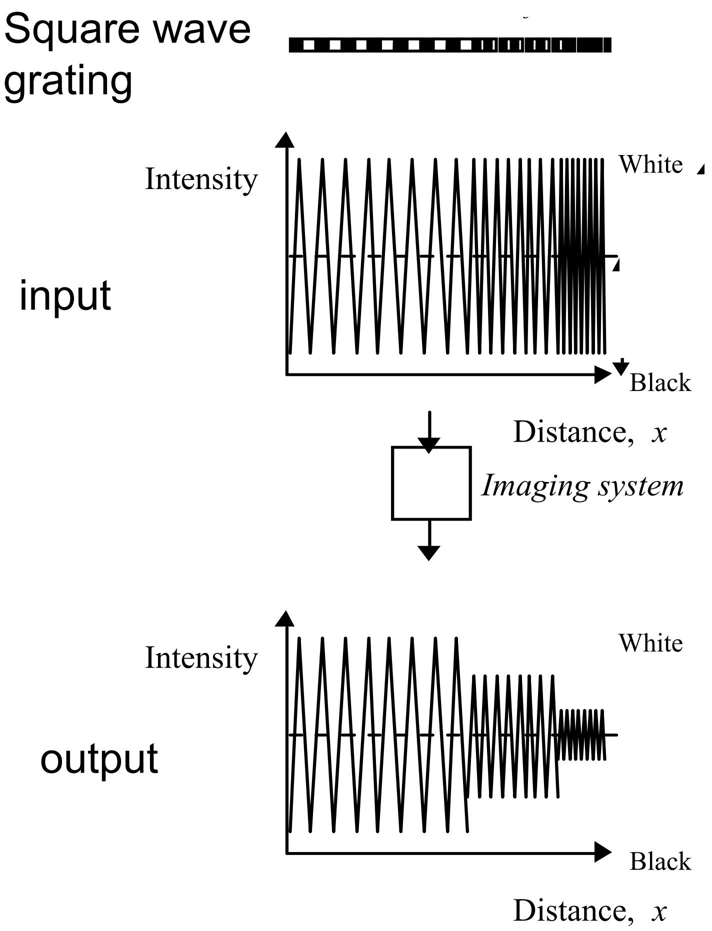




Grating with high spatial frequency

The University Of Sheffield.

Modulation Transfer Function

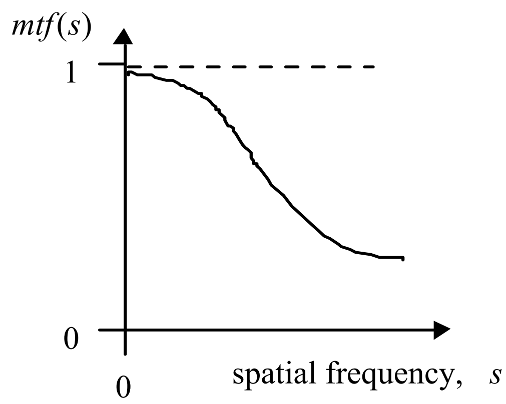


If the light is measured in terms of luminance (L) we can define modulation M for spatial frequency s according to the following equation:

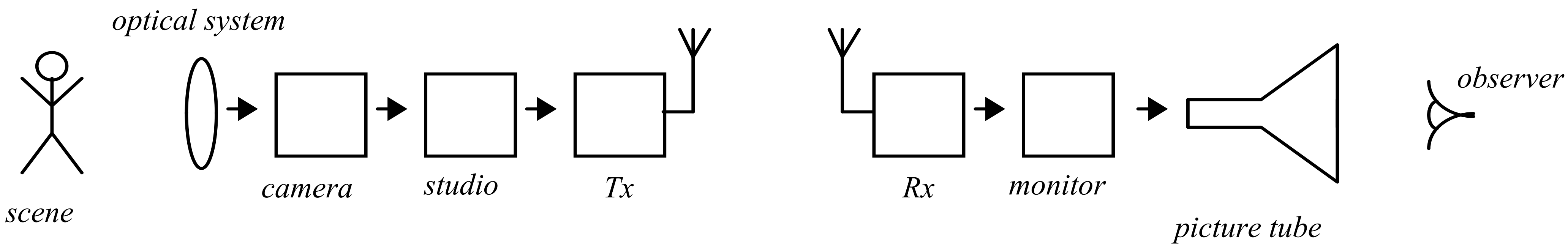
$$M(s) = (L_{max} - L_{min}) / (L_{max} + L_{min})$$

L_{max} and L_{min} are maximum and minimum luminance levels, respectively.

$$\text{Modulation transfer function,} \quad mtf(s) = \frac{M(s)_{out}}{M(s)_{in}}$$



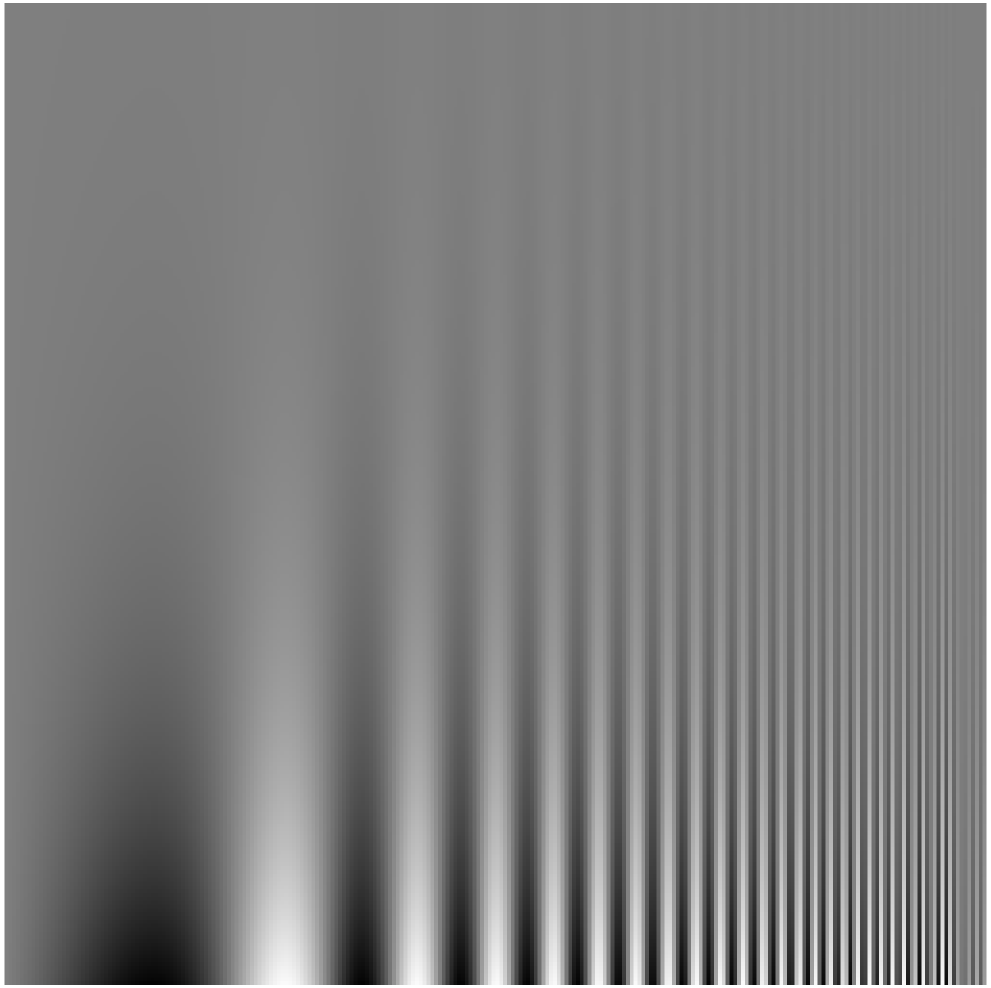
Time domain	*Spatial domain*
temporal frequency	spatial frequency
impulse function	point spread function
amplitude response	modulation transfer function



$$mtf(s)_{system} = mtf(s)_{optics} \text{ x } mtf(s)_{camera} \text{ x } \ldots\ldots\ldots \text{x } mtf(s)_{pic\ tube}$$

Hence we can calculate the overall mtf of any system

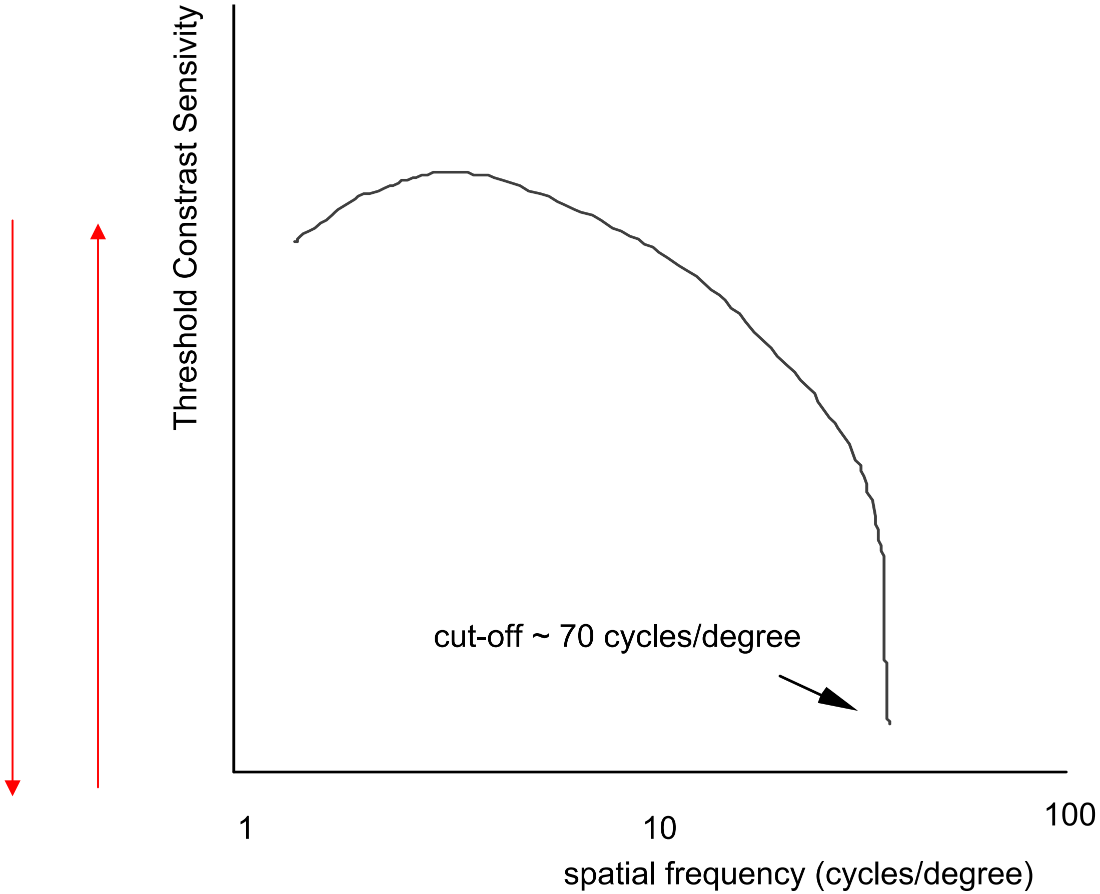
Contrast Threshold Sensitivity Curve

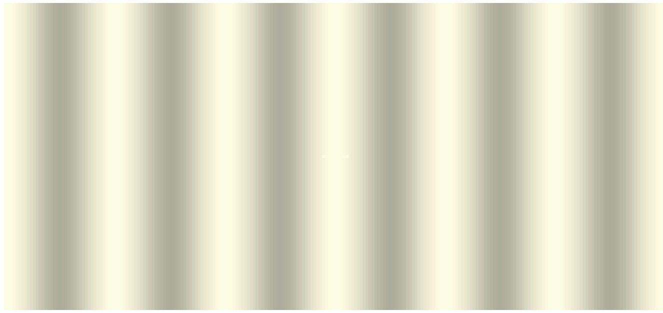


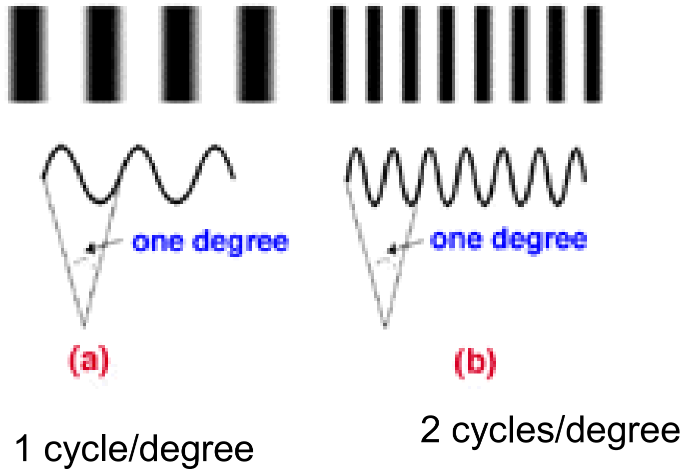
We can measure the MTF to determine the contrast threshold sensitivity.

We call the MTF to model the optics of eye as the Contrast threshold Sensitivity Function (CSF)

Trace the limit of the visibility of the vertical “lines”







Same in both vertical and horizontal spatial directions

Why does the curve peak at about 6 cycles/degree?

The University Of Sheffield.

Temporal sensitivity

Temporal sensitivity refers to our response to timing of visual signals.

Consider a light that is pulsing, such as the blinking cursor on a computer screen. It is easy to discern that the light is blinking because of the low frequency, or rate at which it flashes. Imagine gradually increasing the frequency. As the light blinks faster and faster, it would eventually reach a rate where it would no longer be possible to detect that the light is flashing, it would look like a "solid", or continuous light. We say that our visual system has fused the flicker. The frequency at which that occurs is called the **critical flicker frequency (CFF)**

Chief among the physical factors that determine the rate at which flicker is fused is the location of the stimulus within the visual field. The peripheral receptor fields (comprised of primarily rods) behave quite differently with respect to flicker than do the foveal cones.

A flickering image on the fovea can be detected even if it blinks extremely rapidly, and is only fused when it reaches a rate of **50 to 70 Hz**, depending on adaptation level.

The HVS Model - Summary

The human eye can be considered as a three-dimensional low-pass filter.

Vertical spatial bandwidth – 60-70 cycles/degree
Horizontal spatial bandwidth - 60-70 cycles/degree
Temporal bandwidth - 50-70 Hz

- What resolution is needed for tv or computer displays to match the visual acuity of humans?
- You will need to discover some additional information to answer these questions.

Note

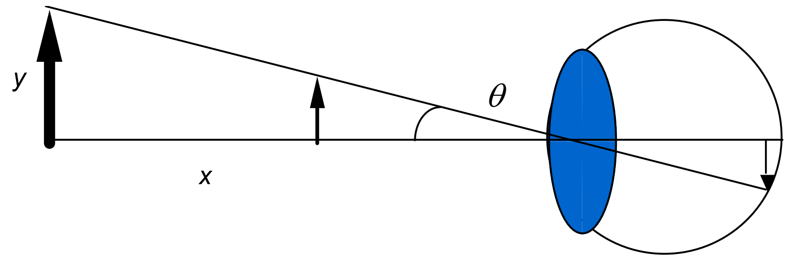
Conventional tv systems have the following specification - how does this fit in with your estimates?

Number of active tv lines = 575
Aspect ratio (picture width:picture height) = 4:3
Frames per second = 25; fields per second = 50

From HVS model studies,

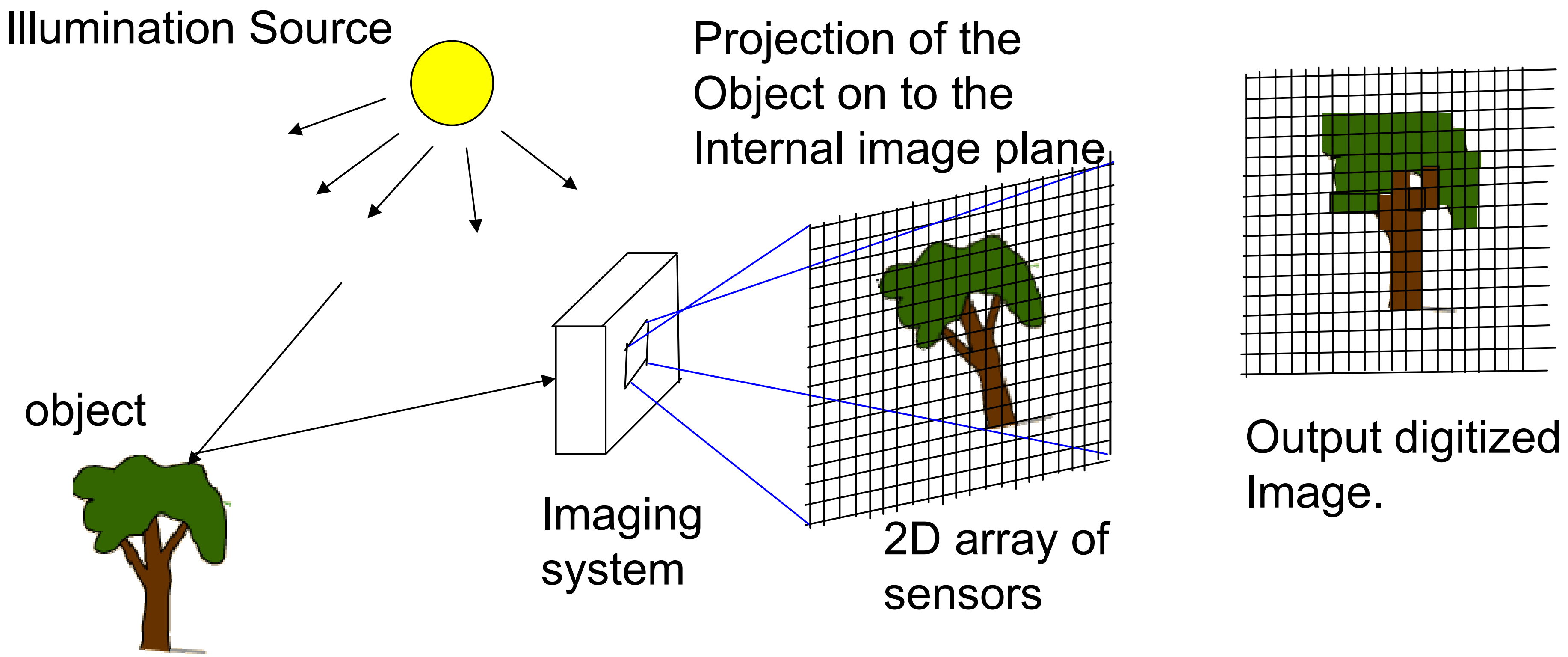
The human eye can be considered as a three-dimensional low-pass filter.

- From the Contrast sensitivity CSF:
  - 1.
  - 2.
- From the Critical flicker frequency (CFF):
  - 1.?
- From the visual acuity:
  - The angle subtended at eye



$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \approx \frac{y}{x}\ if \left(\frac{y}{x}\right)^2 << 1$$

Image Acquisition



- The image acquisition is performed by a 2D array of sensors. This array is coincident with the focal plane of the imaging system.
- At each sensor the incoming illumination energy is transformed into a voltage.
- The output voltage waveform is the response of sensors.
- This waveform is digitized to obtain the digital image.
- What are the two processes involved in digitizing?

Image Acquisition

- We denote images by 2D functions:
  - f(x,y) - This means amplitude of f at coordinates x and y.
- When an image is generated from a physical process, its values are proportional to energy radiated by the physical source.
- Therefore we can say f(x,y) must be non-zero and finite.
  - 0 < f(x,y) < infinity

- f(x,y) is characterized by two components:
  - 1. The amount of source illumination incident on the object – illumination i(x,y)
  - 2. The amount of illumination reflected by the object – reflectance r(x,y)

- Now we can define f(x,y)=i(x,y)r(x,y)
- Usually, 0 < i(x,y) < infinity (theoretical values)
- 0 <= r(x,y) <=1 0 is total absorption 1 is total reflectance. r(x,y) is a characteristic of the imaged object.
- E.g., black velvet r=0.01 stainless steel r=0.65 snow r=0.93

Monochrome images

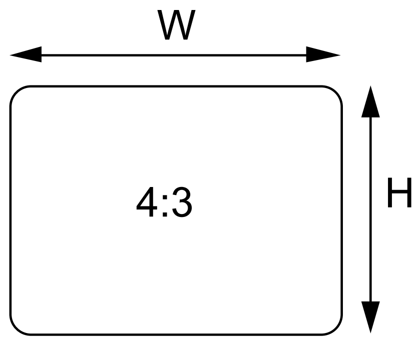
- The intensity of a monochrome image at any coordinate (x.y) is called the gray level (l)
  - we can write $l=f(x,y)$
- We can define the lower and upper bounds for l.
  - $L_{min} <= l <= L_{max}$
  - where $L_{min}=i_{min}r_{min}$ and $L_{max}=i_{max}r_{max}$
  - Typical values L_{min}~10 and L_{max}~1000
- The interval [L_{min} , L_{max}] is called the gray scale.
- It is common to re-map this scale to [0, L-1]
- 0 corresponds to black
- L-1 corresponds to white
- What about all intermediate values in the scale?

Image Representation

- To generate digital images from the sensed data (voltage waveform output)
  - We need to convert the continuous signal into digital form.
- Remember digitisation of data?
- Two processes:
  - 1. ?
  - 2. ?
- In a 2D array of sensors, the spatial sampling is determined by the arrangement of sensors. (we will discuss this later)
- If there is no motion, the output of each sensor is quantised and gray level values are mapped into discrete levels.
- That means we can represent a digital image as a matrix. (An example of image using M x N array of sensors.

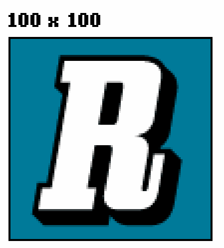
$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}$$

What is image resolution?



- Describes how detail the image is.
- Resolution types in video:
  - Spatial resolution
    - Usually expressed in terms of “pixels”
    - Pixels = Picture-elements (pels)
      - A tiny *dot* that makes up the representation of a picture in a computer's memory.
      - Expressed as W x H
      - Aspect ratio W : H = 4:3 (for standard display) 16:9 for wide screen display
  - Colour depth resolution
    - Defines how many different colours per pixel (C)
    - Usually expressed as how many bits per pixel (N)
  - Temporal resolution
    - Frame rate (frames per second)

Spatial Resolution



The spatial resolution of an image is determined by sampling.

Spatial resolution defines the smallest details that can be discernible.

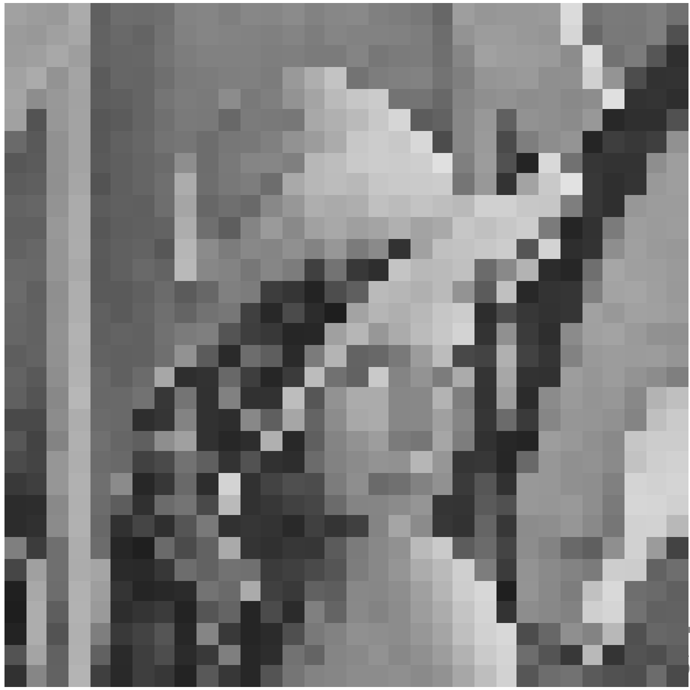
Reduction of spatial resolution (although data size can be reduced) results in loss of details.

256x256

128x128

64x64

32x32



Colour Depth Resolution

- We will first consider gray level resolution:
- Gray-level resolution = the smallest discernible change in gray level.
- We usually consider the number of discrete gray levels as an integer power of 2.
- (Remember this is the quantisation in the A to D conversion)
- Using N bits we can represent 2^N gray levels.
- We have gray levels ranging from 0 to L-1 where L-1 ~ 1000.
- Usually we use 8 to 10 bits per pixel for image representation.
- (Think why using 8 bits is sufficient. $2^{10}=1024$)

- How can you compute the total number of bits required to represent an N-bit image of resolution P x Q.?

- What happens when colour bit-depth resolution is reduced?

Gray-level Resolution example
Lena and Mandrill images

256 levels
8 bits

16 levels
4 bits

4 levels
2 bits

2 levels
1 bit

Gray-level Resolution example
Lena and Mandrill images

- Consider the 8-bit representation:
  - Which regions contain low spatial frequency components?
  - Which regions contain high spatial frequency components?
- Now consider low bit representations:
  - How are the high spatial frequency regions affected?
  - How are the low spatial frequency regions affected?

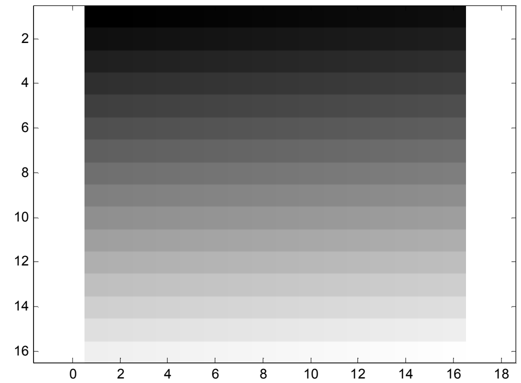
The University Of Sheffield.

Colour Depth Resolution

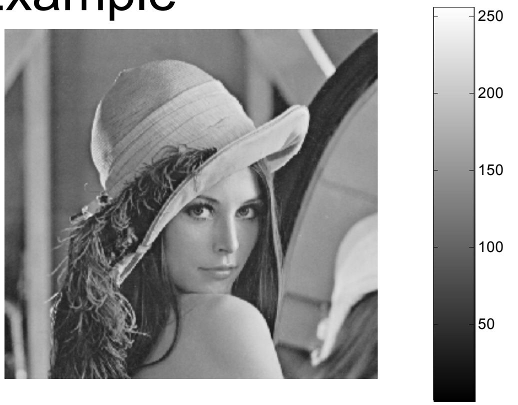
For gray scale: 1 byte per pixel = 8 bits per pixel (bpp)
This can represent $2^8 = 256$ different gray levels.

Numbers 0 – 255 are mapped to different gray shades 0= black 255=white

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95
96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255



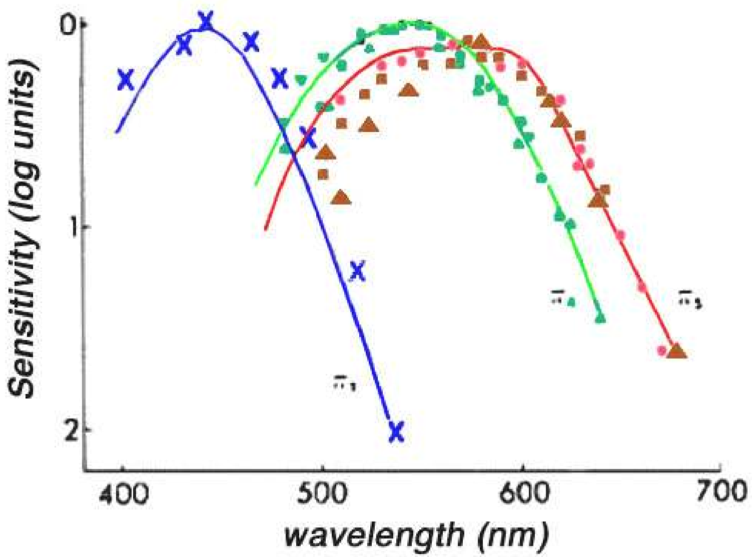
An Example



Digital images are represented as an matrix.
Manipulating numbers → image processing

Colour depth Resolution

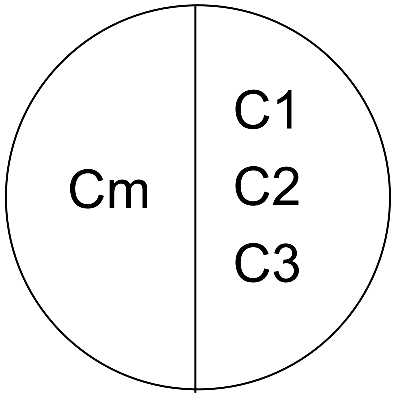
Tristimulus Theory of Colour



We possess three different colour-senstive cones.
Any colour is a combination of the output of these different cones

A colour(Cm) = C1(blue) + C2(green) + C3(red)

Grassman's Law



Match to Cm by varying C1, etc.

Additive Colour Mixing - mixing coloured lights

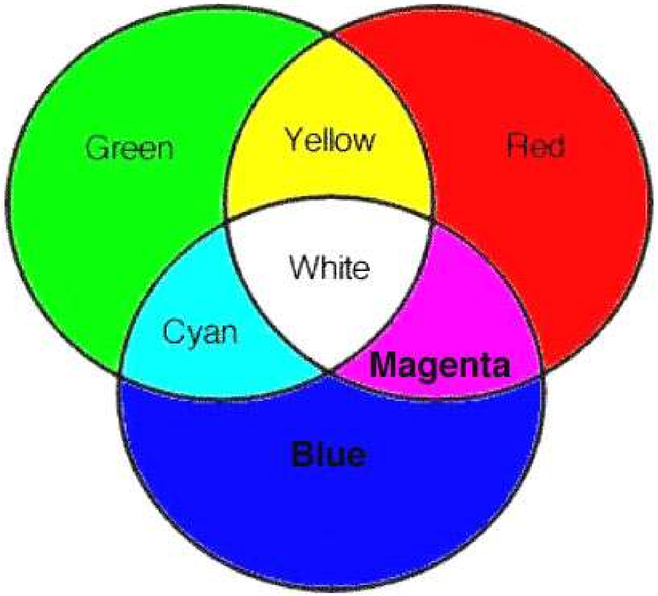


Figure 3. Additive colour mixtures of blue, green and red to produce cyan, magenta, yellow and white.

Display Systems

Class Exercise

- What resolution is needed for TV or computer displays to match the visual acuity of humans?

Note

Conventional TV systems have the following specification:

Number of active tv lines = 575
Aspect ratio (picture width:picture height) = 4:3
Frames per second = 25;
fields per second = 50

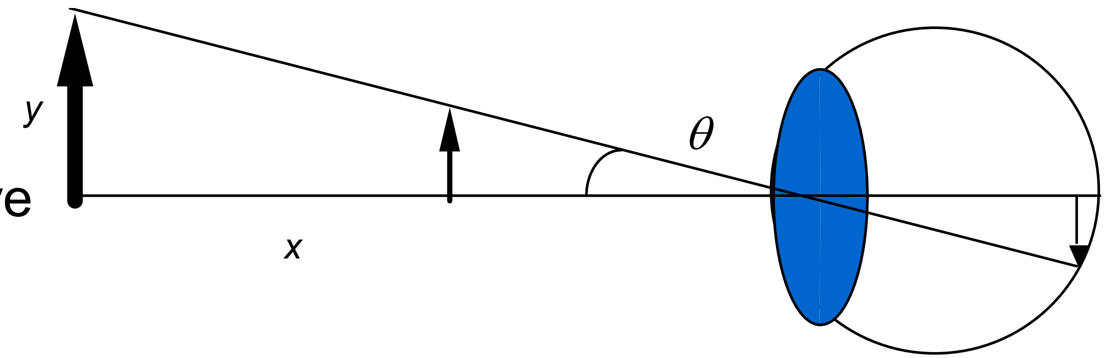
How does this fit in with your estimates?

What do we know from the HVS model?

- From the CSF:
  - 1. ? horizontal cut-off ~60 cycles/degree
  - 2. ? Vertical cut-off ~60 cycles/degree

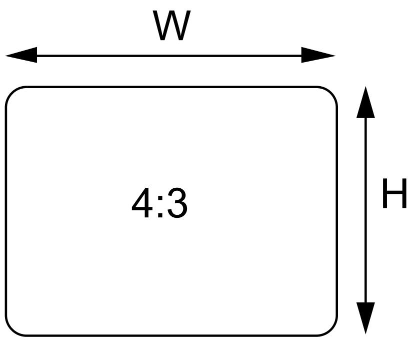
- From the CFF:
  - 1.? Temporal bandwidth 70Hz

- From the visual acuity:
  - The angle subtended at eye



$$\theta = \tan^{-1}\left(\frac{y}{x}\right) \approx \frac{y}{x} \;\; if \left(\frac{y}{x}\right)^2 << 1$$

We also know



The aspect ratio (W:H) = 4:3

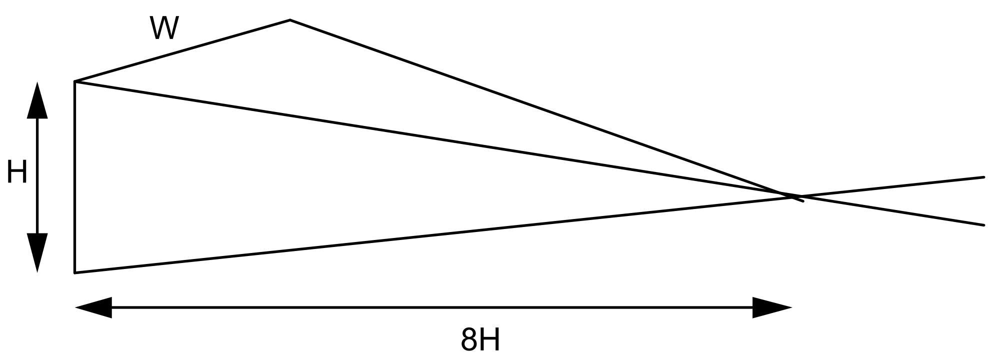
Common to most video standards

Wide screen tv = 16:9

How far do individuals sit in front of their monitor?

Experiments suggest = (8 to 12) x (picture height)

In this problem, we assume individuals sit at a distance of 8 times the height of the monitor



$$\theta_H = \qquad\qquad =$$

$$\theta_W = \qquad\qquad =$$

$$\theta_H = 7^\circ$$

$$\theta_W = 10^\circ$$

With a spatial frequency cut-off (H and W) = ~60 cycles/degree
Now we can compute the number of cycles we the eye can resolve in both directions

No. of cycles/picture height =
No. of cycles/picture width =

Consider picture composed of discrete elements
Picture elements = pixels = pels.
Now we need to convert the above measurement into pixels

Nyquist sampling limit would suggest
2 samples per cycle,

hence
No. of pixels/picture height =
No. of pixels/picture width =

How many pixels?

Though we can register scenes over about 1000 orders of magnitude of illumination We can only perceive about 100-150 distinct grey levels at one time.

How many bits per pixel?

Hence bytes per image (say 1 Mbyte)

We are much more sensitive to colour variations – being able to perceive over 50,000 separate colours (under the optimum lighting conditions)

How many bits per pixel?

Allow 1 byte per primary colour (24 bit colour)
gives 16.8 million colour palette
Now ~3 M byte per image
What is the capacity of a floppy disk?

If the temporal cut-off frequency of human is ~ 70 Hz, then need 140 images per second (remember Nyquist) to give impression of smooth moving images.

This means the bandwidth requirements for Monochrome and colour television are

Bandwidth (monochrome)
(colour).

The bandwidth of conventional analogue terrestrial TV transmission is only 5 MHz. This gives only around 30 Mbps (assuming a 20 dB SNR).

What is the total file size of a 90 minute movie?
What is the capacity of a DVD?

Now you can attempt Q11 parts (i)-(iv).

How can we further reduce the bit rates?

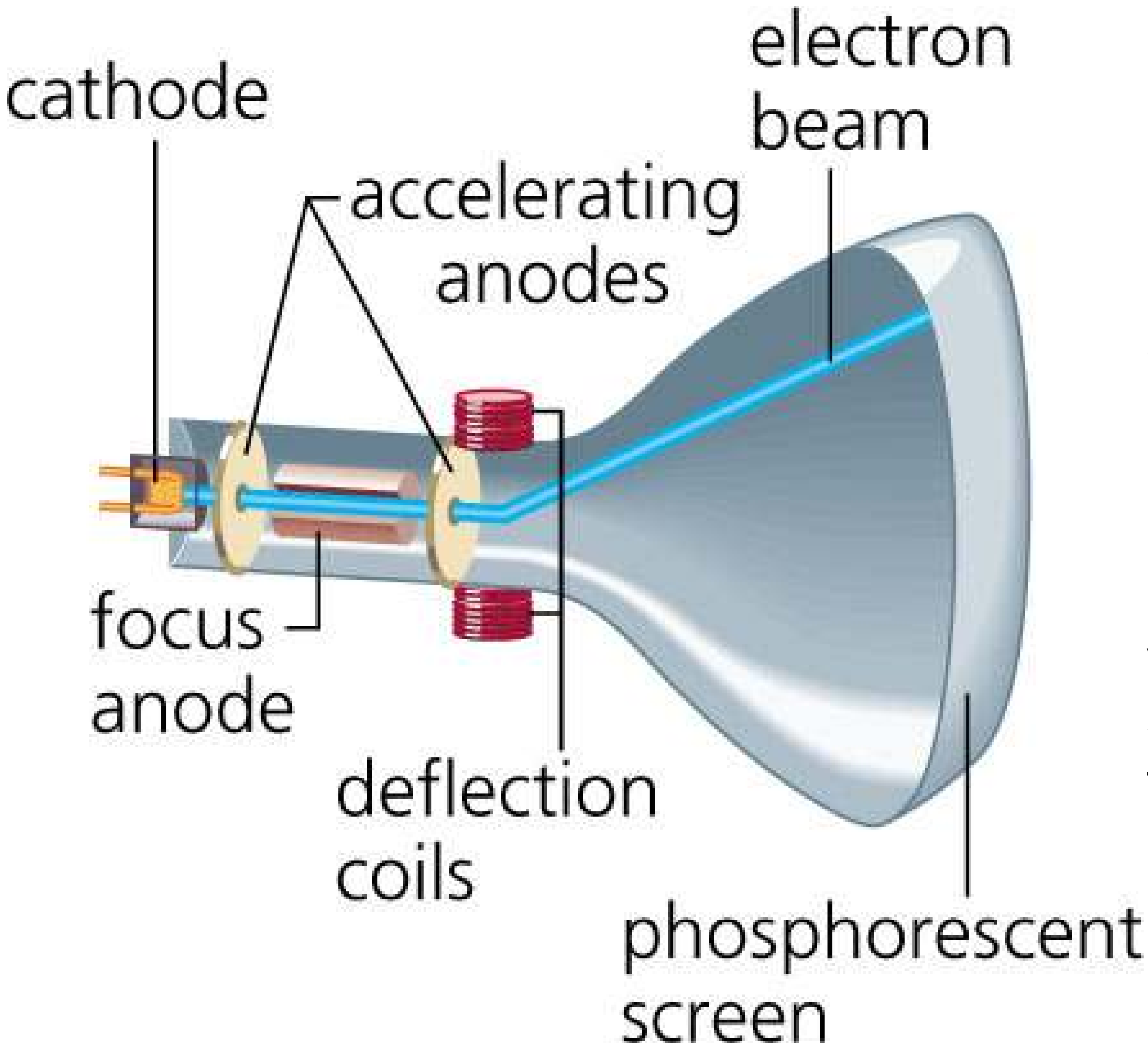
- Our design was a bit conservative. We used Nyquist limit for sampling.
- We can explore
    - the limitations of the HVS
    - the mechanisms of displaying

  to relax the Nyquist limit.

- We will try to re-modify
    - spatial resolution
    - temporal frame rate
    - colour depth resolution

Image display systems

Television Display Tube (Cathode Ray Tube)



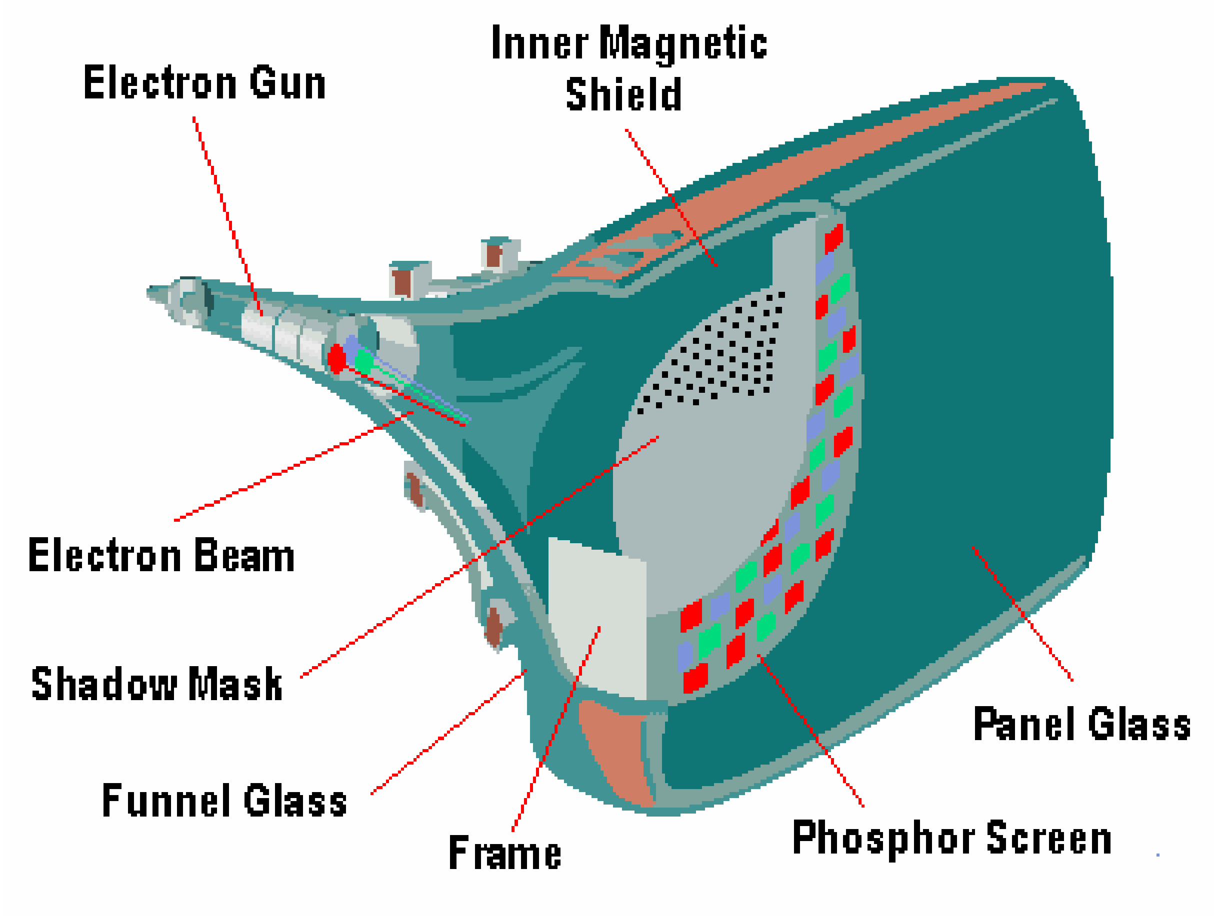
Scans horizontally across the screen

Each line is divided by the number of pixels that it must display

Modulates its intensity by the pixel values and each section is filled in by the electron beam.

Image display systems

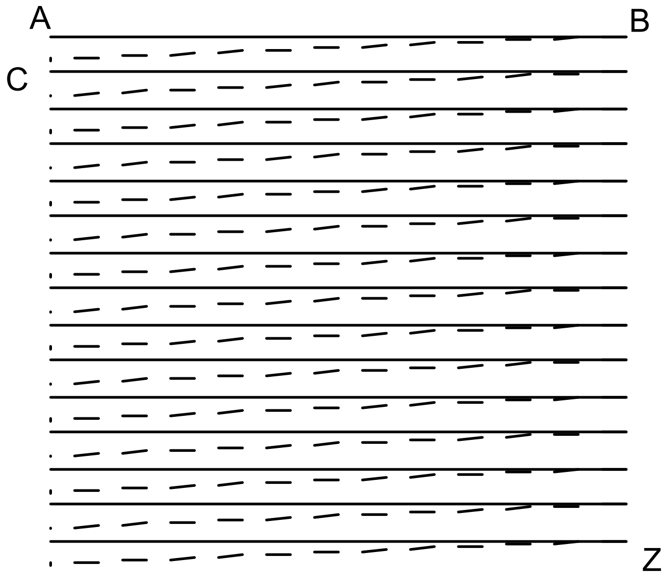
Colour TV Monitor Tube



For colour images,

The scanning electron beam process is triplicated, with a beam for each primary.

Image display systems



Computer monitor/TV display composed of single moving spot - traces out a scanning pattern.

Spot moves horizontally left to right (A to B) on screen with an invisible line fly-back to start of next line (B to C).

At end of scan (a frame) - frame fly-back to start (Z to A)

This type of scan where an image is scanned from left to right and top to bottom (as in raster scans) is called progressive scanning.

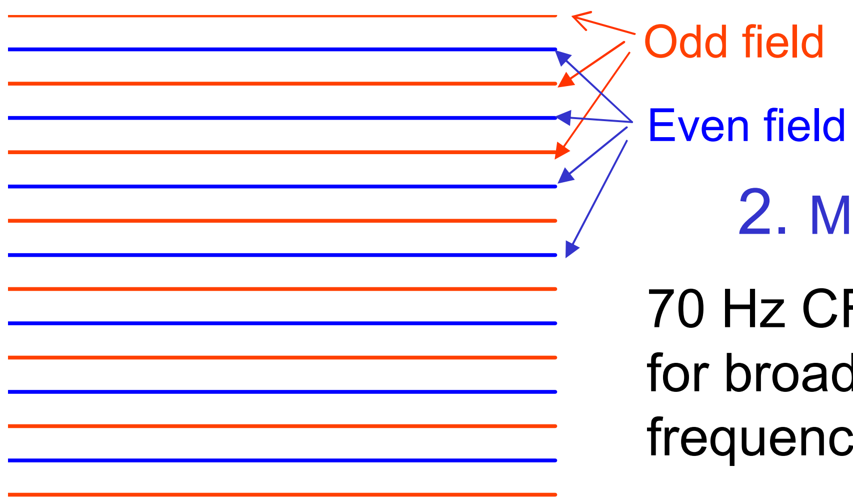
Image display systems

Temporal resolution reduction: 1. Interlaced scanning

We are less sensitive to small-area flicker;
Hence, we can display alternate scan lines as a single “field”.

The afterglow of the phosphor of CRT tubes, and the persistence of vision results in two fields being perceived as a continuous image.

This allows the viewing of full horizontal detail with half the bandwidth which would be required for a full progressive scan.



Further reduction is possible:

2. Mapping the frame rate to mains frequency

70 Hz CFF is a too conservative estimate. Historically for broadcast TV systems advantage to have field frequency at power frequency 50 Hz/60 Hz)

Hence (in UK) 50 fields/s. (25 frames/s)

Image display systems

Spatial resolution reduction – 1. Adjusting using the Kell Factor

We used the Nyquist limit to compute the resolution. That was also a very conservative estimate, when we think of how the display systems work.

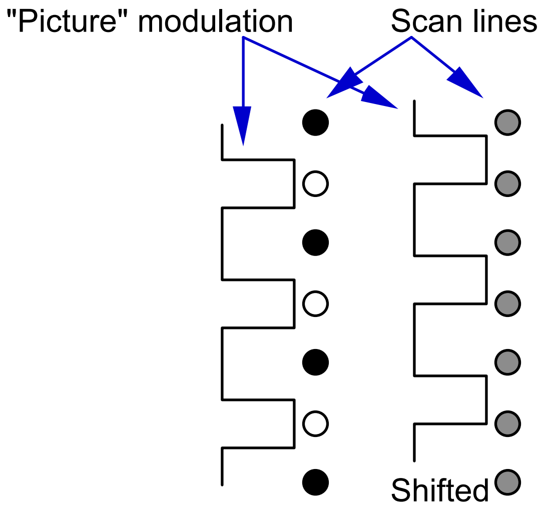
To begin with we relax the Nyquist limit. That means we assume 420 active tv lines/ picture height (as per our example)

Precise position of picture with relative to scan lines will affect visibility of picture. There are also effects due to finite width of scan lines.

The Kell factor is used to determine the effective resolution of the display. Example:
Effective resolution= Kell factor x display resolution

Display resolution = Effective resolution/ Kell factor

Experimental work suggests Kell Factor of ~0.7
Hence actual active tv lines/picture height = 420/0.7 = 600



Hence using Kell factor
We get the actual active tv lines/picture height = 420/0.7 = 600

Using the aspect ratio we get the number of pixels per line

Now what is the picture resolution?

lines/frame x pixels/line = pixels/ frame.

For colour images we need bytes/pixel.

That means, in total we need bytes/pixel x pixels/frame

i.e., bytes/frame.

Now using, mains frequency and interlacing we can have 25 frames/sec frame rate. This means a data rate of bytes/frame x frames/sec

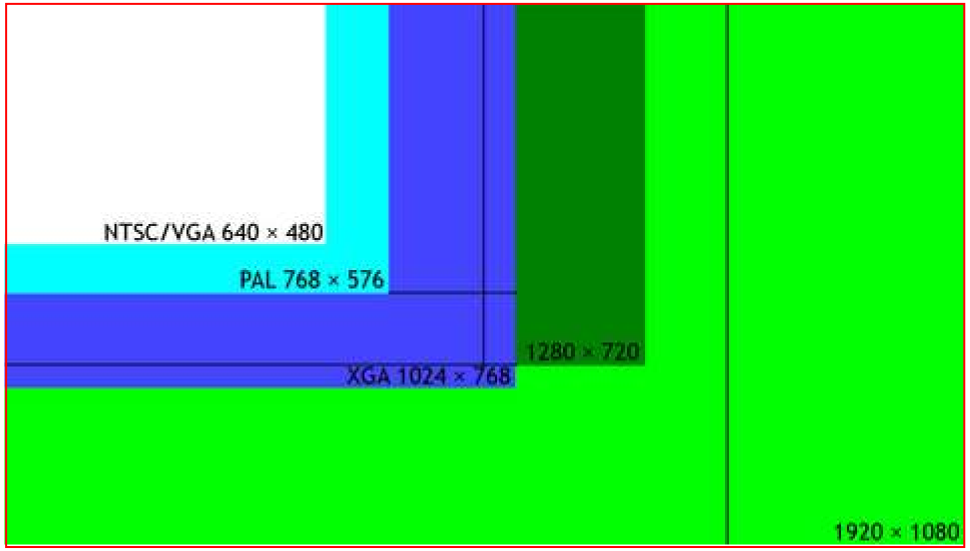
bytes/sec.

Current Spatial-temporal Resolution of display formats

QQCIF	90 x 72	Thumbnail videos for browsing
QCIF	180 x 144	Mobile phone displays
CIF	360 x 288	VCD resolution
4CIF	720 x 576	DVD resolution
SDTV	576 x788 (576i50 (PAL, SÉCAM), 576p25)	
EDTV	576p50, 720i50, 720p24, 720p25	
HDTV	720p50, 1080p24, 1080p25, 1080i50	

Compute Aspect Ratio

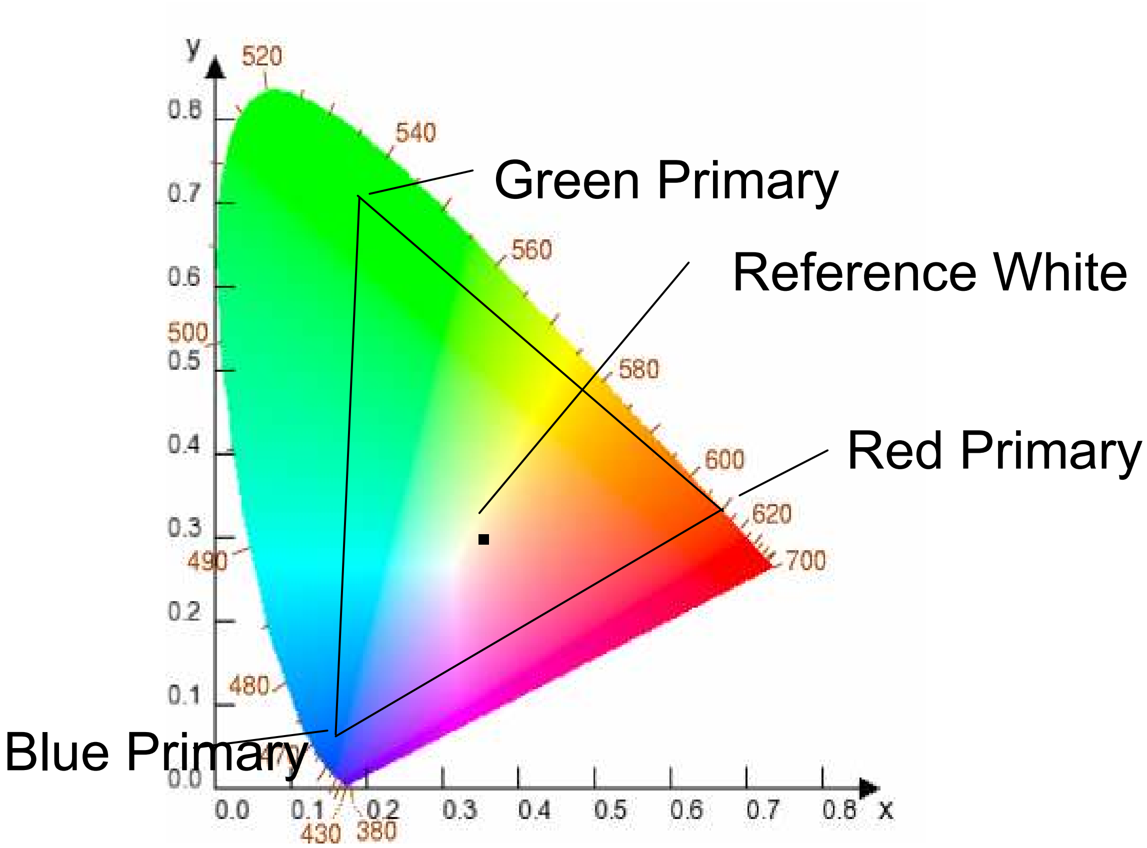
Are they 4:3 or 16:9?



p=progressive
(non-interlace)
i=interlace

- Now you can complete all parts of Q11

That was good improvement. But can we further exploit the eye's sensitivity in colour to reduce data rate?



CIE Chromaticity Diagram

- Can only reproduce colours within the triangle set by the **primary colours**
- Need to specify the **reference white** - taken to be equivalent to daylight

Some properties of colour:

Luminance is the intensity of light - independent of colour
(black - various shades of grey - white).
Hue is the dominant colour
Saturation is the strength or vividness of the colour.

Original

Red

Green

Blue

Luminance Hue Saturation

So far we used the R-G-B additive primary colour space to represent colours.

Due to the position of the primary colours and reference white to produce the effect of white light to the viewer need

$$0.30R + 0.59G + 0.11B$$

We use this relationship to define the luminance of a colour defined by R-G-B.

The luminance signal, Y, of an arbitrary colour is given by

$$Y_s = 0.30R_s + 0.59G_s + 0.11B_s$$

Where R_s, G_s and B_s are magnitudes of the three colour component signals (Red, Green and Blue)

Now we have the **luminance** information, need to code the colour information.

Not practical to supply separate colour signals, hence use colour difference or **chrominance** signals

$$C_b = B_s - Y_s$$

$$C_r = R_s - Y_s$$

No need to supply the third signals – (G_s - Y_s)– since:
We only need three equations with three unknowns.

Get back to the original colour signals by:

$$R_s = C_r + Y_s, \text{ etc.}$$

Why transmit luminance and chrominance signals ($Y\ C_b\ C_r$) rather than separate B_s, G_s and R_s signals?

Separate colour and monochrome information

If the scene is in monochrome, then $B_s = G_s = R_s$

Hence

$C_b = B_s - (0.30B_s + 0.59B_s + 0.11B_s) = 0$, and $C_r=0$

This is called **the Principle of constant luminance**

Also permits monochrome receiver to work satisfactorily on colour signals.

We will learn another advantage soon with respect to reducing the data rate ☺

Colour Representation

Colour space models:

1. RGB (Additive Primaries) – Human eye, image capture in cameras.
2. CMYk (Subtractive Primaries) – In printing industry
3. HSB or HSV (Hue-Saturation-Brightness) – in Computer Graphics
4. YC_bC_r (Luma and 2 Chroma – Blue Chroma and Red Chroma)
   - Used in when colour television broadcast was introduced.

RGB -> YCbCr

$$Y = 0.30\,R + 0.59\,G + 0.11B$$
$$Cb = B - Y$$
$$Cr = R - Y$$

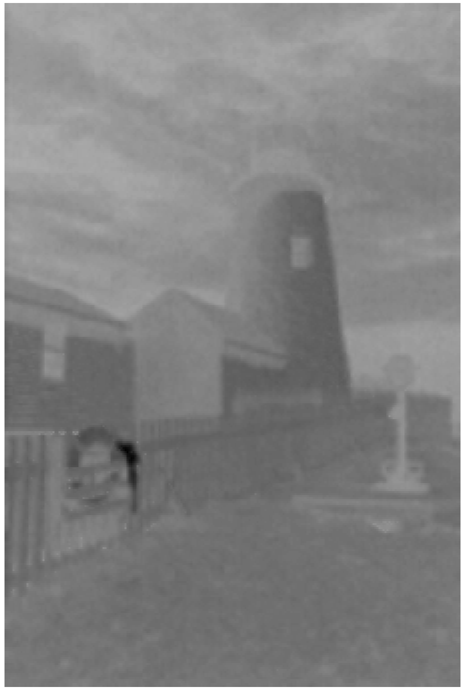
YCbCr -> RGB

$$B = Cb + Y$$
$$R = Cr + Y$$
$$G = (Y - 0.30\,R - 0.11\,B)/0.59$$

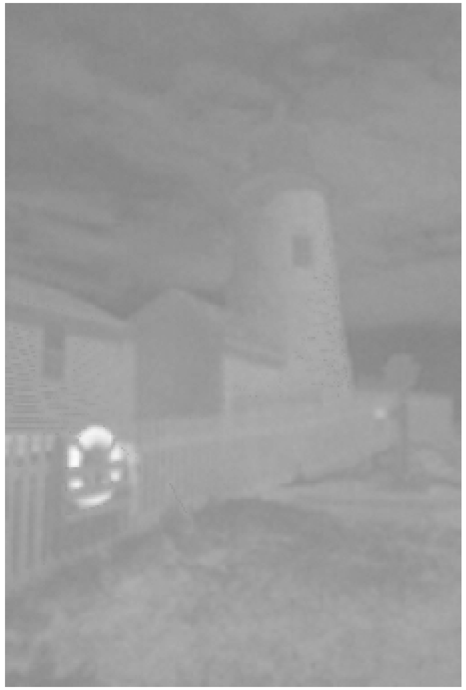
RGB

Y

Cb



Cr



What are your observations on the two chrominance images?

Further reductions of data rate

For analogue – we transmit Luminance (Y) and 2 Chrominance (C_B, C_R) signals.

For digital - could use the separate R, G and B digitised video signals. This requires 3 bytes/pixel memory and then all three signals would have same rate and bits per sample.

But looking at luminance + two chrominance signals, we know that the chrominance signals look smooth, contain fewer finite details and structure is not that visible.

Therefore the human eye is more tolerant to reduced resolution in chrominance channels.

So … down-sample the chrominance channels to reduce data rates.

Chrominance Sub-sampling

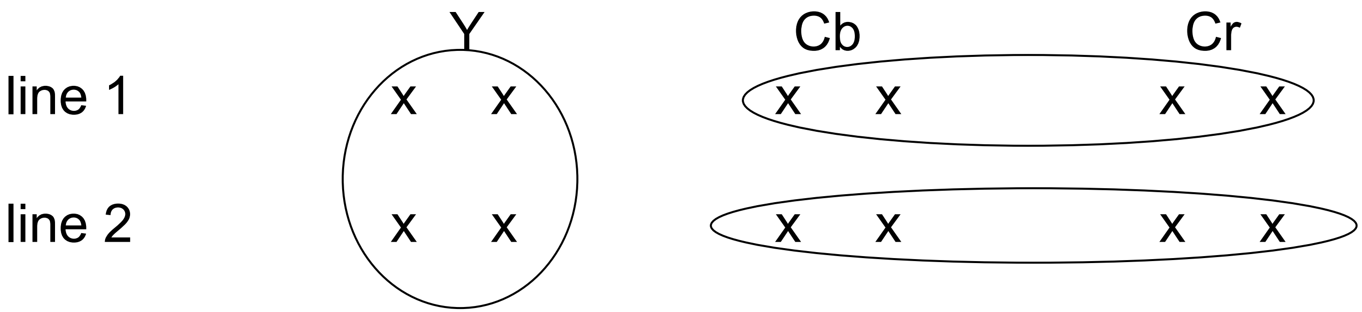
There are three Y Cb Cr formats:

4:4:4 4:2:2 4:2:0

These three numbers A:B:C represent

In a 2x2 group of pixels:

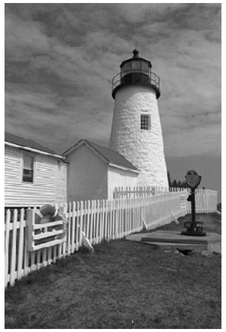
A is the number of Y samples (on both lines).
B is the number of Chrominance (Cb & Cr) samples on line 1
C is the number of Chrominance (Cb & Cr) samples on line 2



W is the Width H is the Height

Y | Cb | Cr

4:4:4



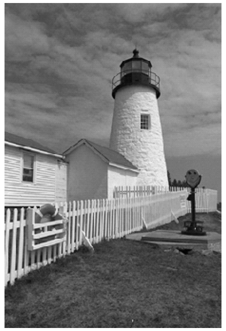
WxH

WxH

WxH

Total bits
=N(WH+WH+WH)
=3NWH

4:2:2



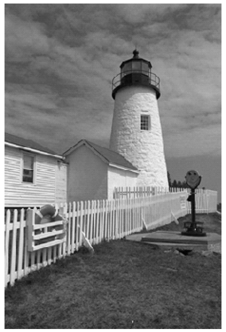
WxH

W/2 x H

W/2 x H

Total bits
=N(WH+WH/2+WH/2)
=2NWH

4:2:0



WxH

W/2 x H/2

W/2 x H/2

Total bits
=N(WH+WH/4+WH/4)
=1.5NWH

N is the bits per colour component

How to find the bit rate?

W pixels/line
H Lines/frame

$H \times W$ pixels/frame (This is the pixels in Y channel)

S Chrominance sub sampling factor
S=3 for 4:4:4 S=2 for 4:2:2 S=1.5 for 4:2:0

N bits/pixel (bpp)

$S \times N \times H \times W$ bits/frame

F frames/sec

$S \times N \times H \times W \times F$ bits/sec

Derive the bit rate for 4:2:2 format 4CIF video!

Y channel resolution of 4CIF video = 576 x 720

Considering N= 8 bits per pixel per colour channel

Cb and Cr resolutions using 4:2:2 sampling S= ?

Memory per frame = ?

Memory to store 90 minutes of video (at 50 frames per sec) = ?

How many DVDs are required to record this programme?
(A single layer DVD can store only about 4.5 Gbytes per disk)

Another Example

Derive bit rate for video transmission over mobile networks using QCIF 4:2:0 format

Mobile phones display resolution

180 x 144

Typical Frame rate is 6.25 fps

Bits/frame ~

At 6.25 frames/s ~

What is the available mobile phone network bandwidth?

What can we do to further reduce the data rate?

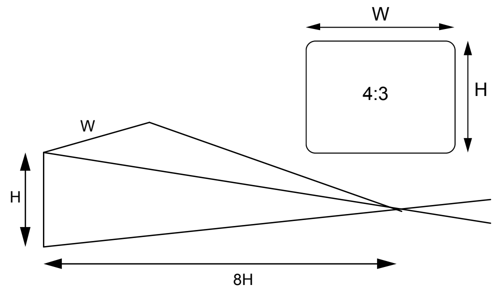
Summary

Three types of resolution in video
1. Spatial
2. Temporal
3. Colour depth

Temporal Resolution:
From CFF limit:
We defined 70 frames/second.
Using Interlace :
Interlace 2 fields.
We can halve the frame rate.
Then using mains frequency (50Hz) as the field rate, we defined
25 frames/second

Colour Depth Resolution:
Defined 1 byte/pixel (for monochrome)
3 bytes/pixel (for colour)

Spatial Resolution:
"Pixel" concept
W x H
Aspect ratio W : H



Used Visual Acuity to find

$$\theta_H = 7^\circ$$

$$\theta_W = 10^\circ$$

CSF limits to find
No. of cycles/picture height = ~420
No. of cycles/picture width = ~600
Using the Nyquist limit we can specify:
1200 x 840 resolution

Relaxed the Nyquist limit and used the Kell factor to correct number of lines.
420/0.7 = 600
what is the new resolution?

What is the bandwidth required for this specification ?

- Now you can attempt Q12