Human Image Perception
ISD M.Sc

Dr. Anil C. Kokaram,

Electronic and Electrical Engineering Dept.,

Trinity College, Dublin 2, Ireland,

anil.kokaram@tcd.ie

Overview

- Need to understand Human Visual System (HVS) to processing results are to be viewed by people.
- Image Comms industry obsessed by picture quality
- Picture Quality assessment v. important
- Objective/Automatic methods of assessment help with Quality of service problems
- Work started since 1920's by Visual Psychologists. David Marr (MIT) published seminal text in 1970's
- Will touch on Intensity Sensitivity, Frequency Sensitivity, Perceptual masking

Measurement of HVS Response

- Notoriously difficult
- Human subject has to separate High Level Visual Processing from Low Level. Its the low level visual processing that is measurable.
- High Level = "That's a circle", "I can't look at another circle"
- Low Level = processing inside cortex or behind retina e.g. spatial frequency response cells, motion response cells
- Lighting conditions must be strictly controlled. Candela/m^2 of display has to be measured

Intensity Sensitivity

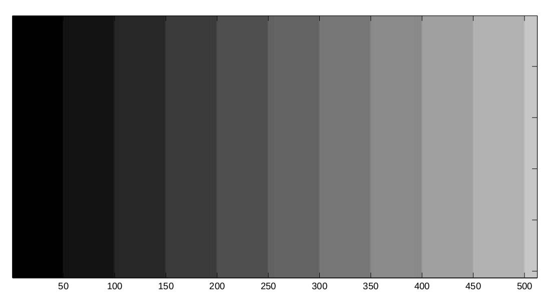
- Weber's law relates the perceived brightness of an object to the brightness of its background. The law can be derived by measuring the 'Just Noticeable Difference' (ΔI) between two visual stimuli.

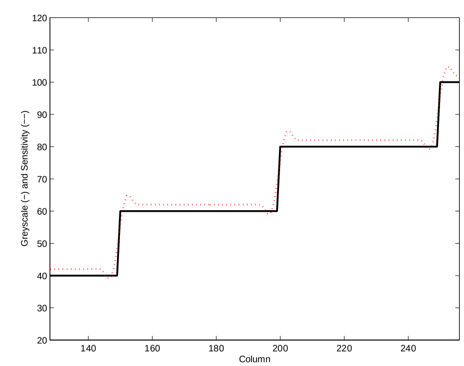
$$\Delta I/I = k \quad \text{Where } k = 0.02, \;\; \Delta I = I_f - I_b \tag{1}$$

- see PPt for a Weber experiment setup
- Weber's law implies that you need more Brightness difference to resolve an object against a bright background than against a dark background. Assuming its bright enough for you to see it in the first place.
- Weber's law is measures *threshold* visibility.

Spatial Sensitivity

Figure shows a staircase image intensity profile.
The HVS perceives that each vertical stripe looks brighter on the left and darker on the right.
This is an effect known as ‘Mach banding’. It is as a direct result of spatial filtering in the visual cortex.
This can be partially understood through the effect of a spatial filter on the image (also shown in figure).





Meaning of Spatial Frequency

- Units are in *cycles per degree* or *cycles per radian*.
- Pictures are being formed on the retina of the eye, and it is there that the visual processing begins. Thus frequencies should be measured in terms of degrees of arc subtended by one luminance cycle on the retina.
- CCIR[a] 500 requires viewing at 5 times the screen height. Viewable height of a monitor $h = \tan^{-1}(h/(5h)) = 0.1974$ rad or 11.3 degrees. PPT pic
- Consider 1024×768 pixel resolution. Resolution of a pixel on the retina is then $\theta = 11.3/768 = 0.0147$ degrees. Sampling frequency of the screen is $1/\theta = 68$ cycles/degree (pixels/degree).
- A vertical sinusoidal grating at .05 cycles per pixel on the screen = frequency of $1/(0.05 \times 0.0147) = 0.3$ cycles per degree on the retina. … …
- It is the cycles per degree which matter for the HVS since it is *at the retina* that processing starts. This explains why the perception of images depends on viewing distance and why International Committees make that a matter for standards.

[a] A regulatory body

Spatial Sensitivity

- The frequency response of the HVS is bandpass.
- Effect demoed in PPT
- The visibility of j.n.d. boundary between the bright and dark vertical lines directly measure the viewer's Modulation Transfer Function (modulated by the printing/photocopying process.).
- If there was no dependence on spatial frequency the boundary between visibility and non-visibility would be a straight line.
- HVS sensitivity is orientation dependent with maximum sensitivity for vertical and horizontal orientations.
- However at most other dirs. are 3dB off peak hence the frequency response can be approximated as isotropic.

Spatial Sensitivity: Main points

- Maximum frequency sensitivity occurs at around 5 cycles/degree. This corresponds to striped patterns with a period of about 1.8 mm at a distance of 1 m ($\sim$ arm's length)
- There is very little response above 100 cycles per degree which corresponds to a stripe width of 0.17 mm at 1m. This implies about 1800 pels per line on a computer display. SVGA of 1024×768 is a little more than half of this. Laptop displays have a pel size of 0.3 mm but are pleasing to view because of the sharpness of the pixels and the lack of flicker.
- Sensitivity to luminance drops off at low spatial frequencies, showing that we are not good at estimating absolute luminance levels *as long as they do not change with time*. The luminance sensitivity to temporal fluctuations (flicker) dos not fall off at low spatial frequencies.

Colour

- The HVS perceives colour using receptors (cones) in the retina which correspond to three broad colour channels in the region of red, green and blue. [ROYGBIV].
- Other colours are perceived as combinations of RGB and thus monitors use RGB to form almost any perceivable colour by controlling the relative intensities of R, G, and B light sources. Hence pixel = [R G B]
- The numerical values used for these intensities are usually chosen such that equal increments in value result in approximately equal apparent increases in brightness.
- In practise this means that the numerical value is approximately proportional to the log of the true light intensity (energy of the wave). This is another statement of Weber's law.

Colour: YUV

- The eye is much more sensitive to luminance (brightness) than to colour changes. Hence B/W pics still make sense. B/W = brightness only.
- The luminance (brightness) of a pel Y may be calculated as $Y = 0.3R + 0.6G + 0.1B$. Approx. values. Diff refs use diff nos.
- The YUV transformation mapping was used in the 1950's so that those with Black and White TV sets could still view colour TV signals.
- RGB representations are usually defined so that if $R = G = B$ the pel is some shade of grey. Thus if $Y = R = G = B$ in these cases, the coefficients used in equation should sum to unity.
- The chrominance of a pel is defined by U and V as below (for PAL[a]).

$$\begin{aligned} U &= 0.5(B - Y) \\ V &= 0.625(R - Y) \end{aligned} \tag{2}$$

Grey pels will always have $U = V = 0$

[a] Phase Alternate Line: the European Colour TV format

Colour: YUV

The transformation between RGB and YUV colour spaces is linear.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \mathbf{C} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad \text{Where} \quad \mathbf{C} = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ -0.15 & -0.3 & 0.45 \\ 0.4375 & -0.3750 & -0.0625 \end{bmatrix} \tag{3}$$

The inverse relationship is had by finding the inverse of $\mathbf{C}$.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \mathbf{C}^{-1} \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \tag{4}$$

Range of Y $= 16 \ldots 255 - 16$; U,V $= \pm 128$. UV are shifted by 128 to allow storage as 8 bit

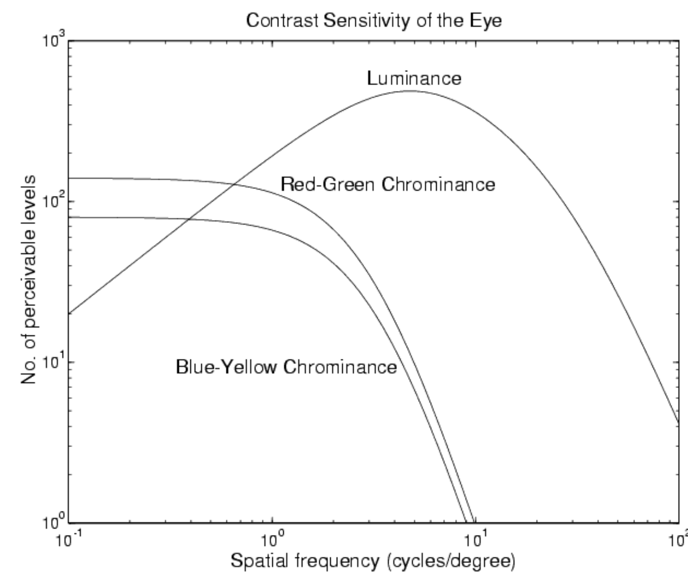
US standard uses slightly different $\mathbf{C}$.

Colour: HSV

- YUV used for DTV broadcasting and Media comms in general
- Not orthogonal w.r.t HVS. HVS sees colour as Hue (redness, greenness etc), Saturation (deep red, light red etc), Value (Brightness/Intensity).
- HSV format better for image analysis
- RGB is a cube. HSV like a cone
- See Matlab demo c.f. YUV and HSV `color-wheels` and `why-hsv`

Colour Sensitivity

- Max chrominance sensitivity < Max Brightness sensitivity
- Chrominance sensitivities fall off over 1 cycle per degree.
- HVS less sensitive to colour
- Don't need full colour resolution: Compression PPT pics



Activity Masking

- Response of HVS depends on what's in the background
- Textural and brightness activity in background can 'mask' HVS responses
- Complicated business depending on similarity between foreground and background patterns
- Rule of thumb: higher the variance of pixels in 8×8 block, lower the contrast sensitivity
- Can use this in noise reduction: reduce noise heavily in flat regions, leave alone near edges
- Difficult to allow for masking quantitatively in processing. (Maths hard).
- PPT pic

Putting some if it together: Compression

- Image data high bandwidth. PAL = 20 MB/sec, DC = 170 MB/sec
- Use image comp. to reduce the amount of data per picture *without* affecting image quality alot
- *Objectionable* degradation depends on usage
- Real time sometimes more important than quality (video over wireless)
- Other times quality is paramount: DTV Digital Cinema
- Same image shown at different formats looks different. e.g. artefacts in format conversion at TV resolution do not appear, but at Film resolution they do. Converting ads from tv→ film is tricky.
- Growing demand for IC IP blocks that implements core compression algos. e.g. PDA's, D-cams

Putting some of it together: What makes compression possible?

- There is alot of **statistical redundancy** in images. For instance, in local image regions say 8×8 blocks, the data tends to be 'flat' or typically homogenous much of the time. This redundancy can be removed without affecting the image substantially, thus reducing the amount of data to be stored.
- The **HVS response** to image stimuli implies that one can introduce artefacts into images *without* them being seen. The colour demonstration accompanying the lecture illustrated this idea with colour subsampling. Thus techniques that remove statistical redundancy can apply that concept heavily in regions where the resulting defects will not be noticed. This further reduces the image data to be stored.
- **Efficient coding techniques** can be used to represent any data as a more compact stream of digits. This technology can be used both for compression and *error-resilience*.

Image Evaluation

- How to actually assess picture quality? Compression: how bad are the artefacts introduced? Restoration: is the picture really better?
- Subjective assessment: 5 point CCIR 500 scale.

  1. Impairment is not noticeable
  2. Impairment is just noticeable
  3. Impairment is definitely noticeable but not objectionable
  4. Impairment is objectionable
  5. Impairment is extremely objectionable

- Lots of subjects, tedious. Ideally want a number to attach to picture which is automatically generated but this is hard. Good effort by Webster et al in 1993.

Image Evaluation: Objective

- Use *error* based schemes to give a number (bit misleading to say *objective* but still ...
- Define an *error* as

$$e(h,k) = \hat{I}(h,k) - I(h,k) \qquad \hat{I} = \text{processed image; } I = \text{original} \tag{5}$$

- e shows how close the processed image is to some 'ideal' $I(h,k)$.
- The *Mean Squared Error* (MSE) and *Mean Absolute Error* are

$$\text{MSE} = \frac{1}{NM}\sum_{\mathbf{x}}(e(\mathbf{x}))^2 \qquad \text{MAE} = \frac{1}{NM}\sum_{\mathbf{x}}|e(\mathbf{x}| \tag{6}$$

where the image size is N rows by M columns, the sum is over all the sites in the image, and $e(\mathbf{x}) = e([h,k])$.

- The **Signal to Noise** ratio is another popular objective measure and it has units of Decibels (dB).

$$\text{SNR} = 10\log 10 \frac{\frac{1}{NM}\sum_{\mathbf{x}} I(\mathbf{x})^2}{\text{MSE}} \tag{7}$$

This is a ratio between the signal power, measured as the sum squared intensities in the original image I, and the 'noise' power measured as the MSE of the error, e.

- PSNR used widely in image compression **Peak SNR** this is the log of the ratio between the peak signal (image) power and the noise power.

$$\text{PSNR} = 10\log 10 \frac{255^2}{\text{MSE}} \quad \text{Units of dB} \tag{8}$$

- Unfortunately, these do not align well with the Human perception of images. Typically large differences match HVS perception ok, small differences say nothing about HVS.

- Matlab demo `snr-mse`

Summary

1. Human visual perception is important for understanding what effects of a processing system will be visible.
2. The perception depends brightness, frequency and the masking effects of the features nearby.
3. Several different colour spaces used. YUV and HSV have applications in DTV and Image Analysis respectively
4. The HVS is less sensitive to colour than brightness (luminance)
5. Perceptual masking is one key to understanding why image compression is possible
6. Automated picture quality assessment difficult because it is difficult to model the HVS and so assign to any arbitrary image an absolute

measure of ‘quality’.

7. CCIR Rec 500 proposes a 5 point subjective evaluation scheme.
8. Objective image evaluation is convenient and MSE, MAE, SNR, PSNR are all used to assess the performance of image processing systems. Comparisons using these measures may not follow human perception of the same images.