GES1810 Perspective in Engineering & Technology

Image Processing

W.K. Cham
Professor
Department of Electronic Engineering
The Chinese University of Hong Kong
**Why Image Processing?**

- Human relies very much on our visual system to collect information about our surrounding. This was the case in the past and will also be true in the future.

- Equipment and software to capture, display, store, edit and transmit images and video are getting cheaper and better quality.

As a result, more & more images and videos are used on multimedia systems and internet.
Topics to be covered

• Human Visual System
• Image Processing
  – Image Enhancement
  – Image Restoration
  – Image Coding
  – Image Understanding
Human Visual System

- Light, Luminance, Brightness and Contrast
- Spatial Frequency Response
- Image Fidelity Criteria
- Color Representation
- Temporal Properties of Vision
Light is the electromagnetic wave that stimulates human visual systems which respond to EM wave of wavelength $\lambda \in [350\text{nm}, 780\text{nm}]$. The light intensity $l$ is measured by illumination (lux) or luminance per unit area (lumen/m$^2$) and is equal to

$$l = k \int_0^\infty c(\lambda) V(\lambda) d\lambda$$

where

(a) $c(\lambda)$ is the power density of light (watt/m$^3$),
(b) $V(\lambda)$ is the relative luminous efficiency function and
(c) $k$ is a constant equal to 685 lumen/watt.

Luminance (lumen or lux m$^2$) and illumination (lux) do not measure human perception of brightness exactly. For example, a projector with 1000 lux does not appear twice as bright as one with 500 lux.
Human Visual System

- Luminance (lumen or lux m$^2$) and illumination (lux) do not measure human perception of brightness exactly. For example, a projector with 1000 lux does not appear twice as bright as one with 500 lux.

- HVS consists of eyes to transform light to neural signals and brain to process neural signals and extract needed information.

- HVS are affected by many factors
  - Adaptation
  - Mach Band Effect
  - Spatial Frequency Response
  - Image Fidelity Criteria
  - Color Representation
  - Temporal Properties of Vision
The retina contains **photoreceptors**, rods & cones.

**Rod**: ~100 millions; sensitive to dim light

**Cone**: ~6.5 million; sensitive to color bright light
Just Noticeable Difference $D$ (j.n.d.) at the intensity $I$ quantifies our ability to resolve two visual stimuli.

Weber's Law: $\frac{D}{I} = \text{constant}$
Intensity Sensitivity affected by Adaptation

The j.n.d. $D$ at the intensity $I$ is affected by the background intensity $I_0$ to which the observer is adapted.
Mach Band Effect: Although the physical brightness is constant across each vertical region, the human observer perceives a brighter left part and a darker right part in each region due to the influence of adjacent regions.
The spatial frequency that gives the maximum response is 10 cycles per degree. With a monitor of distance $d$ cm away, the **10 cycles** occur in $d \tan(1^\circ)$ on the monitor. If $d=50$ cm, then

$$f = \frac{10}{d \tan 1^\circ} = \frac{573}{d} \text{ [cycle/cm]} = 11.46 \text{ cycles/cm}.$$
The physical perception of color is based upon 3 types of cones in the retina. Based upon psychophysical measurements, standard sensitivity curves of these cones have been adopted by the CIE for the standard observer.

A standard observer will perceive 2 color lights the same if they produce the same amount of stimuli on the 3 types of cones. The CIE recommends the use of 3 monochromatic light sources, red (700 nm), green (546.1 nm) and blue (435.8 nm) to reproduce other colors.
Critical Fusion Frequency (CFF): HVS cannot distinguish between a **steady light** and a **flickering light** of frequency above the CFF.

- Different persons have different CFFs which, however, generally do not exceed 60 Hz.
- TV refresh rates are 60 Hz in NTSC and 50 Hz in PAL.

HVS is more **sensitive** to flickering of **high** spatial frequencies than **low** spatial frequencies.

- Monitors of higher resolution need to have higher refresh rates.
- Edges in a video need to be represented by sufficient frames.
Digital Representation of Images

Advantages of digital representation of images

• exact reconstruction of the original from the output of a communication channel or a storage device is possible.

• more error resilient during transmission by error detection or error correction methods.

• can be stored for future uses and processed by computers for different applications (e.g. enhancement, restoration, coding and understanding.)
• For most vidicon tubes $\gamma_v = 0.65$

  For most picture tubes $\gamma_{PT} = 2.2$

• In order to have $E_s \propto E_m$,

  $\gamma_v$ is adjusted to $\gamma'_v$ s.t. $\gamma'_v \gamma_{PT} = 1$, i.e. $\gamma'_v = 0.45$
Histogram Equalization

Original

Enhanced
Histogram Equalization
For most vidicon tubes $\gamma_v = 0.65$

For most picture tubes $\gamma_{PT} = 2.2$

In order to have $E_s \propto E_m$,

$\gamma_v$ is adjusted to $\gamma'_v$ s.t. $\gamma'_v \gamma_{PT} = 1$, i.e. $\gamma'_v = 0.45$
If $\gamma_v \neq 1$, then $E_s$ (illumination input to a camera) is not proportional to $E_m$ (illumination output from a display). The image will appear to be too bright or too dark.

**Gamma correction** is to transform $\{x(i,j)\}$ into $\{y(i,j)\}$ to adjust the brightness by $y = x^\gamma$.

$\gamma > 1$ will make an image darker.

$\gamma > 1$ will make an image brighter.
**Gamma Correction**

Gamma correction is to transform \( \{x(i,j)\} \) into \( \{y(i,j)\} \) to adjust the brightness by \( y = x^\gamma \).

- \( \gamma = 2.2 \): output image darker
- \( \gamma = 0.45 \): output image brighter
Image Restoration

Restoration involves two tasks:

• estimate the information that is available in the idealized model but not in the image
• present the information in the same format as the original image

The idealized models can represent

• deterministic or statistical information of a degradation
• characteristics of the original image
• statistical information about the noise
Image Restoration

• The relation between original image $w$, distorted image $v$, degradation model $G$ and noise $\eta$ can be modeled as

$$v = Gw + \eta.$$

• Difficulties
  – ill-posed inverse problem.
  – no unique solution.
  – requiring *a priori* information about the image, degradation and noise.
Enhancement versus Restoration

**Enhancement**
- Concerning the extraction of image features
- Difficult to quantified performance
- Performance depending on particular image.

**Restoration**
- Concerning the restoration of degradation
- Performance can be quantified
- Performance depending on the ensemble of images.
Image formation models

Motion blur:

$$h(x, y) = \begin{cases} \frac{1}{L} & \text{if } \sqrt{x^2 + y^2} \leq \frac{L}{2} \text{ and } \frac{y}{x} = -\tan \phi \\ 0 & \text{otherwise} \end{cases}$$

Out of focus blur:

$$h(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } \sqrt{x^2 + y^2} \leq R \\ 0 & \text{otherwise} \end{cases}$$

Atmospheric blur:

$$h(x, y) = \frac{1}{2\pi \sigma^2} \exp \left\{ -\frac{x^2 + y^2}{2\sigma^2} \right\}$$
Image formation models

Out of focus blur:

\[ h(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } \sqrt{x^2 + y^2} \leq R \\ 0 & \text{otherwise} \end{cases} \]
JPEG2000 & MPEG4

**Image coding standards**
JPEG (the current most widely used image coding standard), JPEG2000

**Video coding standards**
MPEG1 (VCD), MPEG2 (DVD), MPEG4

**Video conferencing standards**
H.261, H.263
Images coded using JPEG

Original at 8 bpp          JPEG at 0.5 bpp
Images coded using JPEG

JPEG at 0.3 bpp

JPEG at 0.15 bpp
Why JPEG2000

Why do we need a new standard?

- A single standard for different applications such as
  - natural images, video component frame, electronic photography
  - computer generated images
  - medical imaging
  - facsimile, laser print rendering

- Higher compression and better image quality
  - JPEG2000 use Wavelet Transform and Zero-tree Quantization (SPIHT)

0.3 bpp vs 0.1 bpp
JPEG2000 features

JPEG at 0.15 bpp  
SPIHT at 0.1 bpp
MPEG Standards

• MPEG-1 and MPEG-2
  – deal with frame-based video and audio. They allow content to be accessed randomly.
  – make storage and transmission more efficient by compressing the material.

• MPEG-4
  – has all the capabilities of MPEG-1 and MPEG-2
  – video & audio are represented in the form of 'objects' that can be flexibly and interactively used and re-used.
Why MPEG-4 necessary?

− More and more content is **audiovisual**, i.e. containing not only text but also sound, image and video data.
− A growing part of the information is read, seen and heard in **interactive** ways.
− The borders between
  − **telephone** (communication)
  − **PC** (interactive and computing)
  − **TV** (broadcast)

are blurring. MPEG-4 is a standard that supports all of these services and so allows systems to be built with a combination of these services.
What MPEG-4 does?

MPEG-4 defines an audiovisual scene as a coded representation of audiovisual objects that have certain relations in space and time (rather than video frames with associated audio).

- An audiovisual object could be
  - a **video object** such as a car, a dog or the complete background.
  - an **audio object** such one instrument in an orchestra, the barking of the dog, a voice.
- Each object is coded separately.
What MPEG-4 does?