EE 417 Introduction to Computer Vision / EE 569 3D Vision

Instructor: Associate Prof. Mehmet Keskinöz

Communication Theory &Technologies (CTT) Group,
Electronics Engineering Program,
Computer Science and Engineering,
Cyber-Security Program
Faculty of Engineering and Natural Sciences

Email: keskinoz@sabanciuniv.edu

Sabancı Üniversitesi

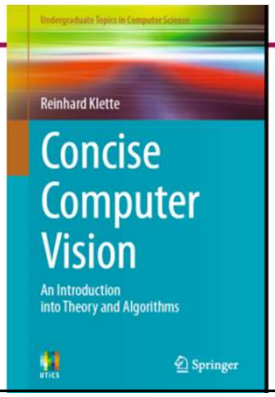
Assoc. Professor Mehmet Keskinöz

Email: keskinoz@sabanciuniv.edu

URL: http://people.sabanciuniv.edu/keskinoz/

- Education:
  - BS from EE Department of Boğaziçi University, 1994
  - MS (in 1997) and PhD (in 2001) from Electrical and Computer Engineering, Carnegie Mellon University (CMU).
- Research Interests
  - High Rate and Robust Communication Receiver Design for Wired and Wireless Technologies such as ADSL, 3G , 4G, 802.11, Wireless Sensor Networks, MIMO etc.
  - Network Coding; Optimal Power Allocation
  - Big Data Analytics& Storage& Communications
  - Information and Decision Fusion; Biometric Fusion
  - Distributed Estimation and Detection
  - Digital Multimedia and Biometric Security; Biometric Hashing and Digital Watermarking Technologies; 2-D Bar Code Design and Its Applications.
  - Control Algorithm Design for Target Detection, Localization and Navigation
  - Correlation Filter Pattern Recognition Theory and Applications ;
  - Error Control Codes (Turbo , LDPC etc) and Their Applications

Text: Concise Computer Vision: An Introduction into Theory and Algorithms, Springer, Series:Undergraduate Topics in Computer Science, by Reinhard Klette, 2014. ISBN: 978-1-4471-6319-0



Reference Books:
Computer Vision: Algorithms and Applications, R. Szeliski, Springer, 2010.
"Fundamentals of Digital Image Processing," by Anil K. Jain, Englewood Cliffs, NJ : Prentice Hall, 1989. ISBN: 0133361659

What is Vision?

"Vision is the act of knowing what is where by looking." --Aristotle

Special emphasis: relationship between 3D world and a 2D image. Location and identity of objects.

What is Computer Vision?

It is related, but not equivalent to:

- Photogrammetry
- Image Processing
- Artificial Intelligence

Computer Vision

- The ability of computers to see
  - Image Understanding
  - Machine Vision
  - Robot Vision
  - Image Analysis
  - Video Understanding

Why study Computer Vision?

- Images and movies are everywhere
- Fast-growing collection of useful applications
  - building representations of the 3D world from pictures
  - automated surveillance (who's doing what)
  - movie post-processing
  - face finding
- Various deep and attractive scientific mysteries
  - how does object recognition work?
- Greater understanding of human vision

Goals and Objectives

- Introduce the fundamental problems of computer vision.
- Introduce the main concepts and techniques used to solve those problems.
- Enable one to implement vision algorithms
- Enable one to make sense of the vision literature

A picture is worth a thousand words.

A picture is worth 1000 words.

A video is worth 1000 sentences?

http://marsrovers.jpl.nasa.gov/gallery/press/opportunity/20040125a.html
JPL Mars' Panorama captured by the Opportunity

- Rich info. from visual data
- Examples of images around us
  natural photographic images; artistic and engineering drawings
  scientific images (satellite, medical, etc.)
- "Motion pictures" => video
  movie, TV program; family video; surveillance and highway/ferry camera

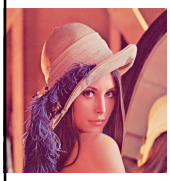
Why Do We Process Images?

- Enhancement and restoration
  - Remove artifacts and scratches from an old photo/movie
  - Improve contrast and correct blurred images
- Composition (for magazines and movies), Display, Printing …
- Transmission and storage
  - images from oversea via Internet, or from a remote planet
- Information analysis and automated recognition
  - Providing "human vision" to machines
- Medical imaging for diagnosis and exploration
- Security, forensics and rights protection
  - Encryption, hashing, digital watermarking, digital fingerprinting …

Digital Data or Media

They are all a sequence (or array) of numbers with finite precision!

Examples of Digital Data



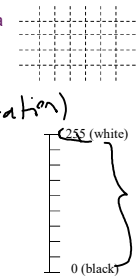
```
50 36 0a 35 31 32 20 35 31 32 0a 32 35 35 0a df P6.512 512.255..
88 7e e0 89 7f e2 8a 80 e1 89 7f e0 88 7c df 87 .~...........|..
7b e0 88 7c e1 89 7d e3 8b 7d e1 89 7b e3 8b 7d {..|..}..}..{..}
df 87 79 d9 80 70 e6 8d 7d de 85 75 dc 83 71 e0 ..y..p..}..u..q.
85 73 de 83 70 e6 8b 76 e1 86 71 dd 83 6b df 85 .s..p..v..q..k..
6d db 81 67 e0 86 6c dd 83 6b e0 86 6e de 83 6e m..g..l..k..n..n
db 80 6b dd 82 70 df 84 72 dd 81 72 dd 82 70 dc ..k..p..r..r..p.
81 6e e1 86 71 d9 7e 69 e3 88 73 de 83 6e e9 8e .n..q.~i..s..n..
79 e2 88 70 e8 8e 76 e5 8d 75 e5 8d 75 e5 8f 74 y..p..v..u..u..t
e8 94 78 ec 98 7c eb 9a 7d e8 99 7a e6 95 77 eb ..x..|..}..z..w.
96 77 ec 95 77 e6 8f 74 ef 98 7d ec 91 7c e8 8c .w..w..t..}..|..
77 e6 87 75 e9 85 75 dd 77 6b dd 75 6c e7 7b 78 w..u..u.wk.ul.{x
d9 6d 6d c8 5d 65 c0 56 62 c2 57 67 af 48 5b a5 .mm.]e.Vb.Wg.H[.
```

Ls.(.A.dministra
tors-.O.n.L.ine)
People who mon
itor and oversee
online chat roo
ms and online ga
mes. In other w
ords, the "onlin
e police."....A

```
4C 73 0A 28 05 41 04 64 6D 69 6E 69 73 74 72 61 ;
74 6F 72 73 2D 05 4F 04 6E 05 4C 04 69 6E 65 29 ;
20 20 50 65 6F 70 6C 65 20 77 68 6F 20 6D 6F 6E ;
69 74 6F 72 20 61 6E 64 20 6F 76 65 72 73 65 65 ;
20 6F 6E 6C 69 6E 65 20 63 68 61 74 20 72 6F 6F ;
6D 73 20 61 6E 64 20 6F 6E 6C 69 6E 65 20 67 61 ;
6D 65 73 2E 20 20 49 6E 20 6F 74 68 65 72 20 77 ;
6F 72 64 73 2C 20 74 68 65 20 22 6F 6E 6C 69 6E ;
65 20 70 6F 6C 69 63 65 2E 22 0A 0A 0A 0A 06 41 ;
```

Sampling and Quantization

- Computer handles "discrete" data.
- Sampling
  - Sample the value of the image at the nodes of a regular grid on the image plane.
  - A pixel (picture element) at (i, j) is the image intensity value at grid point indexed by the integer coordinate (i, j).
- Quantization (a.k.a rounding off operation)
  - Is a process of transforming a real valued sampled image to one taking only a finite number of distinct values.
  - Each sampled value in a 256-level grayscale image is represented by 8 bits.

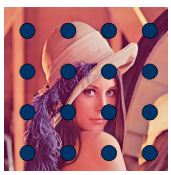


Why Digital?

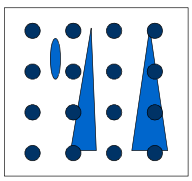
- "Exactness"
  - Perfect reproduction without degradation
  - Perfect duplication of processing result
- Convenient & powerful computer-aided processing
  - Can perform sophisticated processing through computer hardware or software
  - Even kindergartners can do some!
- Easy storage and transmission
  - 1 CD can store hundreds of family photos!
  - Paperless transmission of high quality photos through network within seconds

Image Sampling

- The sampling theorem applies to 2D signal (images) too.



Sampling on a grid

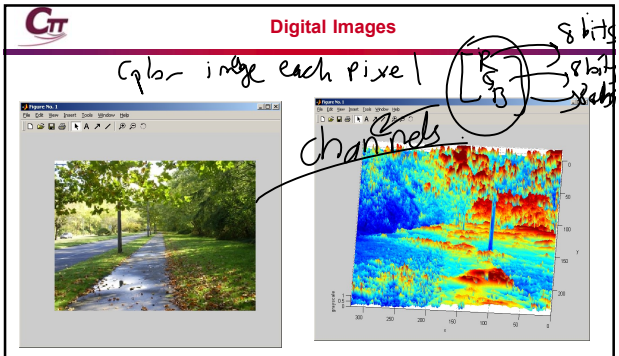


Sampling problem

The image of Barbara

Aliasing due to sampling

Digital Images



An image contains 2D samples of a surface, which can be represented as matrices. Each sample in an image is called a pixel.

Types of Digital Images

- Grayscale image
  - Usually we use 256 levels for each pixel. Thus we need 8bits to represent each pixel (2^8 == 256)
  - Some images use more bits per pixel, for example MRI images could use 16bits per pixel.

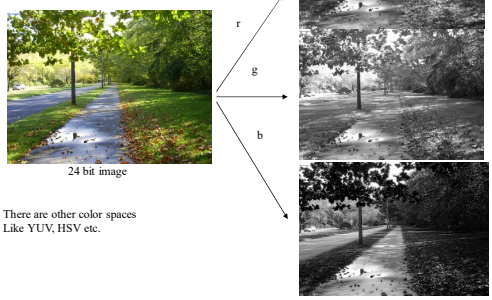
A 8bit grayscale Image.

- Binary Image

A binary image has only two values (0 or 1).

Binary image is quite important in image analysis and object detection applications.

Color Image



24 bit image

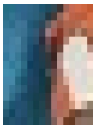
There are other color spaces Like YUV, HSV etc.

What is an image?

What is an image?

A (color) image is a 3D tensor of numbers.

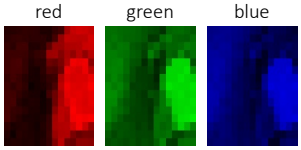
What is an image?



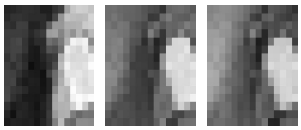
color image patch

How many bits are the intensity values?

red green blue



colorized for visualization



actual intensity values per channel

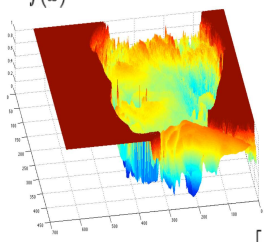
Each channel is a 2D array of numbers.

What is an image?

grayscale image

What is the range of the image function f?

$f(x)$

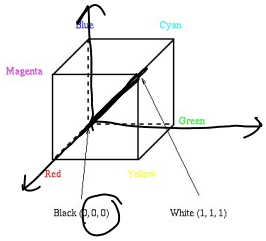


domain $x = \begin{bmatrix} x \\ y \end{bmatrix}$

A (grayscale) image is a 2D function.

RGB Primaries and Color Representation

- Use red, green, blue light to represent a large number of visible colors
- The contribution from each primary is normalized to [0, 1]



The RGB Cube

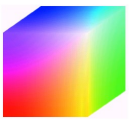
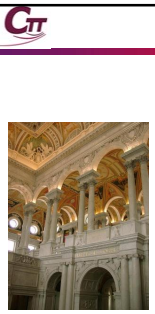
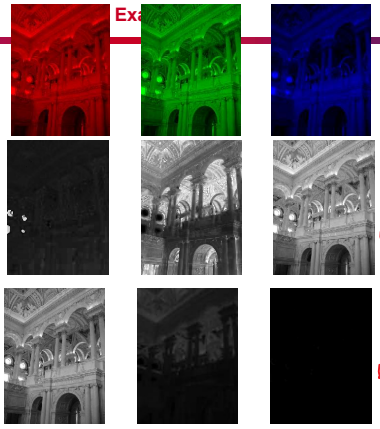


FIGURE 6.8 RGB 24-bit color cube.

Color-cube figures: left figure is from B.Liu ELE330 S'01 lecture notes @ Princeton, right figure is from slides at Gonzalez/ Woods DIP book website

Ex





RGB

HSV

YUV

Color Coordinates Used in TV Transmission

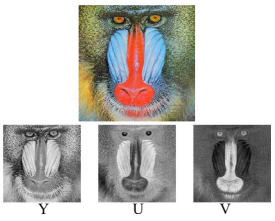
- Facilitate sending color video via 6MHz mono TV channel
- YIQ for NTSC (National Television Systems Committee) transmission system
  - Use receiver primary system (R_N, G_N, B_N) as TV receivers standard
  - Transmission system use (Y, I, Q) color coordinate
    - Y ~ luminance, I & Q ~ chrominance

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R_N \\ G_N \\ B_N \end{bmatrix}. \quad \begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R_P \\ G_P \\ B_P \end{bmatrix}.$$

- YUV (YCbCr) for PAL and digital video
  - Y ~ luminance, Cb and Cr ~ chrominance

Color System in Video

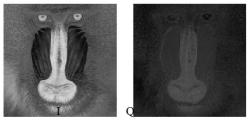
- YUV was used in PAL (an analog video standard) and also used for digital video.
- Y is the luminance component (brightness)
  Y = 0.299 R + 0.587 G + 0.144 B
- U and V are color components
  U = B – Y
  V = R - Y



Color Image

- YIQ is the color standard in NTSC.

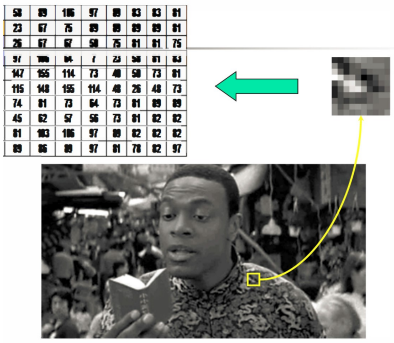
$$\begin{bmatrix} Y' \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.144 \\ 0.595879 & -0.274133 & -0.321746 \\ 0.211205 & -0.523083 & 0.311878 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$



- YCbCr: A color system used in JPEG.

$$\begin{bmatrix} Y' \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

- 2-D array of numbers (intensity values, gray levels)
- Gray levels 0 (black) to 255 (white)
- Color image is 3 2-D arrays of numbers
  - Red
  - Green
  - Blue
- Resolution (number of rows and columns)
  - 128X128
  - 256X256
  - 512X512
  - 640X480



Video

- Sequence of frames
- 30 frames per second

Video Clip

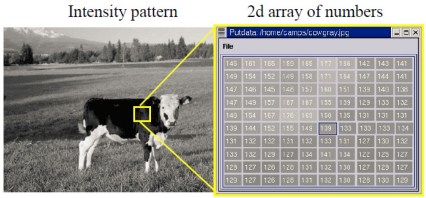
Sequence of Images

Why is Computer Vision Hard?

Visualizing Images

Recall two ways of visualizing an image

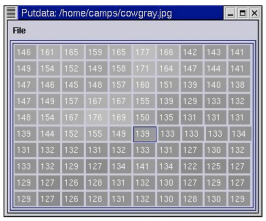
Intensity pattern | 2d array of numbers



We "see it" at this level | Computer works at this level

Why is Computer Vision Hard?

We are trying to infer things about the world from an array of numbers



Shoulder of a cow…

problems: too local; lack of context.
But wait, it's even worse than that...

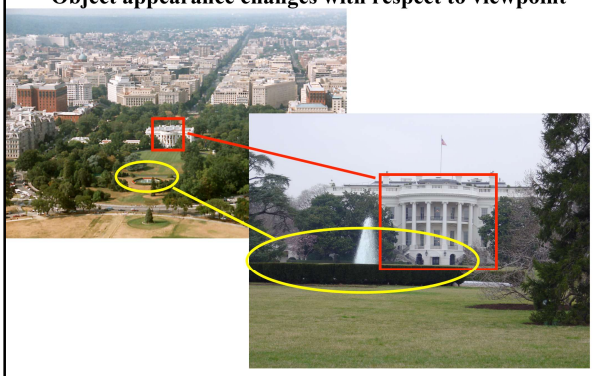
Why is Computer Vision Hard?

If we already know the geometry, surface material and lighting conditions, it is well-understood how to generate the value at each pixel. [this is Computer Graphics]

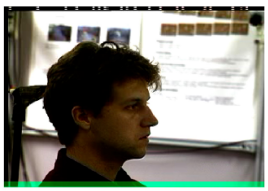
But this confluence of factors contributing to each pixel can not be easily decomposed. The process can not be inverted.

More Difficulties

Object appearance changes with respect to viewpoint



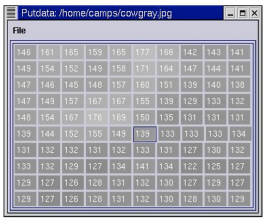
Effects of Lighting



Object appearance also varies with respect to lighting magnitude and direction

Why is Computer Vision Hard?

We are trying to infer things about the world from an array of numbers



Shoulder of a cow…

problems: too local; lack of context.
But wait, it's even worse than that...

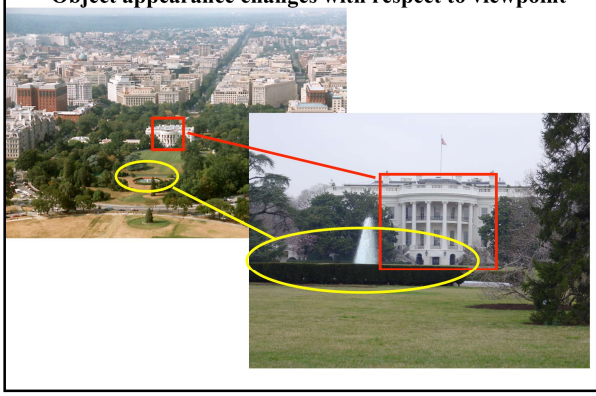
Why is Computer Vision Hard?

If we already know the geometry, surface material and lighting conditions, it is well-understood how to generate the value at each pixel. [this is Computer Graphics]

But this confluence of factors contributing to each pixel can not be easily decomposed. The process can not be inverted.

More Difficulties

Object appearance changes with respect to viewpoint



Effects of Lighting

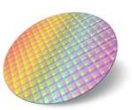
Object appearance also varies with respect to lighting magnitude and direction

plus we might have "occlusion"

Applications of computer vision

Automated visual inspection

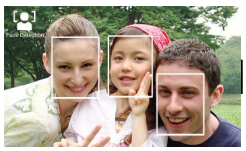
Machine vision



Object Recognition

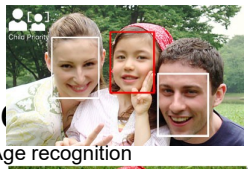
Toshiba Tech IS-910T 2013

DataLogic LaneHawk LH4000 2012

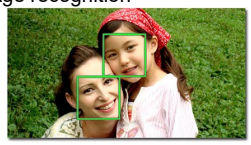


Sony Cyber-shot

Face detection

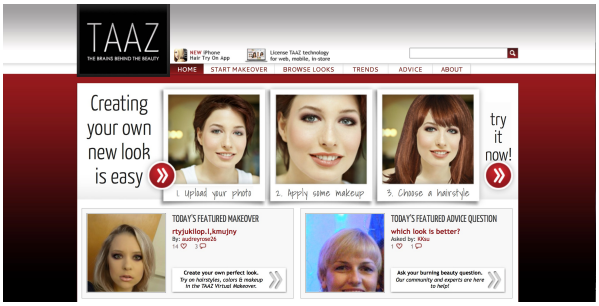


Age recognition



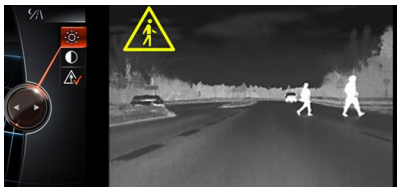
Smile recognition

Face makeovers



BMW 5 series

BMW night vision



Photometry Overview

Source emits photons

Photons travel in a straight line

And then some reach an eye/camera and are measured.

They hit an object. Some are absorbed, some bounce off in a new direction.

Light Transport

Source emits photons

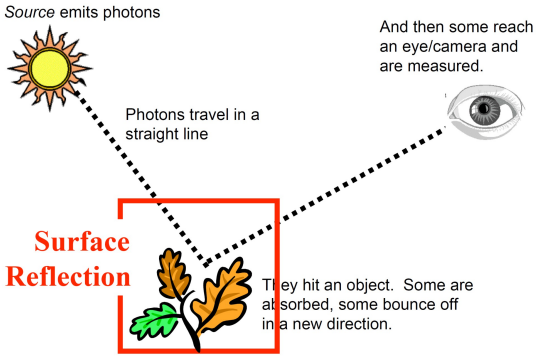
Illumination

Photons travel in a straight line

And then some reach an eye/camera and are measured.

They hit an object. Some are absorbed, some bounce off in a new direction.

Light Transport



Light Transport

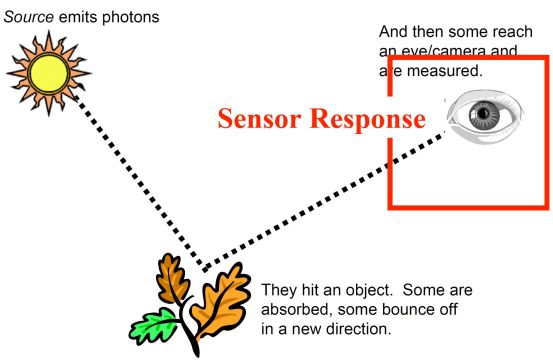
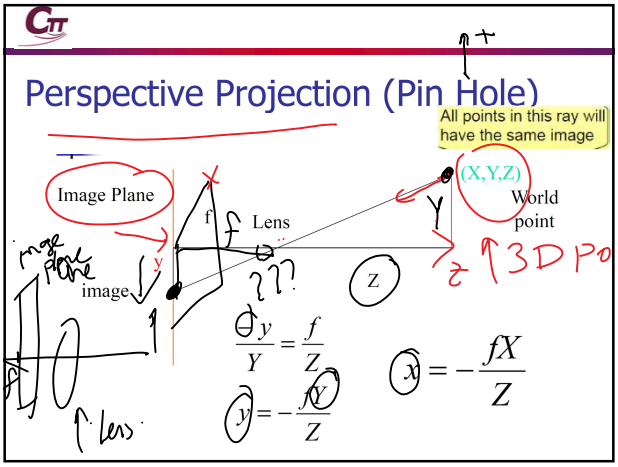


Image Formation

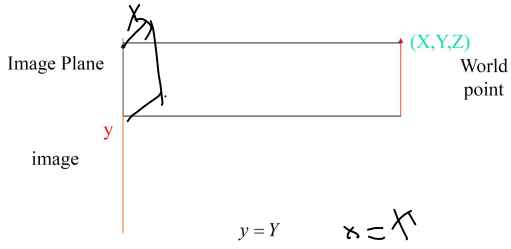
- Light Source
- Camera (extrinsic and intrinsic parameters)
- Scene (Surface reflectance, Surface shape)

Camera Models

Perspective Projection (Pin Hole)

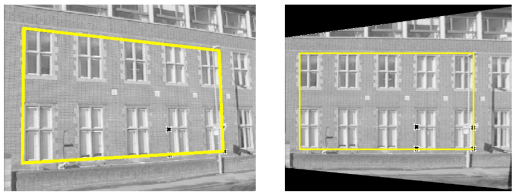


Orthographic Projection



When the object is very far ($Z \to \infty$)
$x = X, y = Y$
Can be used as long as the depth variation within the object is small compared to the distance of the object.

Plane to Plane Mappings

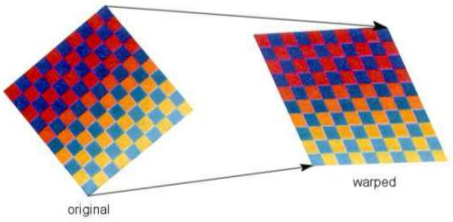


Rigid, Similarity, Affine, & Projective Mappings
Homography Estimation
Image Warping

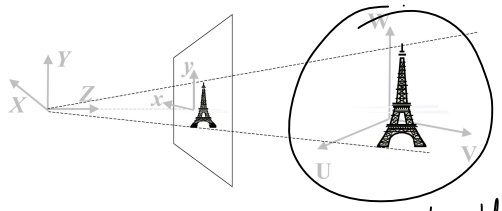
Image Warping

Objective: Change appearance of image by performing geometric transformation, i.e., change the position of a point in the image to a new position.

Example:

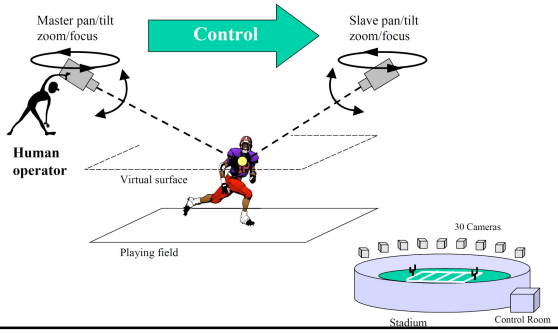


Camera Projection Models



Projection Models
Intrinsic (lens) Paramters → focal length, pixel sizes etc
Extrinsic (pose) Parameters → rotation matrix, translation
Camera Calibration

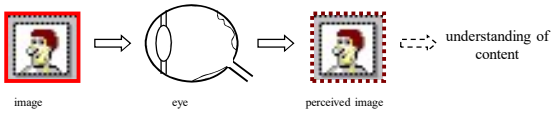
Application: Eyevision System
CAMERA CALIBRATION!



Eyevision : SuperBowl XXXV

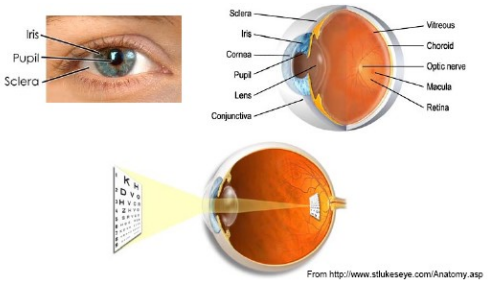
Human Visual Perception

Information Processing by Human Observer

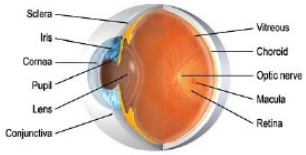


- Visual perception
  - Concerns how an image is perceived by a human observer
    - preliminary processing by eye
    - further processing by brains
  - Important to develop image fidelity measures
    - How to evaluate DIP/DVP algorithms & systems

Eye Anatomy



Eye Versus Camera



Camera components	Eye components
Lens	Lens, cornea
Shutter	Iris, pupil
Film	Retina
Cable to transfer images	Optic nerve send the info to the brain

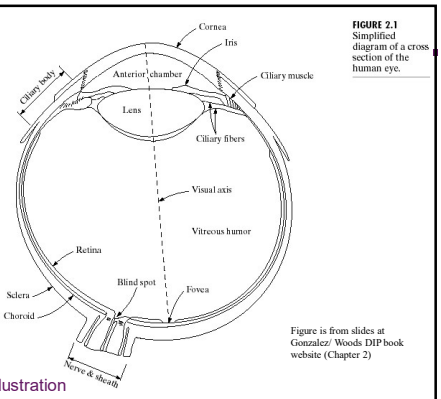


FIGURE 2.1 Simplified diagram of a cross section of the human eye.

Figure is from slides at Gonzalez/ Woods DIP book website (Chapter 2)

- Cross section illustration
- Retina ~ the "film" in eyes to hold our visual sensors

Two Types of Photoreceptors at Retina

- Rods
  - Long and thin
  - Large quantity (~ 100 million)
  - Provide *scotopic* vision (i.e., dim light vision or at low illumination); night vision
  - Only extract luminance information and provide a general overall picture
- Cones
  - Short and thick, densely packed in fovea (center of retina)
  - Much fewer (~ 6.5 million) and less sensitive to light than rods
  - Provide *photopic* vision (i.e., bright light vision or at high illumination); day vision
  - Help resolve fine details as each cone is connected to its own nerve end
  - Responsible for color vision
  - There are three types of cones : Red, Green and blue cones (each type has different frequency response
  - Mesopic vision
    - provided at intermediate illumination by both rod and cones

Human Visual Perception (Monochrome Vision)

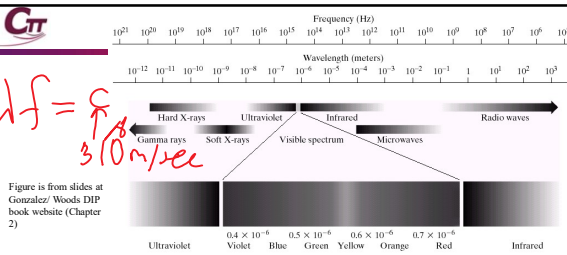


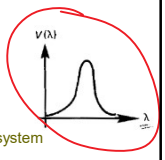
Figure is from slides at Gonzalez/ Woods DIP book website (Chapter 2)

- Light is an electromagnetic wave
  - with wavelength of 350nm to 780nm stimulating human visual response
- Expressed as spectral energy distribution $I(\lambda)$
  - The range of light intensity levels that human visual system can adapt is huge: ~ on 10 orders of magnitude (10^{10}) but not simultaneously
  - Brightness adaptation: small intensity range to discriminate simultaneously

Luminance vs. Brightness

- Luminance (or intensity)
  - Independent of the luminance of surroundings

$$L(x,y) = \int_0^{inf} I(x,y,\lambda)V(\lambda)d\lambda$$

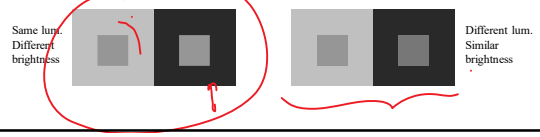


I(x,y,λ) -- spatial light distribution

V(λ) -- relative luminous efficiency function of visual system

(bell shape; different for scotopic vs. photopic vision; highest for green wavelength, second for red, and least for blue)

- Brightness
  - Perceived luminance => Depends on surrounding luminance



Contrast and Weber's Law

- From psychovisual research
  - HVS more sensitive to luminance contrast than absolute luminance
  - Eye-brain response to the % changes in intensity is approx. constant
- Weber's Law: $| L_s - L_0 | / L_s \approx$ const (0.02)
  - Luminance of an object (L_0) is set to be just noticeable from luminance of surround (L_s)
  - For just-noticeable luminance difference (jnd) ΔL:

Define $\Delta C \equiv \Delta L / L \approx d(\log L) \approx 0.02$

- equal increments in log luminance are perceived as equally different

- Empirical Luminance-to-Contrast models

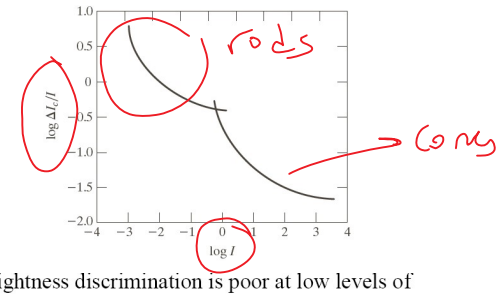
$C = 50 \log_{10} L$ (logarithmic law, widely used)

for $L \in [1, 100]$ and $C \in [0, 100]$

Contrast and Weber's Ratio

- The ability of the eye to discriminate between changes in light intensity at any specific adaptation level is different
- Weber ratio : $\frac{\Delta I}{I}$, ΔI is the increment of illumination discriminable 50% of the time with background illumination I
  - Small $\frac{\Delta I}{I}$: good brightness discrimination
  - Large $\frac{\Delta I}{I}$: poor brightness discrimination

Contrast and Weber's Ratio



- Brightness discrimination is poor at low levels of illumination
- At low levels of illumination, vision is carried out by activity of rods, at high levels, it is by cones

FIGURE 2.7 (a) An example showing that perceived brightness is not a simple function of intensity. The relative vertical positions between the two profiles in (b) have no special significance; they were chosen for clarity.

Figure is from slides at Gonzalez/ Woods DIP book website (Chapter 2)

- Visual system tends to undershoot or overshoot around the boundary of regions of different intensities

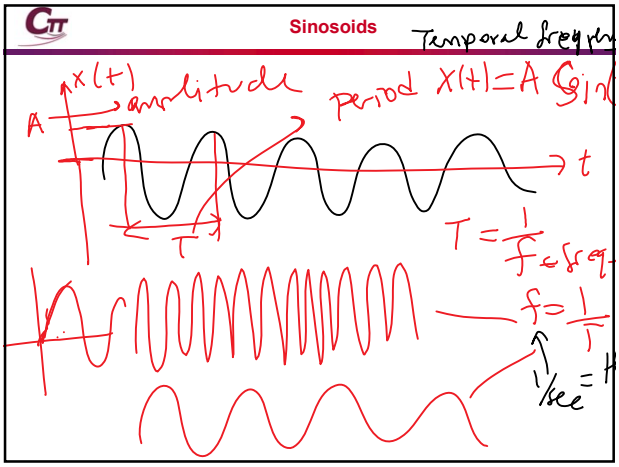
➔ Demonstrates that the perceived brightness is not a simple function of light intensity

Spatial Frequency

- Spatial frequency measures how fast the image intensity changes in the image plane
- Spatial frequency can be completely characterized by the variation frequencies in two orthogonal directions (e.g horizontal and vertical)
  - f_x: cycles/horizontal unit distance
  - f_y : cycles/vertical unit distance
- It can also be specified by magnitude and angle of change

$$f_m = \sqrt{f_x^2 + f_y^2}, \theta = \arctan(f_y / f_x)$$

Sinosoids



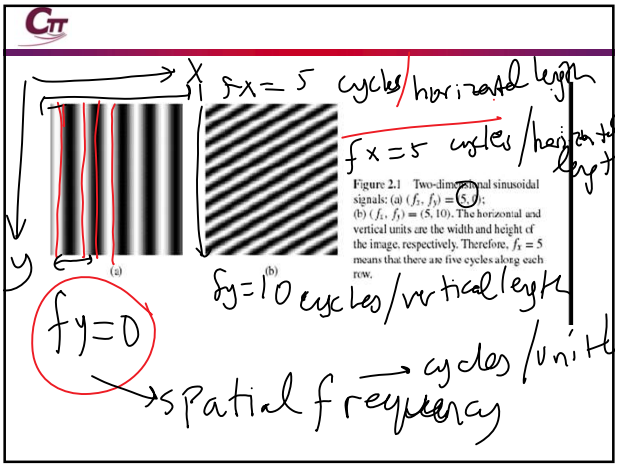
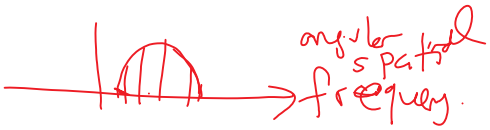


Figure 2.1 Two-dimensional sinusoidal signals: (a) $(f_x, f_y) = (5, 0)$; (b) $(f_x, f_y) = (5, 10)$. The horizontal and vertical units are the width and height of the image, respectively. Therefore, $f_x = 5$ means that there are five cycles along each row.

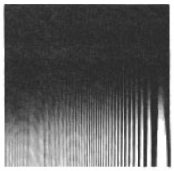
Visual Angle and Spatial Frequency

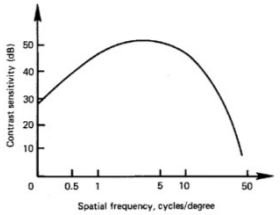
- Angular Spatial Frequency
  - Measures the extent of spatial transition in unit of "cycles per visual degree"
- Visibility thresholds
  - Eyes are most sensitive to medium spatial freq. and least sensitive to high frequencies
    ~ similar to a band-pass filter
  - More sensitive to horizontal and vertical changes than other orientations



Visibility Thresholds at Various Frequencies

- Measuring visibility threshold using sinusoidal grating of varying contrast and frequencies
- Human visual system's frequency response is similar to a band-pass filter





From Jain's Fig.3.7 (pp55) w.r.t. radial spatial freq.

Visibility Threshold at Various Spatial Frequency

- Visibility threshold at different spatial frequency
  - Eyes are most sensitive to mid frequencies, and least sensitive to high frequencies
  - Most sensitive to horizontal and vertical ones than other orientations

Monochrome Vision Models

- Human Visual System (HVS)
  - from experiment with sinusoidal grating of varying contrast
  - similar to a band-pass filter
    - most sensitive to mid frequencies
    - least sensitive to high frequencies
  - also depends on the orientation of grating
    - most sensitive to horizontal and vertical ones
- Overall monochrome vision models
  - How light is transformed by eye to brightness information

Image Fidelity Criteria

- Subjective measures
  - Examination by human viewers
  - Goodness scale: excellent, good, fair, poor, unsatisfactory
  - Impairment scale: unnoticeable, just noticeable, …
  - Comparative measures
    - with another image or among a group of images
- Objective (Quantitative) measures
  - Mean square error and variations
  - Pro:
    - Simple, less dependent on human subjects, & easy to handle mathematically
  - Con:
    - Not always reflect human perception

Mean-square Criterion

- Average (or sum) of squared difference of pixel luminance between two images

$$\mathcal{E}_1 = E[|u - u'|^2] \quad \text{(mean square error)}$$

$$\mathcal{E}_2 = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}|u(m,n) - u'(m,n)|^2 \quad \text{(average squared error)}$$

$$\mathcal{E}_3 = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}E[|u(m,n) - u'(m,n)|^2] \quad \text{(average mean square error)}$$

- Signal-to-noise ratio (SNR)
  - SNR = 10 $\log_{10}$ (σ_s^2 / σ_e^2) in unit of decibel (dB)
    - σ_s^2 image variance
    - σ_e^2 variance of noise or error
  - PSNR = 10 $\log_{10}$ (A^2 / σ_e^2)
    - A is peak-to-peak value
    - PSNR is about 12-15 dB higher than SNR