ECE5554 – Computer Vision
Lecture 1a – Introduction; Image Formation

Creed Jones, PhD

Course updates

- HW1 has been posted
  - due next Thursday, July 13, 11:59 PM

- Check Canvas for information on installing Python and OpenCV
  - Go ahead and install it and try a bit of coding

```python
# example.py    Creed Jones Virginia Tech   ECE5554     July 2, 2022
# a simple Python file to load an image and convert colors a bit

import cv2
import numpy as np

def adjcolors(rawimg):
    img = rawimg.astype(np.double)
    redfactor = 0.8
    greenfactor = 1.0
    bluefactor = 1.5
    img[:, :, 2] *= redfactor   # red pixels are in plane 2
    img[:, :, 1] *= greenfactor
    img[:, :, 0] *= bluefactor
    img = np.minimum(img, 255)
    img = np.maximum(img, 0)
    result = img.astype(np.ubyte)   # convert back to 8-bit
    return result

pathname = "C:/Data/Images/"
filename = "face.png"
inputimg = cv2.imread(pathname+filename)
cv2.imshow("RAW IMAGE", inputimg)
cv2.waitKey()
outputimg = adjcolors(inputimg)
cv2.imshow("ADJUSTED", outputimg)
cv2.waitKey()
```

Today's Objectives

Introduction

- Review the Syllabus
- Review the Course Schedule
- Introduction to the Course
- Definition and Concept of Computer Vision
- What Makes Computer Vision Hard?

Image Formation

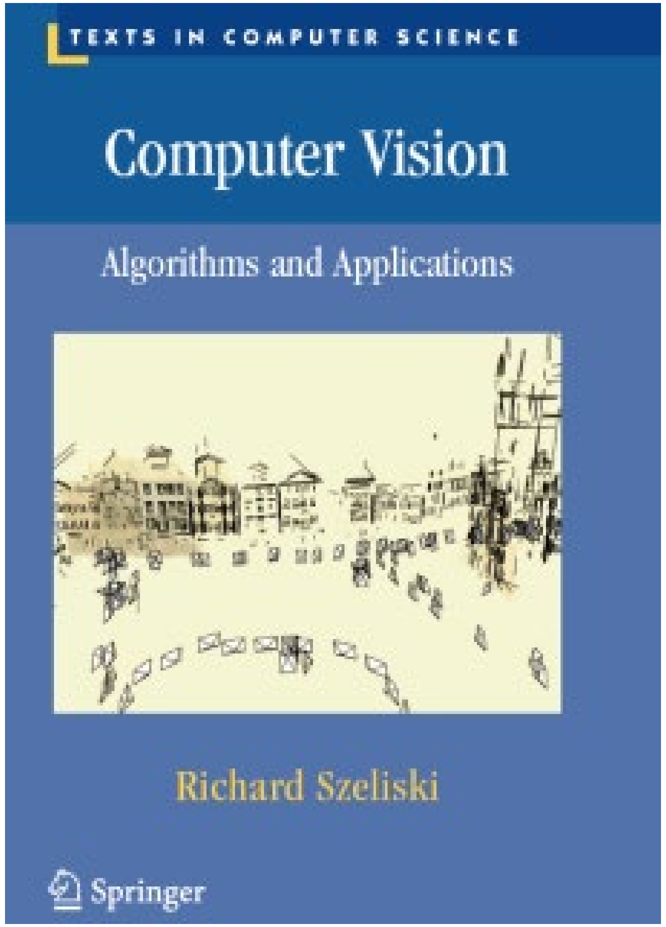
- Light and its interactions with matter
- Components of a simple imaging system
- The human visual system

Required Resources

- **Textbook:**
  Szeliski, *Computer Vision: Algorithms and Applications*, Springer; 2011 edition,
  ISBN-10: 1848829345
  ISBN-13: 978-1848829343

- **Software:**
  Python with OpenCV

- **Class Sessions:**
- July 6 through August 10
- Mon and Wed, 2:00 pm to 5:30 pm
- Zoom ID: 894 6832 3879, password = summer



So what's this course about? Let's look at the catalog description

- Techniques for automated analysis of images and videos. Image formation, feature detection, segmentation, multiple view geometry, recognition, and video processing.

- Computers increasingly require the ability to "see" their surroundings in order to interact with humans and with the three-dimensional world. This course introduces theory and techniques for analyzing the content of images. Applications of computer vision include robotics, autonomous vehicle navigation, industrial automation, content-based search in image databases, face and gesture recognition, and aids for the seeing-impaired.

ECE5554 Course Objectives; what skills should you acquire from this course?

Upon successful completion of this course, students will be able to:

1. contrast common image formation models;
2. implement various ways of extracting features from images;
3. segment an image into meaningful regions;
4. derive the theory behind multi-view geometry;
5. implement various approaches to recognizing objects and scenes in images; and
6. implement techniques for processing video sequences.

Student Assessment

- There will be five homework assignments, a final exam and five short quizzes throughout the semester. Each assignment is due at 11:59 PM on the due date. Assignments will be accepted up to three days late, but will be penalized 10% per day.

Graded Item	# of Items	Points per Item	Total Points	Percentage
Homework Assignments	5	40	200	50%
Final Exam	1	100	100	25%
Quizzes	5	20	100	25%
			400	**100%**

Grading Details

- Homework assignments
  - Due on the Wednesdays of weeks 2-6
  - Typically, Python implementation of methods we are learning about in class
- Quizzes
  - Tuesdays of weeks 2-6
  - Fully online, approximately ten questions, open book and notes
  - 20 minutes in duration; open from 6PM to 11:59PM Eastern time
- Final exam
  - Fully online, open book and notes
  - Thursday, August 11, 7PM to 9 PM, two hour time slot

Academic Integrity

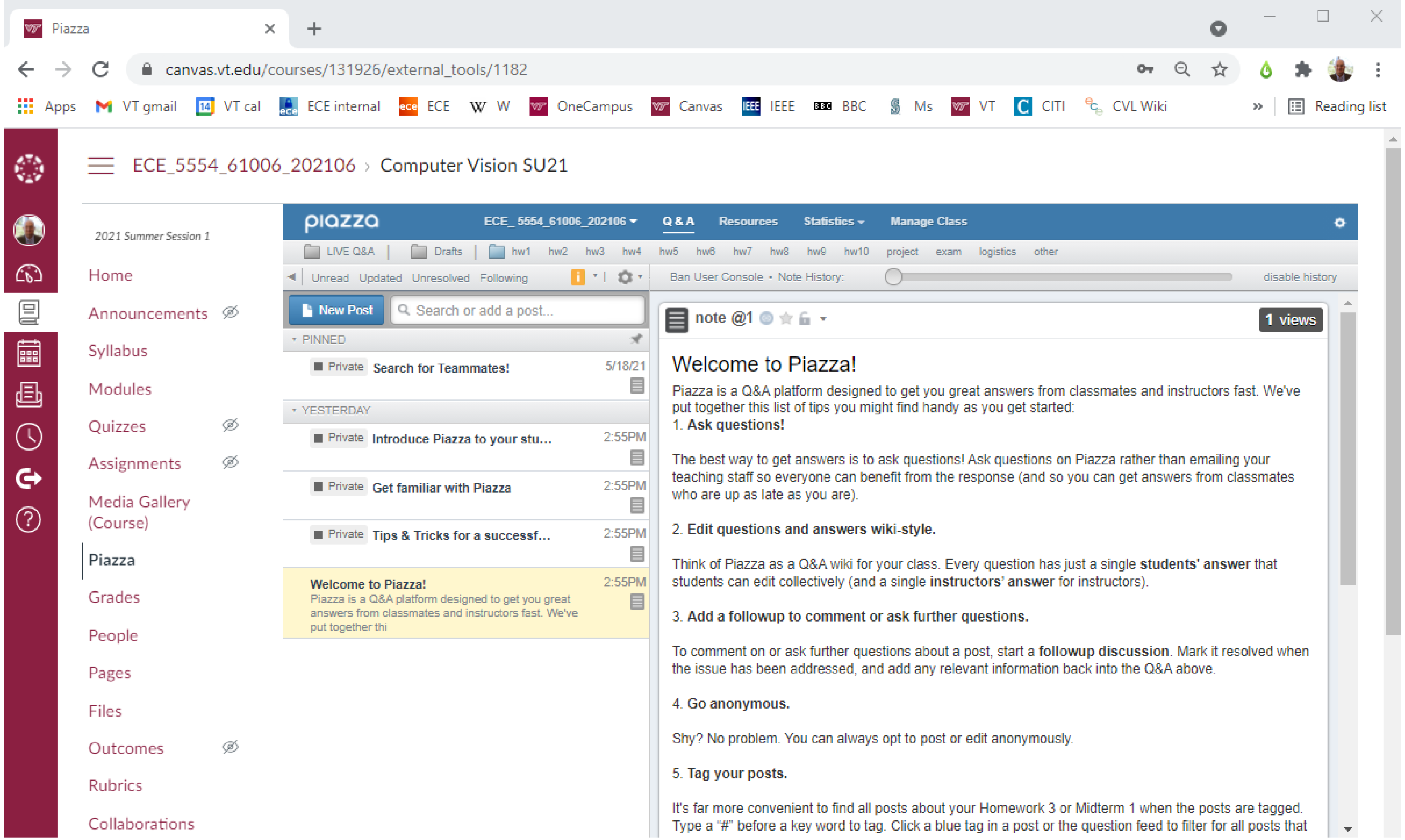
- **<u>All homework assignment and exams must be your own work</u>**

- Cheating: copying another student's work on any assignment
- Plagiarism: quoting or using content from a published source without proper citation

- Sharing answers is cheating

- Cheating or plagiarism will result in a zero on the assignment in question
- The university policies for academic integrity and the honor code will be followed

We will use Piazza to allow questions about lectures or homework; it's accessed in Canvas



ECE5554 SU22 Daily Schedule

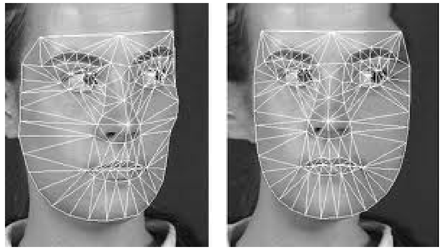
Module	Day	Lec	Reading	Topics	Due
1 - Introduction; Image Processing	6-Jul	1	1.1 - 1.2, 2.1 - 2.3	Introduction; image formation; coordinate transformations	
	11-Jul	2	3.1 - 3.2, 4.2	Pixel operations; filtering; edge detection	
	12-Jul				Quiz 1 (Tue)
	13-Jul	3	3.4 - 3.5	Edges; Fourier transforms; image pyramids	HW1
2 - Features	18-Jul	4	4.1 - 4.3	Patch location; corner detection; Hough and SIFT	
	19-Jul				Quiz 2 (Tue)
	20-Jul	5	8.2 - 8.4	Motion tracking; optical flow; texture	HW2
	25-Jul	6	4.2 - 4.3, 3.3.4	Contours; curvature; region properties	
	26-Jul				Quiz 3 (Tue)
3 - Segmentation	27-Jul	7	3.1, 5.2	Segmentation methods	HW3
	1-Aug	8	5.1 - 5.4	Segmentation by clustering; graph methods; active contours	
	2-Aug				Quiz 4 (Tue)
4 - Structure	3-Aug	9	6.1, 9.1, 3.6.3	Feature-based alignment; image stitching; image morphing	HW4
	8-Aug	10	6.2, 11.1 - 11.5	Pose estimation; stereo vision	
	9-Aug				Quiz 5 (Tue)
	10-Aug	11	12.1	Shape from shading and motion; deep learning; review	HW5
	11-Aug	**FINAL EXAM, 7 PM - 9 PM**			

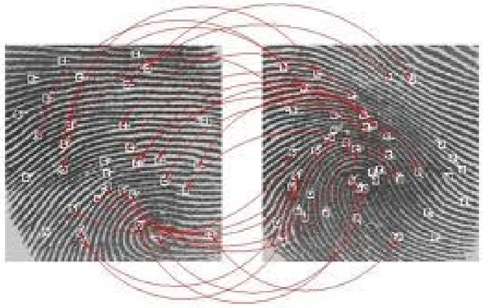
Dr. Creed Jones, Professor

BS & MS in Electrical Engineering from Oakland University (Michigan)

PhD in Computer Engineering from Virginia Tech

Research – Automated Face Recognition





- Worked for:
  - Small technology companies in Image Processing
  - Sagem Morpho in Biometrics (Human Identification)
  - Humana in Pattern Recognition and Analytics
- Taught Computer Science at Seattle Pacific University and California Baptist University
- Hold six patents; authored a number of international standards

crjones4@vt.edu

Here are some of the projects that I have worked on that are related to Computer Vision

- Automated print quality measurement
- Printed circuit film alignment
- Cast parts measurement
- Automotive door hinge inspection
- Automotive radiator inspection
- Wire harness quality inspection
- Glass bottle inspection
- License plate reading
- Cargo container identification
- Inspection of microwaveable meals
- Aluminum ingot identification
- Contact lens inspection
- PLCC package inspection
- Precision drill bit inspection
- Infrared image processing
- High-speed DataGlyph reading
- Infrared military target characterization
- Fingerprint recognition
- Fingerprint classification
- Face Recognition
- Hand geometry recognition
- Image-based monitoring of glaucoma

Computer vision is automated processing that achieves some level of understanding of the contents of an image or video

- *"In computer vision, we are trying to…describe the world that we see in one or more images and to reconstruct its properties, such as shape, illumination and color distributions." [Szeliski, p. 5]*

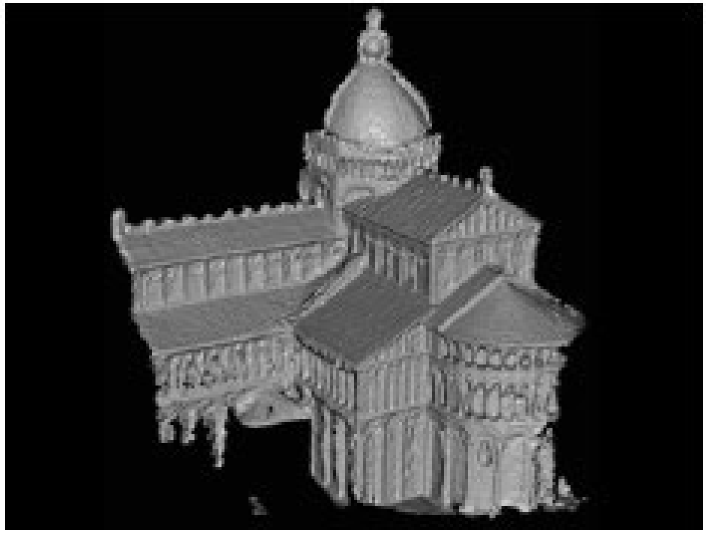
- Little a priori information about the specific image

- Often we have recourse to many other images about which we know something
  - Learning systems

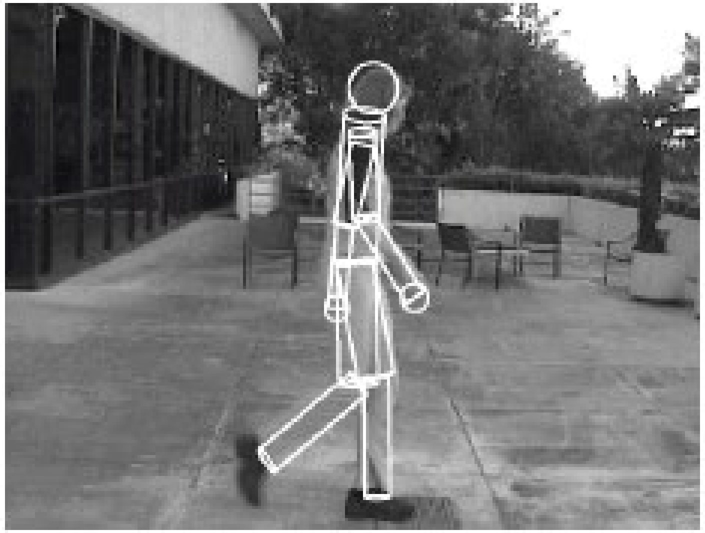
Examples of Computer Vision:

(a) Structure from Motion
(b) Stereo Matching
(c) Person Tracking
(d) Face Detection

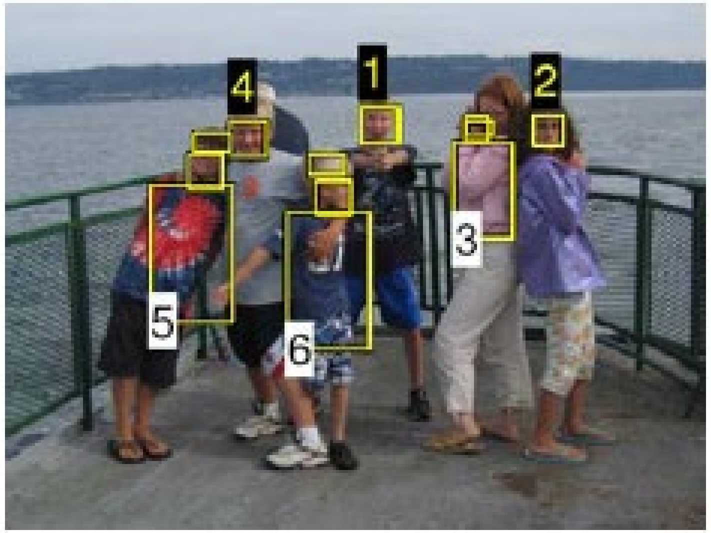
(a)



(b)



(c)



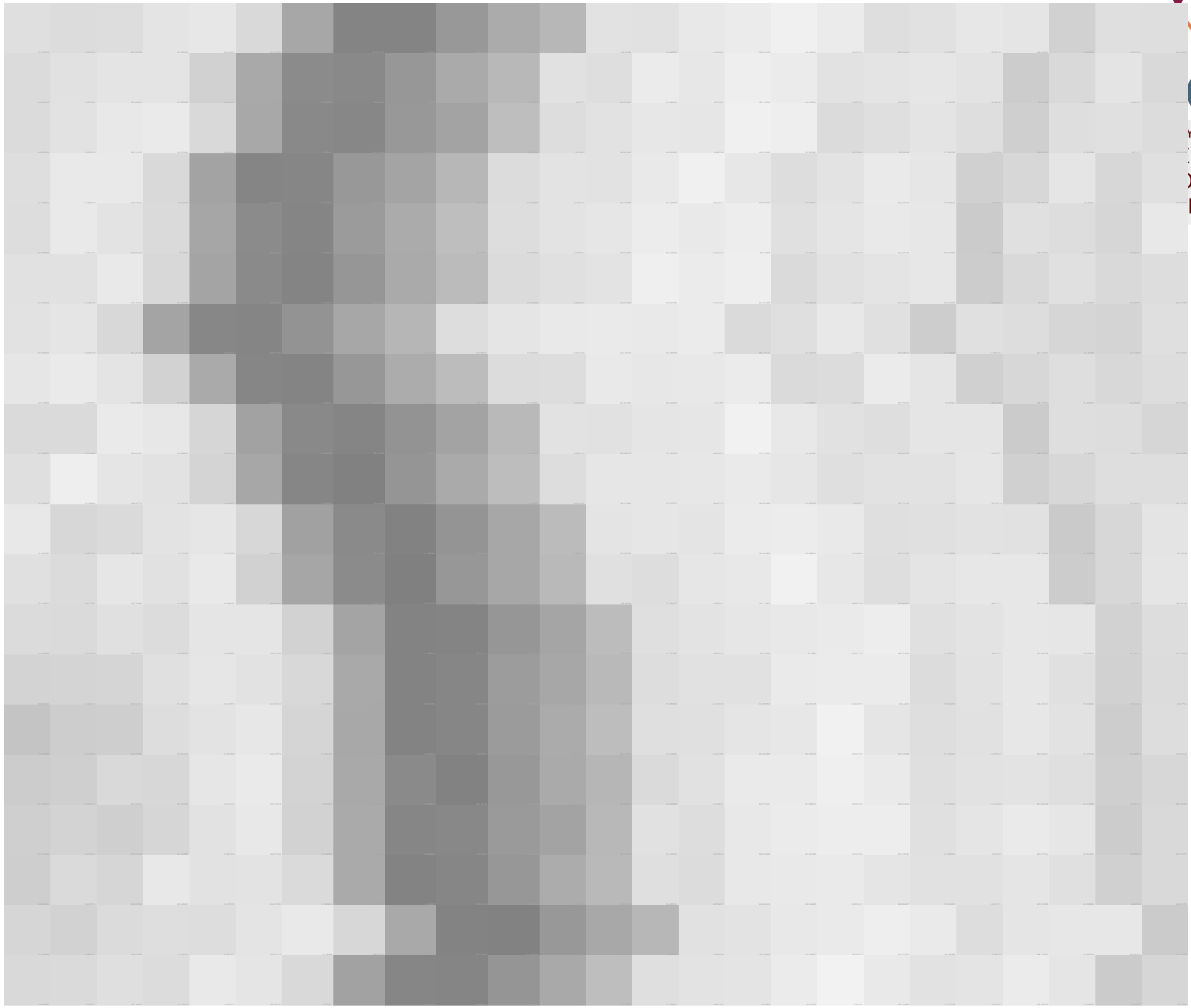
(d)

Computer Vision, Szeliski

Here is an old photograph. What can you say about it?

Images are displayed as brightness matrices – because there's so much data

(this is zoomed in)



The very same data can be seen as a matrix of data values; of course, this is what algorithms deal with

145	147	152	158	159	140	101	63	62	87	102	120	155	158	164	166	163	163	157	157	156	162	136	152	153
153	158	163	162	141	99	68	67	81	97	122	150	157	163	163	171	164	156	160	159	156	138	154	150	145
152	158	155	157	145	95	69	62	83	99	119	153	151	160	162	169	161	150	159	162	153	132	153	156	144
152	160	160	139	94	68	61	84	103	117	158	156	157	163	171	162	154	153	159	156	139	146	158	153	151
160	158	163	146	95	66	66	87	97	118	158	157	163	165	167	165	152	151	159	160	137	146	158	145	162
152	159	156	144	101	63	58	79	98	119	150	159	156	168	170	166	150	151	161	154	136	146	159	152	158
162	156	145	101	64	61	77	97	118	153	160	163	168	165	165	148	151	156	154	137	153	151	151	146	152
159	158	162	146	99	63	64	82	95	121	155	152	162	162	162	160	154	158	157	160	136	147	158	143	151
150	147	155	163	140	95	66	63	78	96	113	152	158	164	166	169	164	152	160	156	157	141	149	149	149
157	164	159	158	142	93	68	67	81	96	113	151	154	158	160	172	168	150	153	156	160	133	146	158	152
158	145	153	155	163	141	98	69	61	85	102	121	155	155	164	162	170	159	150	152	161	152	136	149	153
152	158	157	159	160	144	97	61	67	82	93	117	156	153	156	168	165	166	149	153	161	153	140	150	155
153	140	151	147	163	164	145	94	64	66	82	101	118	157	157	158	167	163	167	154	157	156	153	136	147
139	149	150	155	164	157	138	98	68	64	79	94	120	152	156	159	162	171	167	150	158	163	154	136	154
126	135	136	155	158	159	144	92	70	62	78	101	117	149	152	157	163	164	165	150	158	165	161	135	154
140	137	145	151	155	160	143	96	71	62	80	100	116	155	159	155	160	172	165	154	158	159	157	133	151
139	143	144	146	156	159	147	97	69	60	81	102	117	149	159	158	161	166	161	151	153	158	155	138	145
141	143	152	156	162	161	139	101	64	63	83	95	117	152	158	159	169	163	166	151	159	162	153	133	145
145	141	143	154	151	154	157	147	93	66	66	85	96	117	148	159	160	163	163	163	154	154	156	152	140
142	153	155	157	158	158	138	99	62	66	77	99	115	155	153	160	169	169	167	153	160	156	158	135	148

We want to develop the ability to think of images in both ways, for visualization, understanding and reasoning

150	149	155	161	163	146	95	68	65	87	93	117	157	160	159	168	162	160	154	160	163	153	133	150	159
150	150	164	156	144	94	63	61	85	98	118	153	150	159	164	172	161	150	151	165	159	136	150	150	143
148	159	160	161	147	99	63	58	78	101	116	155	157	156	168	168	164	157	151	161	157	139	151	152	150
153	160	158	147	98	64	66	82	97	113	154	158	164	163	164	158	150	153	158	152	136	147	151	144	150
152	163	156	145	95	64	66	78	94	122	149	155	158	160	164	165	154	158	159	153	140	152	153	151	155
150	162	162	144	100	71	63	82	101	122	151	152	157	166	167	162	150	155	155	154	135	147	152	153	157
155	161	145	98	67	67	77	99	120	158	159	161	167	164	164	152	160	161	156	136	154	154	144	148	153
152	157	158	143	94	66	65	79	98	116	154	155	159	159	164	165	152	159	162	158	134	145	158	151	153
154	150	162	165	139	101	68	61	84	96	121	152	155	161	161	162	159	149	152	164	154	136	152	150	148
162	161	156	158	142	92	67	59	78	94	119	155	154	157	164	165	160	155	153	157	158	135	145	153	144
158	151	155	162	156	143	93	66	65	85	94	122	153	151	158	168	167	165	150	153	161	155	132	150	155
152	150	158	156	163	145	94	70	65	85	98	120	156	156	161	159	171	166	154	160	162	155	134	152	155
151	140	148	146	163	159	141	101	65	66	80	98	113	151	150	161	163	165	166	153	158	160	158	133	149
142	147	145	152	160	156	142	96	69	63	83	102	120	152	160	161	163	169	167	155	154	161	156	133	150
126	142	138	146	162	163	145	93	69	67	84	96	113	152	151	161	167	168	161	147	160	157	157	134	145
135	140	146	147	160	161	144	102	63	67	82	94	113	149	154	158	161	168	158	154	159	155	162	138	149
136	136	138	146	159	163	142	97	69	58	86	102	120	151	158	160	160	163	159	153	151	158	160	136	148
138	142	153	154	157	156	143	99	62	68	86	93	115	150	154	160	163	163	159	154	155	161	155	134	153
146	139	152	156	151	157	163	146	95	69	58	81	99	118	152	150	159	164	167	165	151	156	162	157	140
143	150	150	154	161	157	144	102	61	67	81	93	114	156	152	163	168	168	165	149	156	160	159	141	148

Why is vision difficult?

- **Too much information!**
  - Possibly 640 × 480 pixels/image × 8 bits/pixel
  - Possibly 3 color values per pixel
  - Possibly 30 images/second
  - Could use 2 or more cameras simultaneously

- **Not enough information!**
  - Each pixel represents light intensity (usually)
  - The physical properties that we want to measure (distance, orientation, reflectance)
    are not available directly

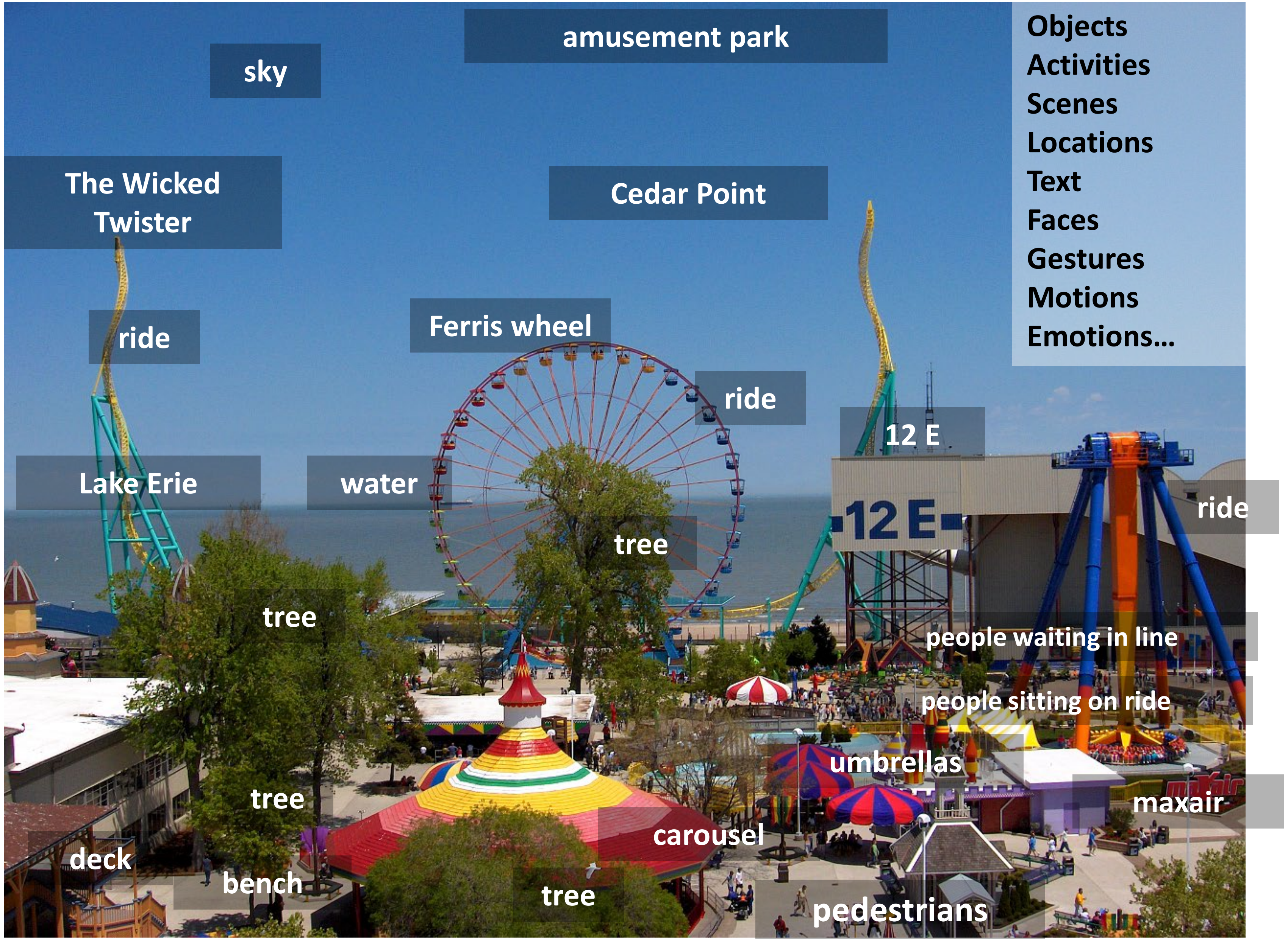
Broad problem areas of Computer Vision

- **Recognition**
  - *"What is it?"*
  - *Identify 3D objects, people, scenes, activities, . . .*
- **Range estimation**
  - *"Where is it?"*
  - *Compute properties of the 3D world from visual data*
  - *Related to biological **depth perception***
  - *Pose = Position + Orientation*
  - *Shape analysis*
- **Motion analysis**
  - *What direction is it moving? How fast?*
  - *Is the object rigid?*
  - *Is a collision imminent?*
- **Visual database search**
  - *CBIR = Content-Based Image Retrieval*

Vision for recognition, image interpretation

Slide credit: Grauman

Vision for recognition, image interpretation

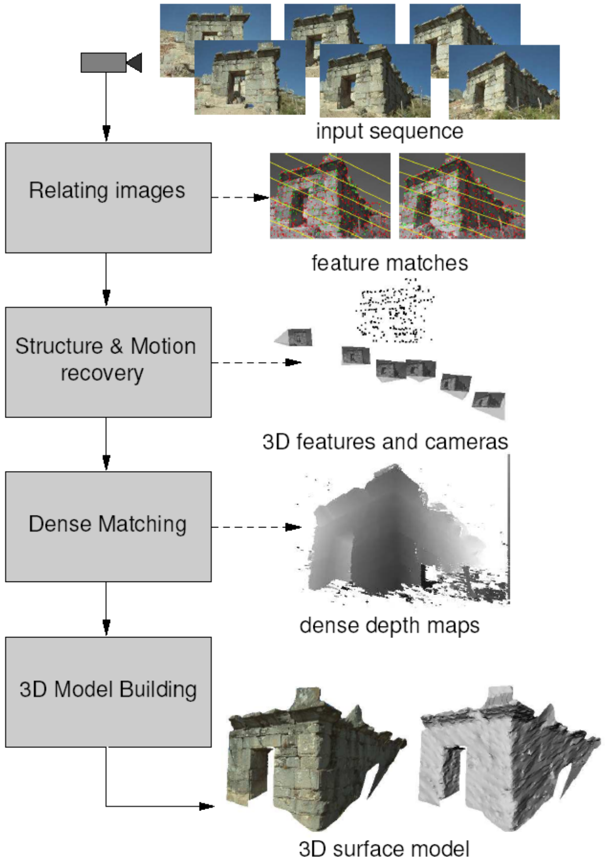


Objects
Activities
Scenes
Locations
Text
Faces
Gestures
Motions
Emotions…

Slide credit: Grauman

Motion analysis

Structure from motion

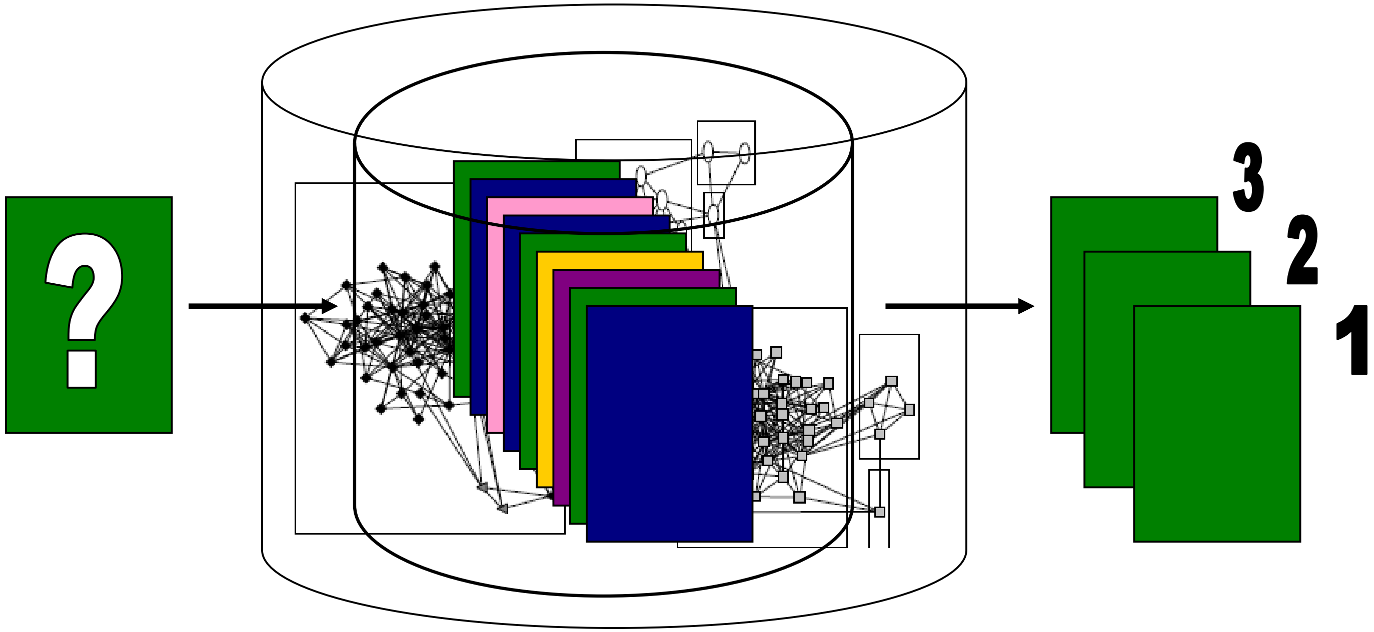


Snavely, et al.

Tracking

Demirdjian et al.

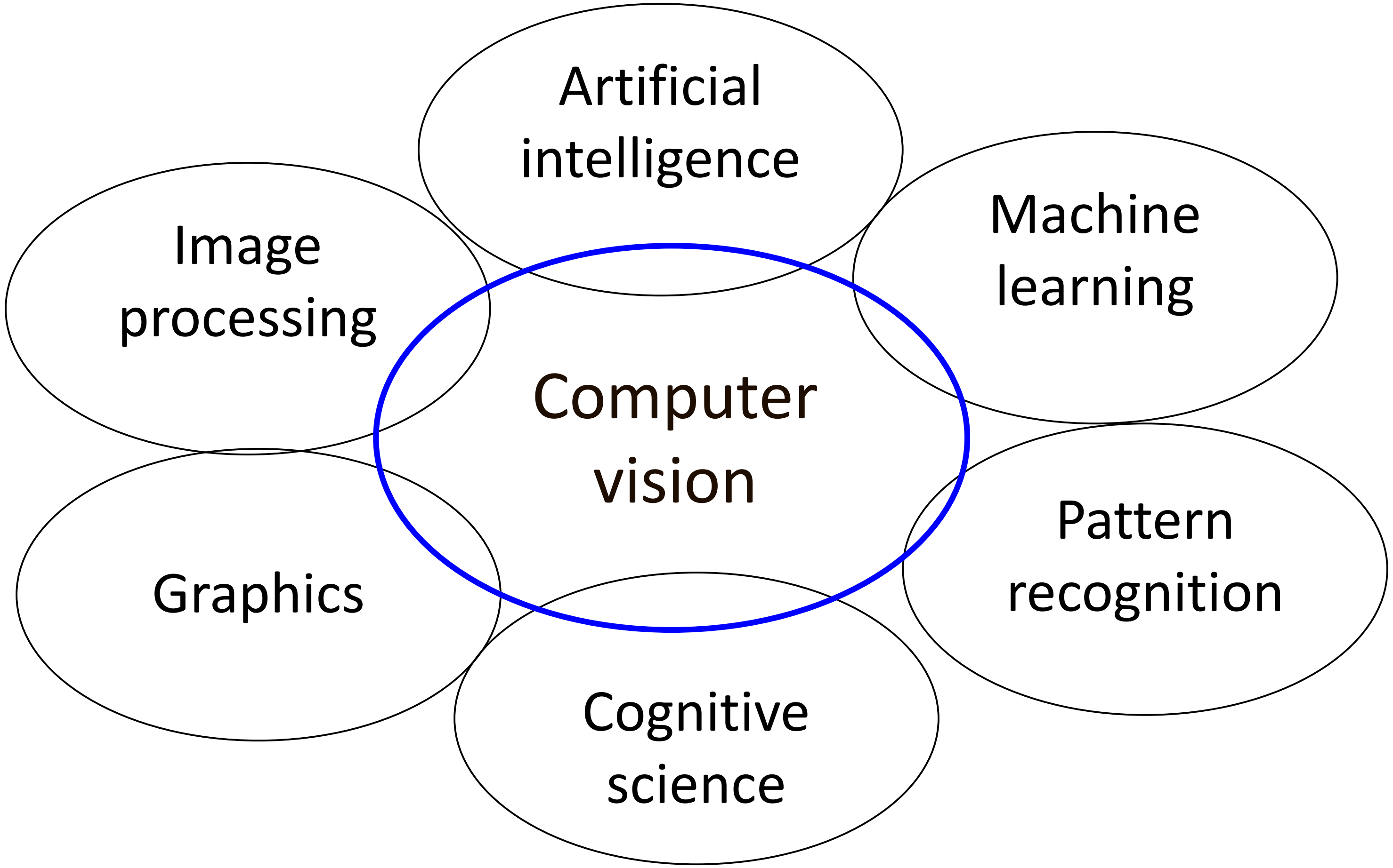
Visual database search, organization



Query **Image or video archives** **Relevant content**

Slide credit: Grauman

Related disciplines

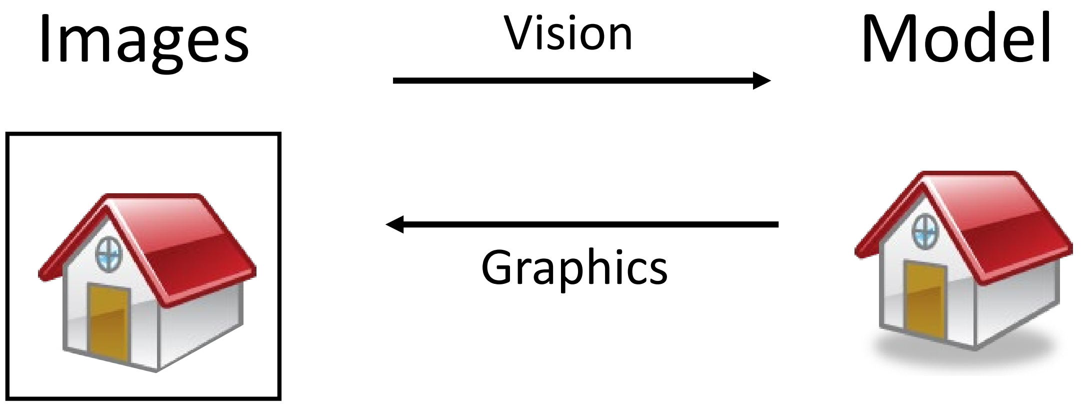


Slide credit: Grauman

Related fields (a simplistic view)

• *Image Processing*	Image →	Computer	→ Image
• *Computer Vision*	Image →	Computer	→ Description of image contents
• *Computer Graphics*	Description of image contents →	Computer	→ Image
• *Pattern Recognition*	Feature vector →	Computer	→ Class identifier
• *Photogrammetry*	Image →	Person	→ 3D distance measurements

Vision and graphics



Analysis and synthesis are inversely related

Slide credit: Grauman

A few **applications** of computer vision

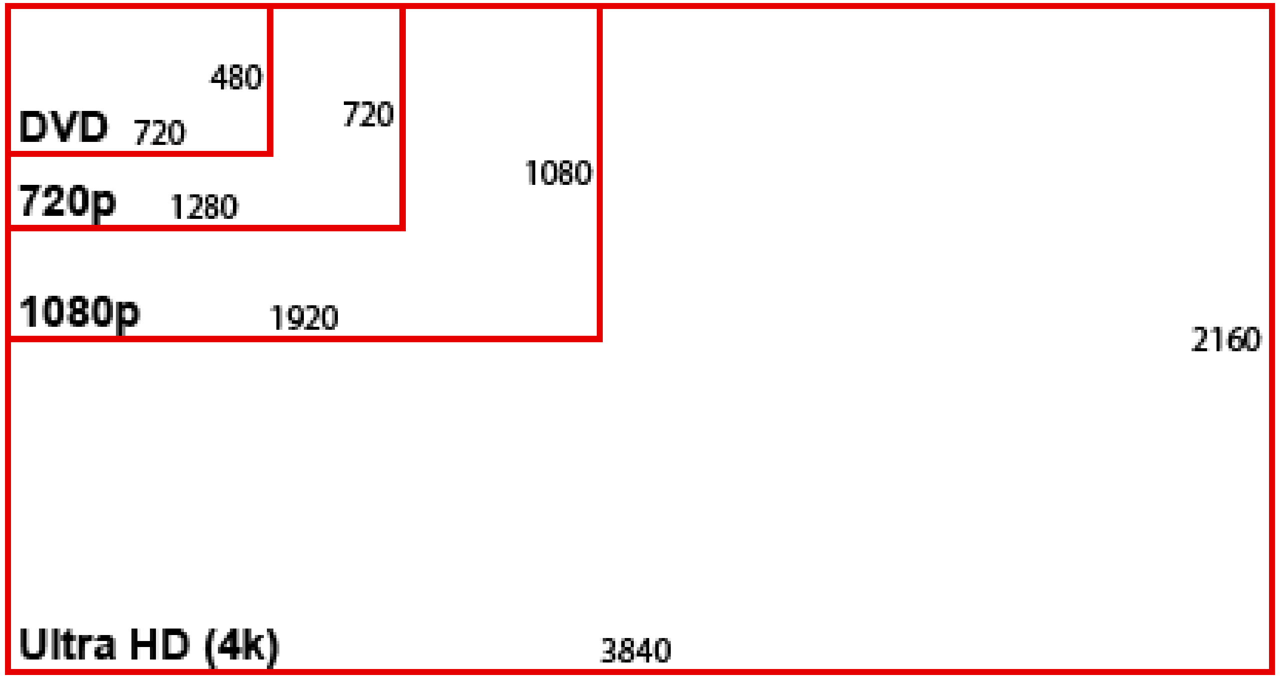
- Industrial inspection
  - Quality assurance
  - Metrology (= measurements)
- OCR (Optical Character Recognition)
- Document analysis
- Biometrics
  - Identification
    (Who is this person?)
  - Authentication
    (Is this person who he says he is?)
- Law enforcement
  - Surveillance
  - Reconstruction/visualization of crime scenes
  - . . .
- Remote sensing
- Robotics
- Autonomous vehicles
- Assistive Technology
- Medicine
  - Biomedical image analysis
  - Image-guided surgery
- Improved image/video compression
- Social media
  - Organizing and sharing photo collections
  - Tagging acquaintances
- Computational photography

In the simplest case, an image is a two-dimensional array of intensity values

- Each element of the array is called a **pixel** (picture element)
- Common sizes are 640x480, 1280x720, 1920 x1080, etc. (*columns* x *rows*)

  The Nikon D810: 7360 x 4912 (~36M pixels)

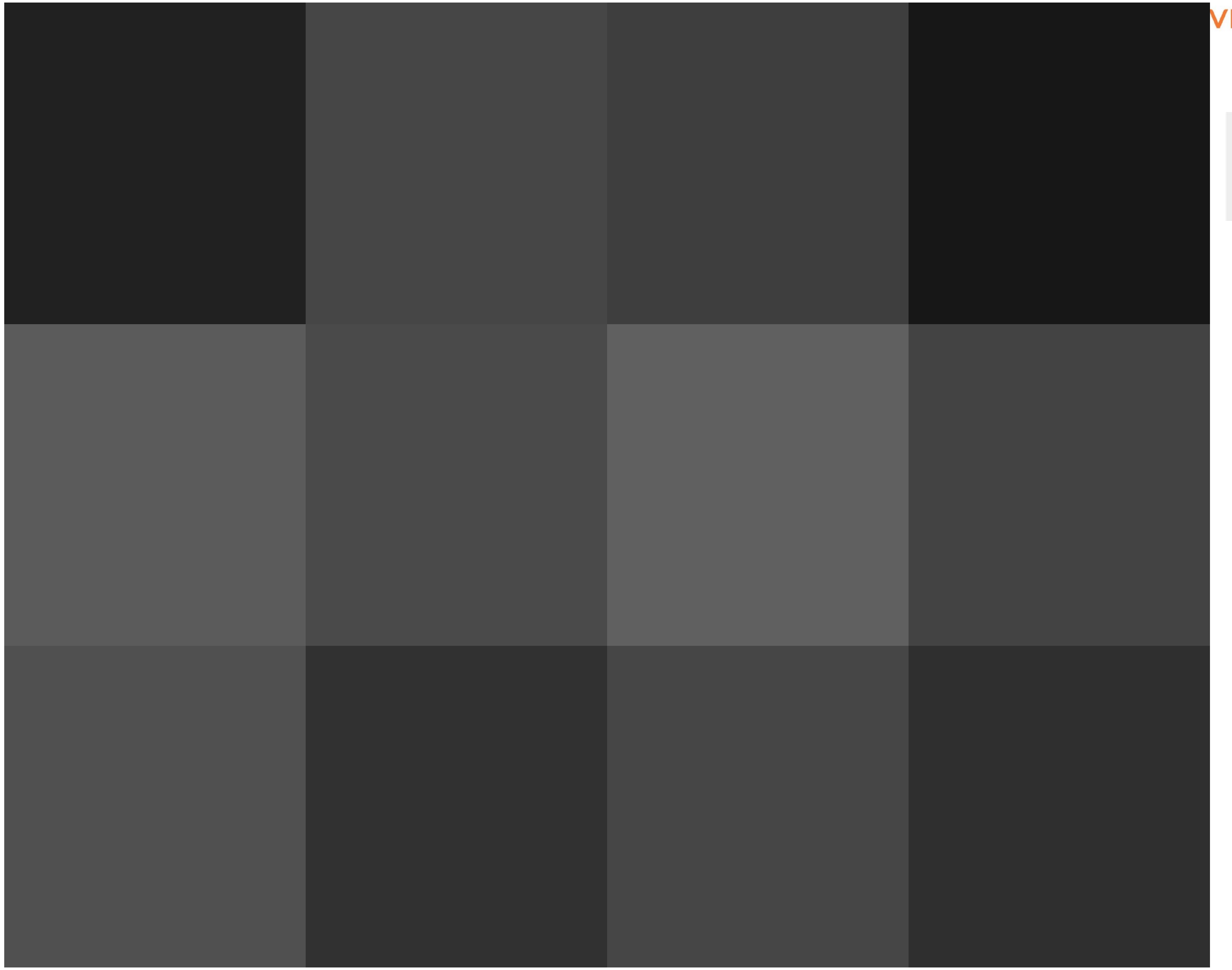
- More pixels: *potentially* more detail



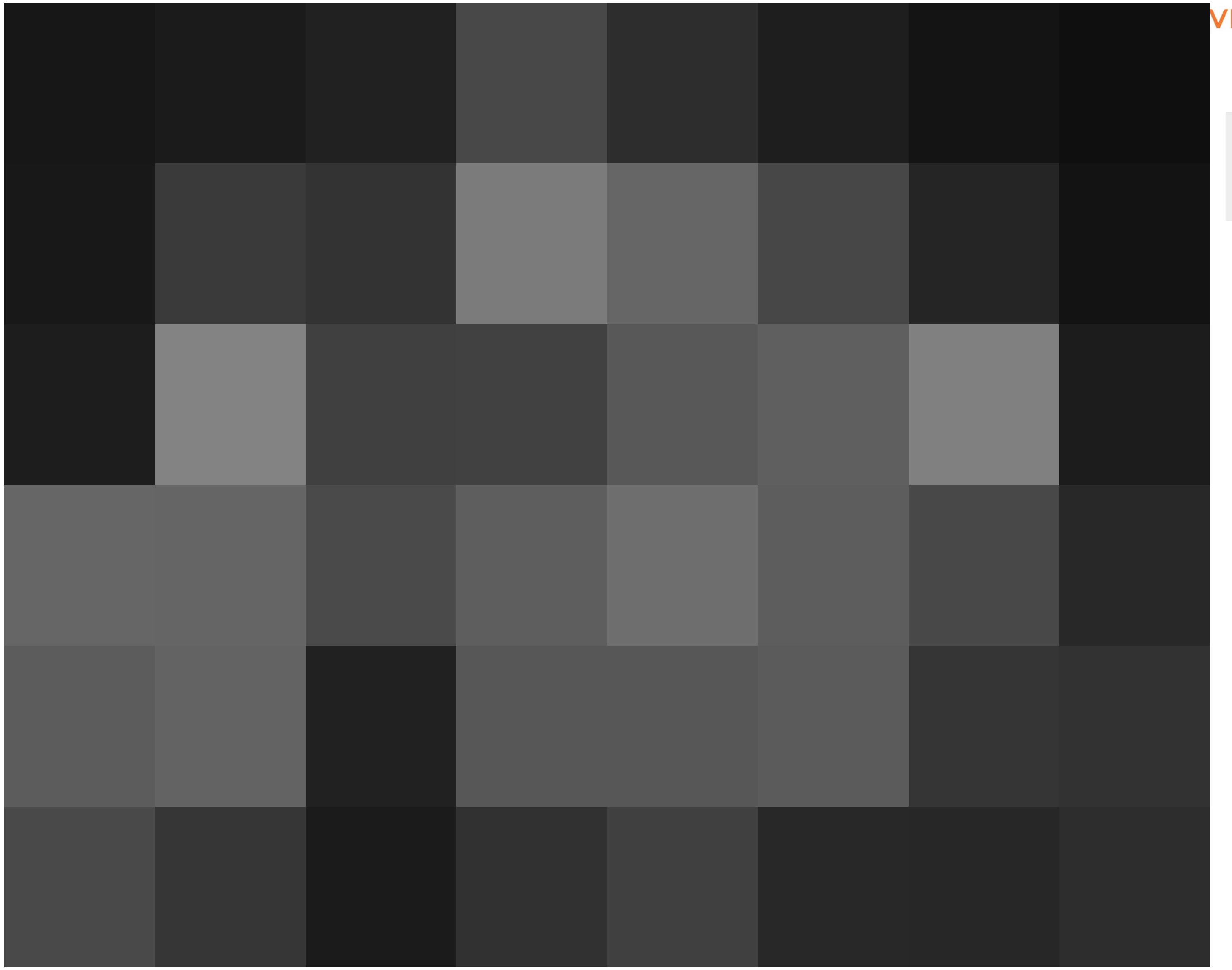
One pixel

2x2

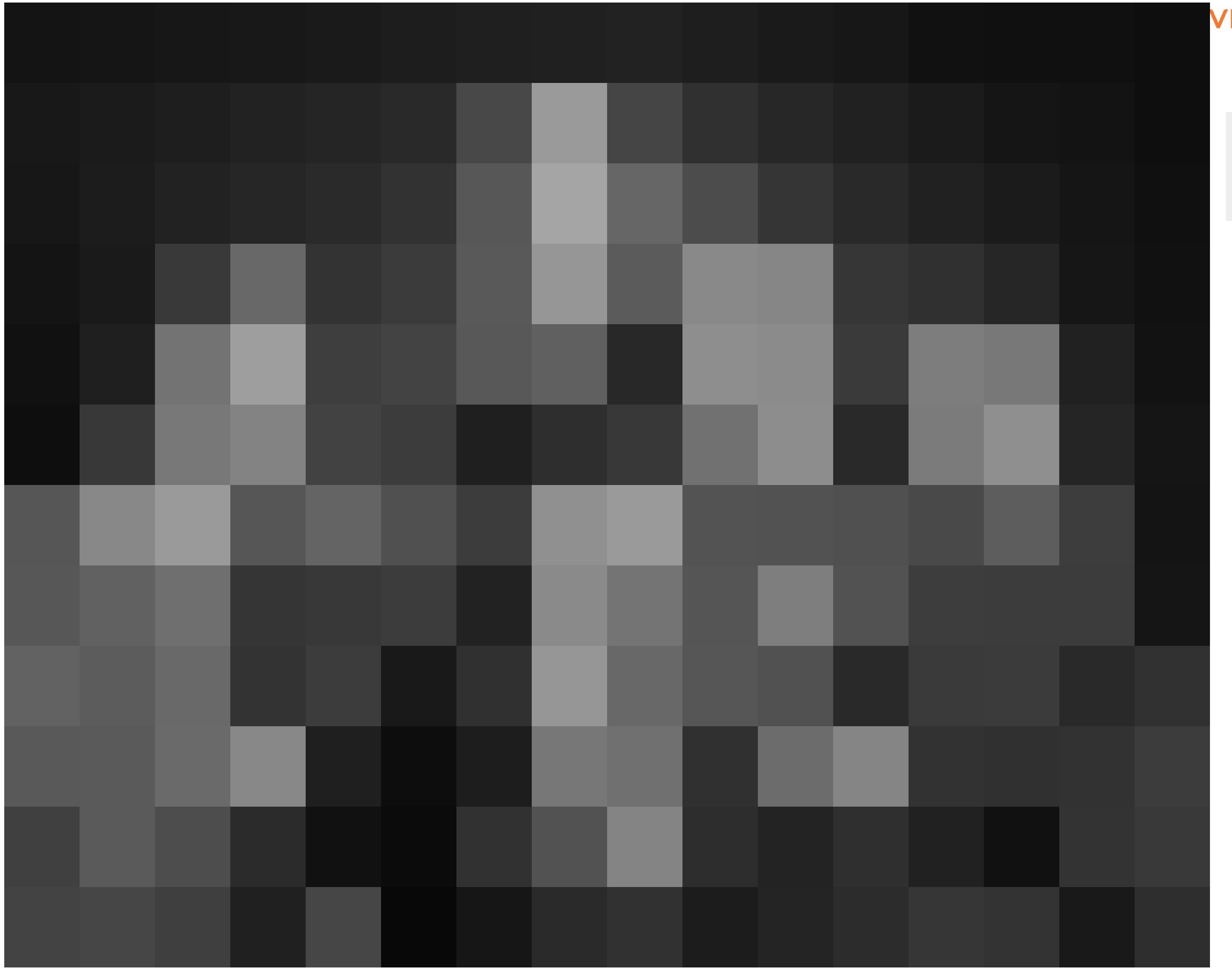
4x3



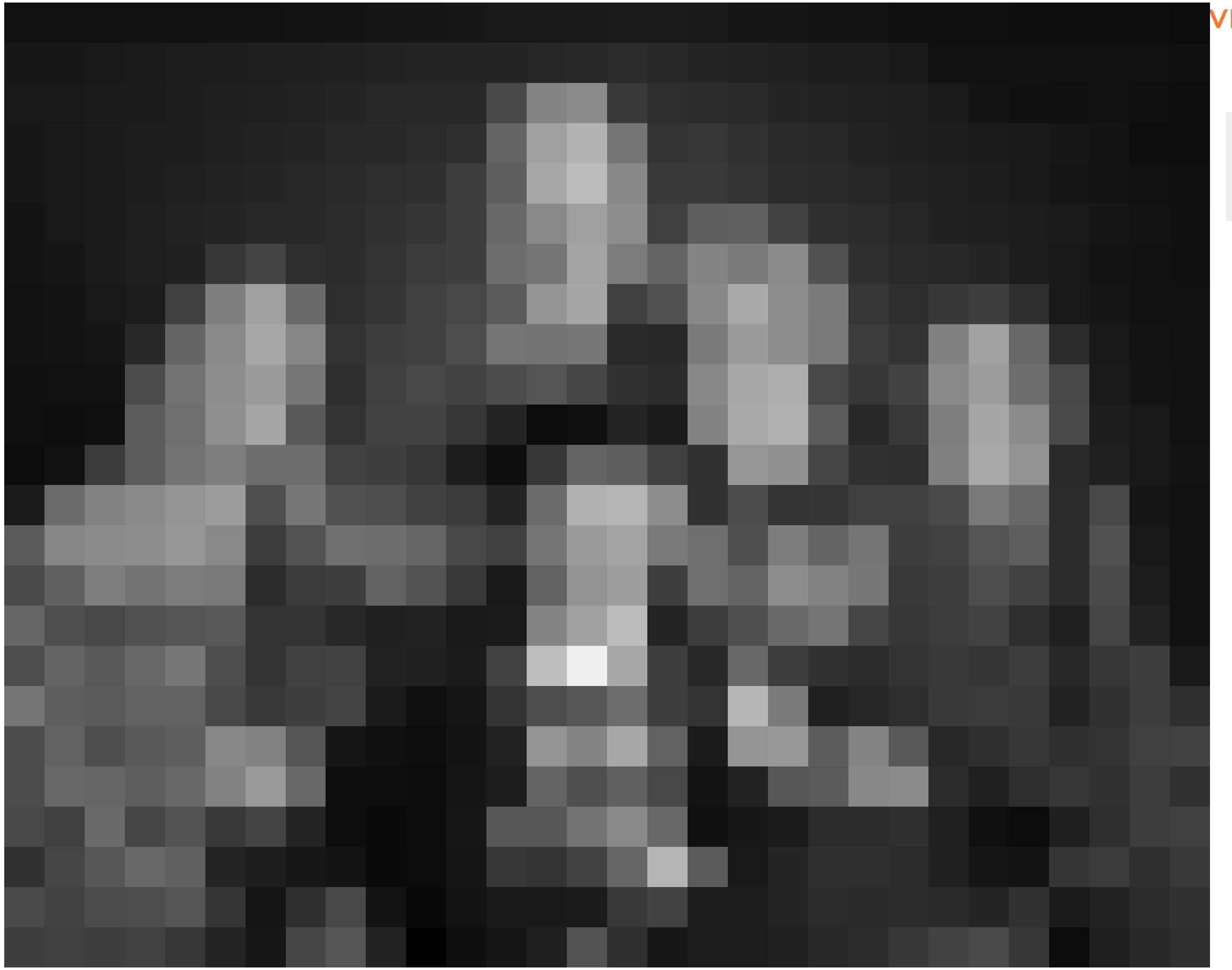
8x6



16x12



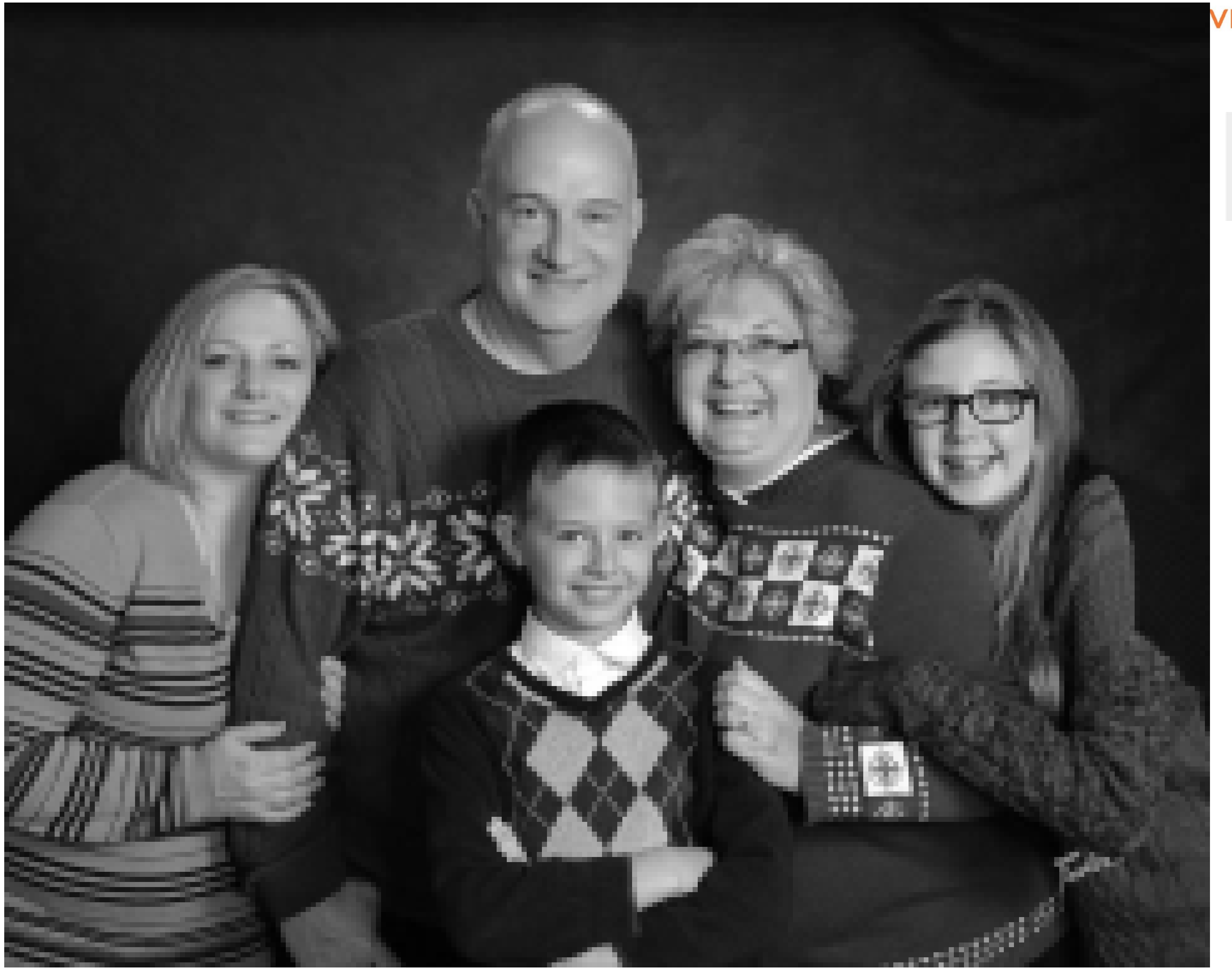
30x24



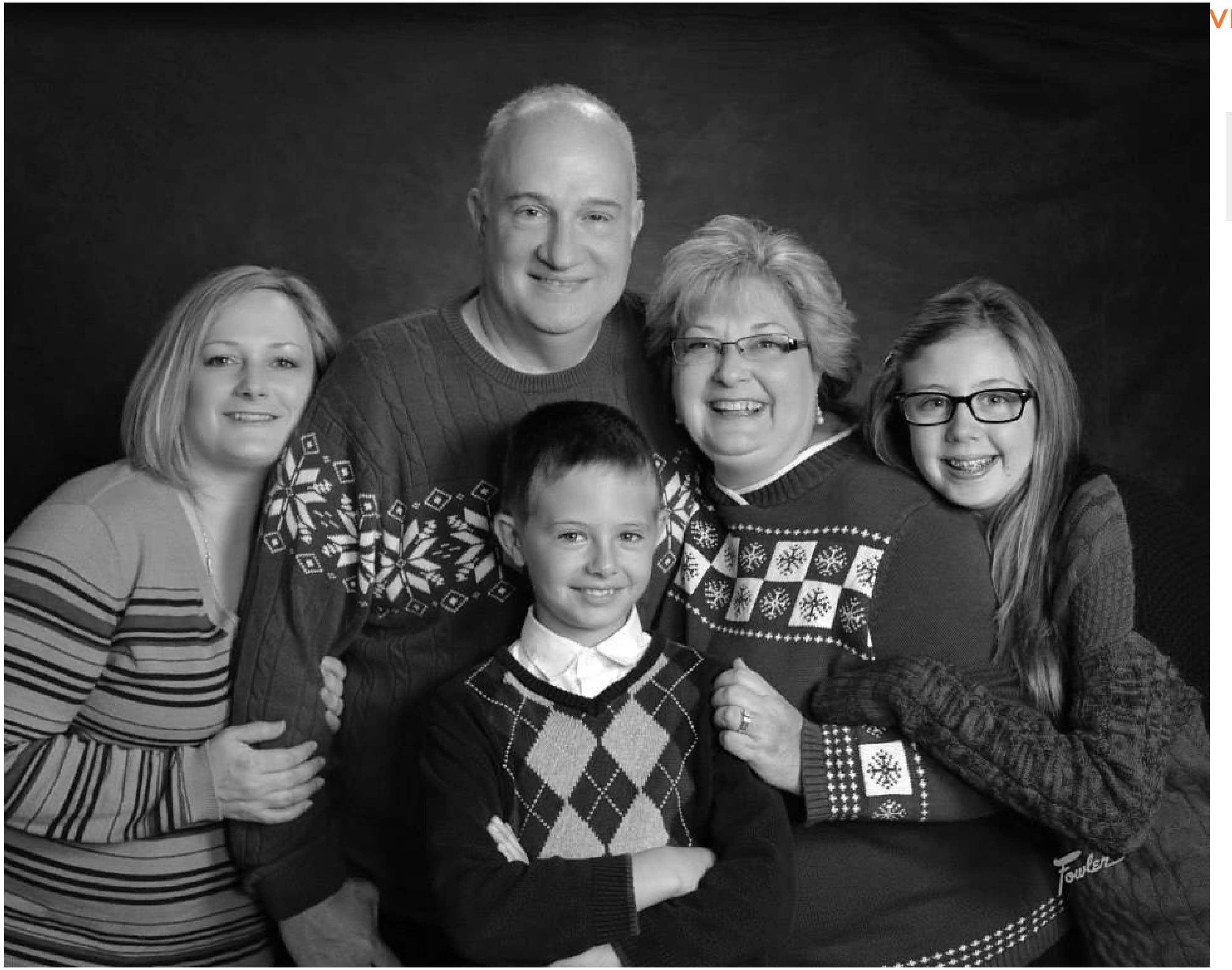
60x48

120x96

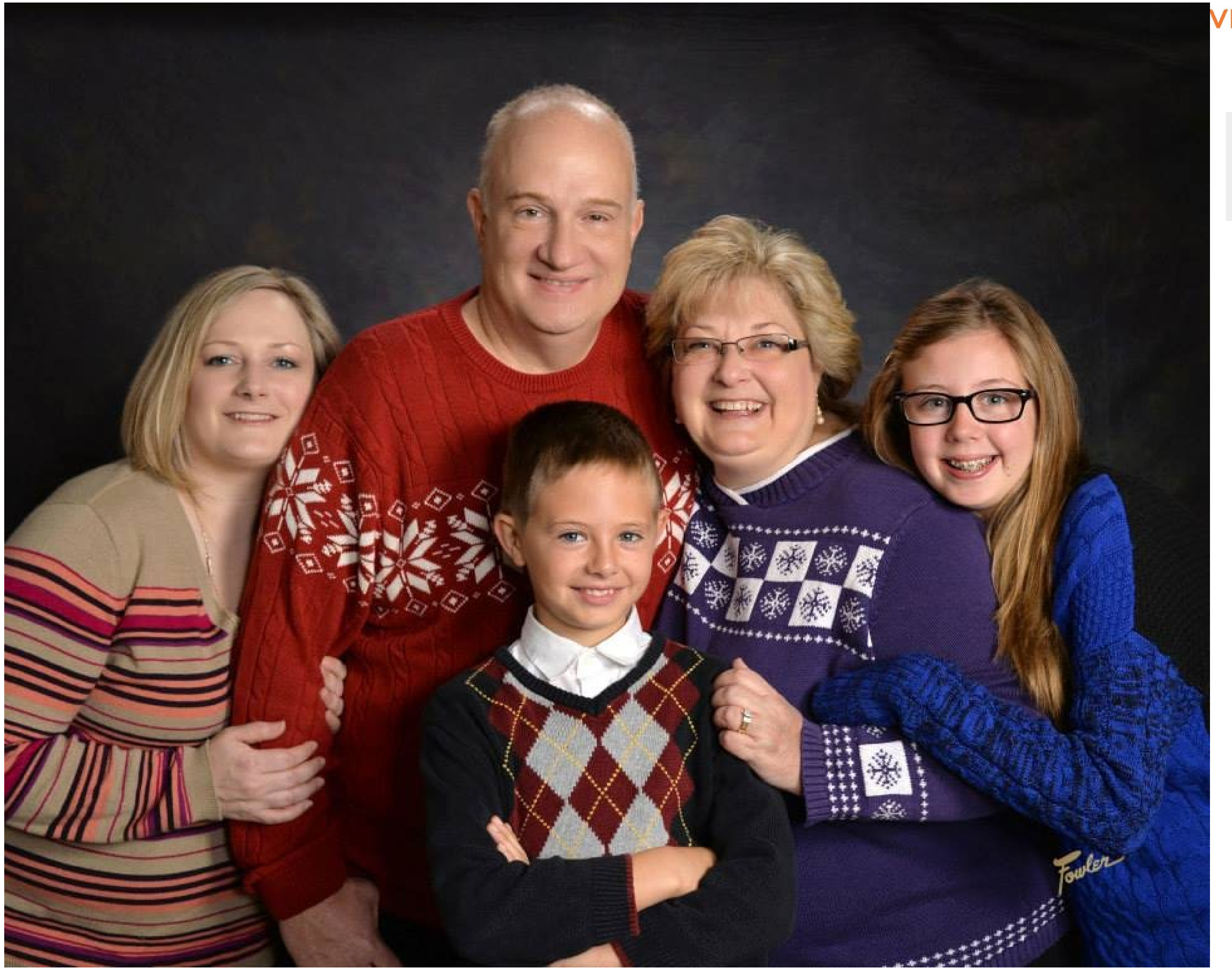
240x192



480x384



960x768



Basics of image representation

- For the <u>monochromatic</u> ("black and white") case, a typical image contains 8-bit pixels, so that intensity values range from 0 (black) to 255 (white)
- "Intensity values" are often called "pixel values" or "brightness values" or "gray levels"
- For a color image, we typically have separate information for red, green, and blue (RGB) wavelengths

RGB color

8-bit gray

3-bit gray

2-bit gray

1-bit gray (binary)

Image understanding

Object recognition

Surface reconstruction

. . .

Preprocessing

Noise reduction

Image enhancement

Feature extraction

Image formation

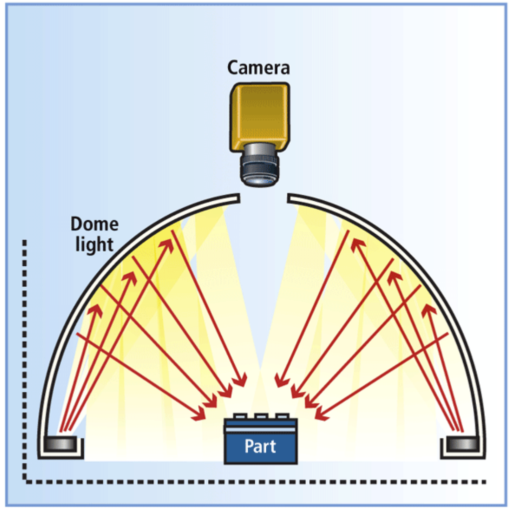
Optics, geometry, radiometry

Why study vision?

- To relieve humans of boring, easy tasks
- To enhance human abilities
  - Human-computer interaction
  - Virtual reality, augmented reality
- To use imaging devices and databases that seem to be available everywhere
- To develop “perceptual abilities” for machines
- To gain a better understanding of human perception
- It’s fun!

LIGHT

The first step in (most) image acquisition is light: either a controlled or uncontrolled source illuminates an object, and part of the light is sensed by a camera with associated optics to form a digital image



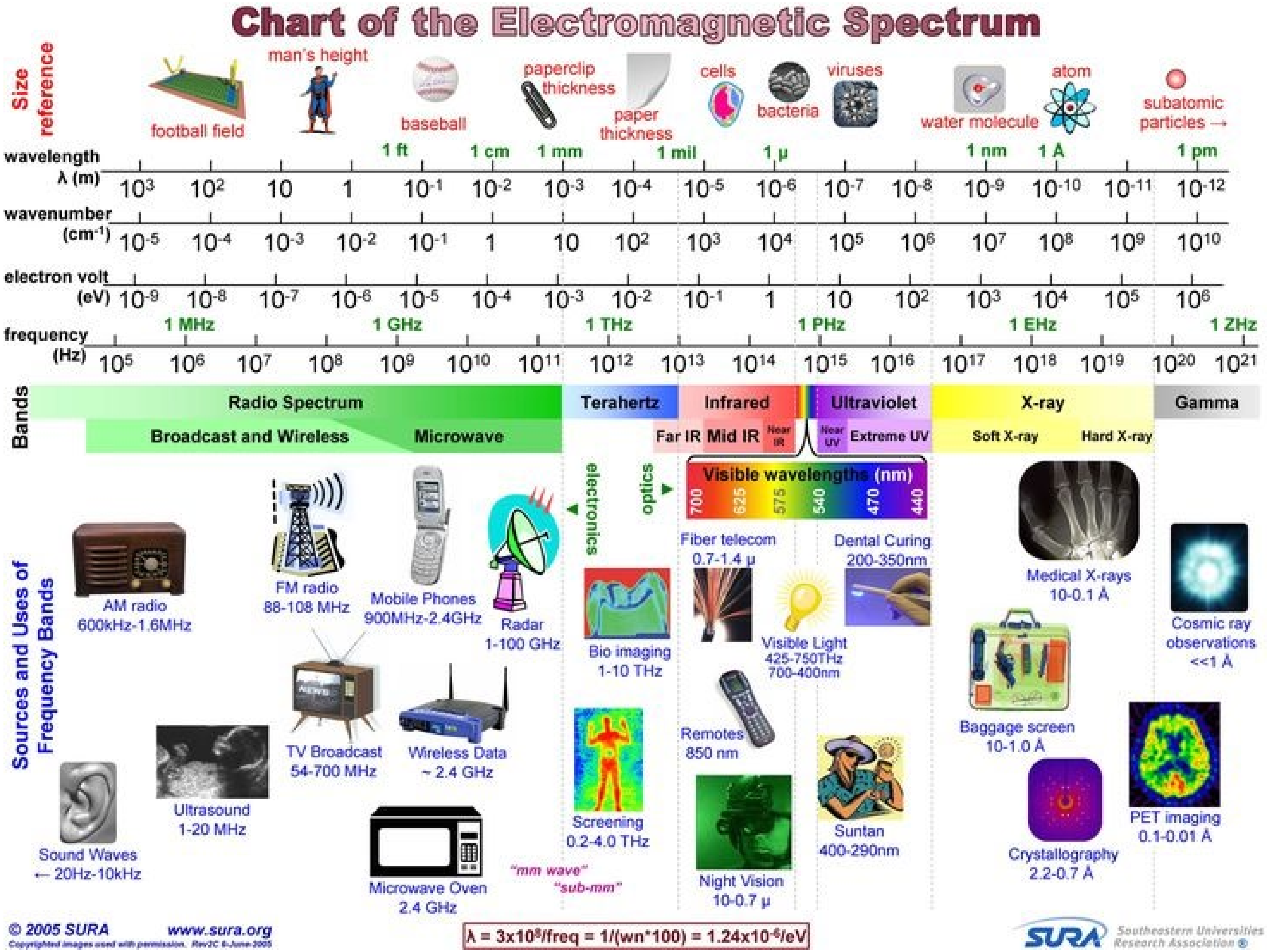
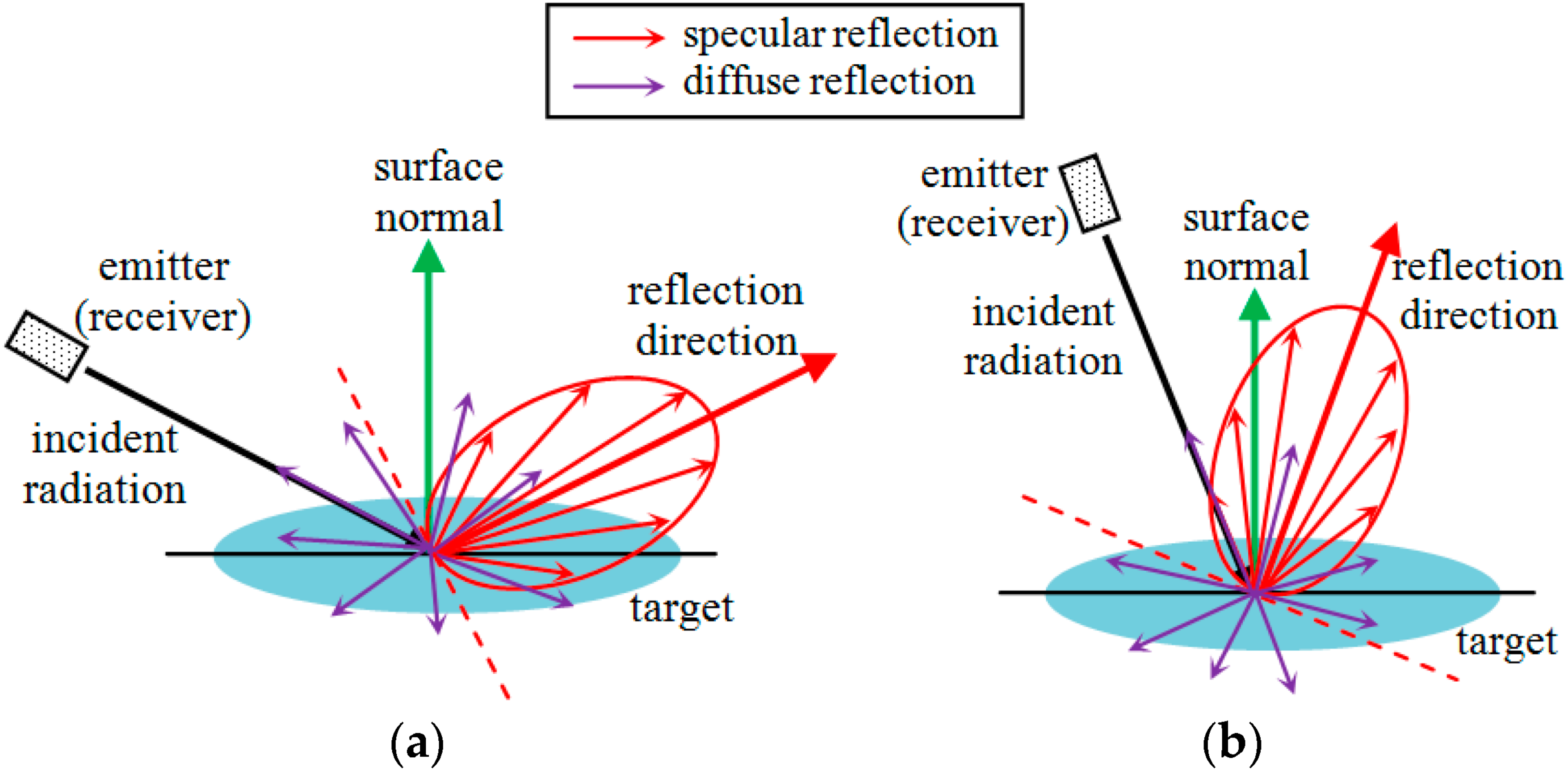


Chart of the Electromagnetic Spectrum
Size reference
football field
man's height
baseball
paperclip thickness
paper thickness
cells
bacteria
viruses
water molecule
atom
subatomic particles →
wavelength λ (m)
1 ft
1 cm
1 mm
1 mil
1 μ
1 nm
1 Å
1 pm
10^3 10^2 10 1 10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 10^{-9} 10^{-10} 10^{-11} 10^{-12}
wavenumber (cm^{-1})
10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1} 1 10 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^{10}
electron volt (eV)
10^{-9} 10^{-8} 10^{-7} 10^{-6} 10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1} 1 10 10^2 10^3 10^4 10^5 10^6
frequency (Hz)
1 MHz
1 GHz
1 THz
1 PHz
1 EHz
1 ZHz
10^5 10^6 10^7 10^8 10^9 10^{10} 10^{11} 10^{12} 10^{13} 10^{14} 10^{15} 10^{16} 10^{17} 10^{18} 10^{19} 10^{20} 10^{21}
Bands
Radio Spectrum
Broadcast and Wireless
Microwave
Terahertz
Infrared
Far IR
Mid IR
Near IR
Ultraviolet
Near UV
Extreme UV
X-ray
Soft X-ray
Hard X-ray
Gamma
electronics
optics
Visible wavelengths (nm)
700 625 575 540 470 440
Sources and Uses of Frequency Bands
AM radio
600kHz-1.6MHz
FM radio
88-108 MHz
Mobile Phones
900MHz-2.4GHz
Radar
1-100 GHz
Bio imaging
1-10 THz
Fiber telecom
0.7-1.4 μ
Dental Curing
200-350nm
Medical X-rays
10-0.1 Å
Cosmic ray observations
<<1 Å
Visible Light
425-750THz
700-400nm
TV Broadcast
54-700 MHz
Wireless Data
~ 2.4 GHz
Remotes
850 nm
Baggage screen
10-1.0 Å
Sound Waves
← 20Hz-10kHz
Ultrasound
1-20 MHz
Screening
0.2-4.0 THz
Suntan
400-290nm
PET imaging
0.1-0.01 Å
Crystallography
2.2-0.7 Å
Microwave Oven
2.4 GHz
"mm wave"
"sub-mm"
Night Vision
10-0.7 μ
© 2005 SURA
www.sura.org
Copyrighted images used with permission. Rev2C 6-June-2005
λ = 3x10⁸/freq = 1/(wn*100) = 1.24x10⁻⁶/eV
SURA Southeastern Universities Research Association ®

REFLECTION OF LIGHT

Light incident on an opaque surface is subject to several phenomena: absorption (not shown), diffuse reflection and specular reflection



Specular vs. Diffuse reflection is often determined by the texture of the surface (a mirror reflects in a specular manner; a sheet of white paper in a diffuse manner)

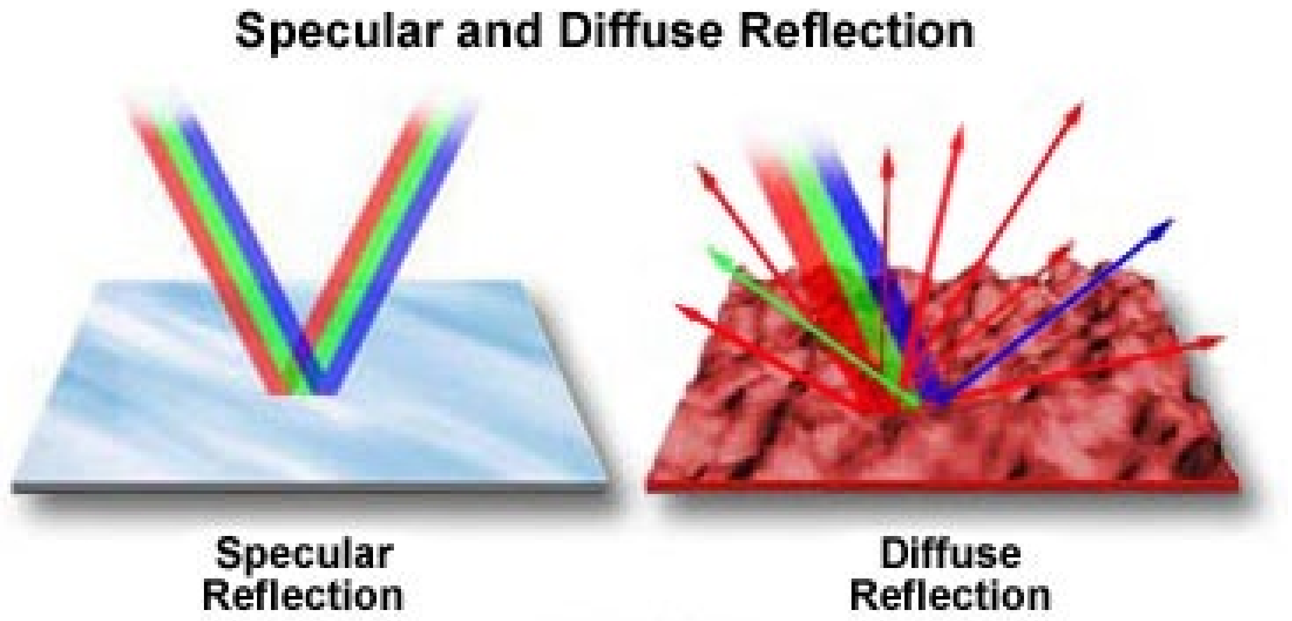
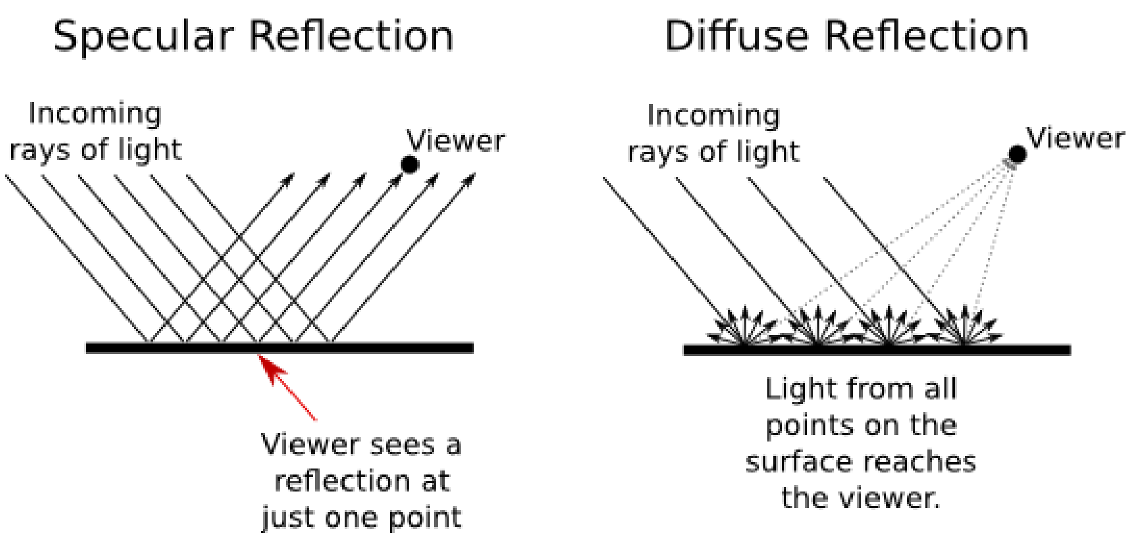
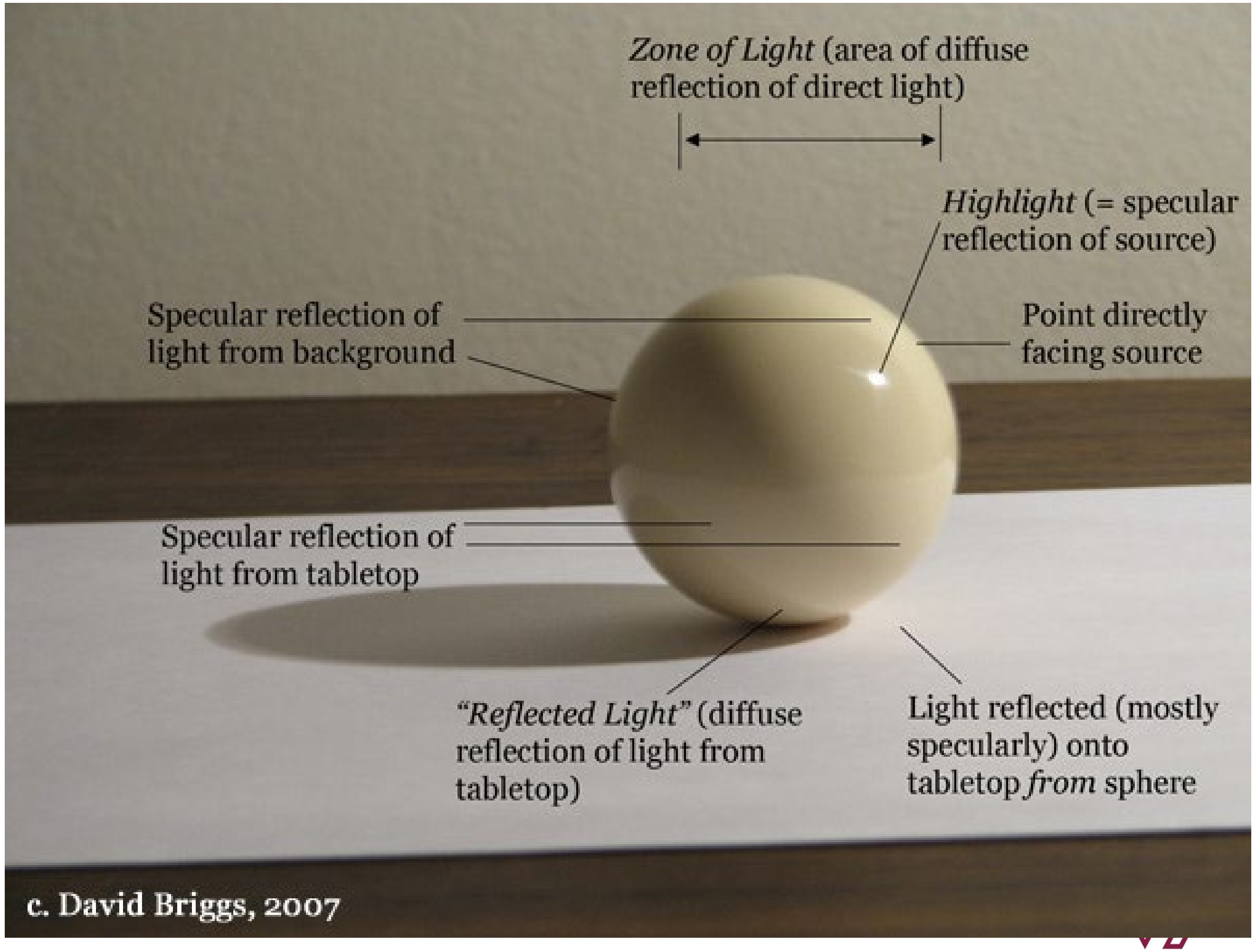


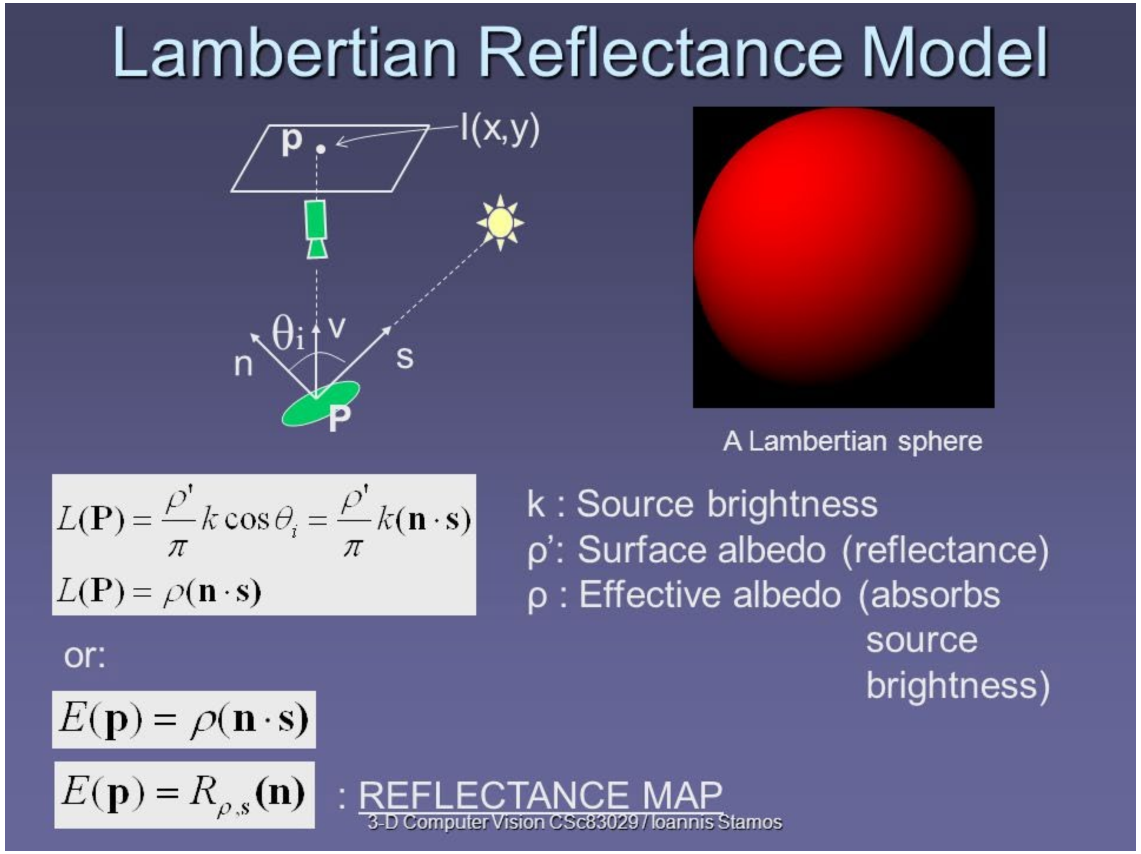
Figure 2

A viewer will only see specular reflection at (close to) the complementary angle from the light source; diffuse reflection will be seen at many different angles and locations on the object





Lambertian reflectance is a model for a diffuse (matte) surface



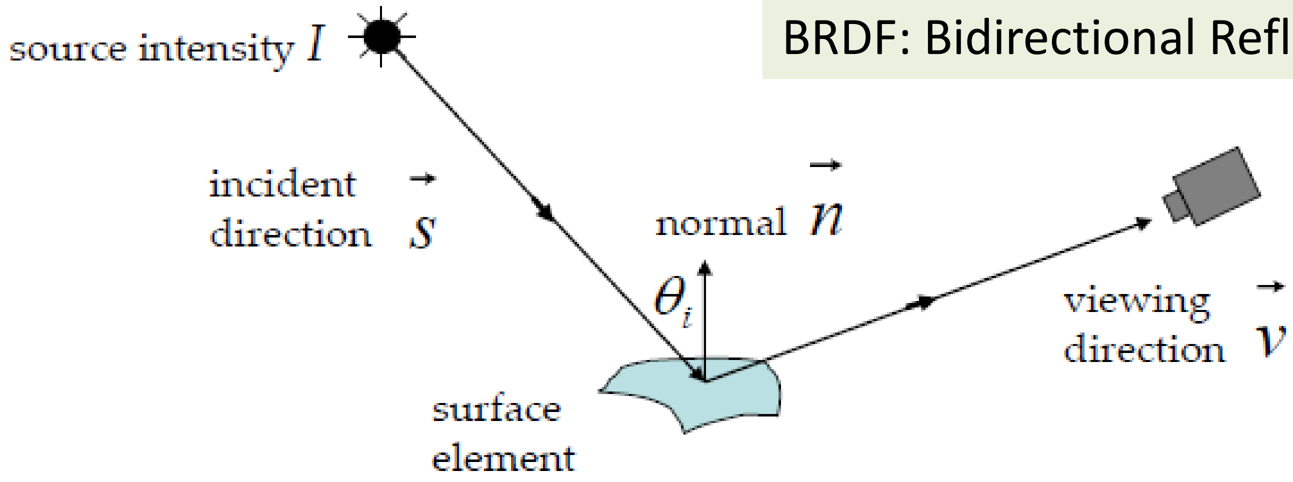
- The reflected intensity of light can be calculated as

$$I_R = \vec{L} \cdot \vec{N} \rho I_I$$

- where
  - I_R is the reflected intensity
  - $\vec{L}$ is a direction vector from the surface to the light source
  - $\vec{N}$ is a surface normal vector
  - ρ is the albedo (related to color of th surface)
  - I_I is the intensity of the illuminating light, measured at the surface

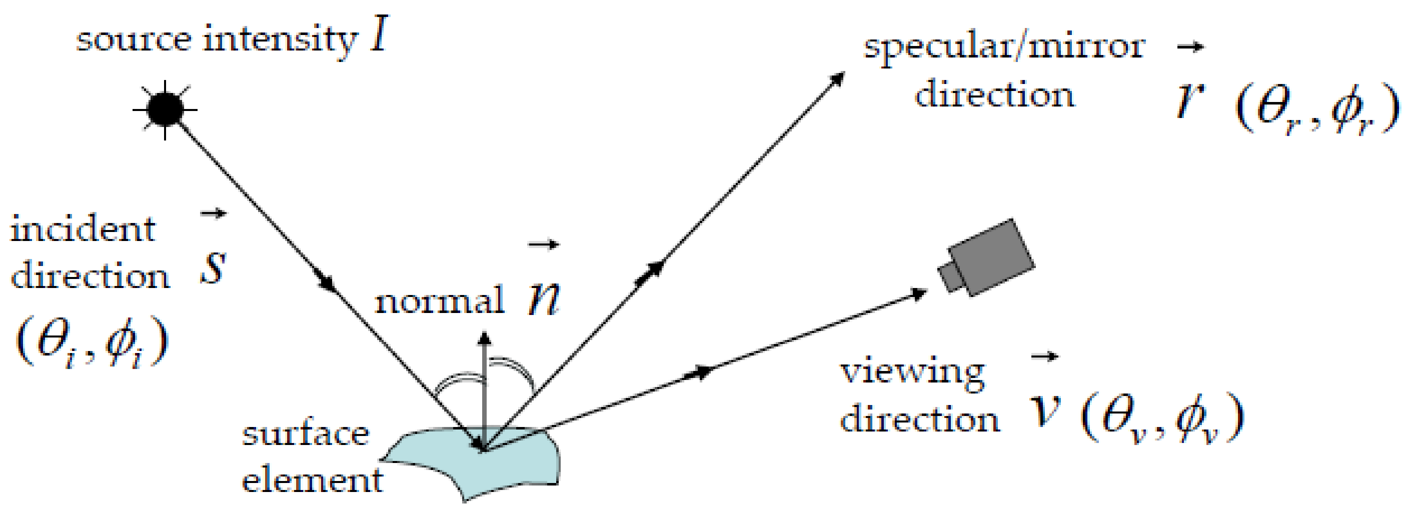
Diffuse Reflection and Lambertian BRDF

BRDF: Bidirectional Reflectance Distribution Function



- Surface appears equally bright from ALL directions! (independent of $\vec{v}$)
- Lambertian BRDF is simply a constant : $f(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{\rho_d}{\pi}$ albedo
- Surface Radiance : $L = \frac{\rho_d}{\pi} I \cos\theta_i = \frac{\rho_d}{\pi} I \; \vec{n} \cdot \vec{s}$ source intensity
- Commonly used in Vision and Graphics!

Specular Reflection and Mirror BRDF



- Very smooth surface.
- All incident light energy reflected in a SINGLE direction. (only when $\vec{v} = \vec{r}$)
- Mirror BRDF is simply a double-delta function :

specular albedo

$$f(\theta_i, \phi_i; \theta_v, \phi_v) = \rho_s\ \delta(\theta_i - \theta_v)\ \delta(\phi_i + \pi - \phi_v)$$

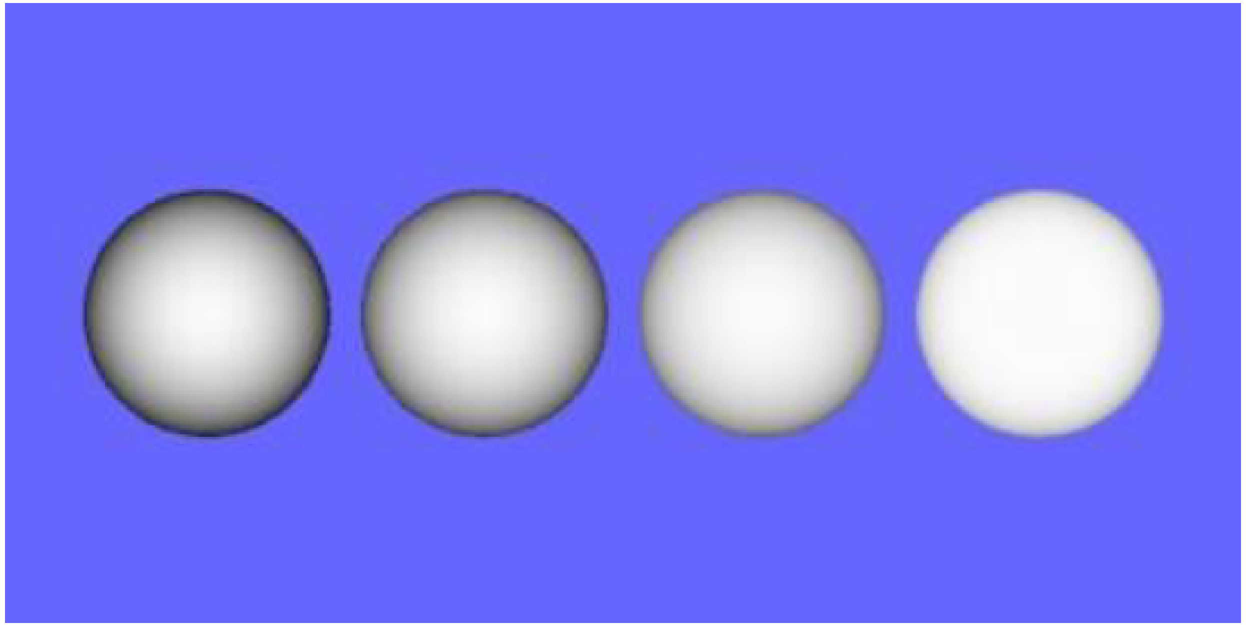
- Surface Radiance : $L = I\ \rho_s\ \delta(\theta_i - \theta_v)\ \delta(\phi_i + \pi - \phi_v)$

Why does the Full Moon have a flat appearance?

- The moon appears matte (or diffuse)

- But still, edges of the moon look bright (not close to zero) when illuminated by earth's radiance.

Surface Roughness Causes Flat Appearance



Increasing surface roughness →

Lambertian model

Valid for only SMOOTH MATTE surfaces.

Bad for ROUGH MATTE surfaces.

Take a minute and look at a few objects around you

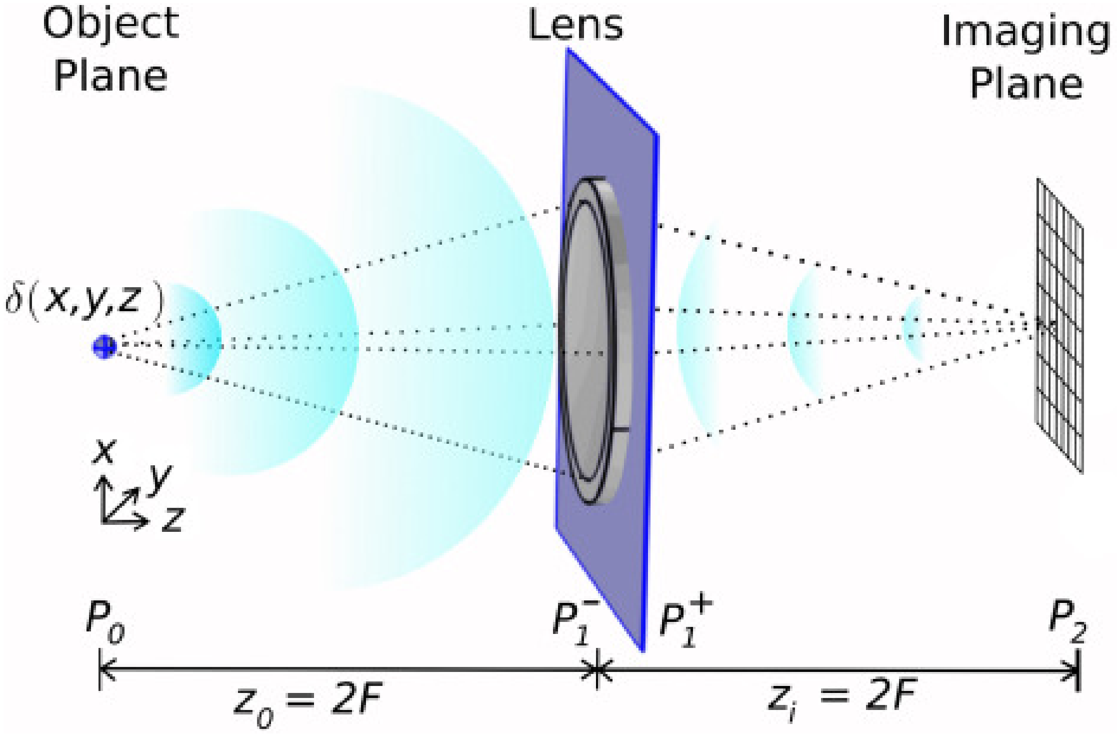
- What color do they appear?
  - So, what colors do they *absorb?*

- Is the surface rough, smooth, or textured?
  - How can you tell from looking at it?
  - What is the difference between *rough* and *textured?*

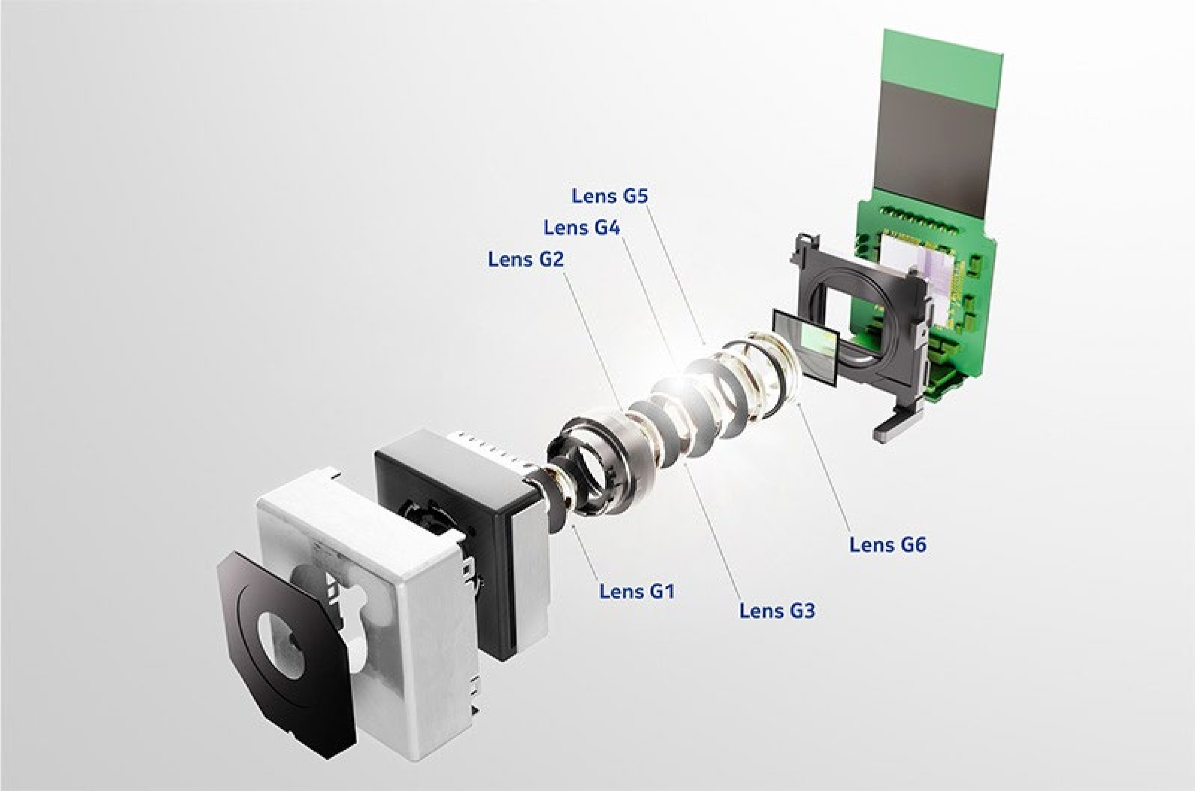
- Can you tell from looking at the object where the light sources are?
  - How?

IMAGING COMPONENTS

Since light emitted or reflected by an object diverges, we use a lens to focus the image onto an *imaging plane* where we can sense the pattern of light



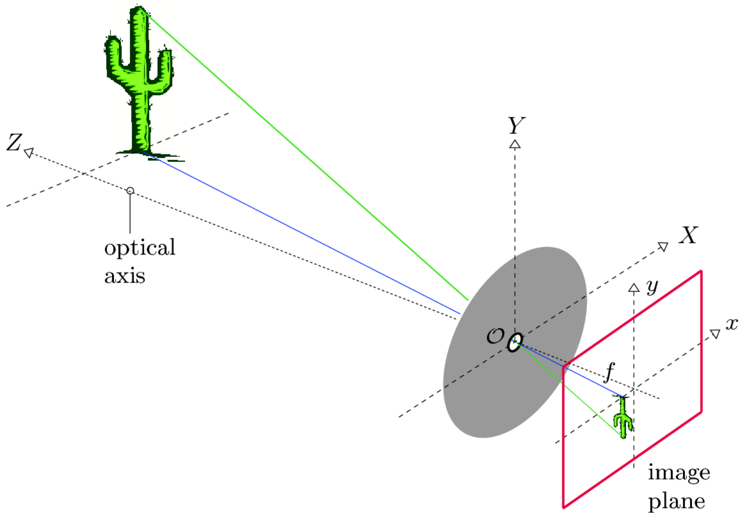
https://www.sciencedirect.com/science/article/pii/S2213597917300216



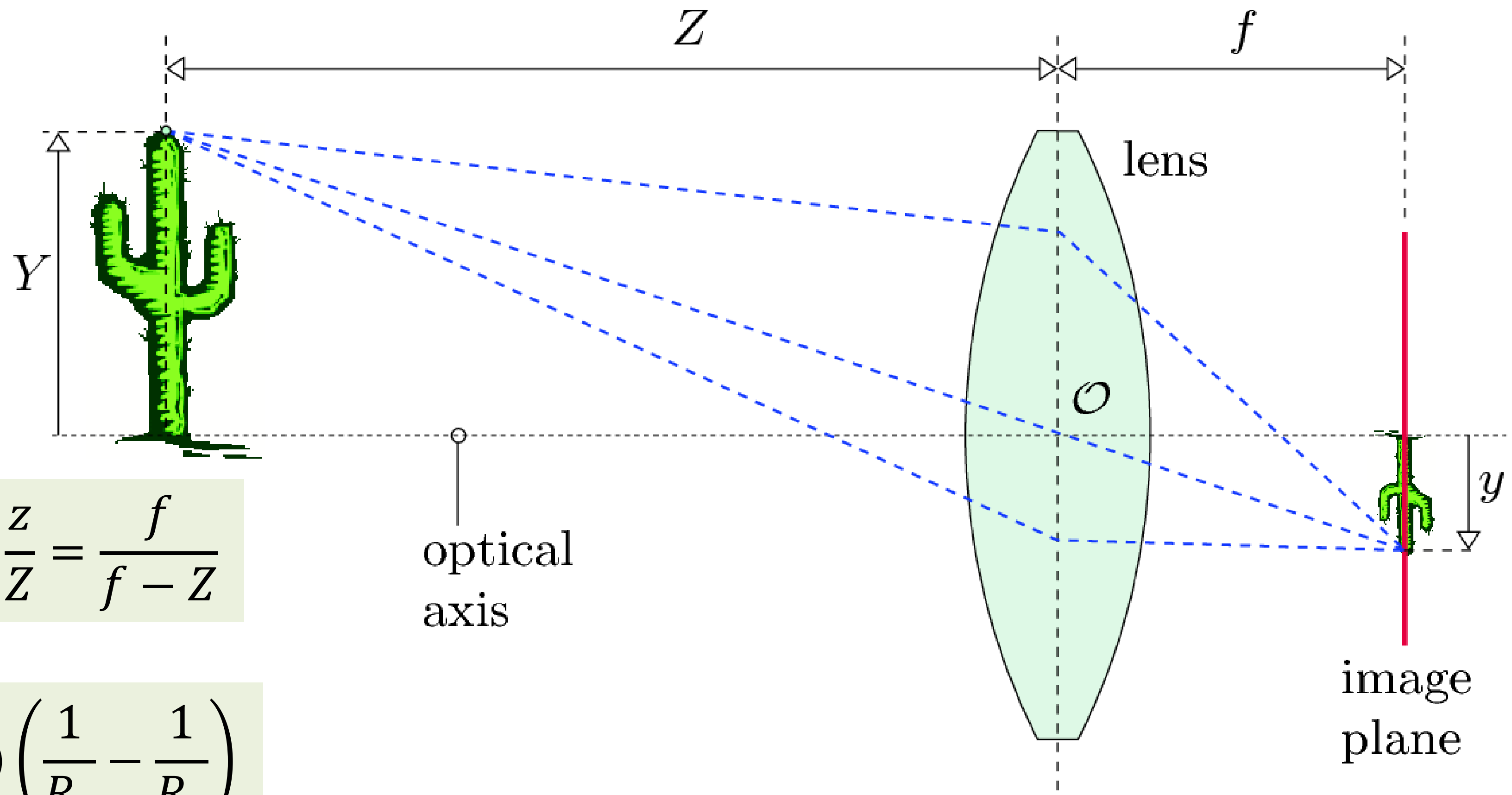
https://www.techspot.com/guides/850-smartphone-camera-hardware/

Ray tracing of a simple pinhole camera shows how light from an object creates an "image" in the focal plane;

ray tracing is based on the rectilinear propagation of light through a single medium (ignoring diffraction and bending by gravity)



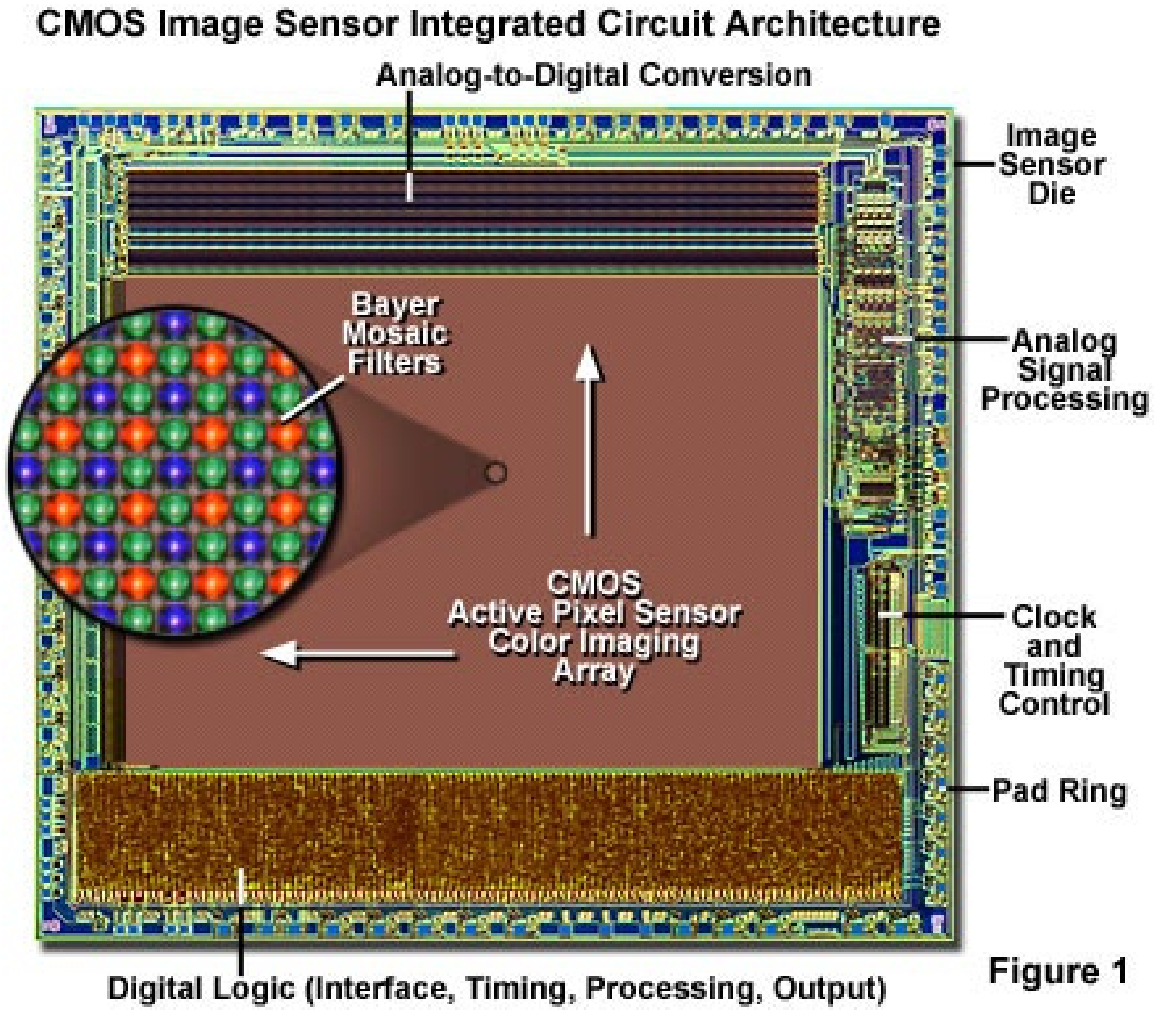
A "thin-lens" system uses the refractive properties of the boundary between glass and air to bend and focus the light from an object



$$M = \frac{y}{Y} = -\frac{z}{Z} = \frac{f}{f - Z}$$

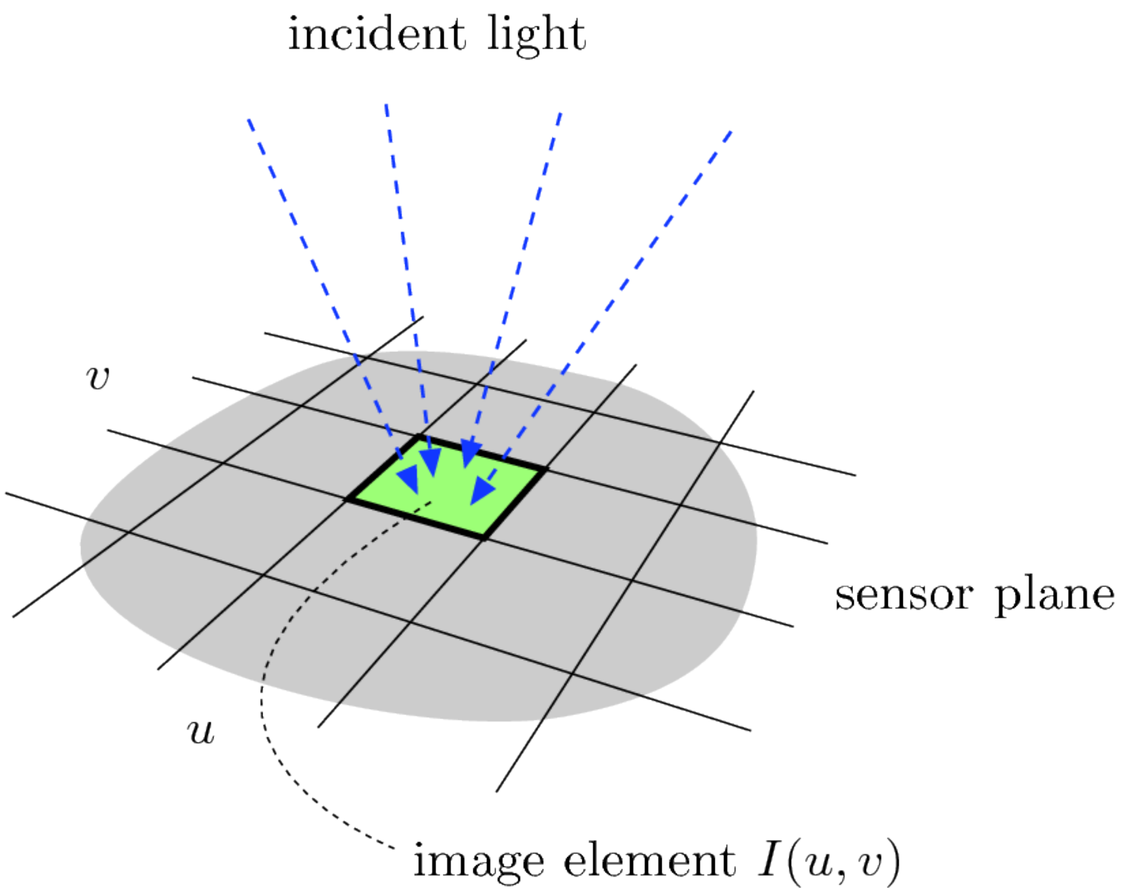
$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)$$

A CMOS image sensor sits in the image plane to sense the intensity (and color) of the light at different places in the image formed by the optics

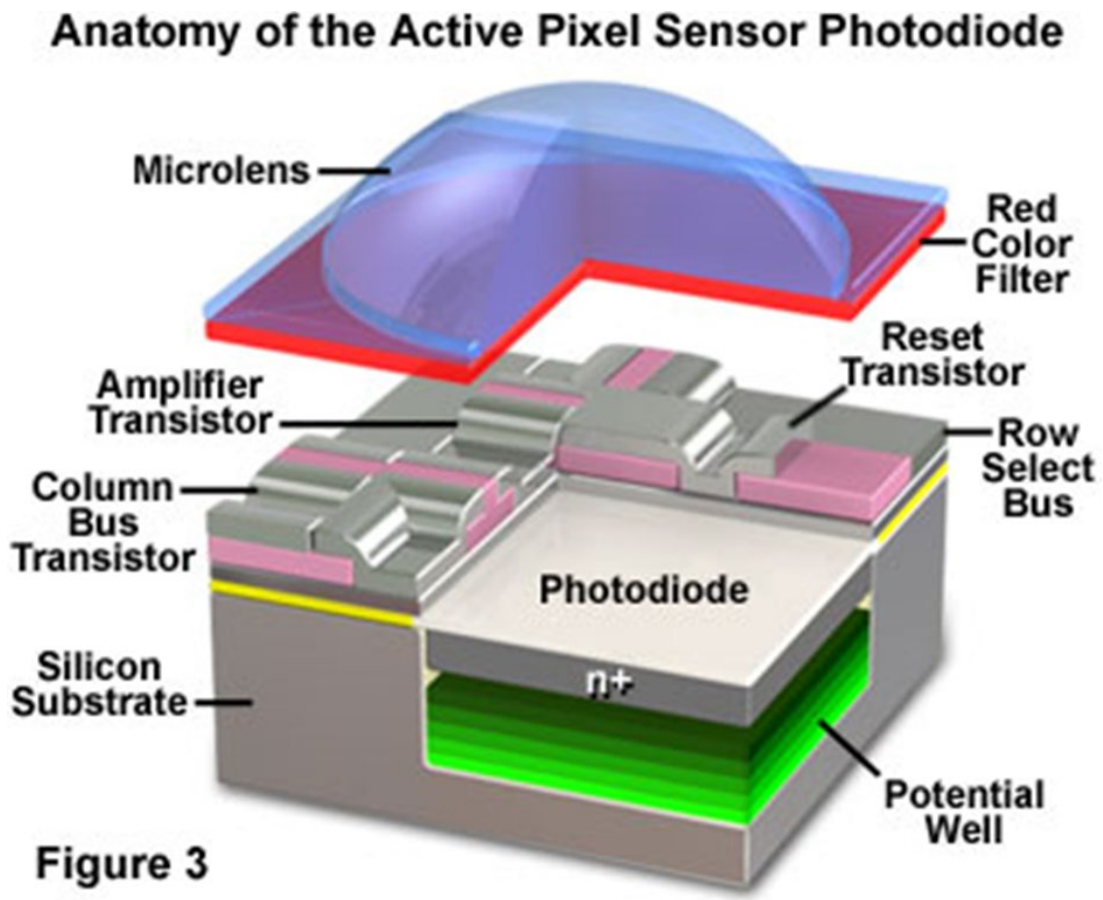


https://micro.magnet.fsu.edu/primer/digitalimaging/images/cmos/cmoschipsfigure1.jpg

Each location on an image sensor collects light from all directions, which causes a fuzzy image; it's the role of the optics to focus the incoming rays



Each photo sensing site (pixel) in the array senses one location in the image – if the sensor is in the focal plane



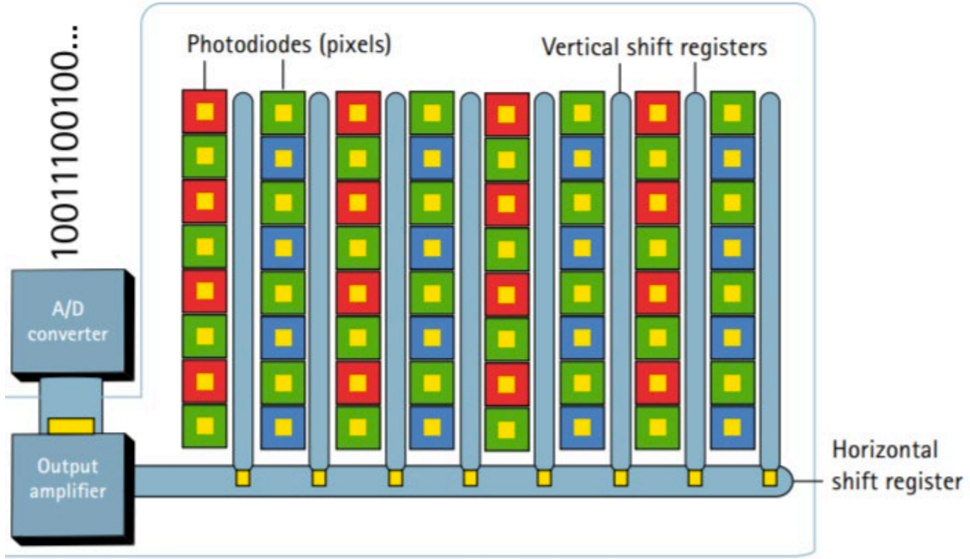
- How much energy is collected by the sensor?
- Assuming that the image intensity at the sensor is $0.01\frac{W}{m^2}$, the energy in one 60th of a second frame is:

$$0.01\frac{W}{m^2}\cdot 13\mu m^2\cdot\frac{1}{60}sec = 2.82\times 10^{-14}J$$

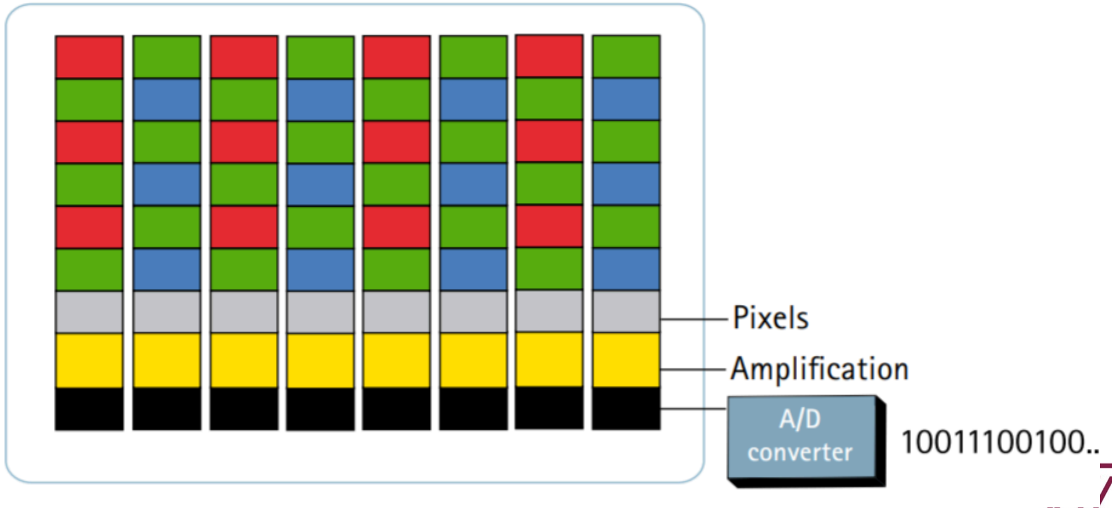
- Even less, actually, due to:
  - Inefficiency of the lens
  - Filtering (only certain colors)
  - Inefficiency of the silicon photodiode

Image sensor devices are usually either CMOS (complementary metal-oxide semiconductor) or CCD (charge-coupled device)

- CCDs are the older technology
  - Image data shifts out through the ends of rows or columns
  - Pixel sites are a bit larger and more well-behaved



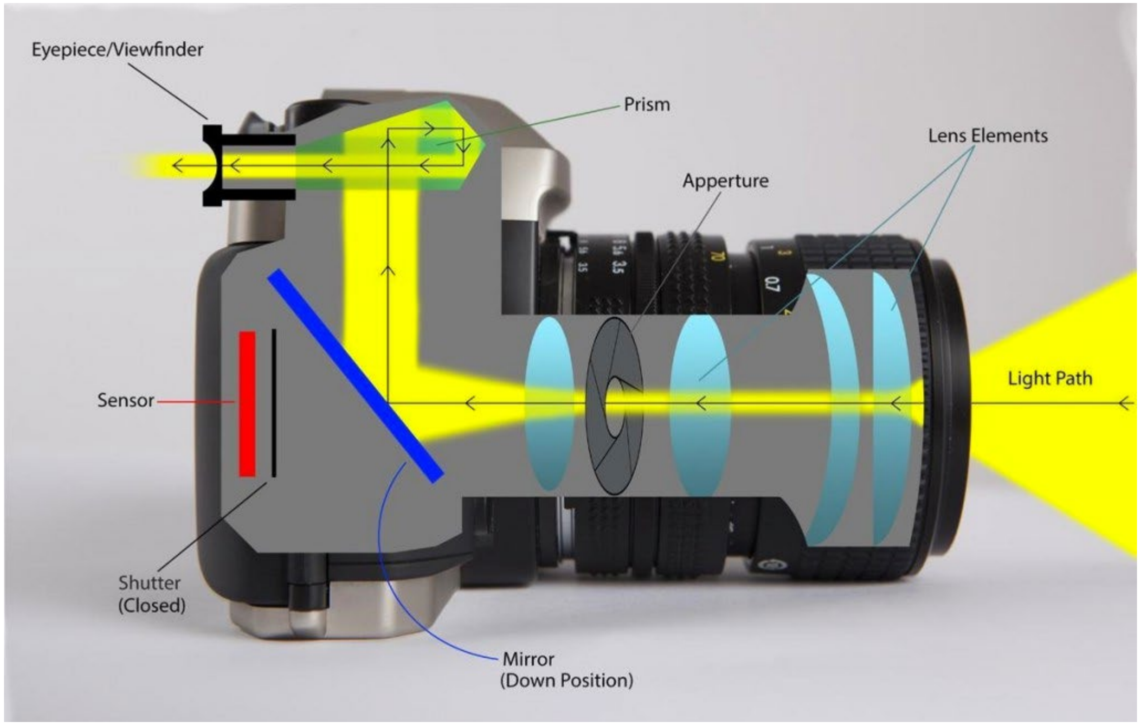
- CMOS has now improved its response and noise performance to be on par
  - Each site converts from analog to digital
  - There may be more “fixed-pattern” noise



Digital images are the 2-dimensional collection of values captured by the image sensor; they can be manipulated as a data set in the computer

Even the image focused on the sensor is “reduced” in three important ways:

- Spatial sampling
- Pixel quantization
- Temporal sampling



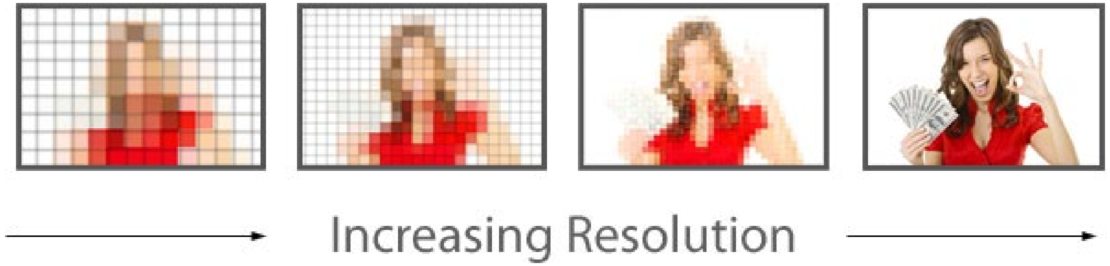
Spatial sampling is needed because the sensor has a limited number of sensing sites; resolution expresses the “fineness” of detail being captured

VIRGINIA TECH.
BRADLEY DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

- Consider a current imaging sensor chip from OmniVision:
  - “The 1/2.6-inch OV10823 uses advanced 1.4-micron OmniBSI-2™ pixel architecture to capture full-resolution 10.5-megapixel (4320 x 2432 pixels) video at 30 fps…”
  - fps is “frames per second”
- The image focused on the sensor will generate a 2D matrix of pixel values 4320 wide by 2432 high
  - This is really good
- Each pixel is 1.4 μm in size
  - so the sensing area is: 4320(1.4) by 2432 (1.4) = 6mm by 3.4mm

Resolution can be confusing – we just calculated the pixel size at the sensor, but we more likely care about the size of the pixel at the object plane

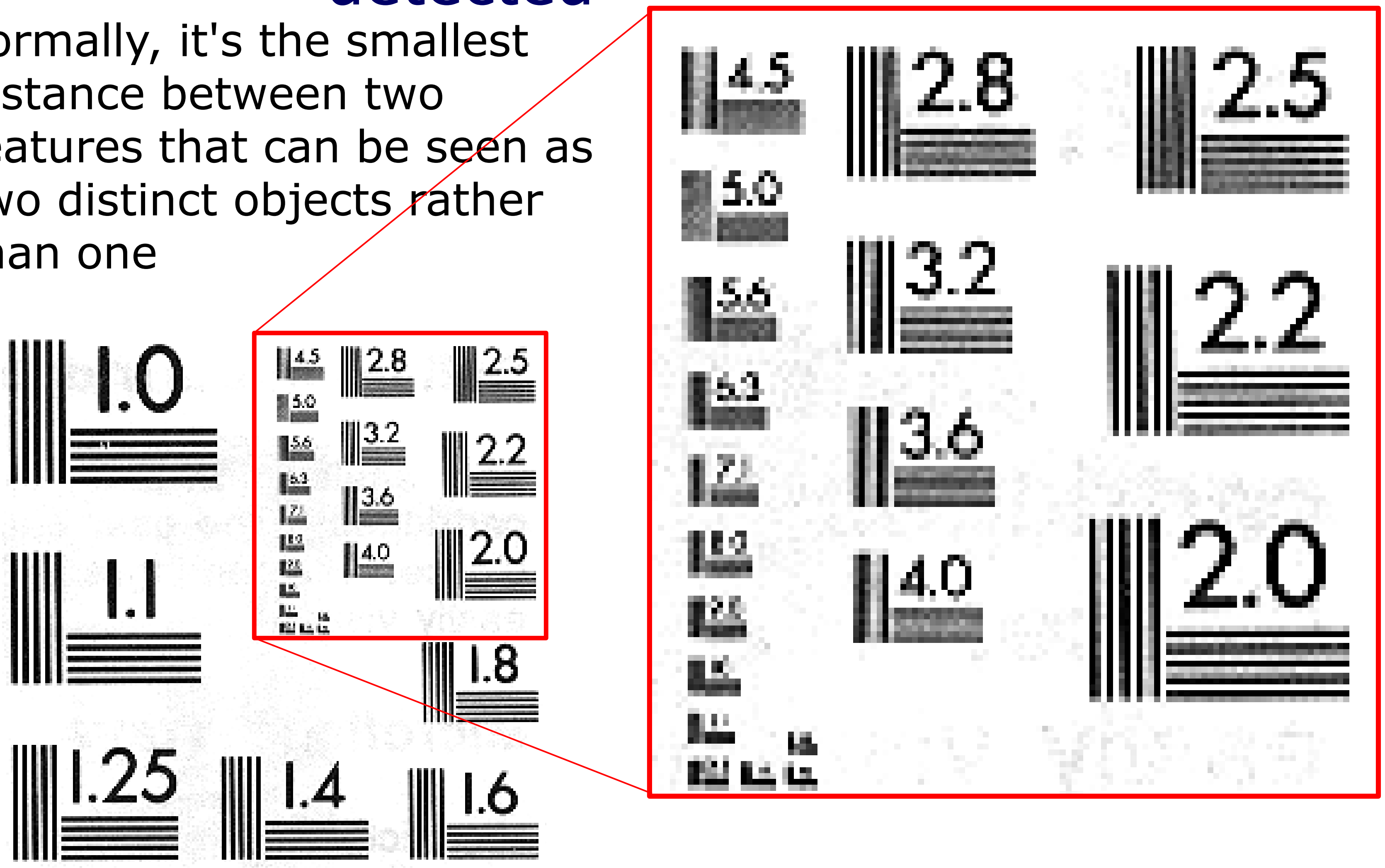
- Let's use the OV10823 to image an object that's 5 meters across
  - and set the optics so that it just fills the image
- So, at the object, we have 4320 pixels across five meters
- Each pixel in the image represents an area of the object that is $r = \frac{5m}{4320\ pix} = 0.00115741\ \frac{m}{pix} = 1.15741\ \frac{mm}{pix}$
  - The pixels are about 1 mm in size at the object
  - We call this 1mm object resolution (or 1 mm resolution)
- Higher resolution means more pixels, therefore smaller pixels
- Lower resolution means fewer pixels and larger pixels



http://www.ovt.com/products/category.php?id=37

Spatial Resolution of an imaging system is the smallest feature (in real-world units) that can be detected

- Formally, it's the smallest distance between two features that can be seen as two distinct objects rather than one



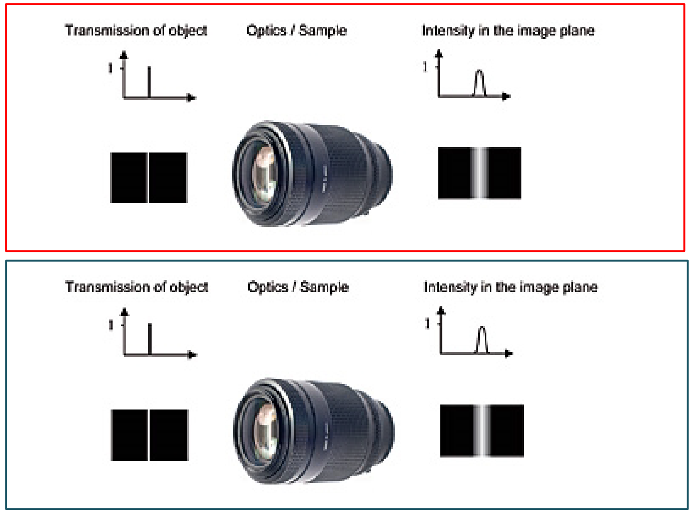
The Line Spread Function (LSF) is a representation of how a sharp, thin line in the physical world is imaged by a given system

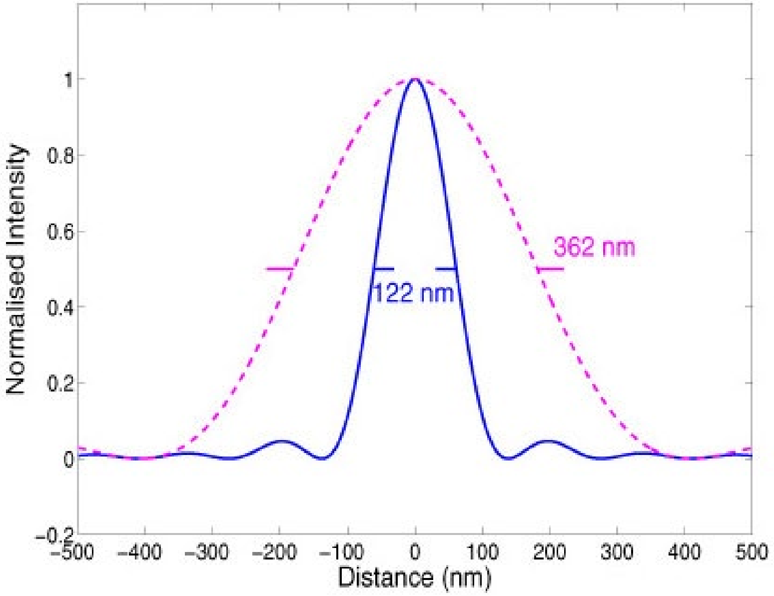
- Effects like blurring, limited resolution and limited electronic bandwidth tend to "round off" edges
- The resulting image data is roughly Gaussian
- The shape (characterized by the distribution's Full Width at Half Maximum) is used to quantify a system's degradation of a thin line

The LSF of a better imaging system (as measured by the FWHM) is smaller than the LSF of a worse system

worse system (**RED**)

better system (**BLUE**)



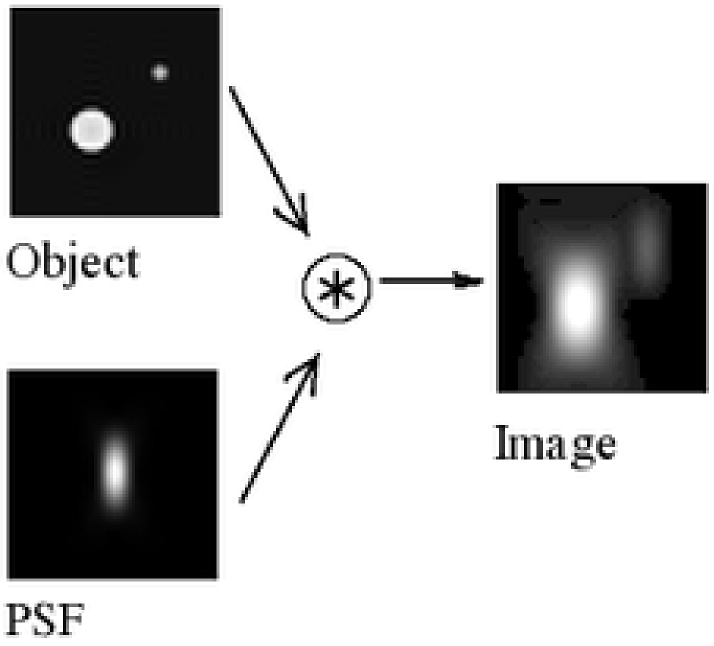


- The **RED** system has a wider LSF (362nm) than the **BLUE** system (122nm)
  - Note that the LSF, the response to imaging a thin line, looks more true to the original for the blue system than the red.

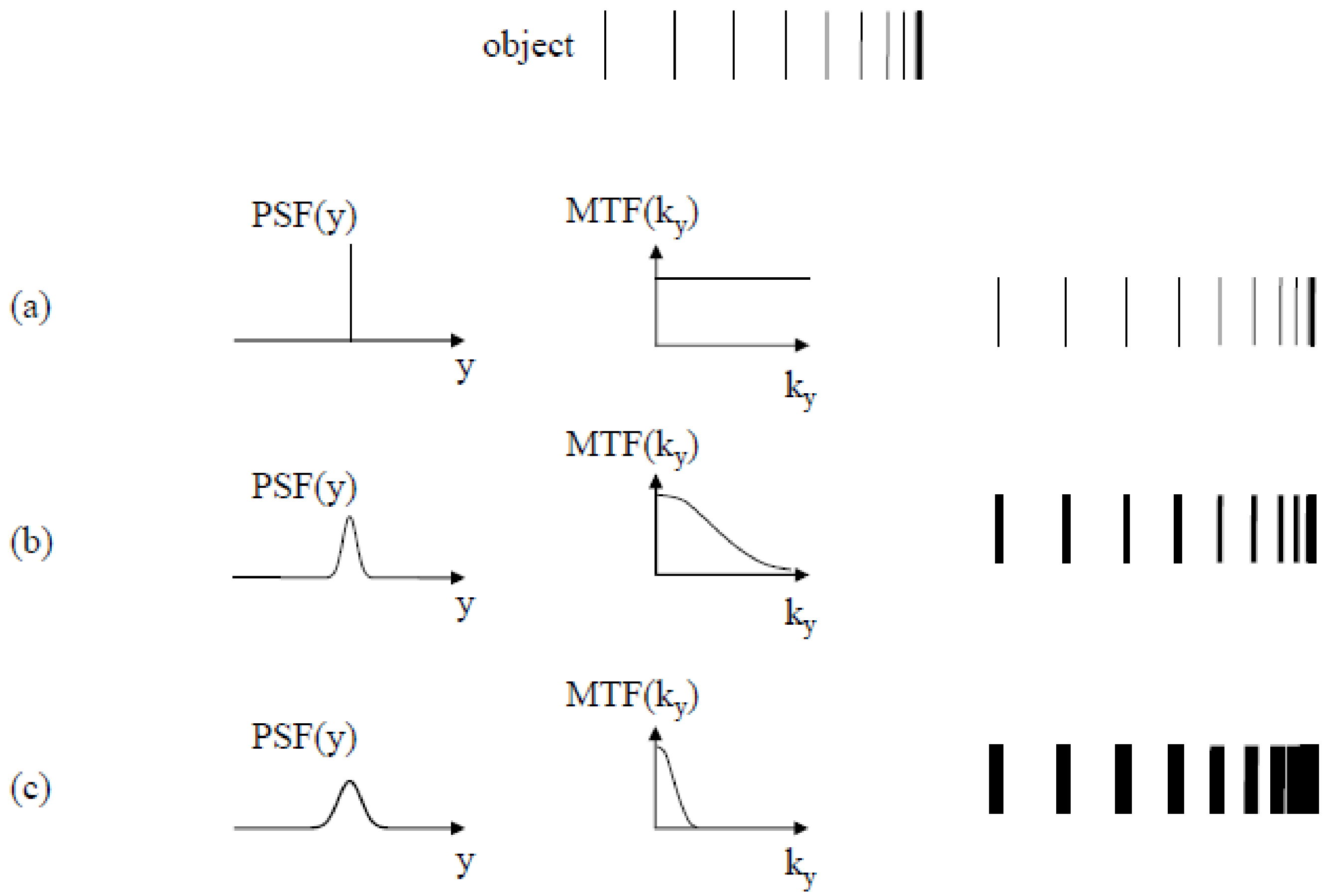
The point spread function takes the same idea into multiple dimensions; often, the optical properties aren't the same (especially in medical imaging equipment)

- The PSF (often written as $h(x, y, z)$) is convenient because we can convolve it with the shape of the object to calculate the image that object will produce:

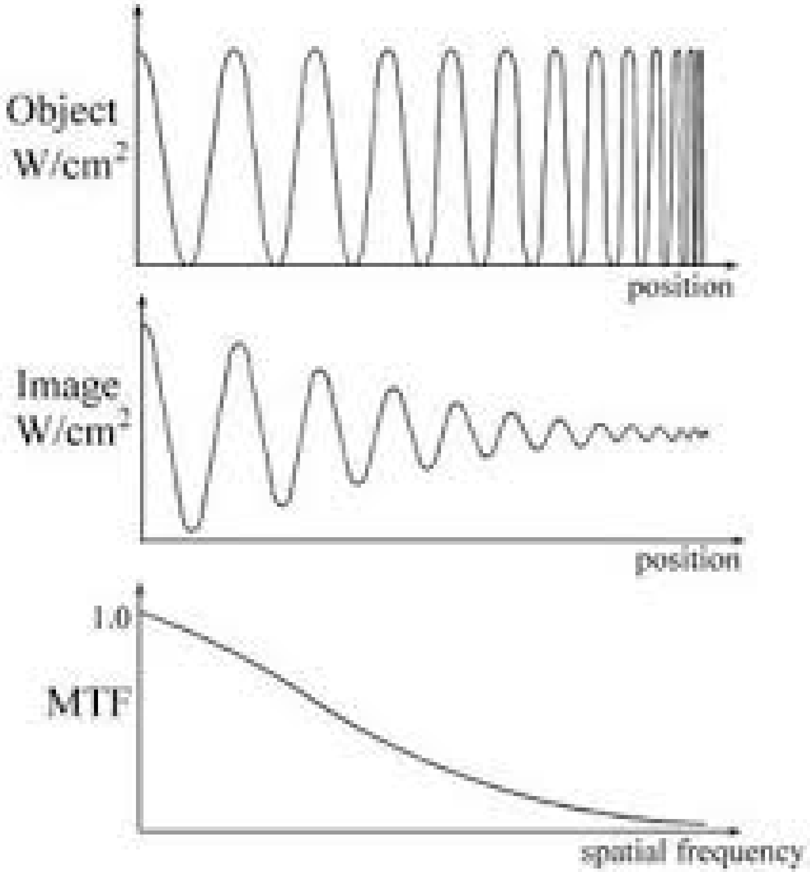
$$I(x, y, z) = O(x, y, z) * h(x, y, z)$$



The Modulation Transfer Function (MTF) is the Fourier transform of the PSF; it's the most common way to express optical system quality

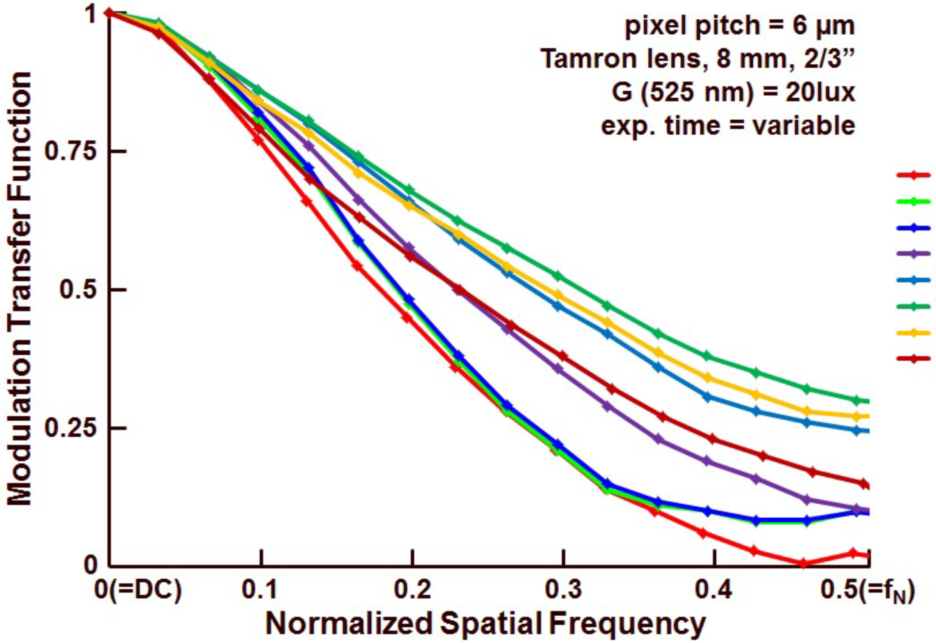


The MTF falling off to the right indicates that the system is poorer at resolving details at higher spatial frequencies



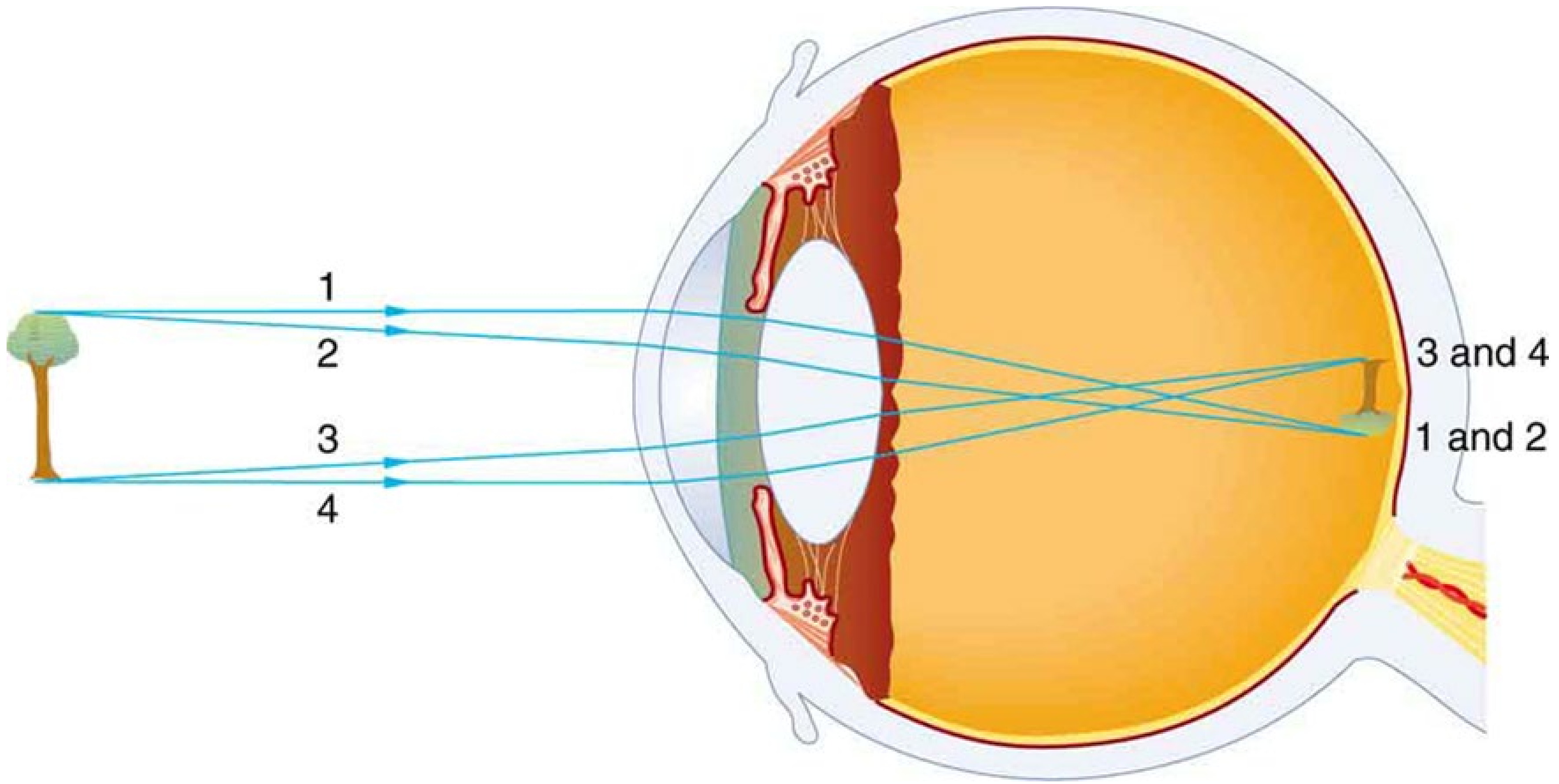
MTFs have some convenient properties

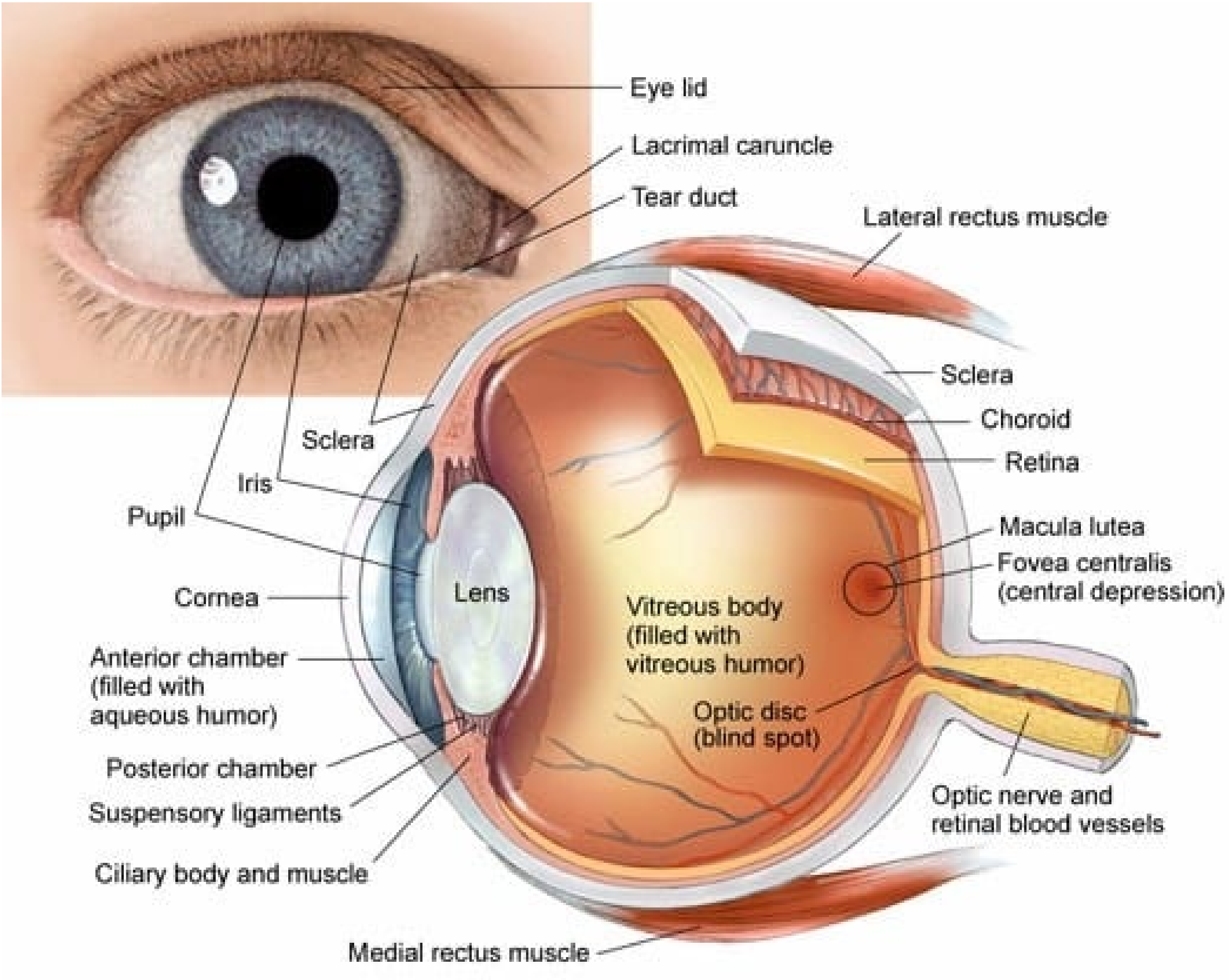
- The MTF of an overall system is the product of the MTFs of separate components
  - Equal to convolving the PSFs of the separate stages
  - Therefore, a single component with a narrow (limiting) MTF will reduce the quality of the whole system
- Wider MTFs are better (like frequency responses)
- Often, more pixels don't help, if the rest of the system has a narrower MTF



THE HUMAN VISUAL SYSTEM

The same process takes place in the eye to create, on the retina, images of the world around us; the cells of the retina sense the image information for transmission to the visual centers of the brain





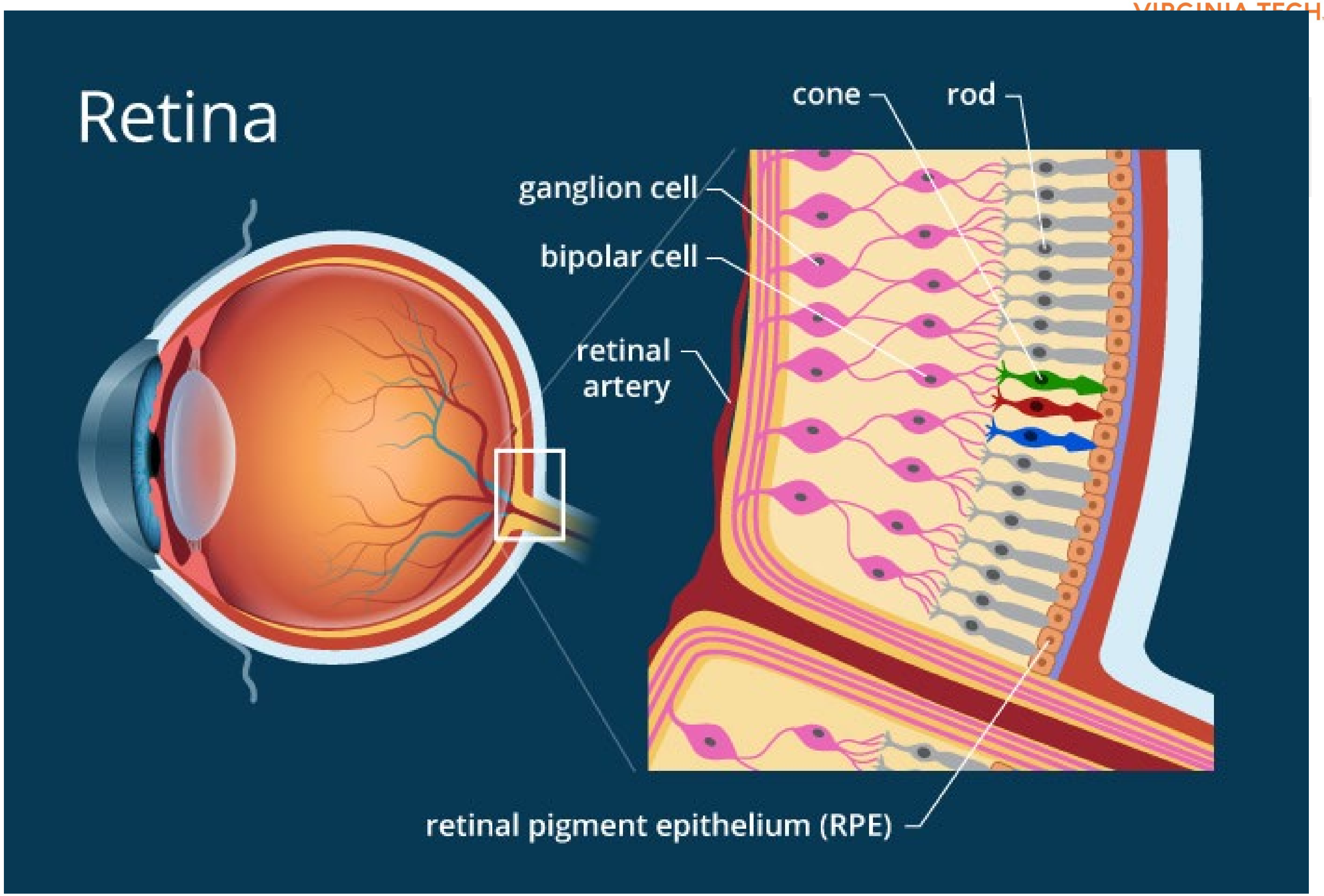
Right Eye (viewed from above)

https://columbiaeyeclinic.com/anatomy-eye/

The retina contains two type of imaging cells:

- cones come in three colors (red, green and blue, approximately)

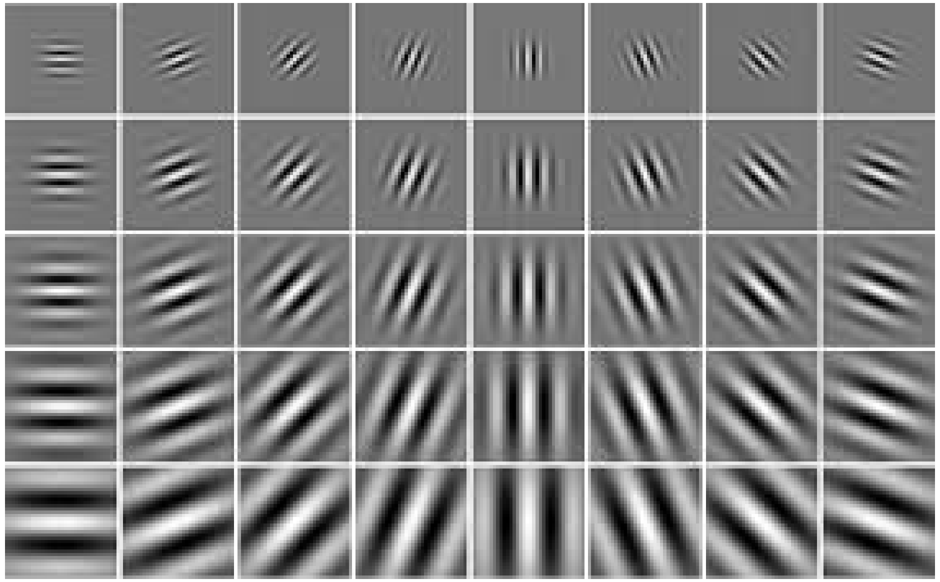
- rods are more sensitive but only monochromatic



https://www.findlight.net/blog/2018/03/16/artificial-photoreceptors/

The simple concept is that the retina is the camera and the brain is the computer – but there’s actually a lot of processing that goes on in the cells that carry retinal information to the brain

- Tomasi, C., Early Vision, Encyclopedia of Cognitive Science, 2006
- Color information is probably encoded in “opponents” – light/dark, red/green, blue/yellow
- Sets of filters sense changes in oriented spatial change and temporal change:
- Correspondence and difference between the two eyes determines depth
- Active area of research in neuroscience



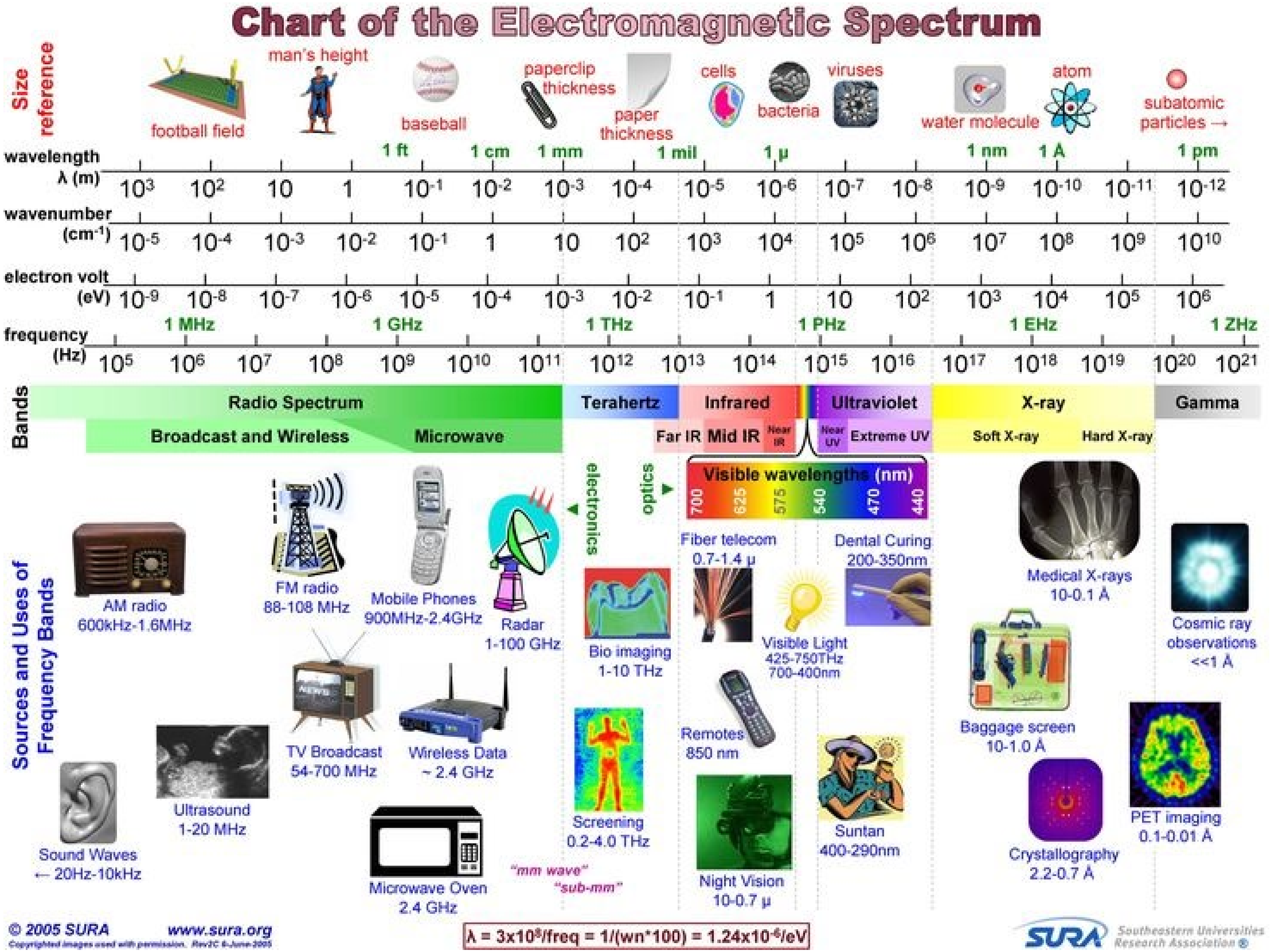
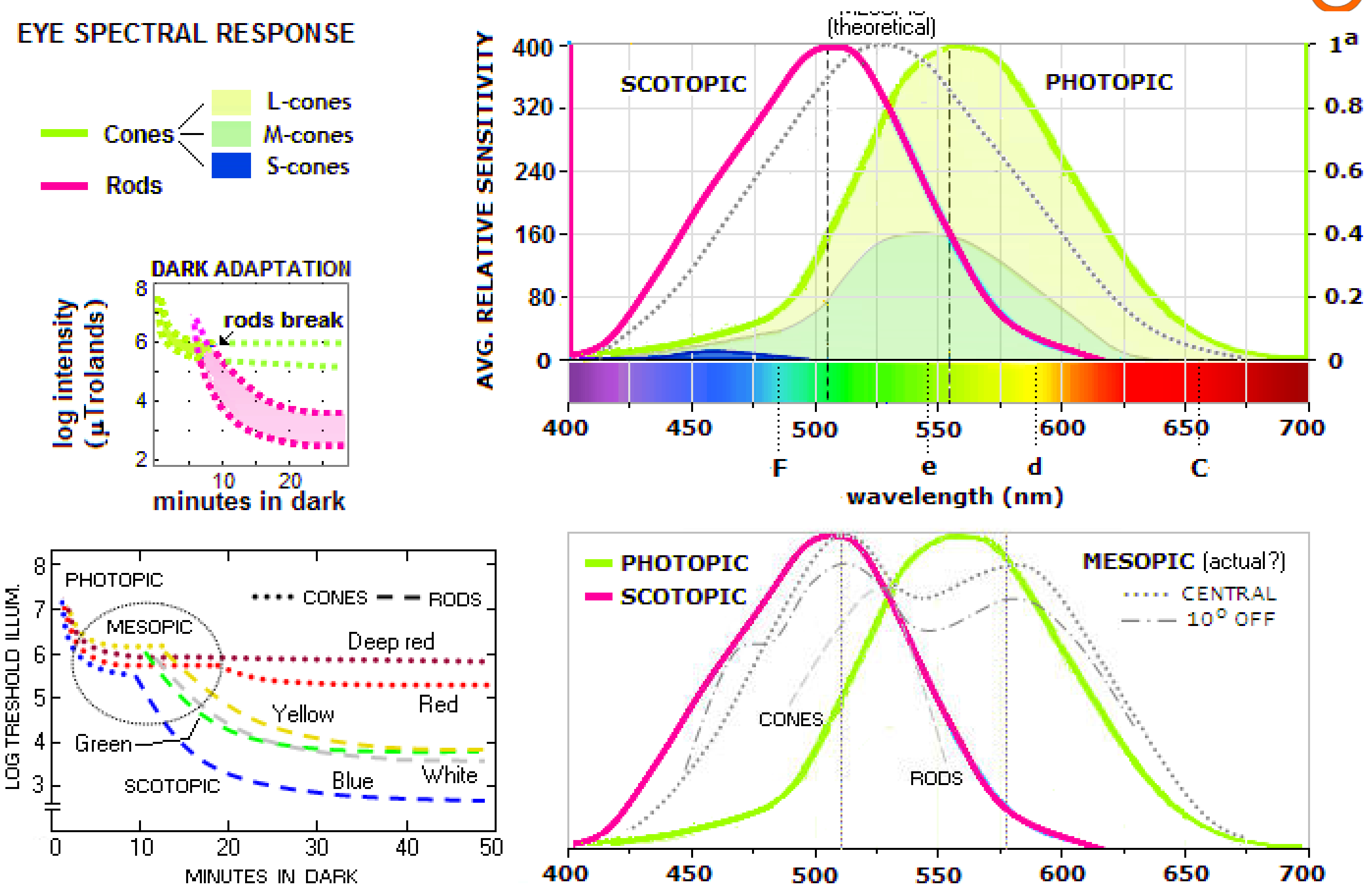


Chart of the Electromagnetic Spectrum
Size reference
football field
man's height
baseball
paperclip thickness
paper thickness
cells
bacteria
viruses
water molecule
atom
subatomic particles →
wavelength λ (m)
1 ft
1 cm
1 mm
1 mil
1 μ
1 nm
1 Å
1 pm
10^3 10^2 10 1 10^-1 10^-2 10^-3 10^-4 10^-5 10^-6 10^-7 10^-8 10^-9 10^-10 10^-11 10^-12
wavenumber (cm^-1)
10^-5 10^-4 10^-3 10^-2 10^-1 1 10 10^2 10^3 10^4 10^5 10^6 10^7 10^8 10^9 10^10
electron volt (eV)
10^-9 10^-8 10^-7 10^-6 10^-5 10^-4 10^-3 10^-2 10^-1 1 10 10^2 10^3 10^4 10^5 10^6
frequency (Hz)
1 MHz
1 GHz
1 THz
1 PHz
1 EHz
1 ZHz
10^5 10^6 10^7 10^8 10^9 10^10 10^11 10^12 10^13 10^14 10^15 10^16 10^17 10^18 10^19 10^20 10^21
Bands
Radio Spectrum
Terahertz
Infrared
Ultraviolet
X-ray
Gamma
Broadcast and Wireless
Microwave
Far IR
Mid IR
Near IR
Near UV
Extreme UV
Soft X-ray
Hard X-ray
electronics
optics
Visible wavelengths (nm)
700 625 575 540 470 440
Sources and Uses of Frequency Bands
AM radio 600kHz-1.6MHz
FM radio 88-108 MHz
Mobile Phones 900MHz-2.4GHz
Radar 1-100 GHz
Bio imaging 1-10 THz
Fiber telecom 0.7-1.4 μ
Dental Curing 200-350nm
Medical X-rays 10-0.1 Å
Cosmic ray observations <<1 Å
Visible Light 425-750THz 700-400nm
TV Broadcast 54-700 MHz
Wireless Data ~ 2.4 GHz
Remotes 850 nm
Baggage screen 10-1.0 Å
Sound Waves ← 20Hz-10kHz
Ultrasound 1-20 MHz
Screening 0.2-4.0 THz
Suntan 400-290nm
PET imaging 0.1-0.01 Å
Microwave Oven 2.4 GHz
"mm wave"
"sub-mm"
Night Vision 10-0.7 μ
Crystallography 2.2-0.7 Å
© 2005 SURA
www.sura.org
Copyrighted images used with permission. Rev2C 6-June-2005
λ = 3x10^8/freq = 1/(wn*100) = 1.24x10^-6/eV
SURA Southeastern Universities Research Association

The spectral response of the eye – sensitivity vs. wavelength of light – is different for cones (photopic) and rods (scotopic)



The mesopic response is the daylight sensitivity of both rods and cones working together:

Domain	Luminance (cd/m²)
Photopic	>3
Mesopic	0.03–3
Scotopic	<0.03

Today’s Objectives

- Light and its interactions with matter
- Components of a simple imaging system
- The human visual system