Cameras, Images, and Low-Level Robot Vision

RSS Lecture 5

Wednesday, 16 Feb 2011

Prof Rus

Based on Lecture Notes by Prof. Teller

Siegwart and Nourbakhsh § 4.1.8

Today's Lecture

- Brief historical overview
  - From early cameras to digital cameras
- Low-level robot vision
  - Camera as sensor
  - Color representation
  - Object detection
  - Camera calibration
- Putting it all together
  - Visual servo lab (next week)

Eyes

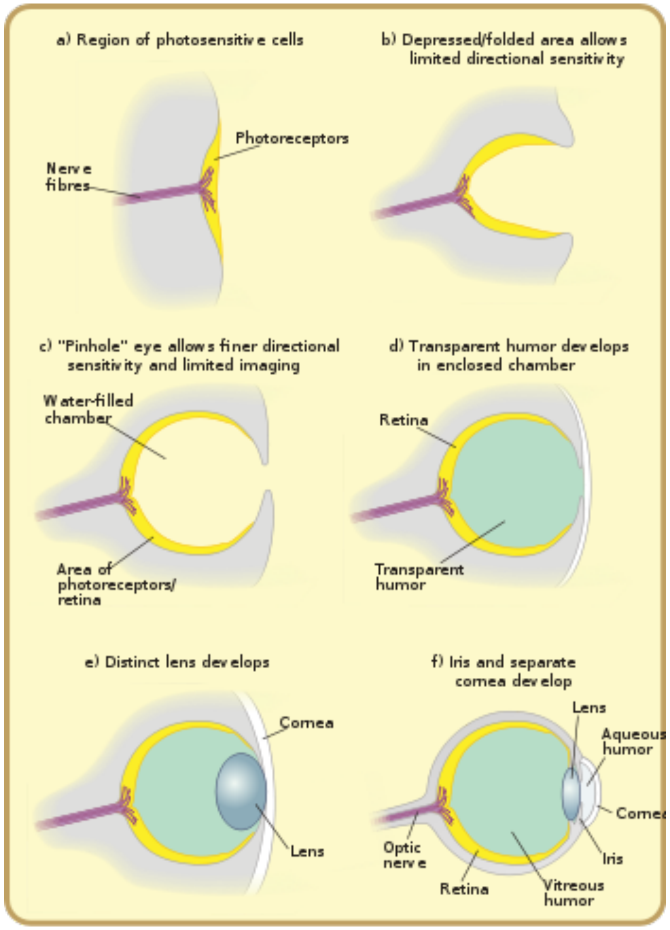


Image source: wikipedia.org

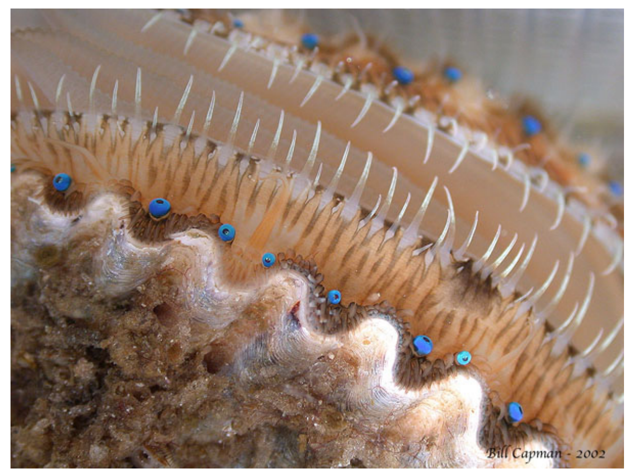
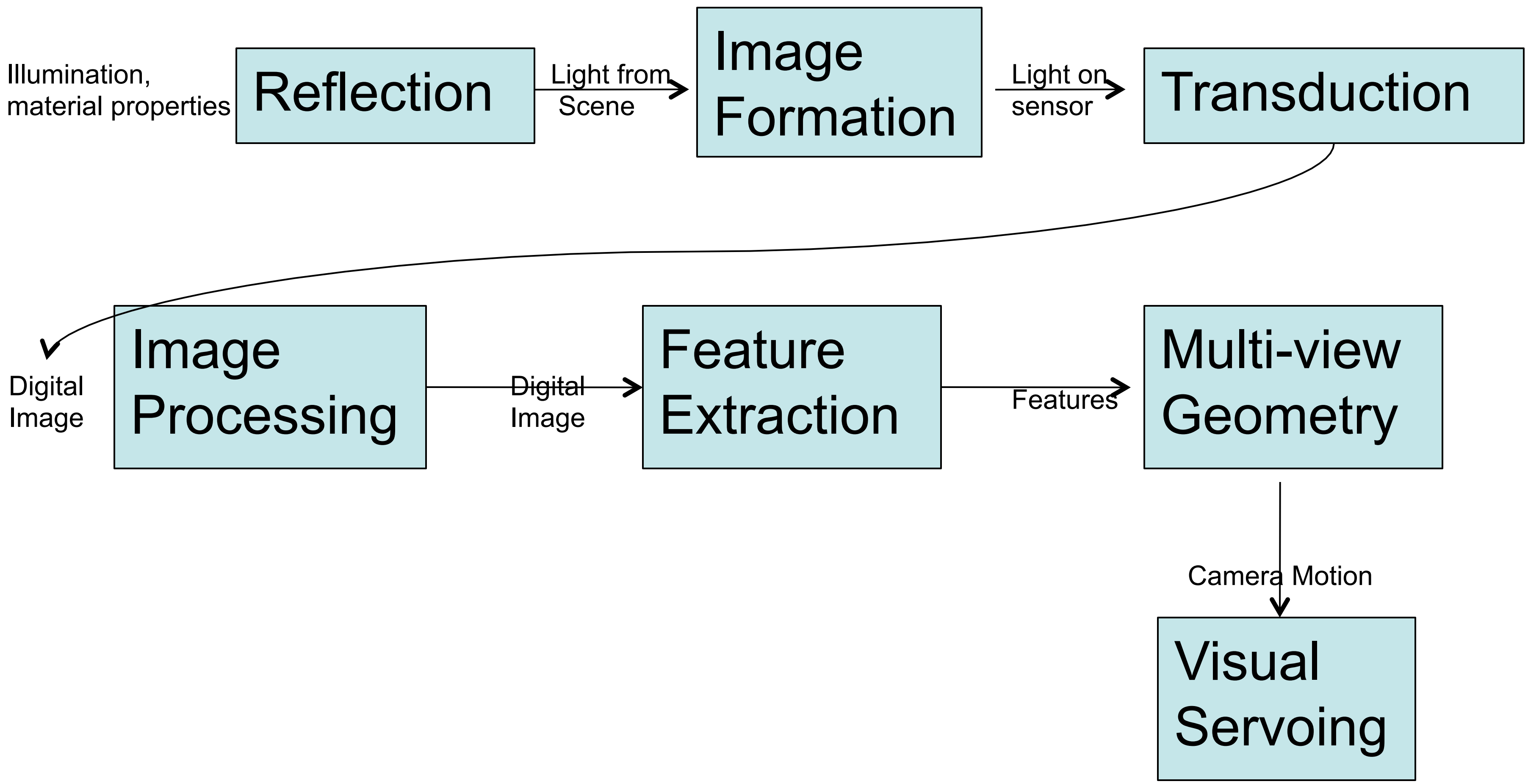


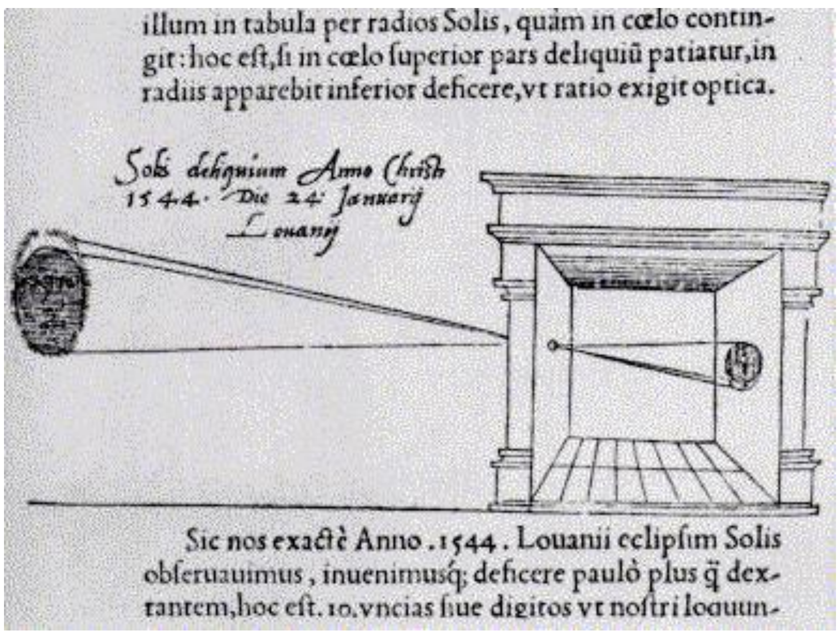
Image source: wikipedia.org

Image Processing



Camera Obscura

- Mo Ti, Chinese philosopher, 5^{th} Century B.C.
  - Described linear light paths, pinhole image formation
- Leonardo da Vinci (1452-1519)
  - Demonstrated camera obscura (lens added later)
  - Etymology: camera + obscura = “dark room”


illum in tabula per radios Solis, quàm in cœlo contingit: hoc est, si in cœlo superior pars deliquiū patiatur, in radiis apparebit inferior deficere, vt ratio exigit optica.

Sic nos exactè Anno .1544. Louanii eclipsim Solis obseruauimus, inuenimusq; deficere paulò plus q̃ dextantem, hoc est. 10. vncias siue digitos vt nostri loquun-

Frisius (1544)

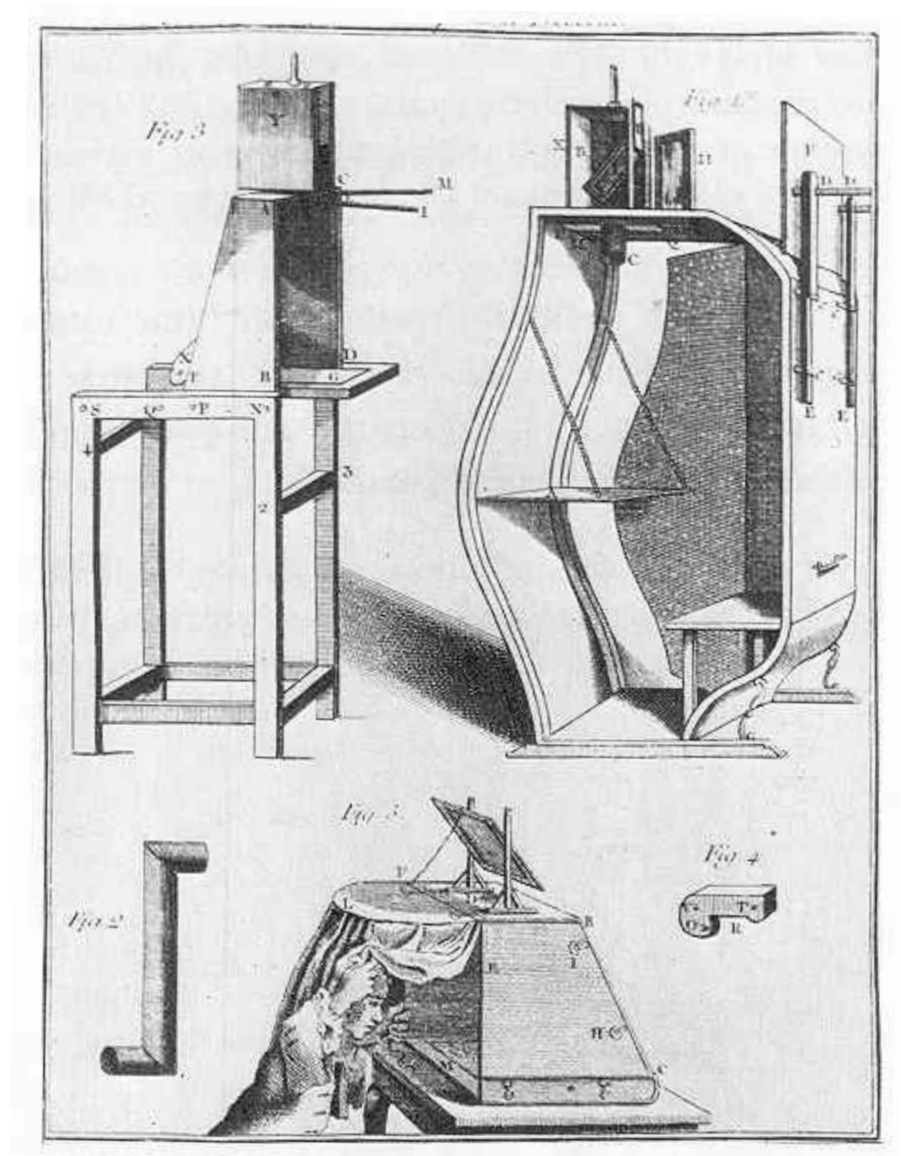
Photograph of camera obscura interior:

Portmerion Village, North Wales

David Hockney: Secret Knowledge

Toward Photography

- People sought a way to “fix” the images at the back of the camera obscura
- Pursued decades of experimentation with light-sensitive salts, acids, etc.
- First photograph produced when?



First Photograph

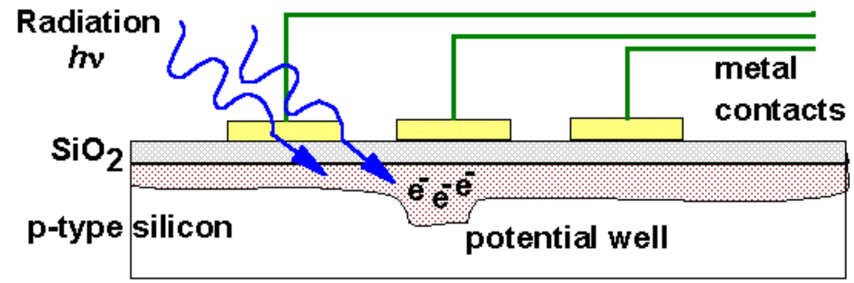
Harry Ransom Center

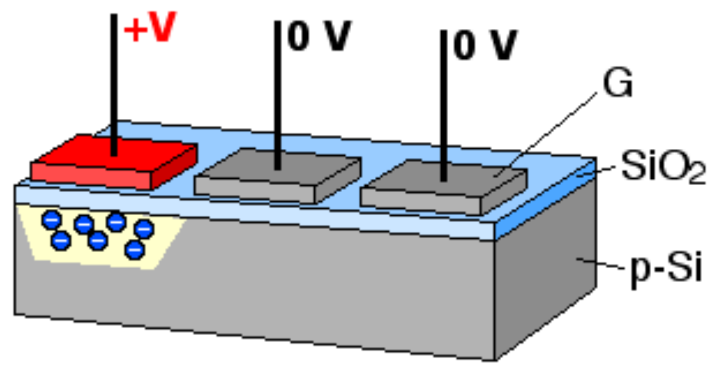
Kodak (reproduction)

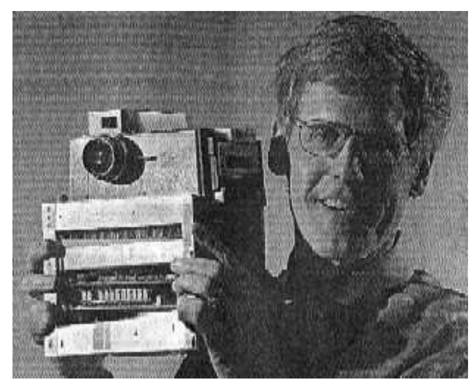
- Joseph Nicéphore Niépce (pronounced “Neeps”), "View from the Window at Le Gras,” c.
- Aluminum plate coated with light-sensitive material
- Why are buildings illuminated on both sides?
- Etymology: “photo”+ “graph” (also: photogene, heliograph)

Digital Camera Precursors

- Basis: photoelectric effect (Hertz 1887; Einstein 1905)
  - As light *frequency* increases?
  - As light *intensity* increases?
- Also: advent of CCDs as shift registers (late 1960’s)
- Integration with photoelectric sensors (early 1970’s)
- First electronic CCD still-image camera (1975):
  - Fairchild CCD element
  - Resolution: 100 x 100 x 1-bit b&w … a whopping *0.1 Megapixels*!
  - Image capture time: *23 seconds*, mostly writing to cassette tape
  - And another 23 seconds to display to a nearby television
  - Total weight: 8-1/2 pounds







Kodak, c. 1975

Kodak

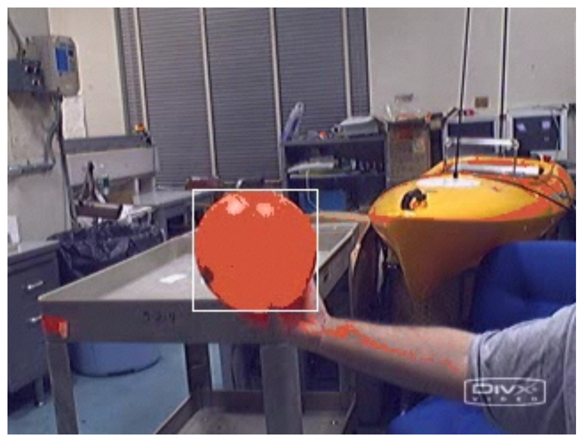
Miniaturization, price point

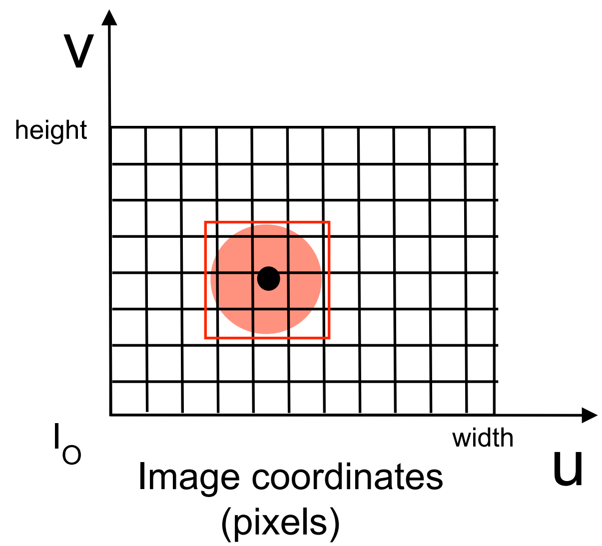
- In 2011, twenty dollars buys a camera with:
  - 640 x 480 pixel resolution at 30Hz
  - 1280 x 960 still image resolution
  - 24-bit RGB pixels (8 bits per channel)
  - Automatic gain control, white balancing
  - On-chip lossy compression algorithms
  - Uncompressed image capture if desired
  - Integrated microphone, USB 2 interface
  - Limitations
    - Narrow dynamic range
    - Narrow FOV (field of view)
    - Fixed spatial resolution
    - No actuation or active vision capabilities

Logitech C250

Digital image contents

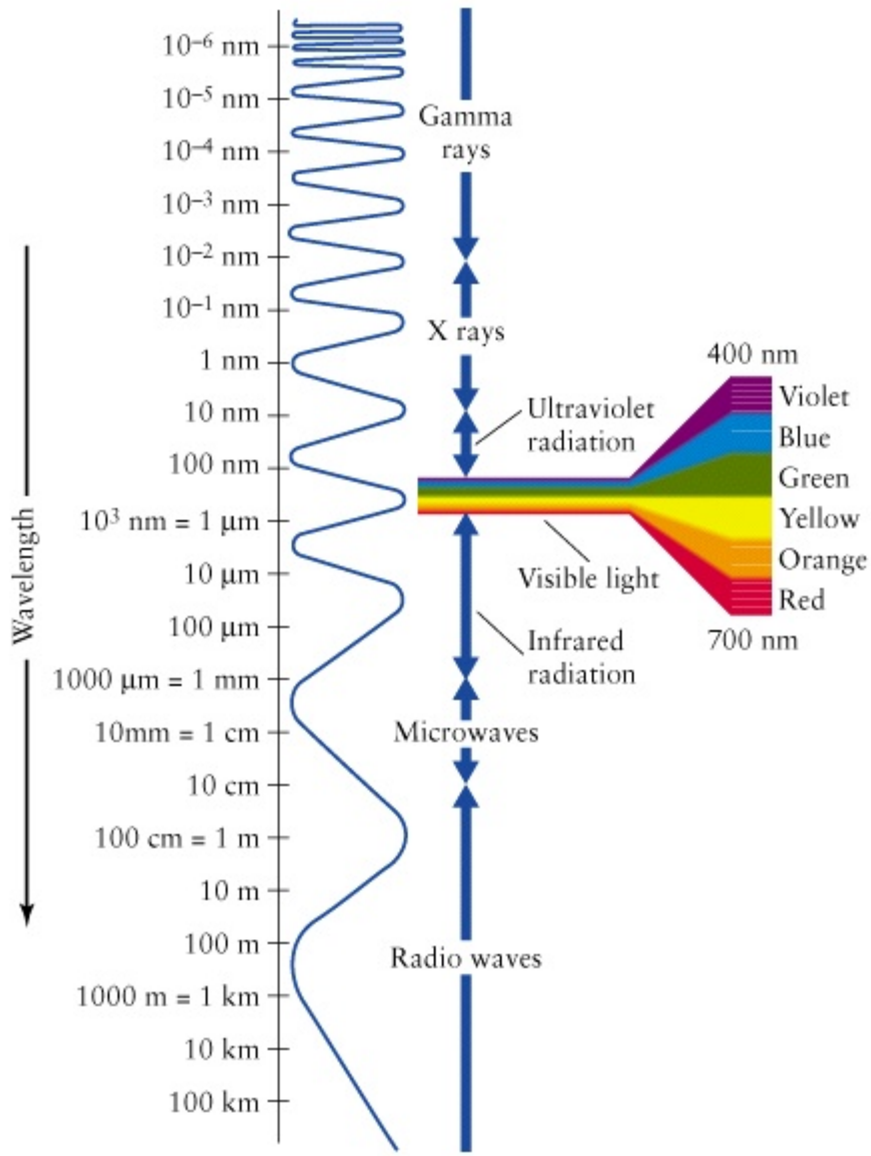
- Why are pixels represented as “RGB”?
  - Is world made of red, green, and blue “stuff”?



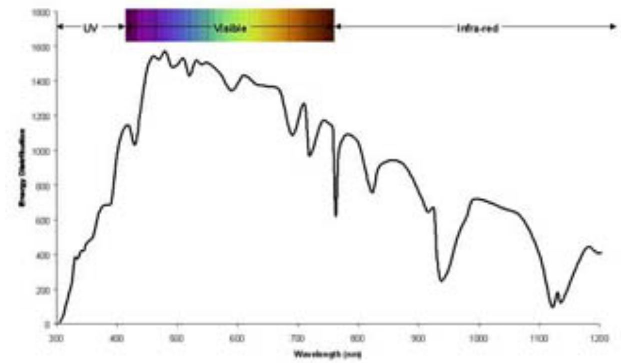


- … Answer requires two brief digressions about human vision & cameras as sensors

Visible light spectrum

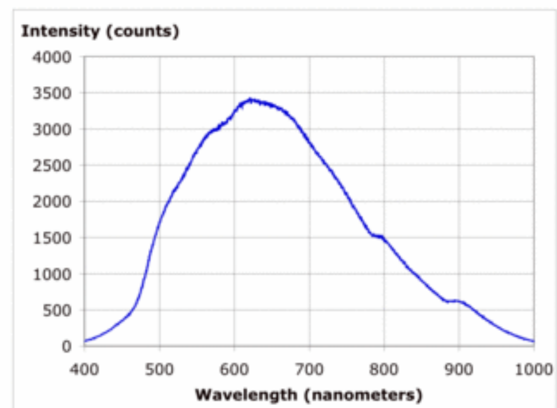


(Freedman & Kaufmann, Universe)



(ECI, Oxford)

Solar spectrum



(Wikipedia)

Incandescent spectrum

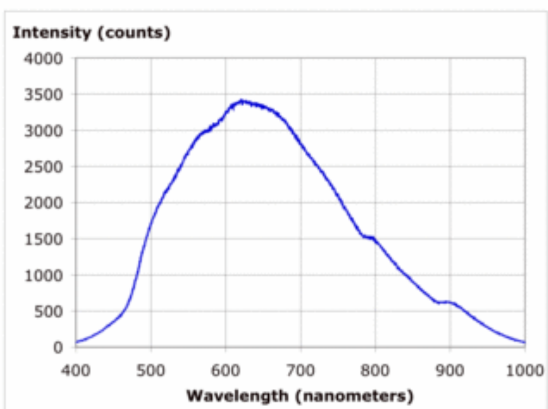
Image as measurement

- What does eye/camera actually *observe*?

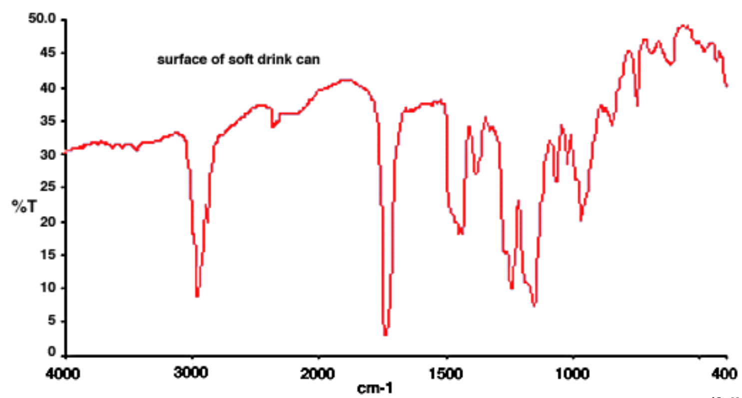
  … the *product* of illumination spectrum

  with absorption or reflection spectrum!

= (at each image point)



X



(IJVS)

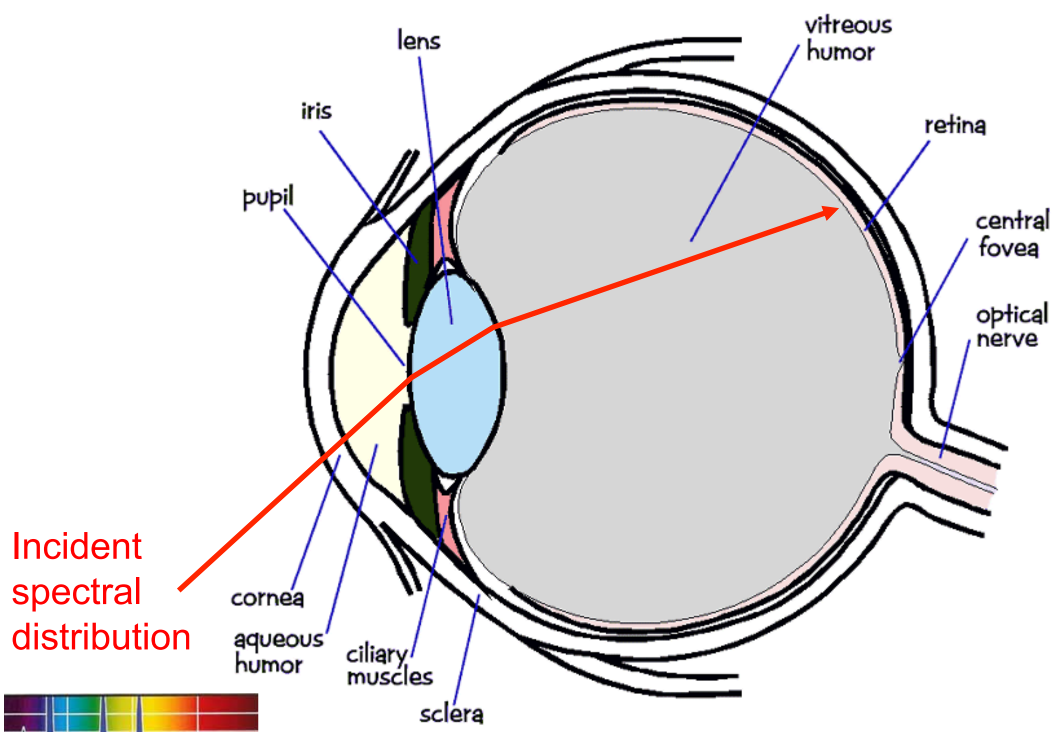
Illumination spectrum

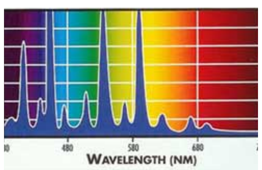
Reflection spectrum

Human eye anatomy

- Spectrum incident on light-sensitive *retina*

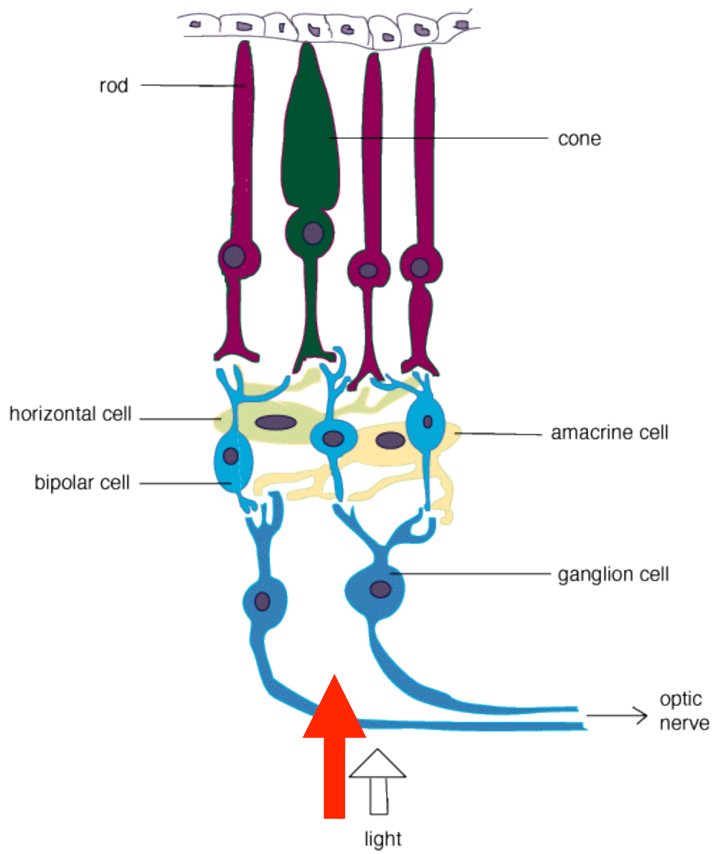
(View of R eye from above)





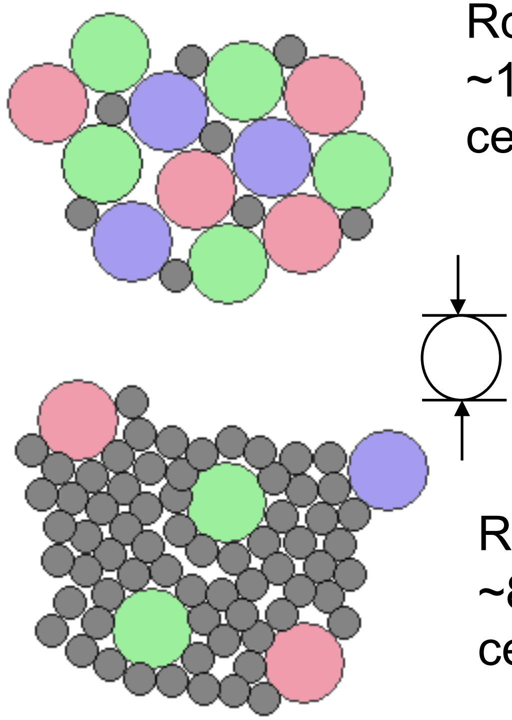
After Isaka (2004)

Rods and cones



Cone sensitivities

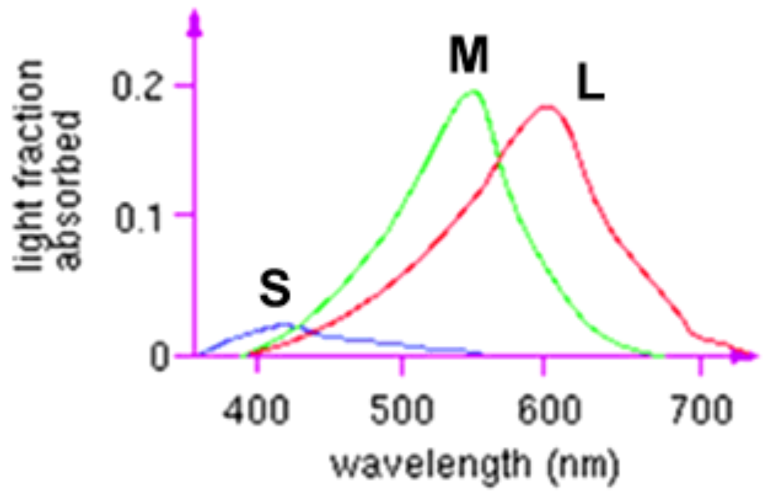
- Three cone types (S, M, and L) are roughly blue, green, and red sensors, respectively. Their peak sensitivities occur at ~430nm, 560nm, and 610nm for an "average" human.



Rods & cones, ~1.35 mm from center of fovea

Rods & cones, ~8 mm from center of fovea

(IJVS)

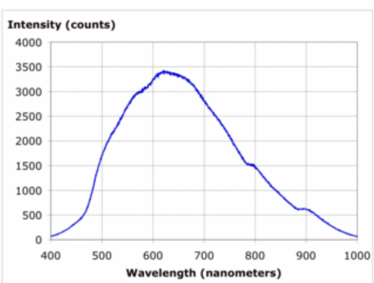


Cone sensitivities as a function of wavelength

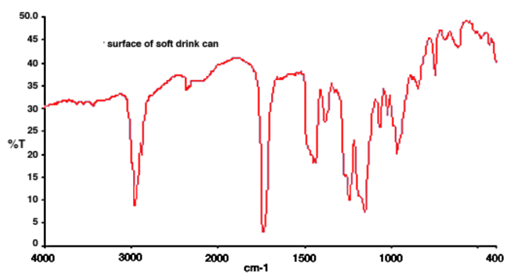
Color perception

- The cones form a spectral "basis" for visible light; incident spectral distribution differentially excites S,M,L cones, leading to color vision

= (at each cone site)

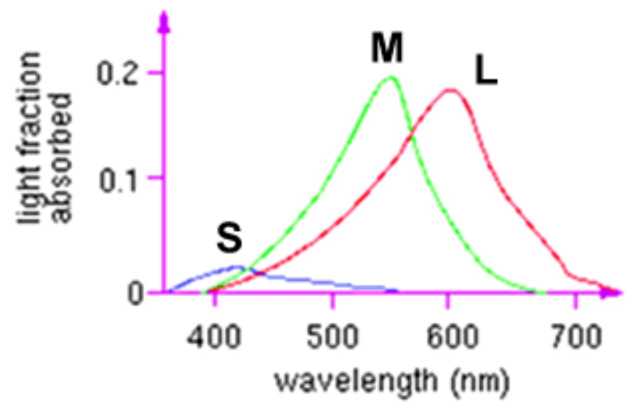


x



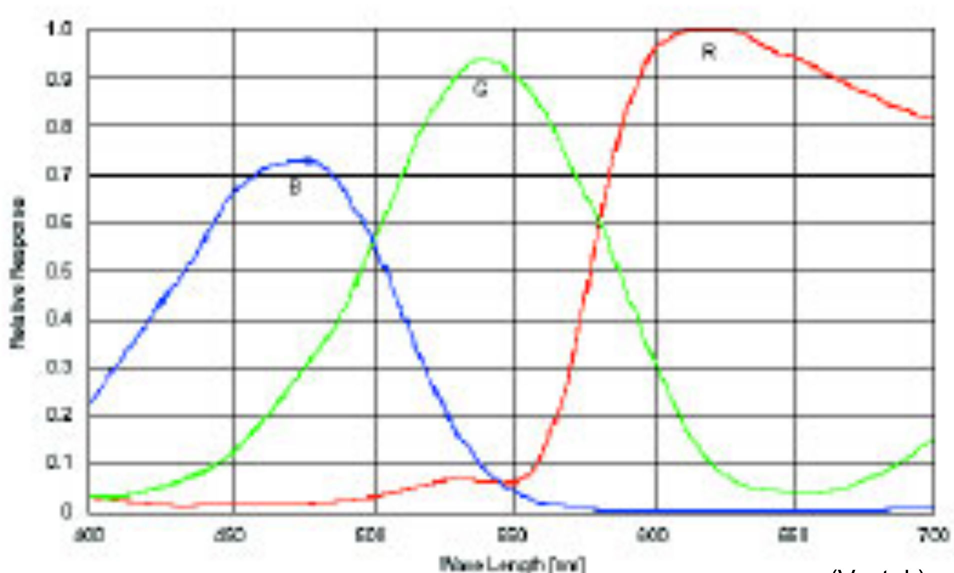
(IJVS)

x



Origin of RGB CCD sensors

- So, in a concrete sense, CCD chips are designed as RGB sensors in order to emulate the human visual system



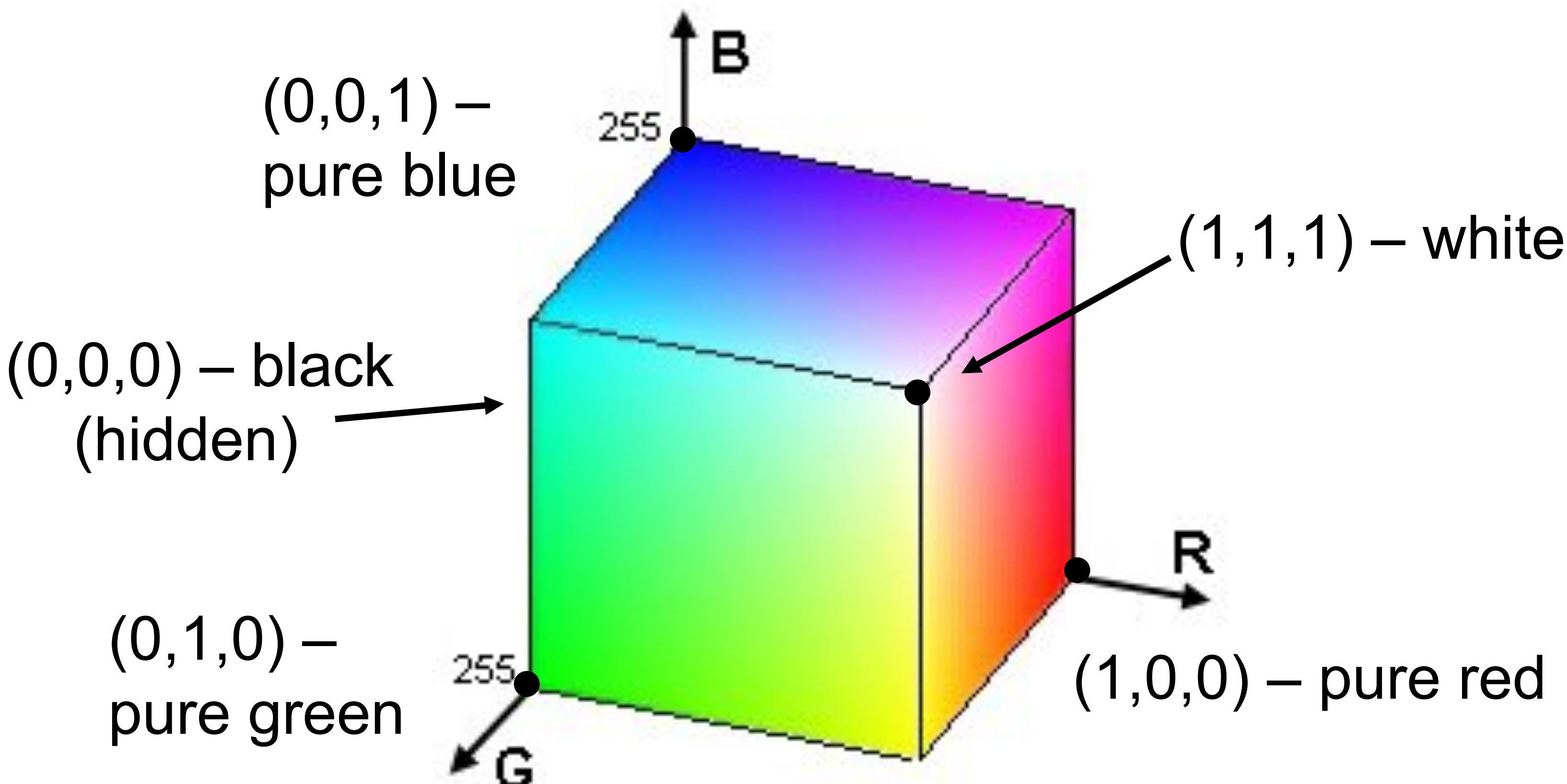
(Vaytek)

CCD with Bayer Filter, Relative Spectral Response Curve

- … End of digressions

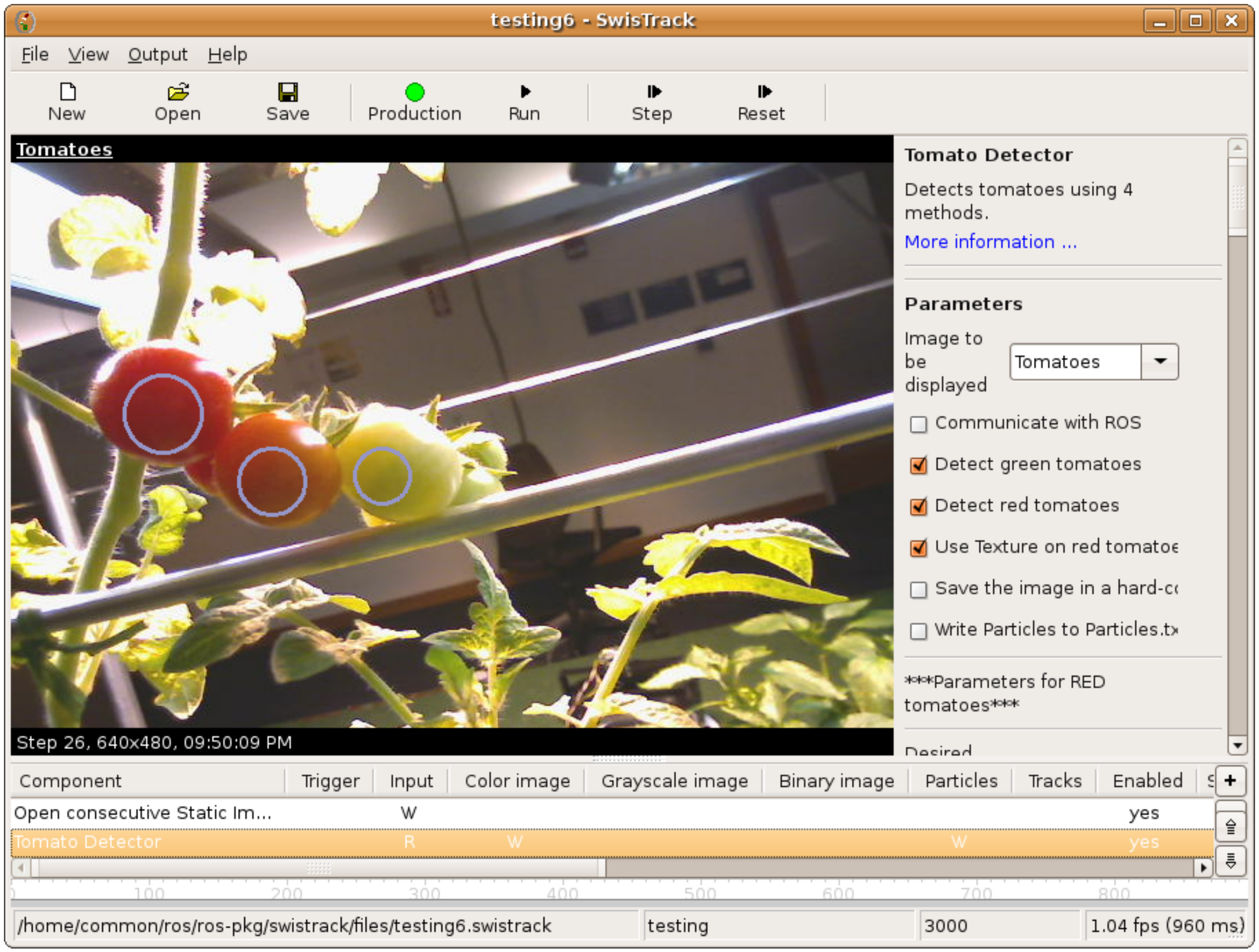
RGB Color Model

- Think of R, G, B as a kind of “color orthobasis”



(Foley & van Dam)

Perception


testing6 - SwisTrack
File View Output Help
New Open Save Production Run Step Reset
Tomatoes
Step 26, 640x480, 09:50:09 PM
Tomato Detector
Detects tomatoes using 4 methods.
More information ...
Parameters
Image to be displayed
Tomatoes
Communicate with ROS
Detect green tomatoes
Detect red tomatoes
Use Texture on red tomatoe
Save the image in a hard-c
Write Particles to Particles.t
Parameters for RED tomatoes
Component Trigger Input Color image Grayscale image Binary image Particles Tracks Enabled
Open consecutive Static Im... W yes
Tomato Detector R W W yes
/home/common/ros/ros-pkg/swistrack/files/testing6.swistrack
testing
3000
1.04 fps (960 ms)

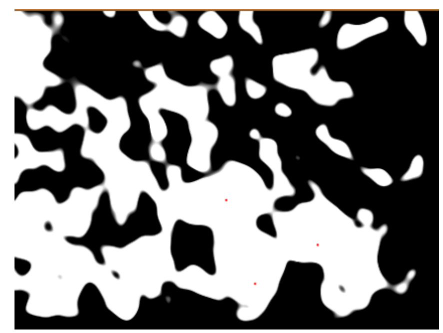


Image Processing

- Object Recognition
  - Compare against a database
  - Blob Detection
  - Edge Detection
  - Corner Detection
  - Feature detection and matching

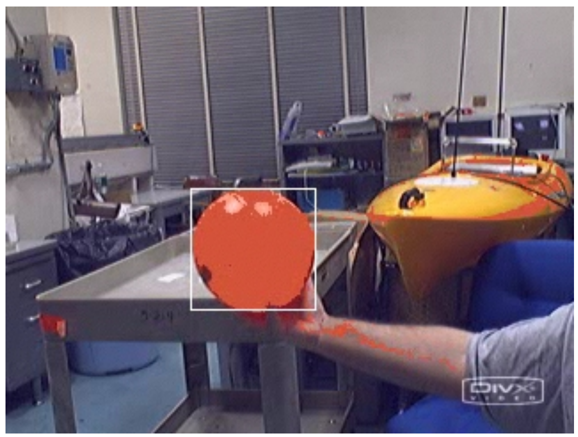
- How can an object look different from image to image?

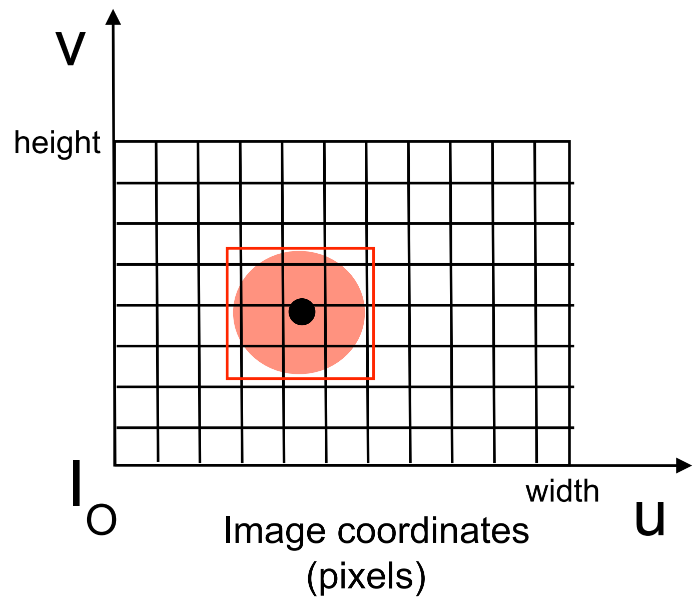
Image Processing

- Object Recognition
  - Compare against a database
  - Blob Detection
  - Edge Detection
  - Corner Detection
  - Feature detection and matching
- How can an object look different from image to image?
  - Change in lighting
  - Change in viewing direction
  - Change in Size

Object detection

- Suppose we want to detect an object (e.g., a red ball) in camera's field of view





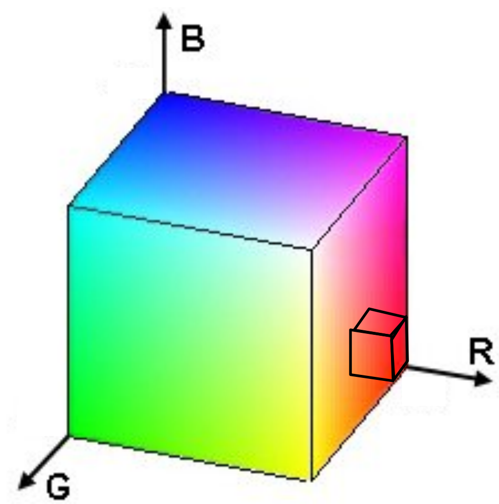
- We simply need to identify all pixels of some specified color in the image … right?

Naïve object detector

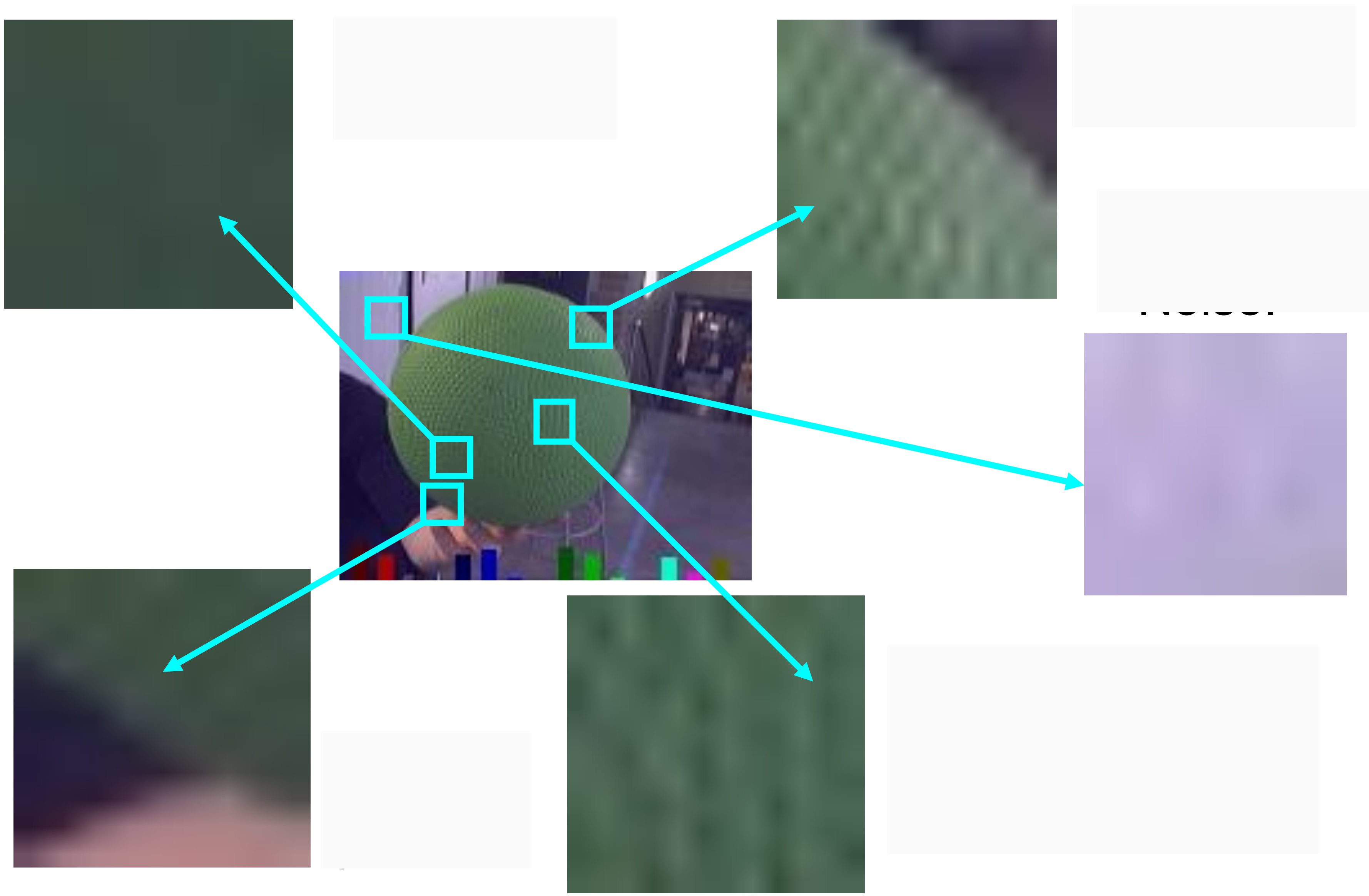
```
set objectPixels = Ø; // empty set

// look for red ball in image
for i = 0 to width-1
   for j = 0 to height-1
      if ( isRed( pixel[i, j] ) ) // classifier
         objectPixels U= {(i, j)};

if ( isBall ( objectPixels ) ) // detector
   // do something in response to ball
```

Pixel classification

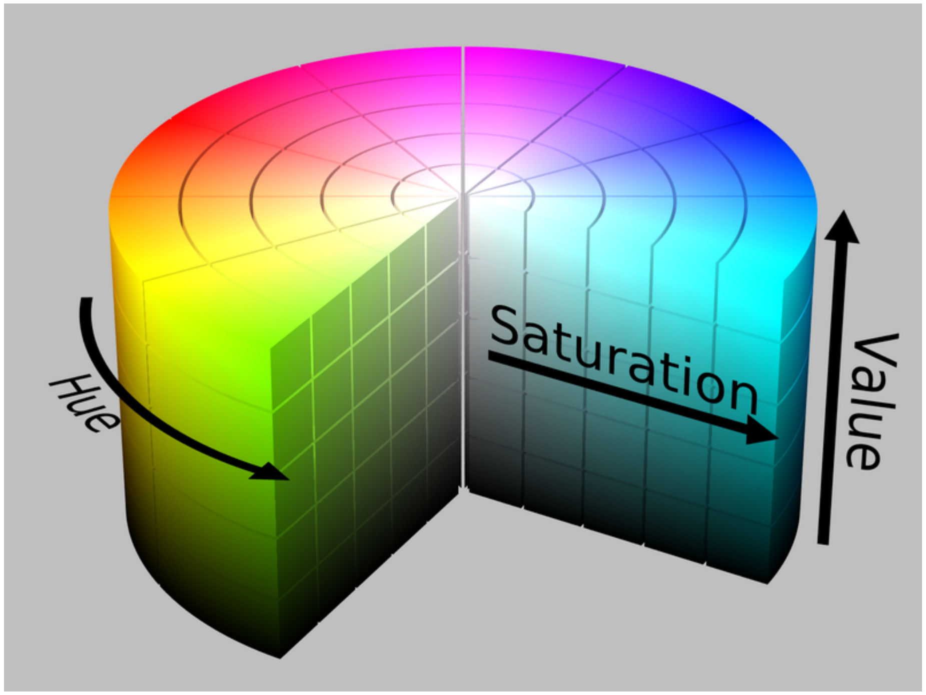
```
Boolean isRed ( pixel p ) {
   if (       p.red >= 0.8     // where do 0.8,
           && p.green < 0.2   // 0.2 come from?
           && p.blue < 0.2 )
      return true;
   else
      return false;
}

// Will this do what we want?
```

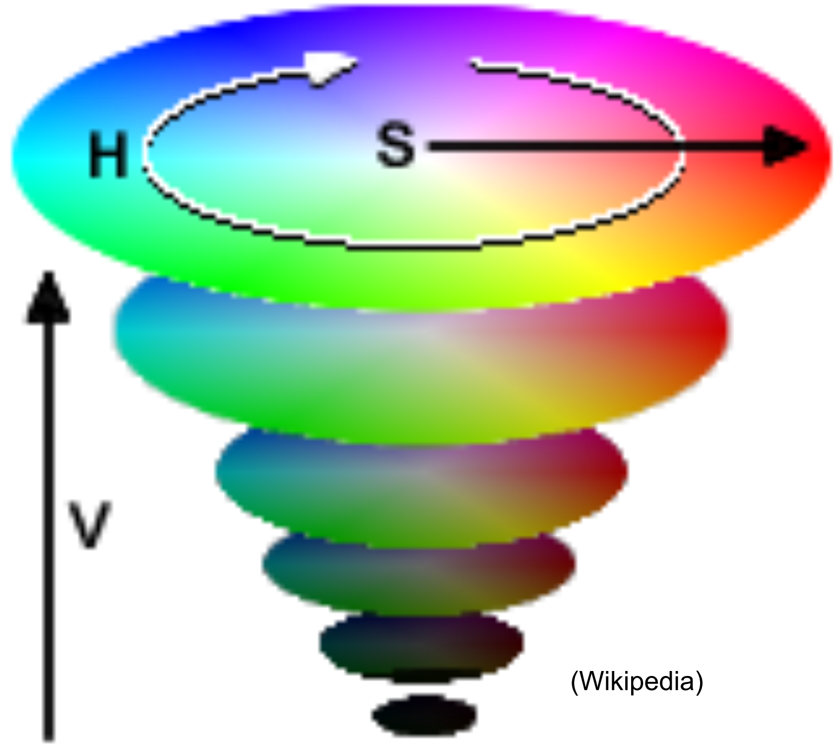


Confounding effects: Real-world images



Alternative: HSV Color Model



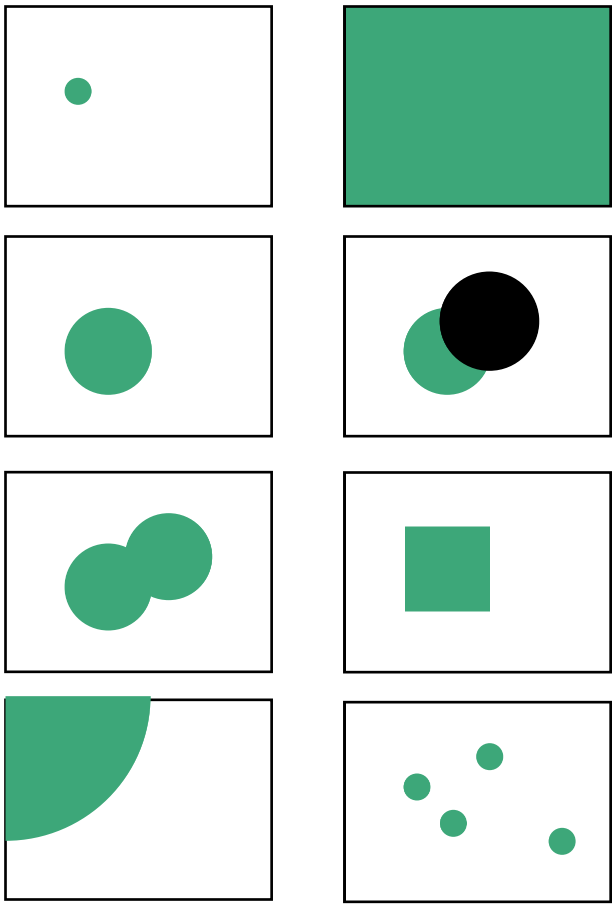


(Wikipedia)

- More robust under illumination changes (why?)
- Still must confront noise, specularity etc.

Naïve object detection

```
Boolean isBall ( set s ) {
   if ( |s| > 0.1 * W * H )  // area threshold
         return true;
   else
         return false;
}

// how might this fail?
```



Etc…

(Slightly) improved detection

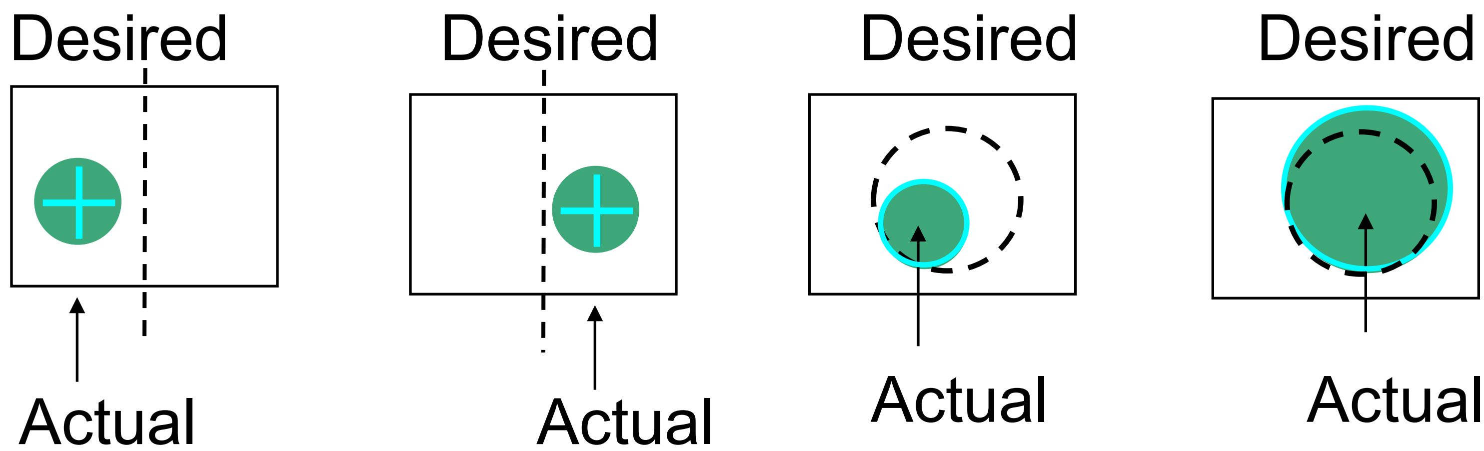
```
Boolean isBall ( set s ) {
   if ( |s| > 0.1 * W * H  // area threshold
        && s is “ball-shaped” ) {
      return true;
   else
      return false;
}

// how might this fail?
```

Doing something useful

- Report *presence* of ball in image
  - As function return, message dispatch, etc.
- Estimate *attributes* of (presumed) object
  - Color
  - Size
  - … how?
  - Centroid
  - … how?

- How / when might these estimates be poor?

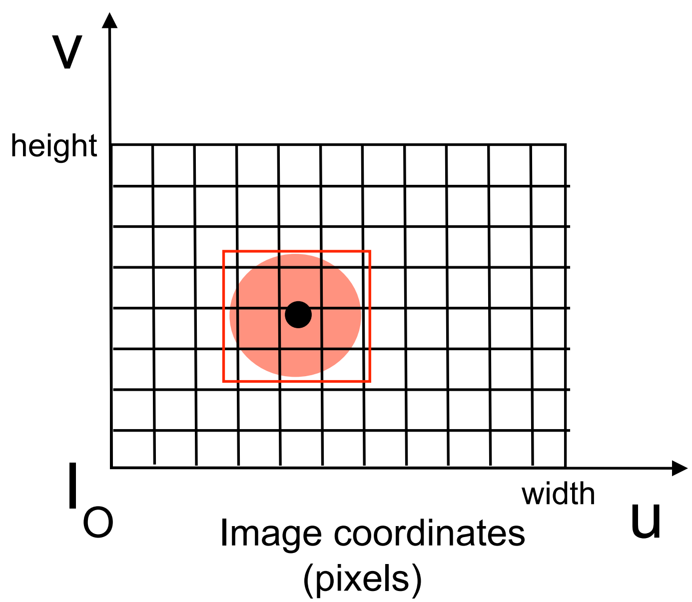
Size, centroid estimation

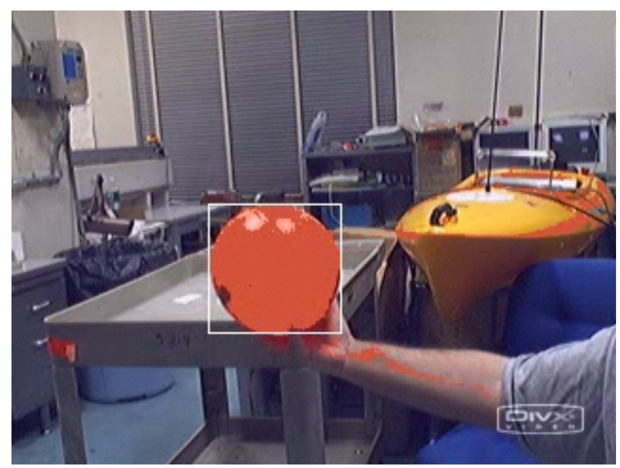


- Can use estimators as inputs to motion controller!
- Suppose we want a 1-meter frontal standoff
  - How do we compute *desired* size in image?
  - Instance of *camera calibration*;
    more examples to come later in term

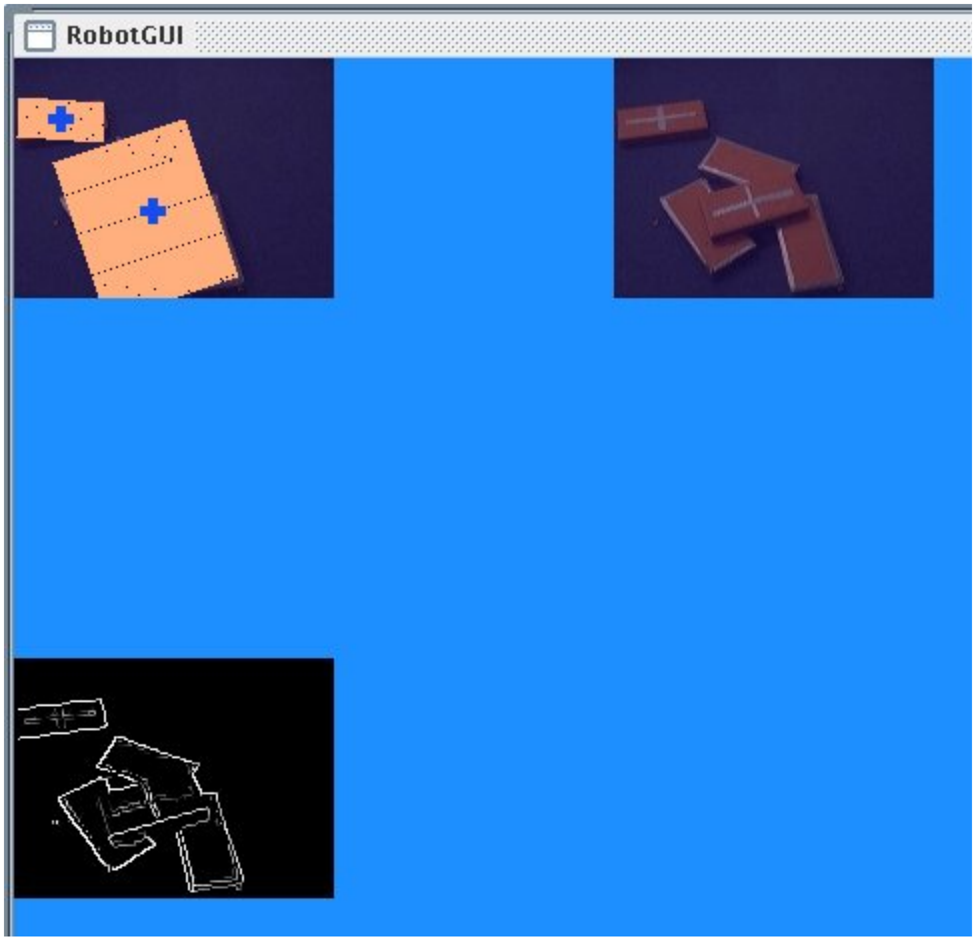
Application: Visual Servoing (Lab 4)

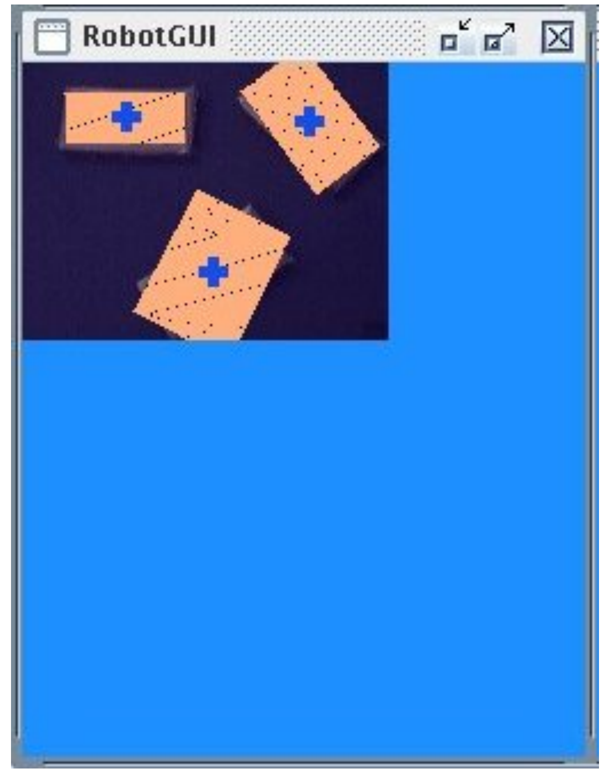
- Write a "blob detector" in integer (u, v) pixel coordinates
  - Transform pixels from (r,g,b) to chrominance, luminance
  - Given a target hue (e.g., red) and error tolerance, find large connected components of pixels with that hue
  - Estimate the area and centroid of largest detected blob
- We will supply several "fiducial objects" (colored balls)
- Issue translation, rotation control so that robot "servos" to the ball, facing it frontally at desired standoff distance



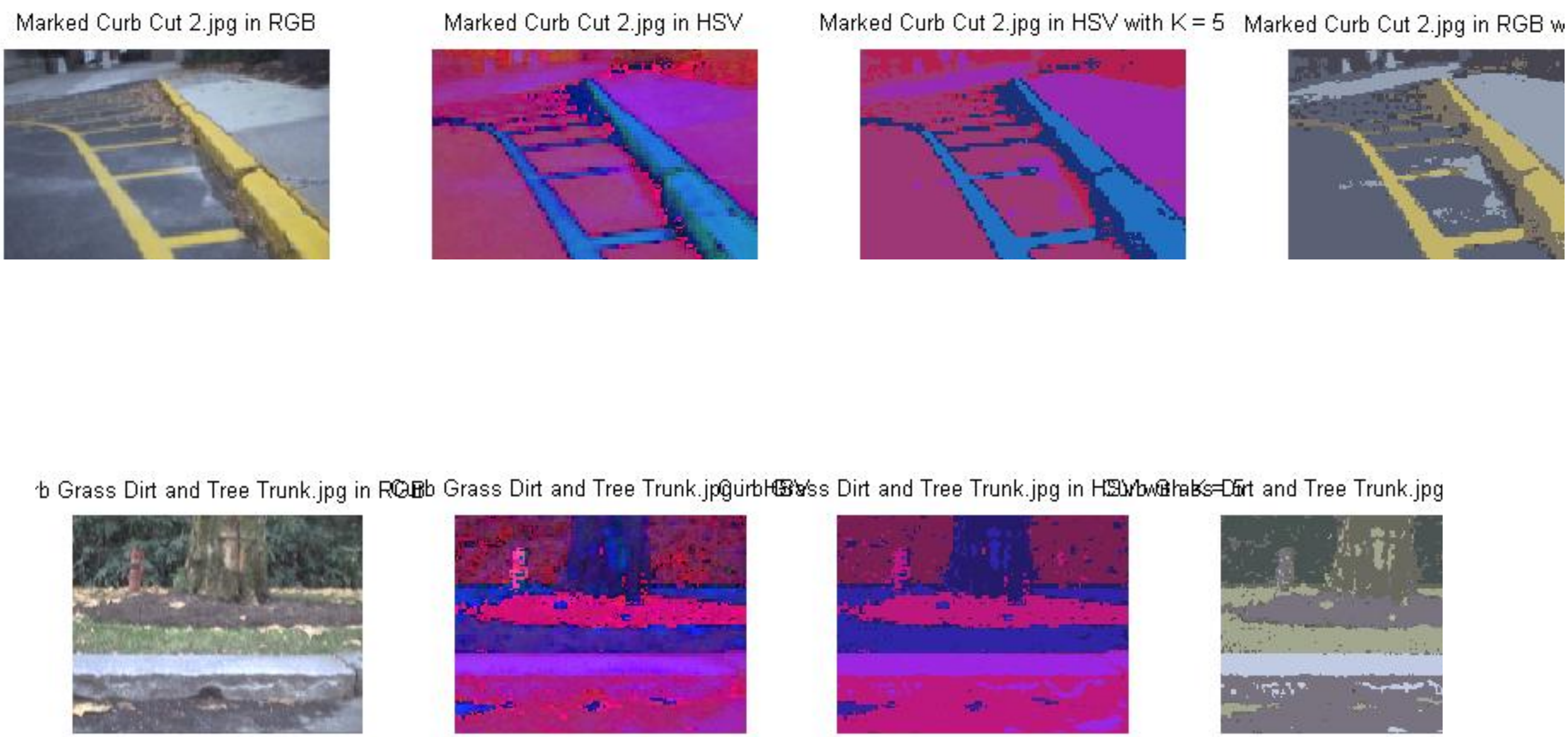


RSS student results



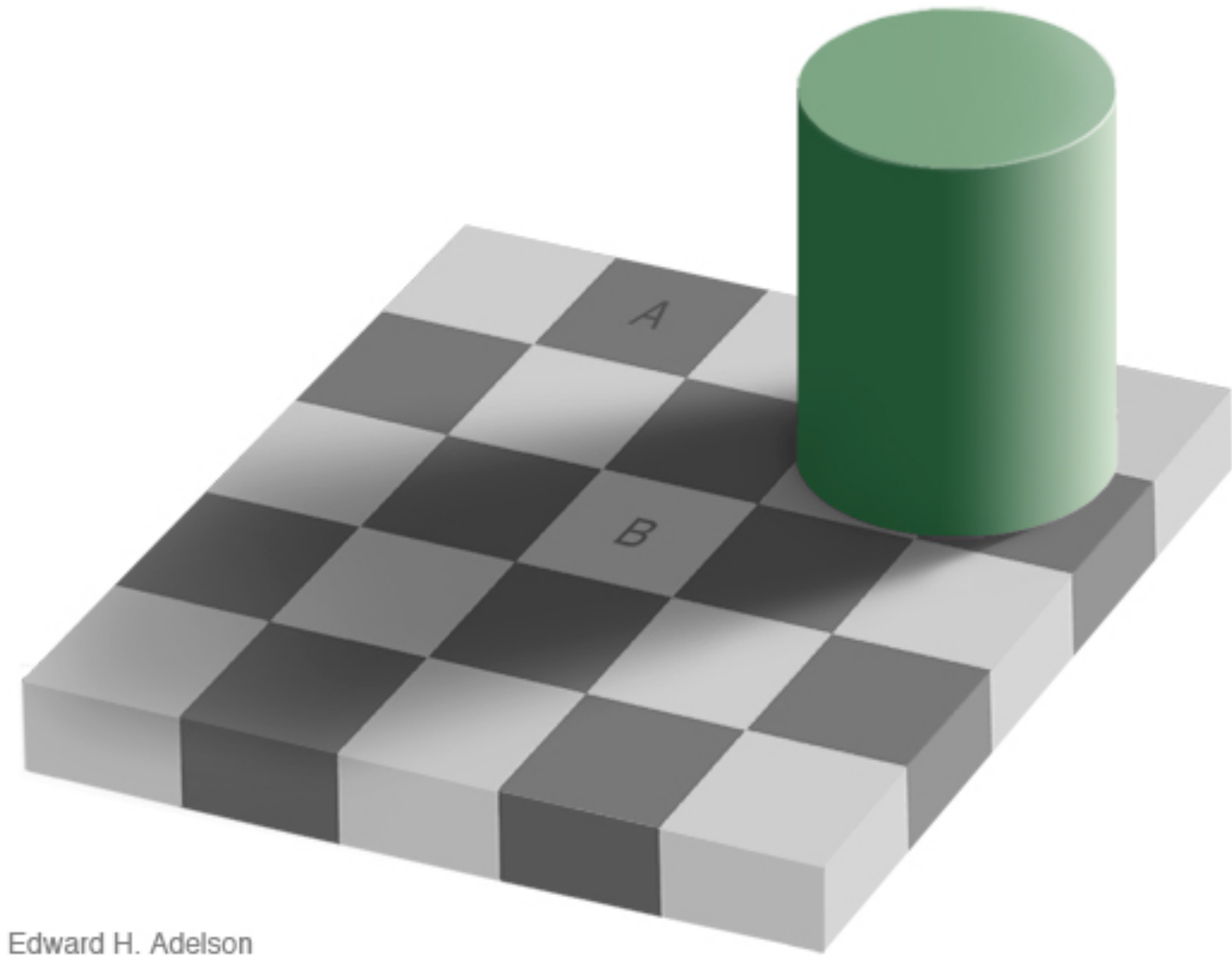


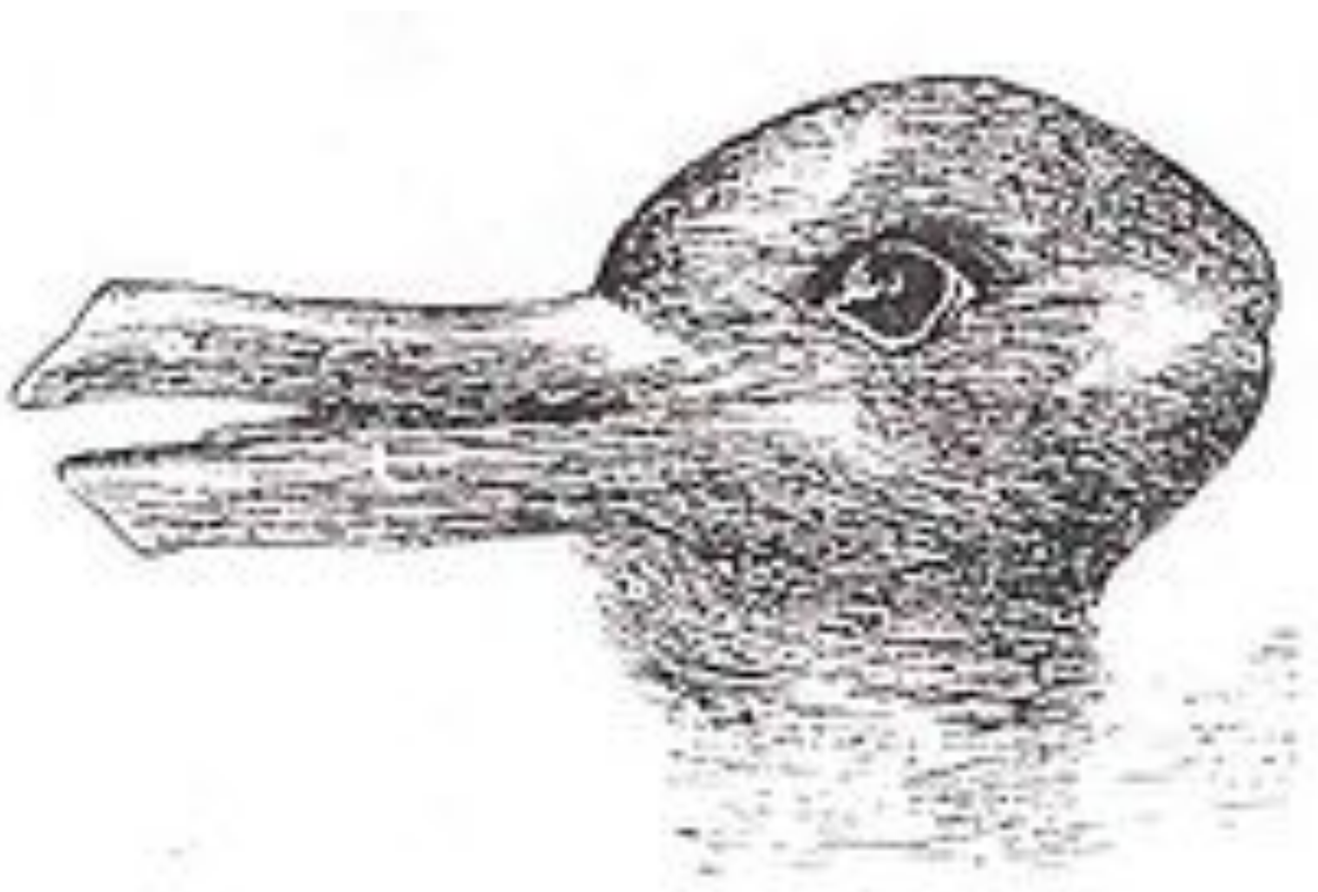
RSS student results



Human Visual System

- Adapts over both short and long time scales
  - Squinting (< 1 sec)
  - Pupil adjustment (~ 4 sec)
  - Retinal chemistry (~ 30 min)
- Adapts spatially
  - Color
  - Surround
  - Gestalt (completion) effects
- Variable resolution
  - Fovea
  - Periphery
  - Mix of color, intensity receptors
- Active
  - Saccading (20-200ms *joint* eye motions at ~500° per second)
  - Smooth pursuit (visual target tracking at up to 100° per second)
  - Closed-loop stabilization (vestibulo-ocular reflex)


A
B
Edward H. Adelson



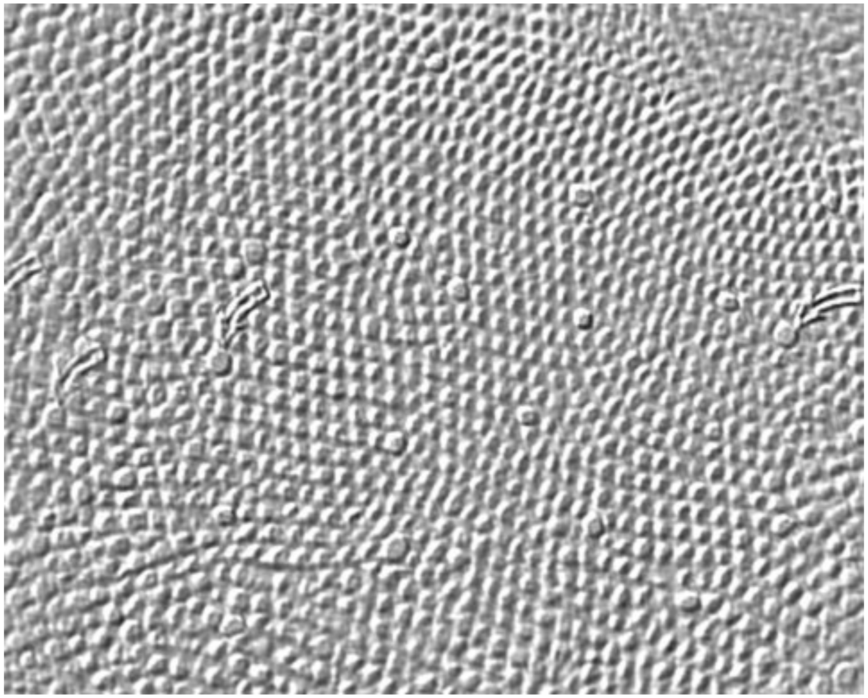
RSS I (6.141 / 16.415) S09

Coming up in RSS:

- Today:
  - Lab 2 (Motor and Motion Control) due
  - Lab 3 (Braintenberg Behaviors begins)
- Friday:
  - No CI-M lecture
  - Individual Project Architecture Reports due (Remember to post as **PDF**, email staff a **link**)
- Next Tuesday (virtual Monday):
  - Lecture 6: CARMEN robot control package
- Next Wednesday:
  - Lecture 7: pizza and DARPA Urban Challenge
  - Lab 3 briefings, wiki materials due
  - Lab 4 (Visual Servoing) begins

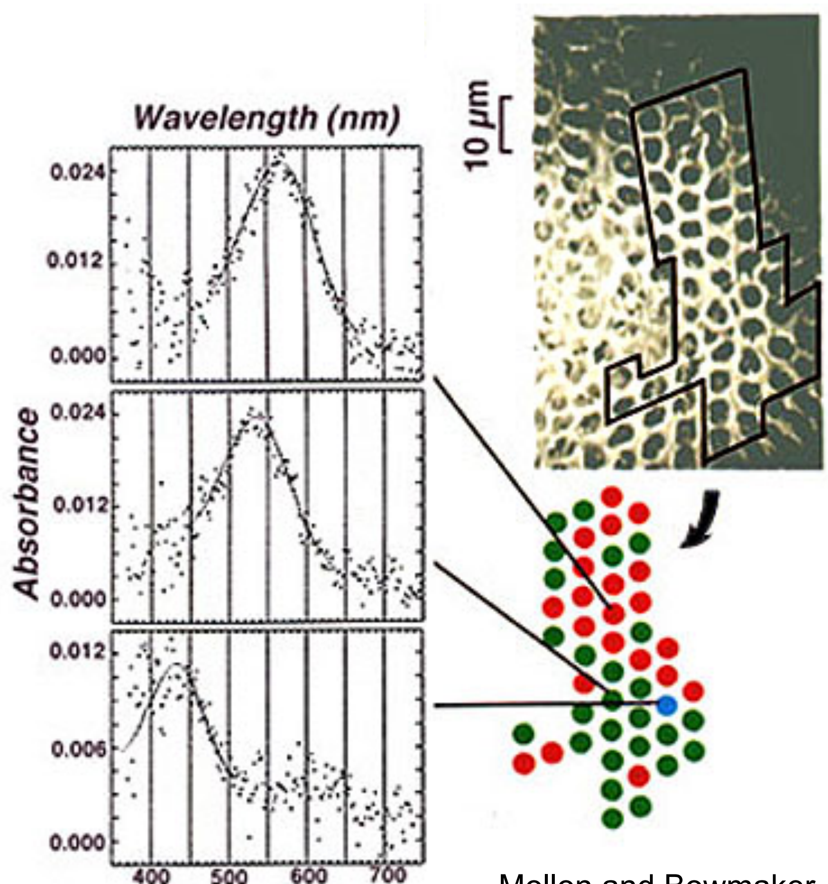
Foveal cone distribution

- Densely packed in fovea, less so in periphery



Visual discrimination of *1 minute of arc* [corresponds roughly to] the center-to-center spacing (3 µm) of the cones of the central mosaic in the foveola (retina.umh.es).

What does “1 minute of arc” mean?



Mollon and Bowmaker

RSS student results