Computer Vision Introduction
http://www.ugrad.cs.jhu.edu/~cs461
or
http://cirl.lcsr.jhu.edu/Vision_Syllabus

Professor Hager

http://www.cs.jhu.edu/~hager

Outline for Today

- Outline and Organization of the course.
- Some examples of what we will cover during the course.
- Q&A

Course Information

- Intended to be an introductory course on Computer Vision
  - Juniors, Seniors, Grad students
  - Both 361 and 461 version
    - 361 is a subset of 461

- Background needed
  - Some programming background (Data Structures)
  - Calculus and linear algebra
  - Basic physics

- Course structure
  - 2 lectures/week
  - Homework every 2 weeks, both written and programming (50%)
  - 1 Exam roughly late October (25%)
  - 1 Final project (25%)

Course Information

- Use the course WEB site
    - http://www.ugrad.cs.jhu.edu/~cs461 (old and unmaintained) or
    - http://cirl.lcsr.jhu.edu/Vision_Syllabus (new and improved?)
- What you need
    - One of the two recommended texts:
        - Computer Vision, Forsyth & Ponce
        - Introductory Techniques for 3D Computer Vision, Trucco and Verri
    - access to Matlab + Image Processing Toolbox
        - CS computing lab
        - Your own PC and the student edition (purchase online at mathworks.com; cost = approx. $160)
        - Any other matlab-capable computer

The Final Project

- The goal of the final project is to use vision technology to solve a "relevant" problem.

- Example 1:
  - I will provide a canned set of typical digital images with the following general properties:
    - Both indoor and outdoor
    - Multiple copies with large overlap and differing quality
    - Similar objects appearing in several

  - Your job is to provide tools to help automate organizing the photos:
    - Spot similar composition and put in a common folder
    - Link photos containing similar objects
    - Categorize as indoor or outdoor

The Final Project

- The goal of the final project is to use vision technology to solve a "relevant" problem.

- Example 1:
  - I will provide a canned set of typical digital images with the following general properties:
    - Both indoor and outdoor
    - Multiple copies with large overlap and differing quality
    - Similar objects appearing in several

  - Your job is to provide tools to help automate organizing the photos:
    - Spot similar composition and put in a common folder
    - Link photos containing similar objects
    - Categorize as indoor or outdoor

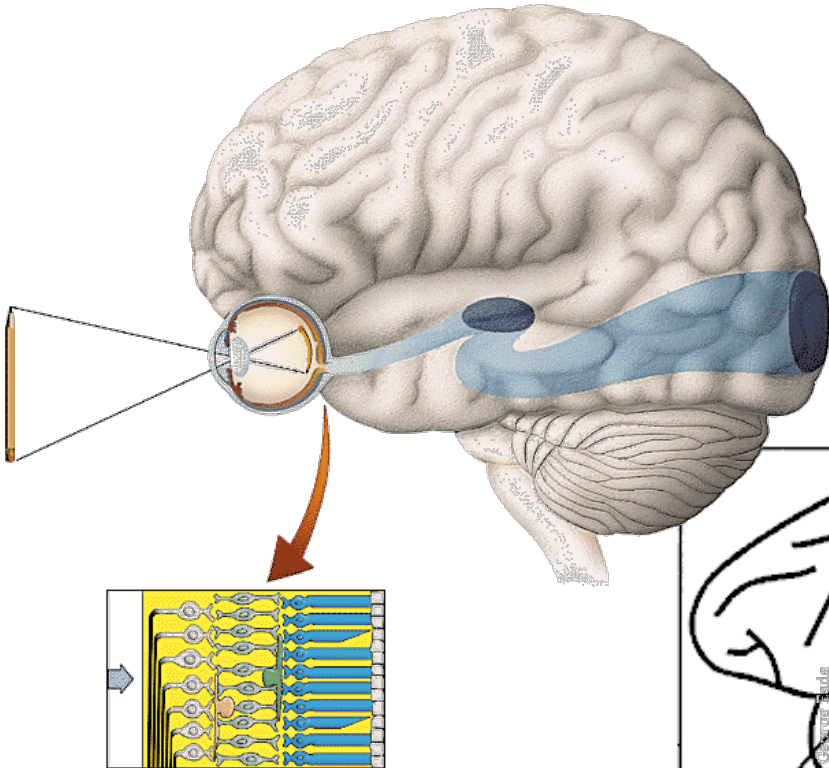
The Final Project

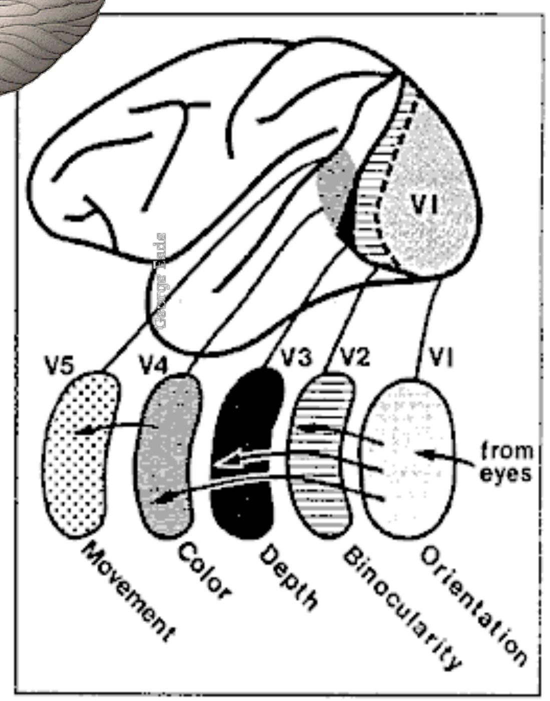
- The goal of the final project is to use vision technology to solve a "relevant" problem.

- Example 2:
  – If you have access to a laptop webcam
    - Develop a face detector for "logging in"
    - Differentiate different users

- Example 3:
  – Develop an interactive video game
    - Animate a ball (or similar icon)
    - Detection color/motion/pattern cues of a paddle
    - Create a multi-player video game

Vision: The Benchmark





More than half the brain is devoted to visual processing.

Processing is highly modularized

Oddly, we are better at subject rather than objective processing e.g. the right segmentation seems obvious to us but measuring the exact distance between objects is difficult

What is Computer Vision?

- **Trucco and Verri**
  - *computing properties of the 3D world from one or more digital images*

- **Stockman and Shapiro**
  - *To make useful decisions about real physical objects and scenes based on sensed images*

- **Ballard and Brown**
  - *The construction of explicit, meaningful description of physical objects from images*

- **Forsyth and Ponce**
  - ***...**extracting descriptions of the world from pictures or sequences of pictures*

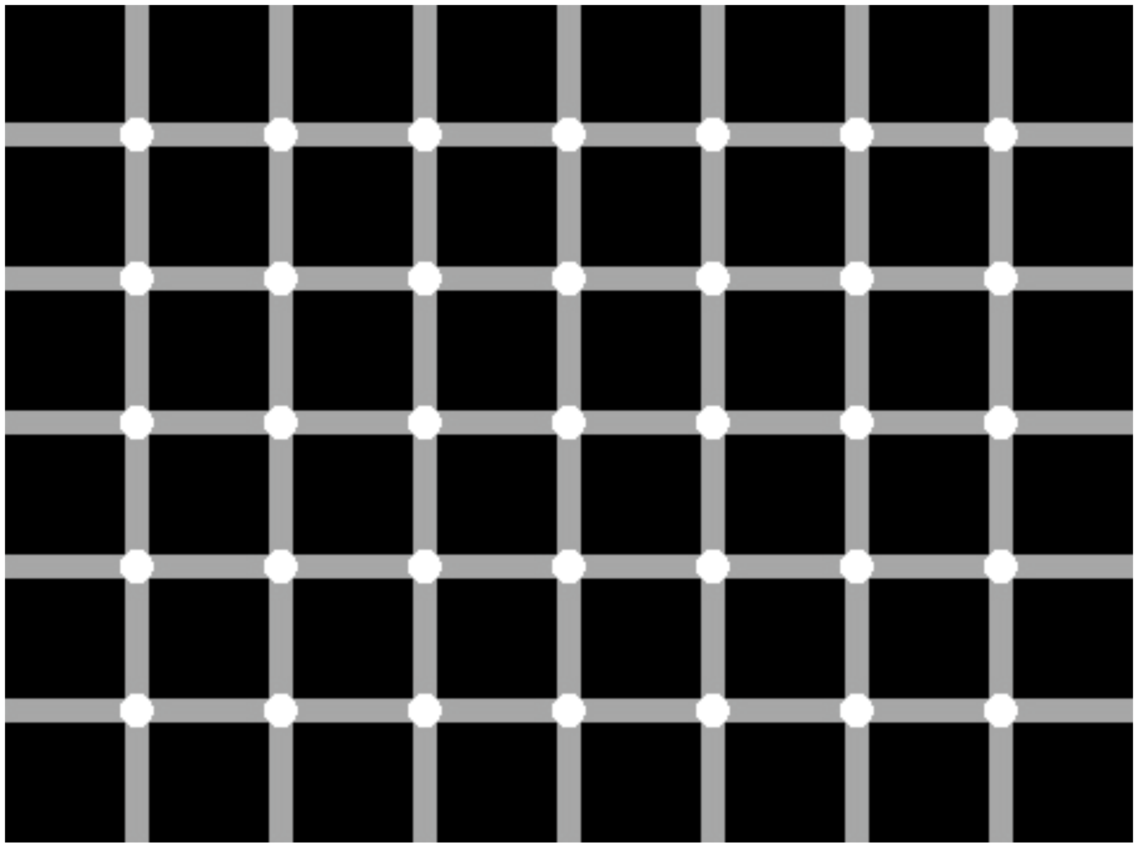
Some Related Terms

- **Image Processing:** the study of the properties of operators that produce images from other images
  - we will touch on image filtering and related operators from image processing
- **Machine Vision:** a somewhat outdated term which now tends to refer to industrial vision applications where (usually) a single camera is used to solve a structured inspection task
  - the "reverse CAD" model
- **Pattern Recognition:** typically refers to the recognition of structures in 2D images (usually without reference to any underlying 3D information).
- **Photogrammetry:** the science of measurement though non-contact sensing, e.g. terrain maps from satellite images. Usually is more focused on accuracy issues than interpretation.

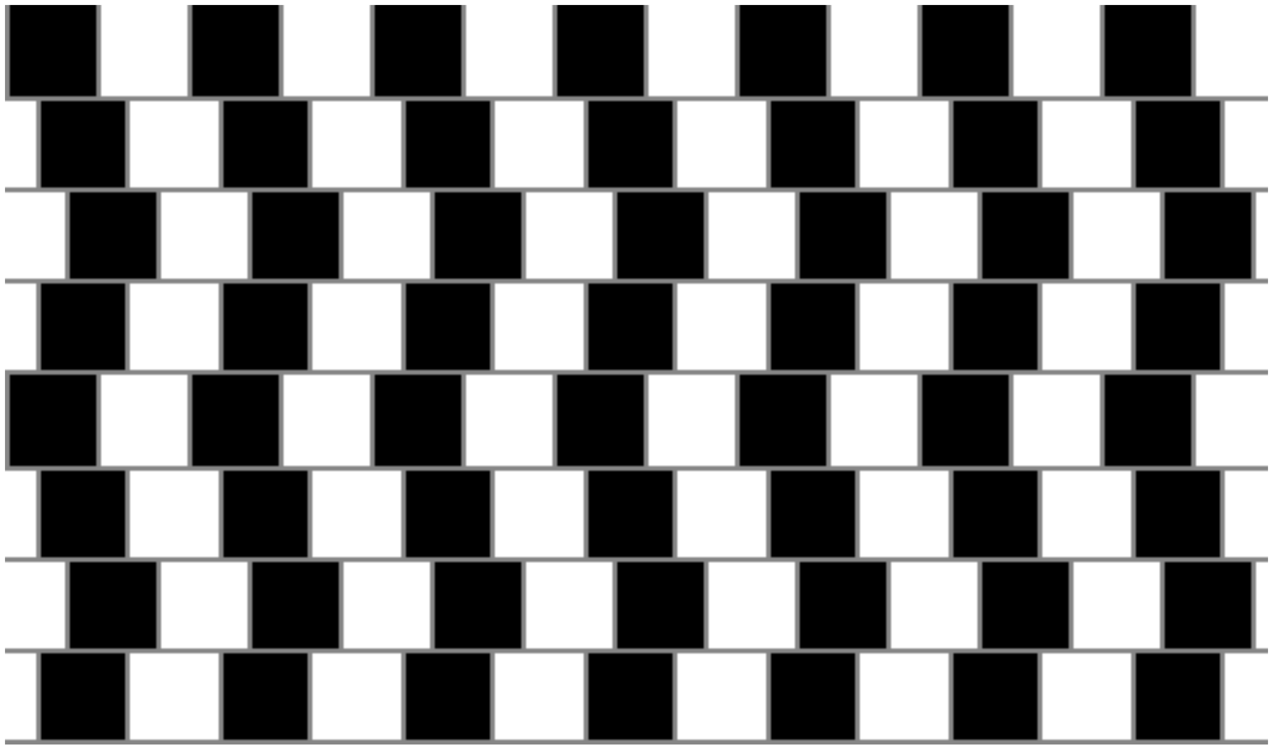
Why Is Vision Hard?

Why Is Vision Hard?

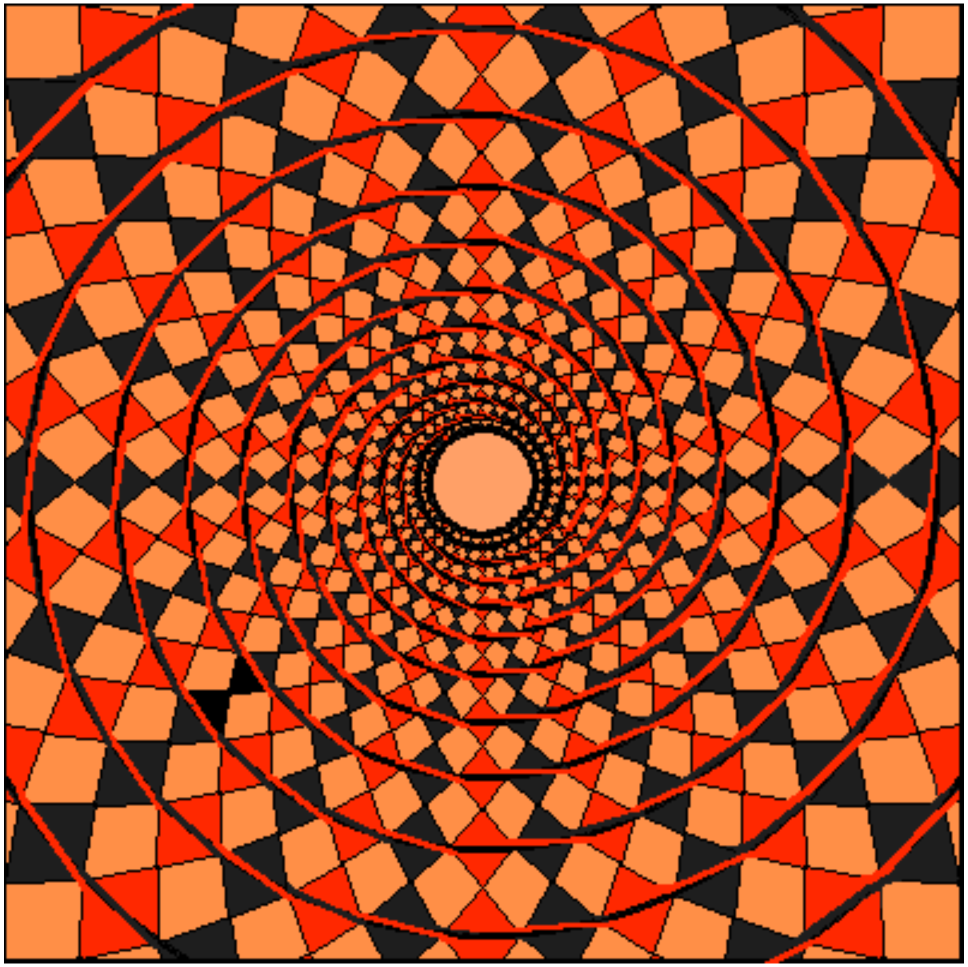
Illusions: What Do They Tell Us?



Illusions: What Do They Tell Us?



Illusions: What Do They Tell Us?



Illusions: What Do They Tell Us?

Illusions: What Do They Tell Us?

Illusions: What Do They Tell Us?

Illusions: What Do They Tell Us?

Why Is Vision Hard?

- “Context” counts for as much as appearance
- Huge amounts of prior knowledge (learned and innate)
- AI complete (hard to solve a cleanly defined “simple” problem without invoking unrealistic assumptions)
- Lack of a clear metric for success
- The diversity of the natural world

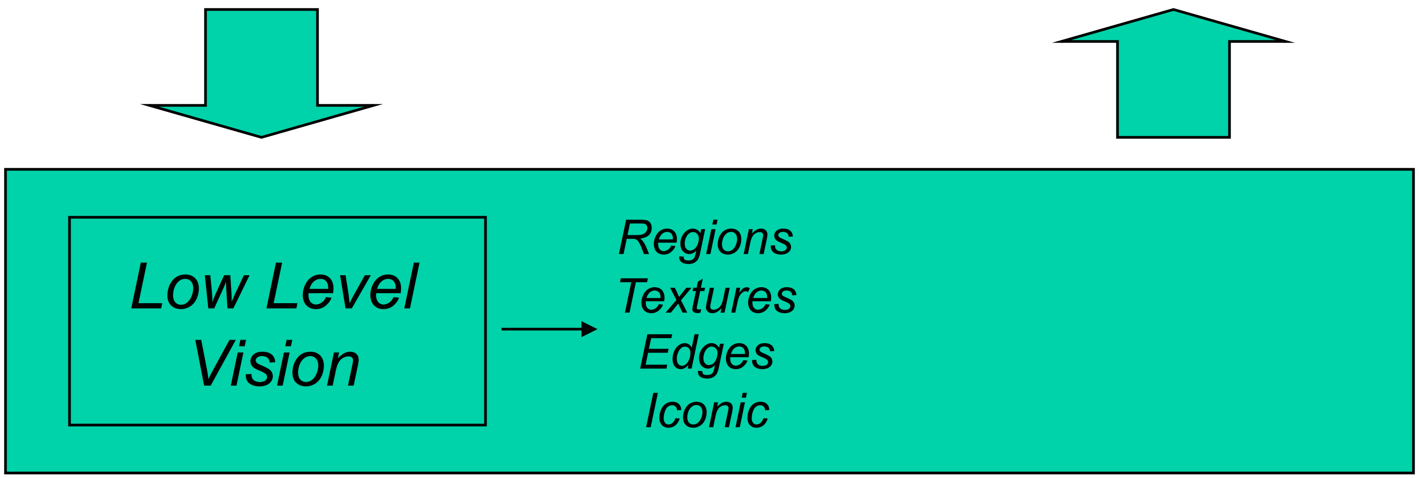
A Model for Vision

Geometry
Objects
Motion
Texture
Lighting
Movement
Activity....

? Vision Processing ?

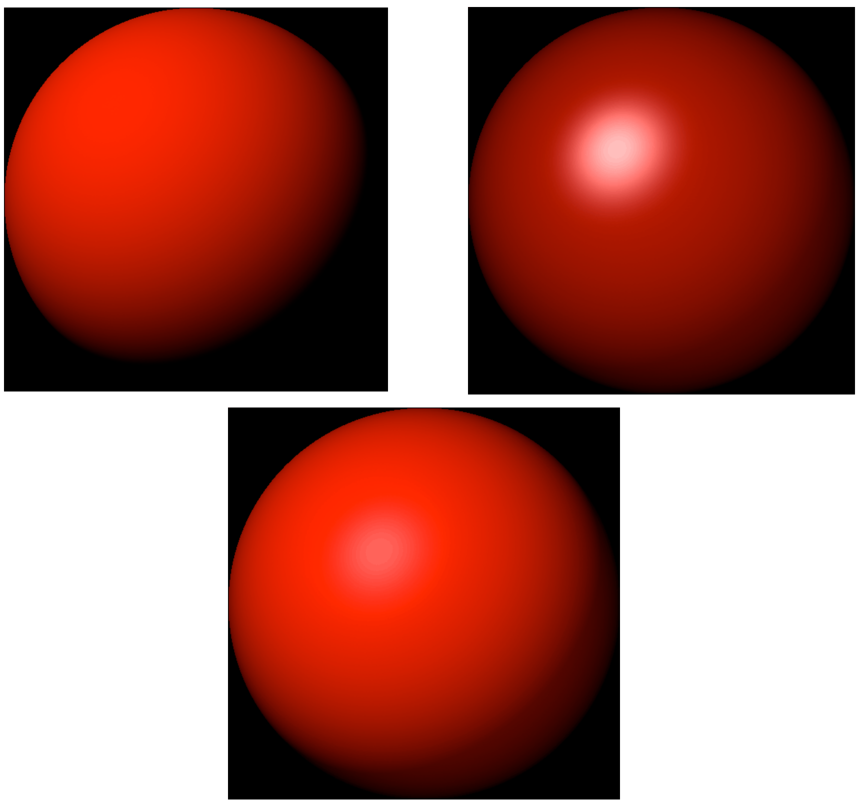
A Model for Vision

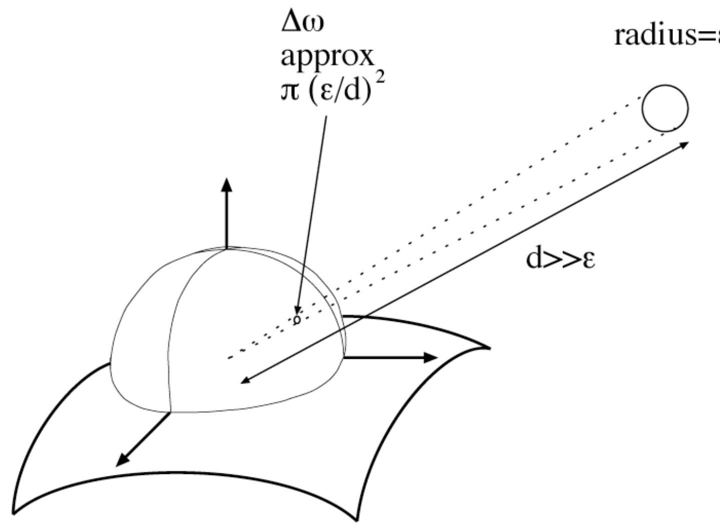
Geometry
Objects
Motion
Texture
Lighting
Movement
Activity....

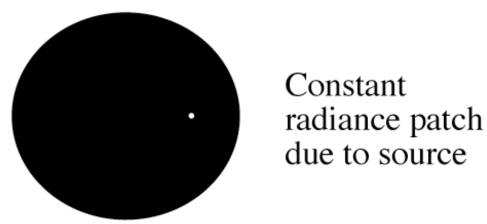


Problems of Computer Vision: Modeling

What are the physical and geometric processes that govern (digital) imaging?

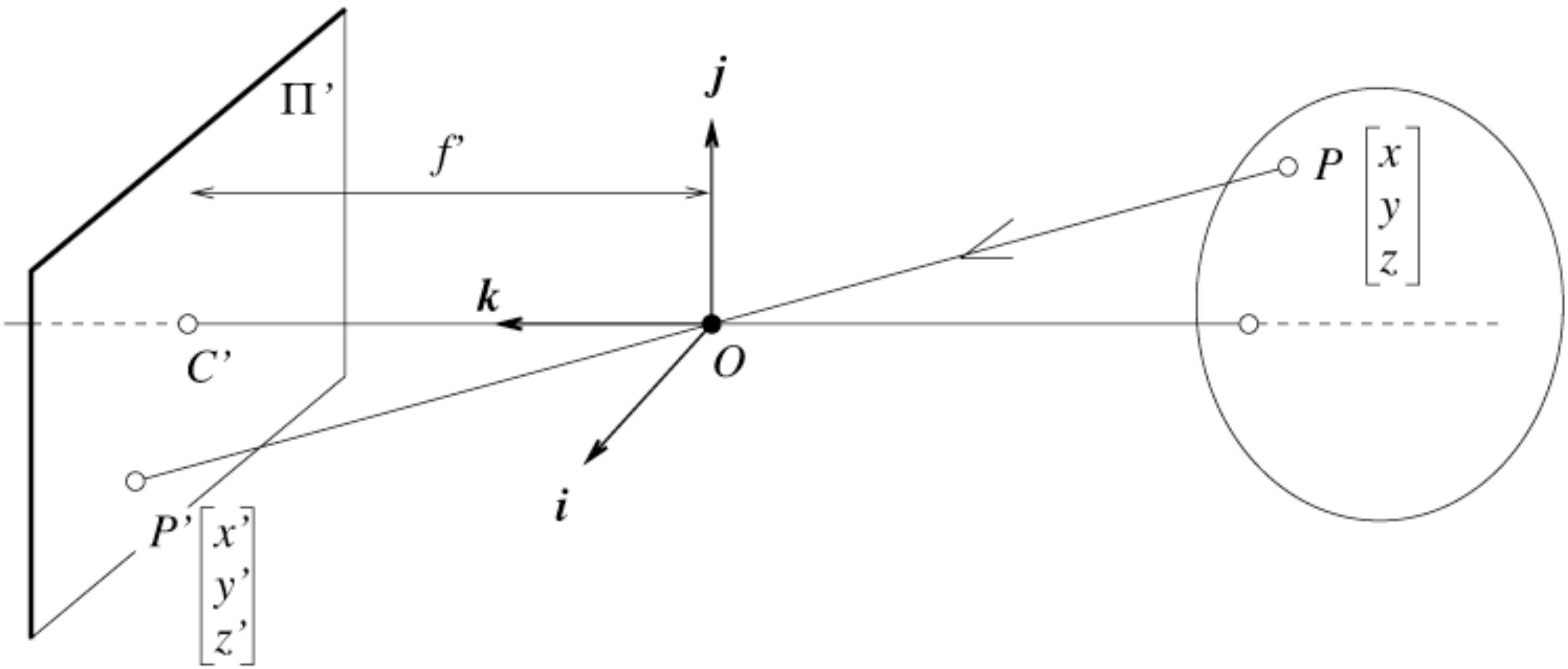






Problems of Computer Vision: Modeling

What are the physical and geometric processes that govern (digital) imaging?



General Rules

If you can't understand (i.e. model) the forward process, you will have a hard time solving the inverse!

A related point: the best way to test vision algorithms is almost always to implement the forward model to test the (inverse) solution.

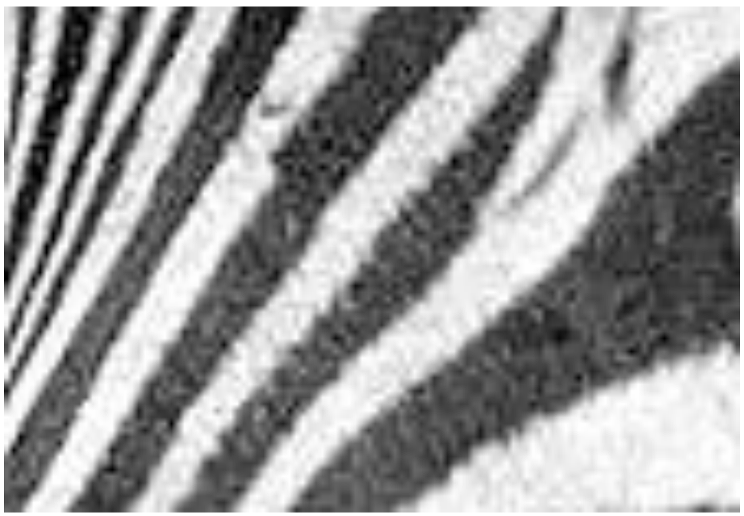
Computer Vision vs. Graphics

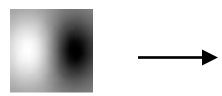
Is Vision the "Inverse" of Graphics?

- Computer Graphics
  - Produce "plausible" images
  - You choose the models, conditions, imaging parameters, etc.
- Computer Vision
  - Given real images with noise, sampling artifacts …
  - Estimate physically quantities
  - Ill-posed ---- what is the minimum world knowledge we need?

Problems of Computer Vision: Image Enhancement and Feature Extraction

What are the "informative" areas of an image and how do we detect them?





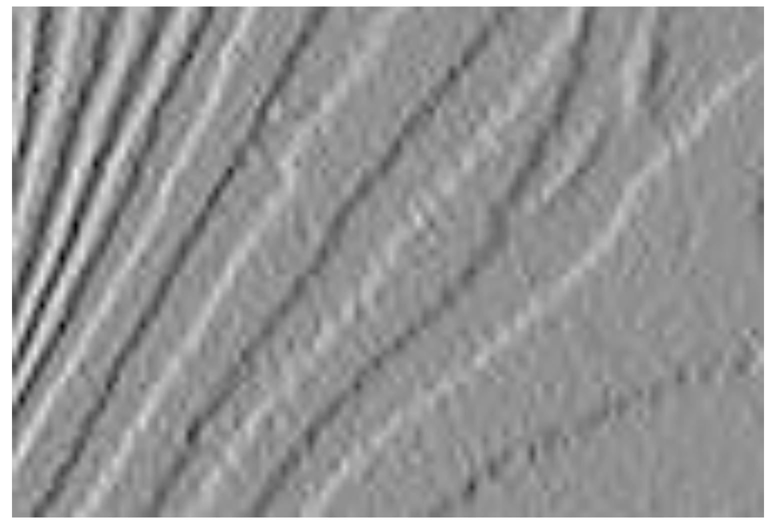


Image Filter Result

Feature Extraction: Edge Detection

Thresholding suppresses “non-feature” areas of the image

Feature Extraction for Object Recognition

(David Lowe)

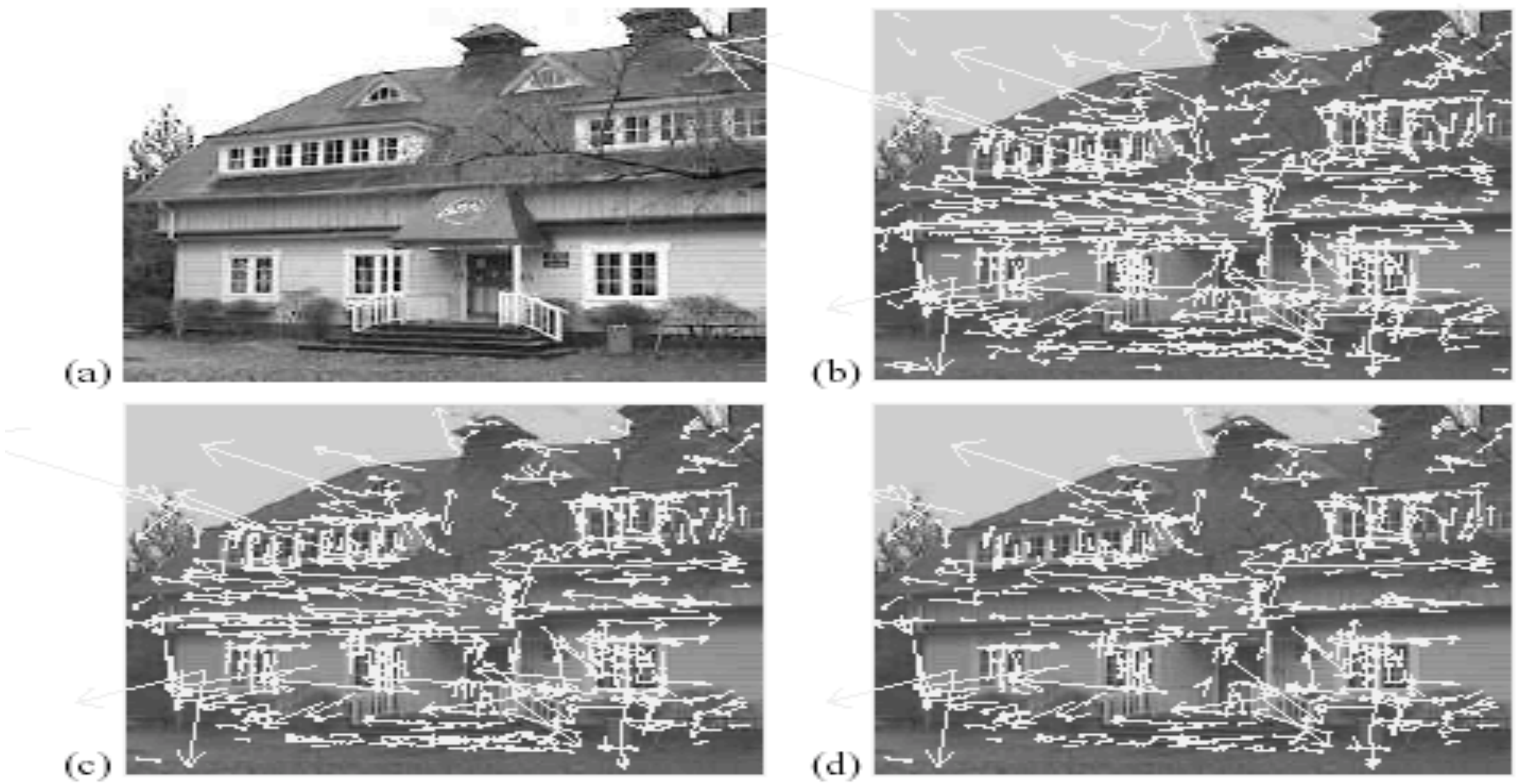


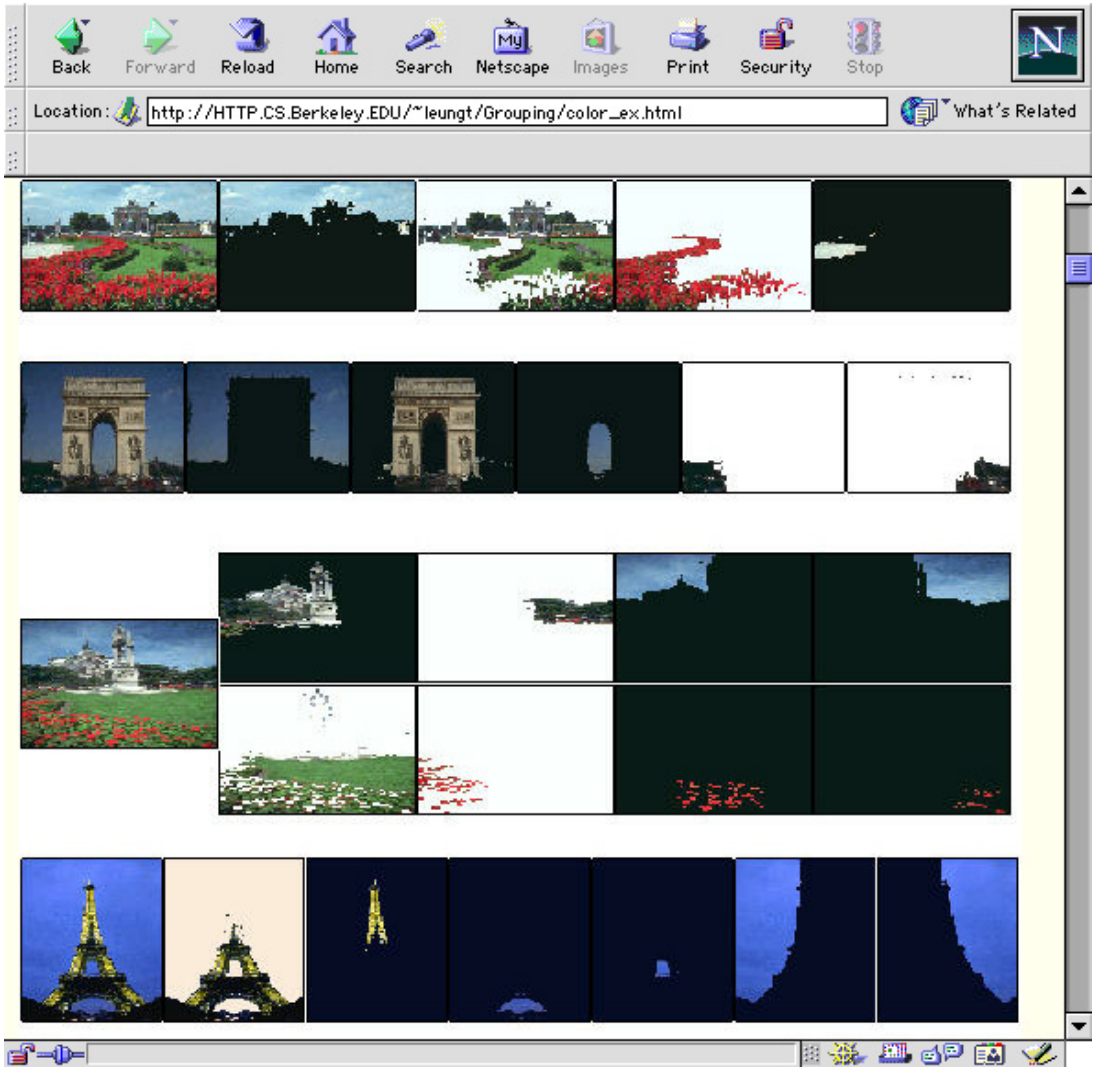
Figure 5: This figure shows the stages of keypoint selection. (a) The 233x189 pixel original image. (b) The initial 832 keypoints locations at maxima and minima of the difference-of-Gaussian function. Keypoints are displayed as vectors indicating scale, orientation, and location. (c) After applying a threshold on minimum contrast, 729 keypoints remain. (d) The final 536 keypoints that remain following an additional threshold on ratio of principle curvatures.

Computer Vision vs. Image Processing

- Image Processing
  - Mostly concerned with *image-to-image* transformations
    - Filtering
    - Enhancement
    - Compression

- Computer Vision
  - Concerned with how images *reflect the 3D world*
    - Filtering *for feature extraction*
    - Enhancement *for recognition/detection*
    - Compression *that preserves geometric information in images*

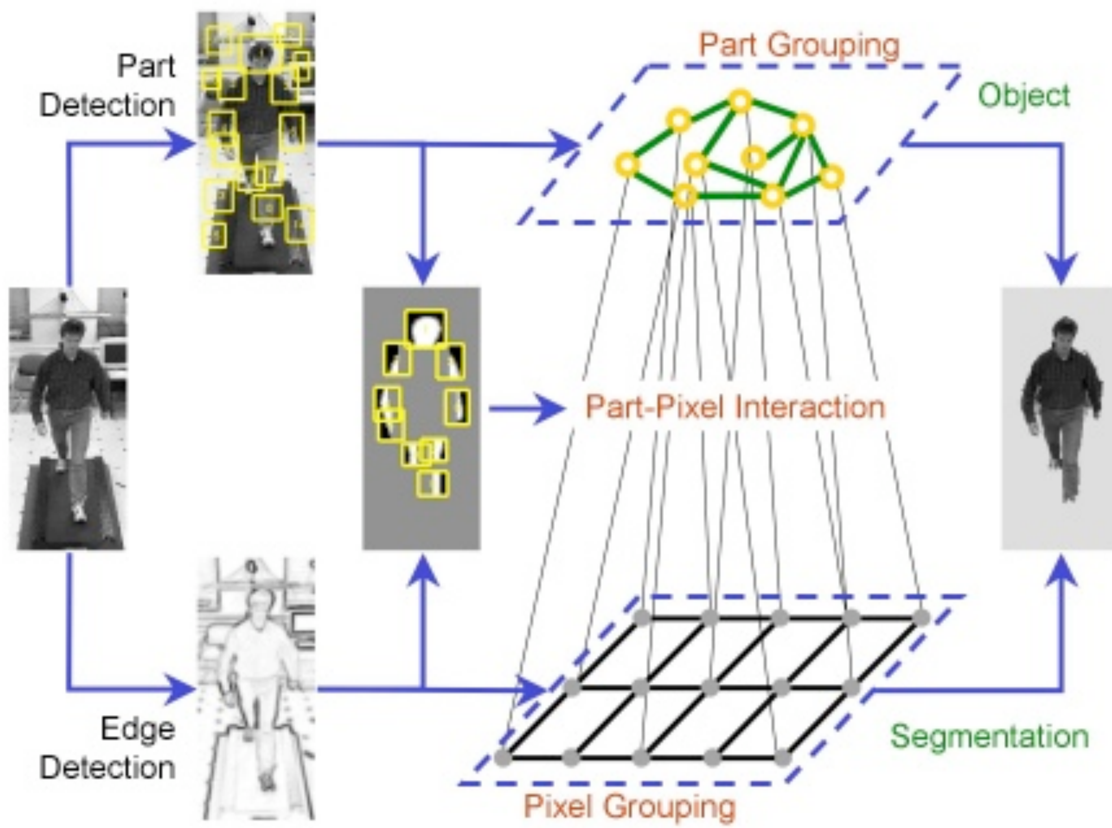
Problems of Computer Vision: Segmentation and Grouping



What portions of an image pertain to one another and to relevant physical phenomena?

Problems of Computer Vision: Segmentation and Grouping

(Yu, Gross, Shi)



What portions of an image pertain to one another and to relevant physical phenomena?

Problems of Computer Vision: Segmentation and Grouping (Lu, Hager)

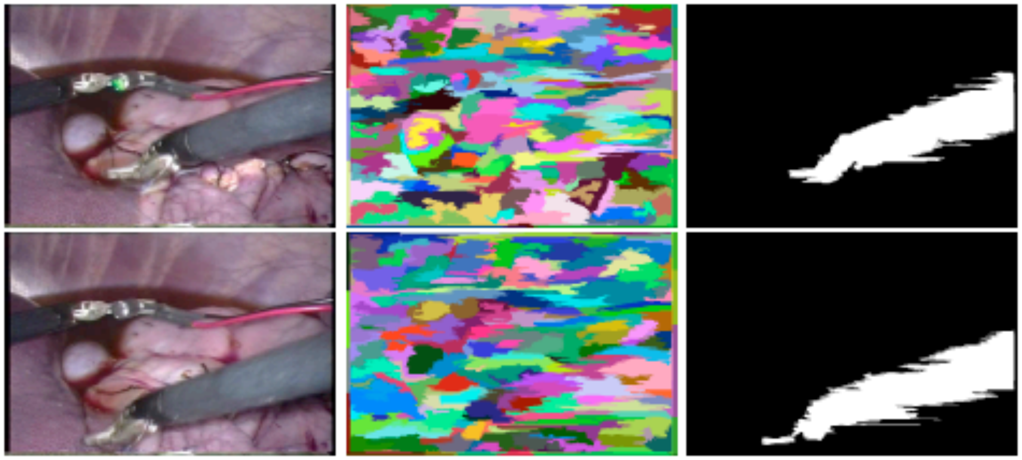


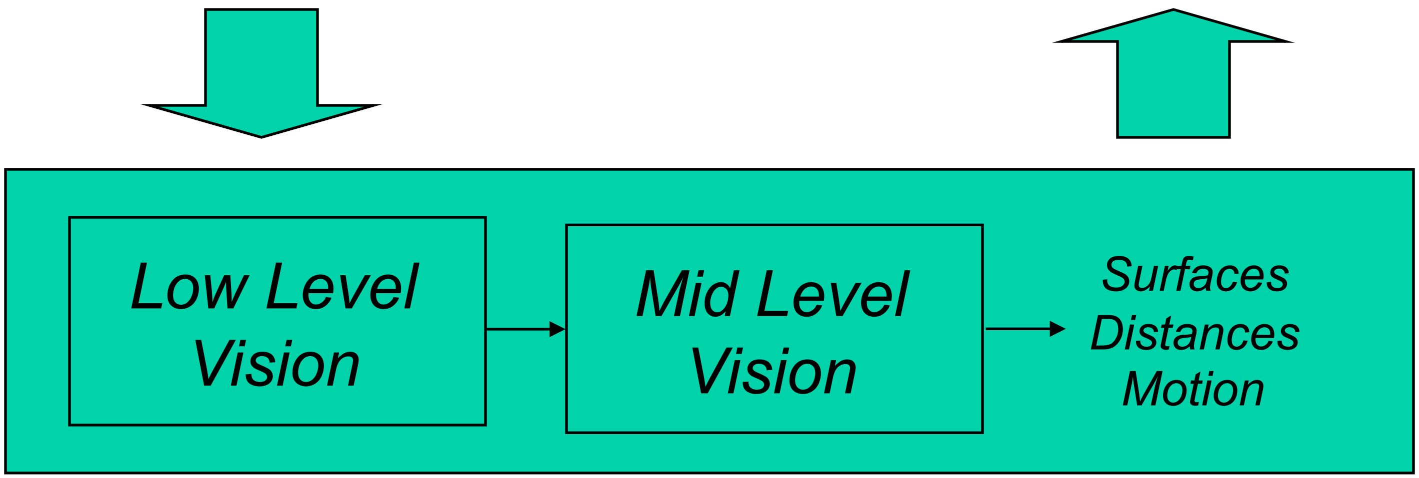
Figure 5: **Top Left:** An image for learning the foreground/background appearance model; **Top Middle:** Its segmentation; **Top Right:** Its labelling mask (White is foreground; black is background); **Bottom Left:** Another image for testing the appearance model; **Bottom Middle:** Its segmentation; **Bottom Right:** Its detected foreground/background mask. We use the patch based raw RGB intensity vector matching and the nearest neighbor matching. Notice the motions between 2 images. Image resolution is 720 by 488 pixels.

(a) 12# (b) 91# (c) 155# (d) 180# (e) 221# (f) 257# (g) 308# (h) 329#

Figure 6: Eight example frames (720 by 480 pixels) from the video sequence *Karsten.avi* of 330 frames. The video is captured using a handheld Panasonic PV-GS120 in standard NTSC format. Notice that the significant non-rigid deformations and large scale changes of the walking person, while the original background is completely substituted after the subject turned his way. The red pixels are on the boundary of segments; the tracked image segments associated with the foreground walking person is coded in blue.

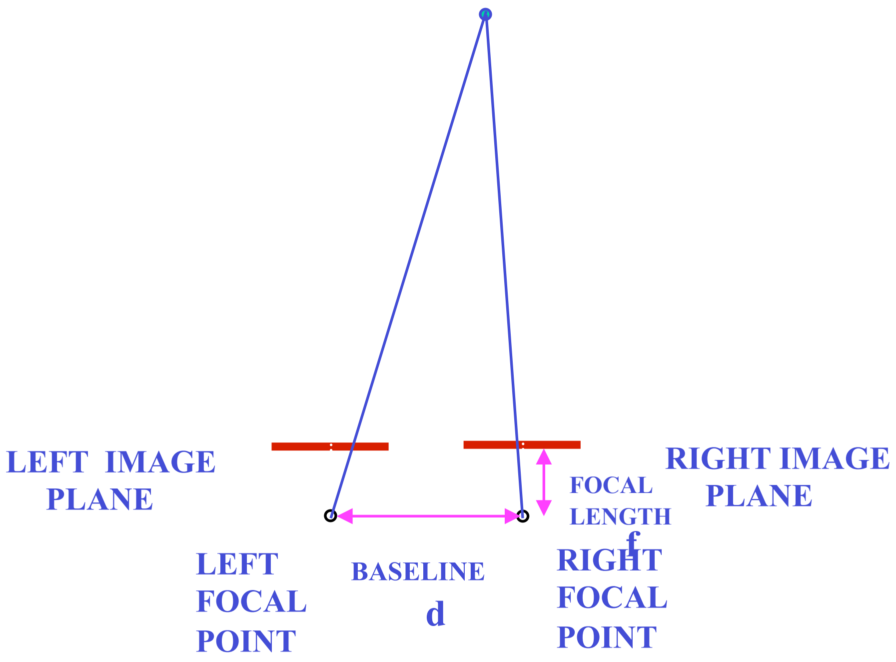
A Model for Vision

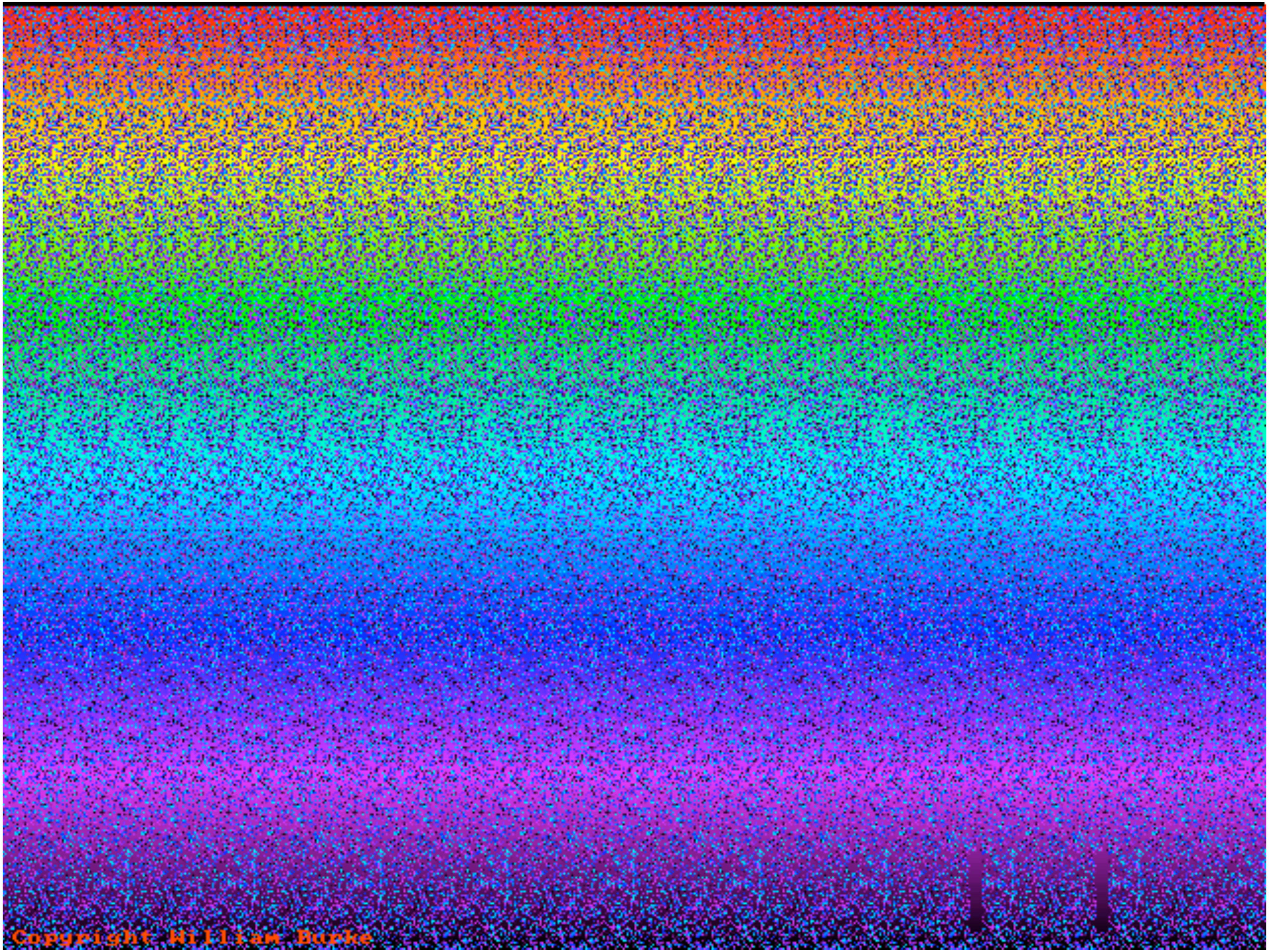
Geometry
Objects
Motion
Texture
Lighting
Movement
Activity….



Problems of Computer Vision: Stereo Vision

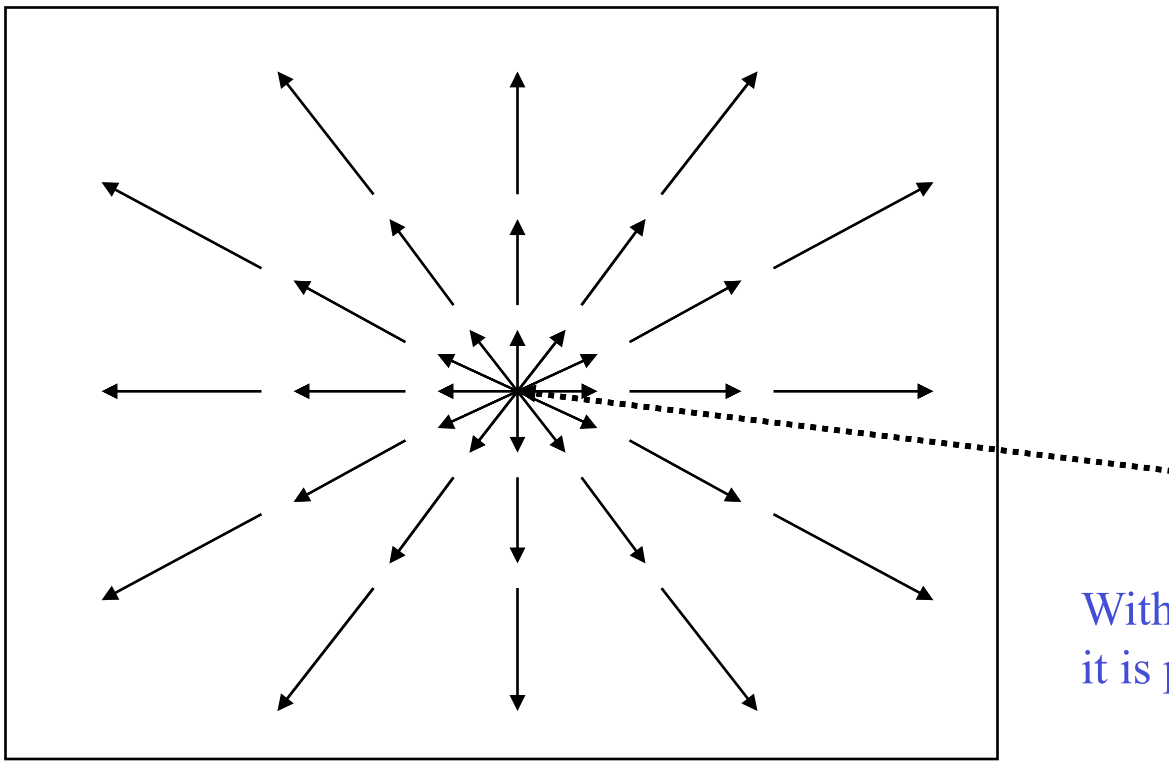
From two (or more) images, determine the geometry of the scene by *matching* corresponding areas of the images




Copyright William Burke

THE MOTION FIELD

The "instantaneous" velocity of points in an image



LOOMING

The *focus of expansion*

With just this information it is possible to calculate:

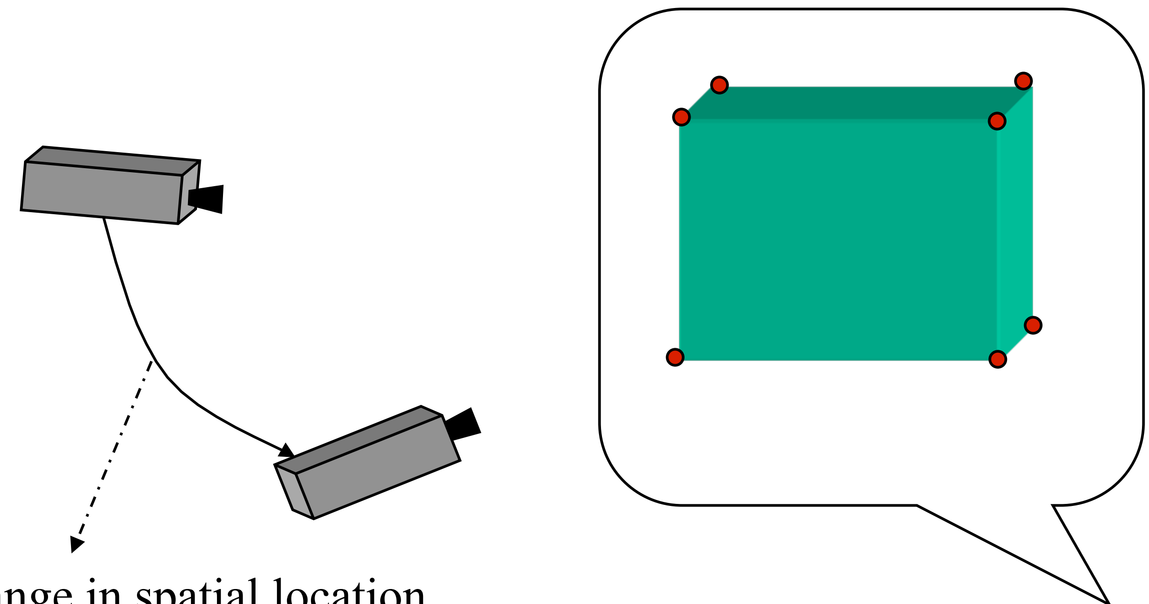
1. Direction of motion
2. Time to collision

Examples

Template Tracking (Hager)

Multiple Target Tracking (Okuma et al.)

MOVING CAMERAS ARE LIKE STEREO



The change in spatial location
between the two cameras (the "motion")

Locations of
points on the object
(the "structure")

Examples

(Courtesy Carlo Tomasi)

(Courtesy Marc Pollefeys)

Another Example

An automatically generated panorama (Matthew Brown)

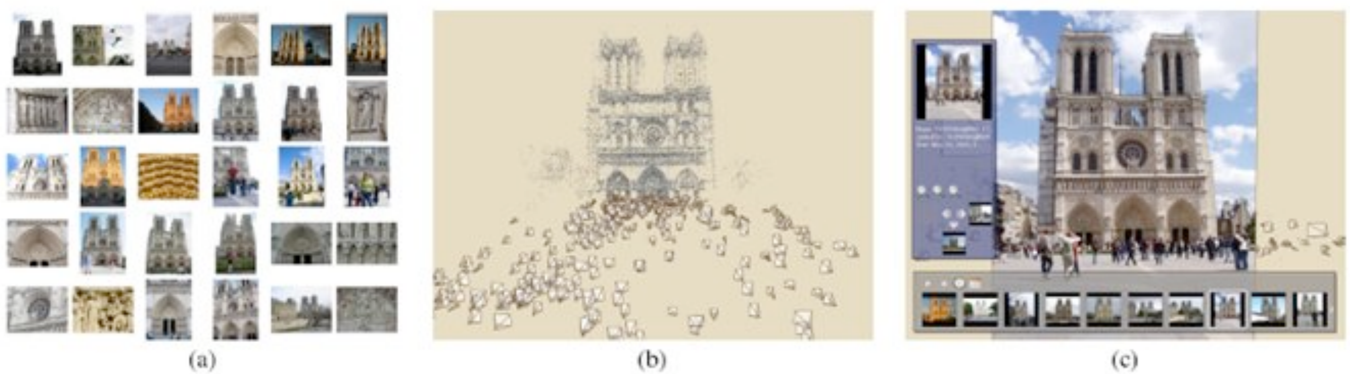
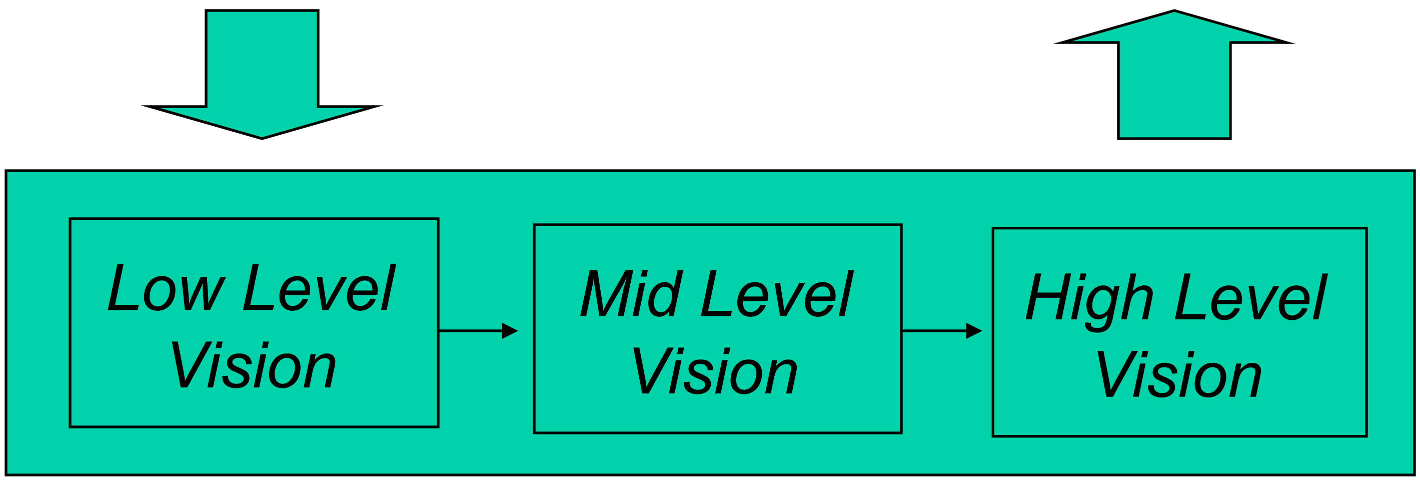


Figure 1: Our system takes unstructured collections of photographs such as those from online image searches (a) and reconstructs 3D points and viewpoints (b) to enable novel ways of browsing the photos (c).

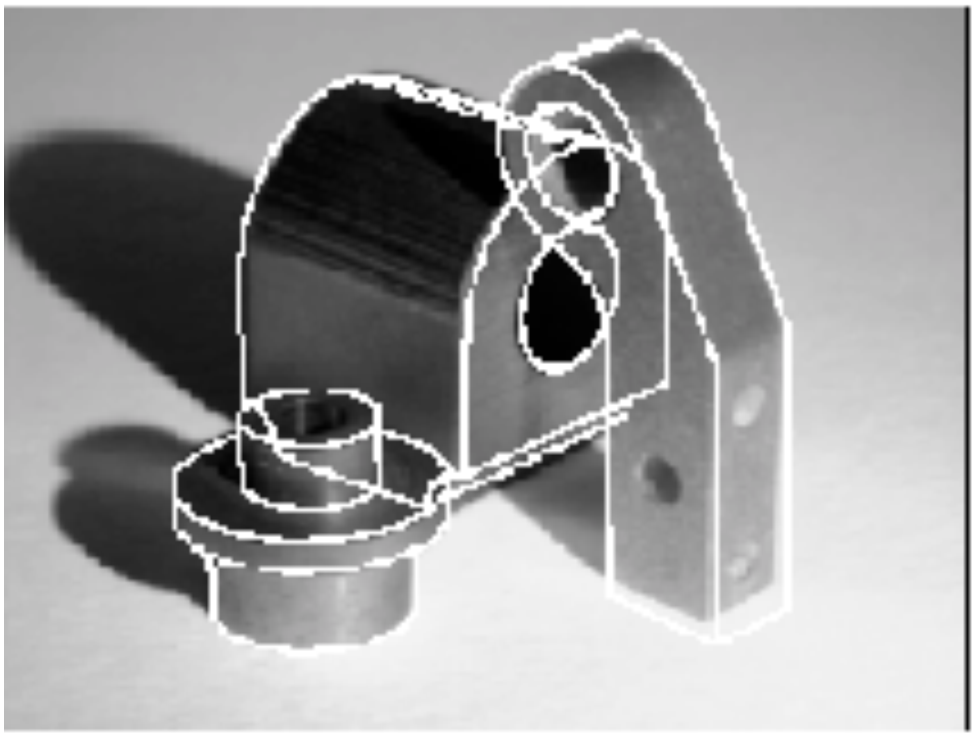
The MSR Photosynth Project *http://photosynth.net/Default.aspx*

A Model for Vision

Geometry
Objects
Motion
Texture
Lighting
Movement
Activity....

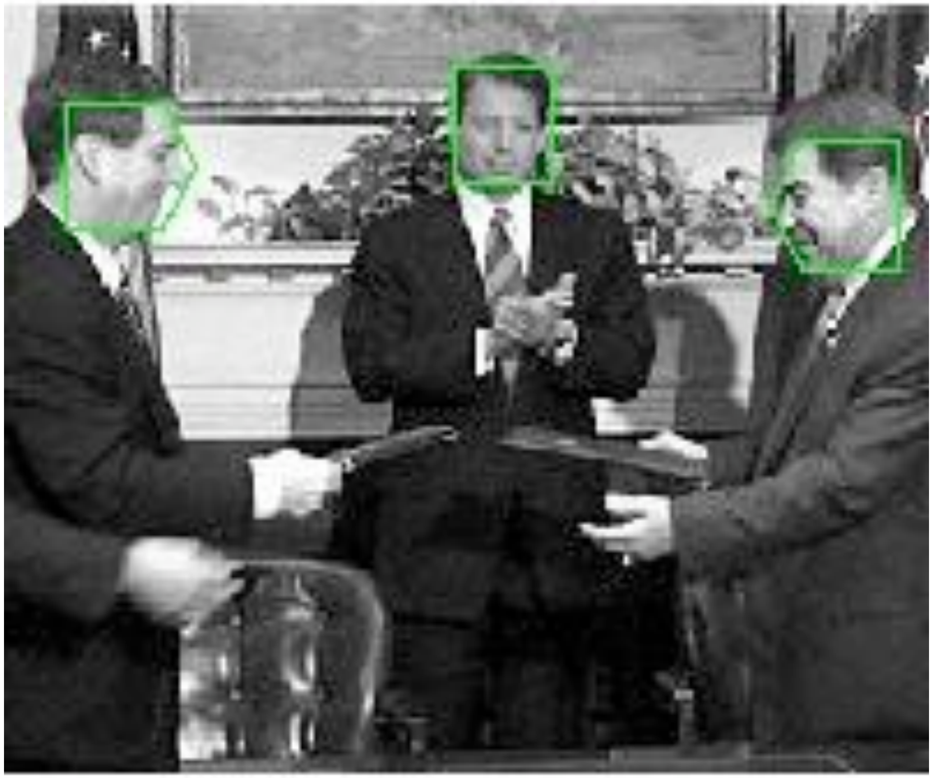


Problems of Computer Vision: Recognition

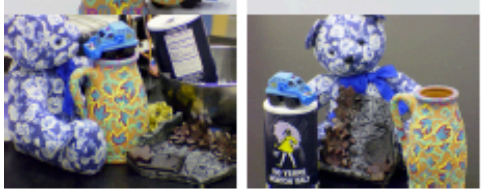


Given a database of objects and an image determine what, if any of the objects are present in the image.

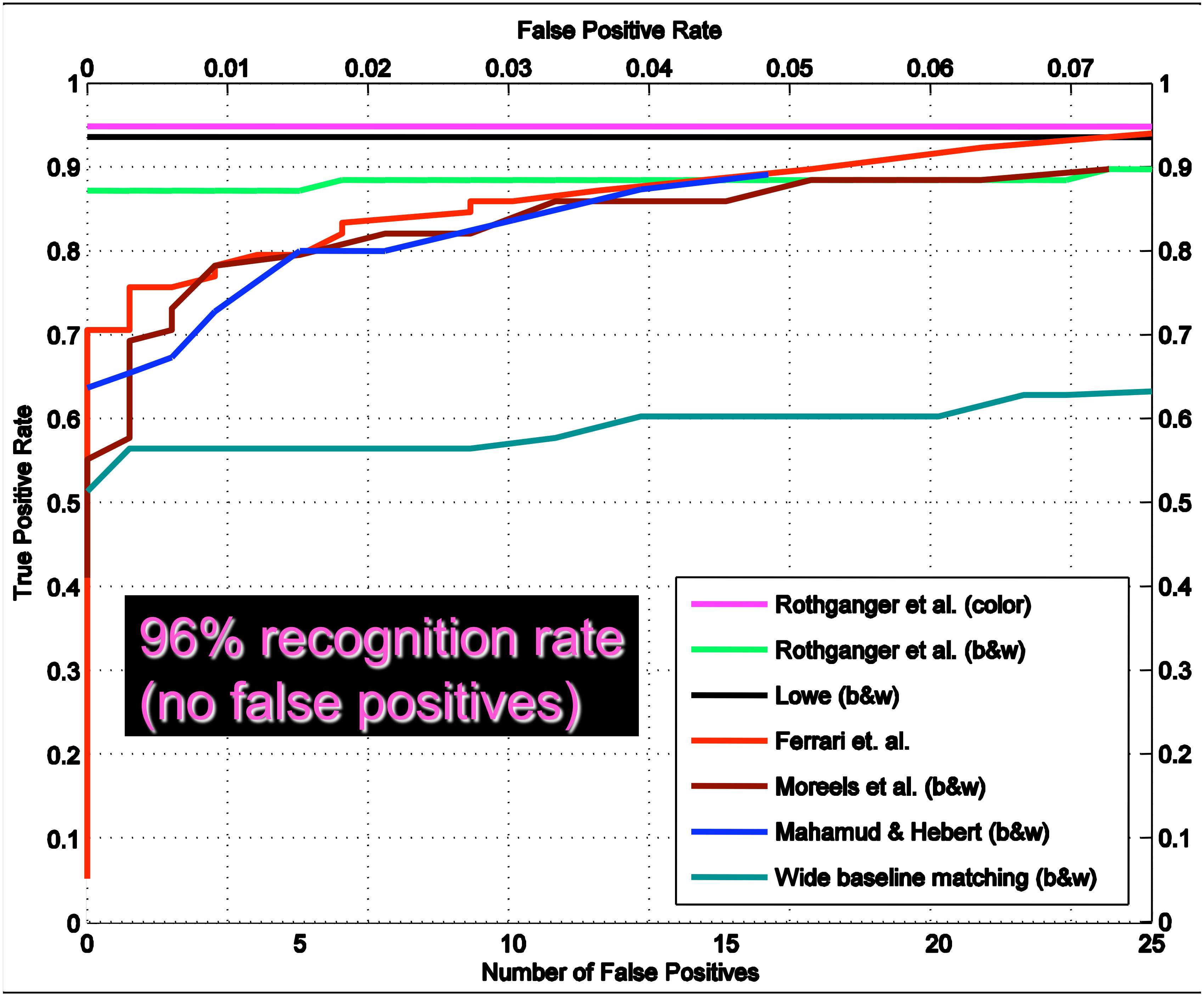
Problems of Computer Vision: Recognition



Given a database of objects and an image determine what, if any of the objects are present in the image.



Dataset: 51 test images with 1 to 5 of the 8 objects present in each image.


False Positive Rate
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
True Positive Rate
96% recognition rate
(no false positives)
Rothganger et al. (color)
Rothganger et al. (b&w)
Lowe (b&w)
Ferrari et. al.
Moreels et al. (b&w)
Mahamud & Hebert (b&w)
Wide baseline matching (b&w)
0
5
10
15
20
25
Number of False Positives

Can We Ever Make Vision Work?

Biology: We have eyes, as do many animals. Here is an extreme example:

Stomatopod eyes are unusual:

- they have stereo vision with just one eye;
- each eye is on a stalk, with a wide range of motion
- stomatopods have up to 16 visual pigments
- stomatopods can also see ultra-violet and infra-red light
- some can even see polarized light

Moving On ...

- Computer vision is still far from most biological systems, but ...
- After several decades of often highly experimental and anecdotal progress ...
- The previous decade saw huge advances in understanding geometric issues in vision, and ever more practical problems attacked....
- With the growth of computing power, it is possible to perform more and more complex processing on more and more images....
- Now, recent approaches/advances take advantage of heavily data-driven approaches
- We will touch on all of these points in more or less detail, starting from the bottom up ...

Your "Homework"

- Get a book
- Make sure you have access to a computer with matlab
- Try accessing the WEB site
- Don't stay up too late!

A Word On Computer-Imaging

- Video imaging has gone from an exotic technology to everyday commodity.

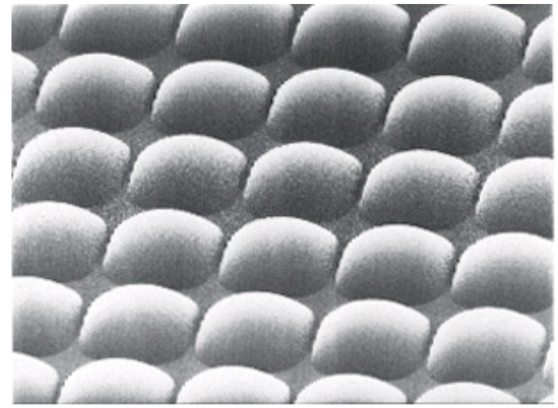
- Originally (since ~1930) NTSC standard
  - 480 x 640 YUV
  - Interlaced

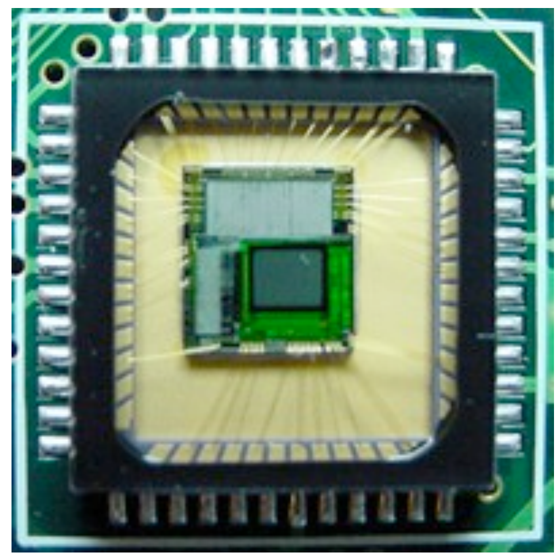
- Now, a wide variety of resolutions and quality
  - VGA (= NTSC)
  - SVGA (= 600x800)
  - XVGA (= 768x1024)
  - SXGA (=1024x1280)
  - UGA (= 1200x1600)

How Cameras Produce Images

- Basic process:
  - photons hit a detector
  - the detector becomes charged
  - the charge is read out as brightness

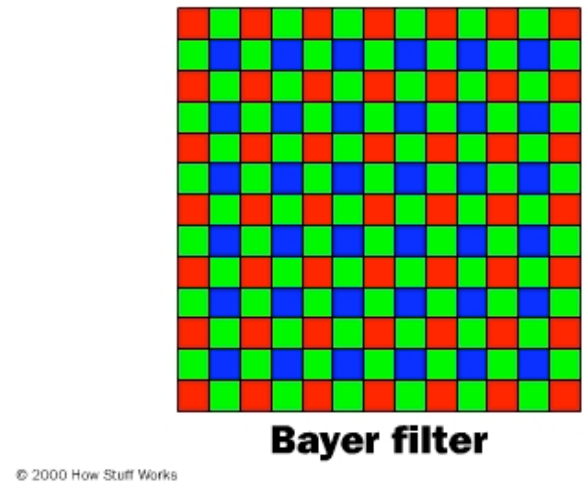
- Sensor types:
  - CCD (charge-coupled device)
    - most common
    - high sensitivity
    - high power
    - cannot be individually addressed
    - blooming
  - CMOS
    - simple to fabricate (cheap)
    - lower sensitivity, lower power
    - can be individually addressed

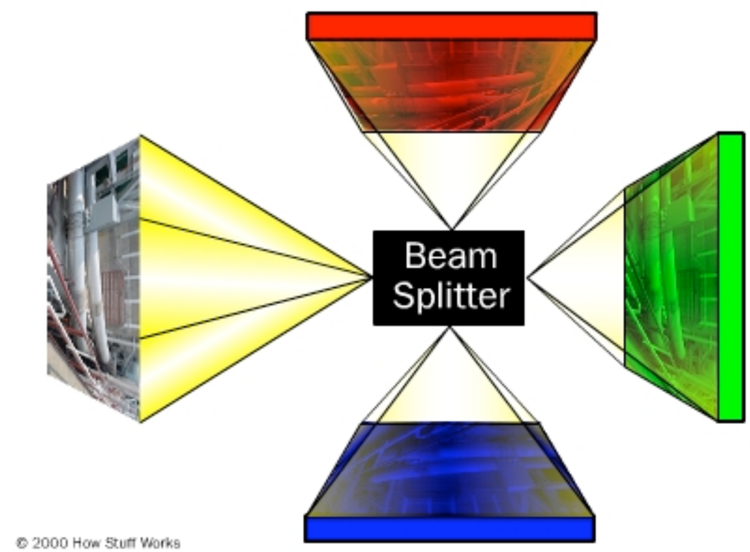




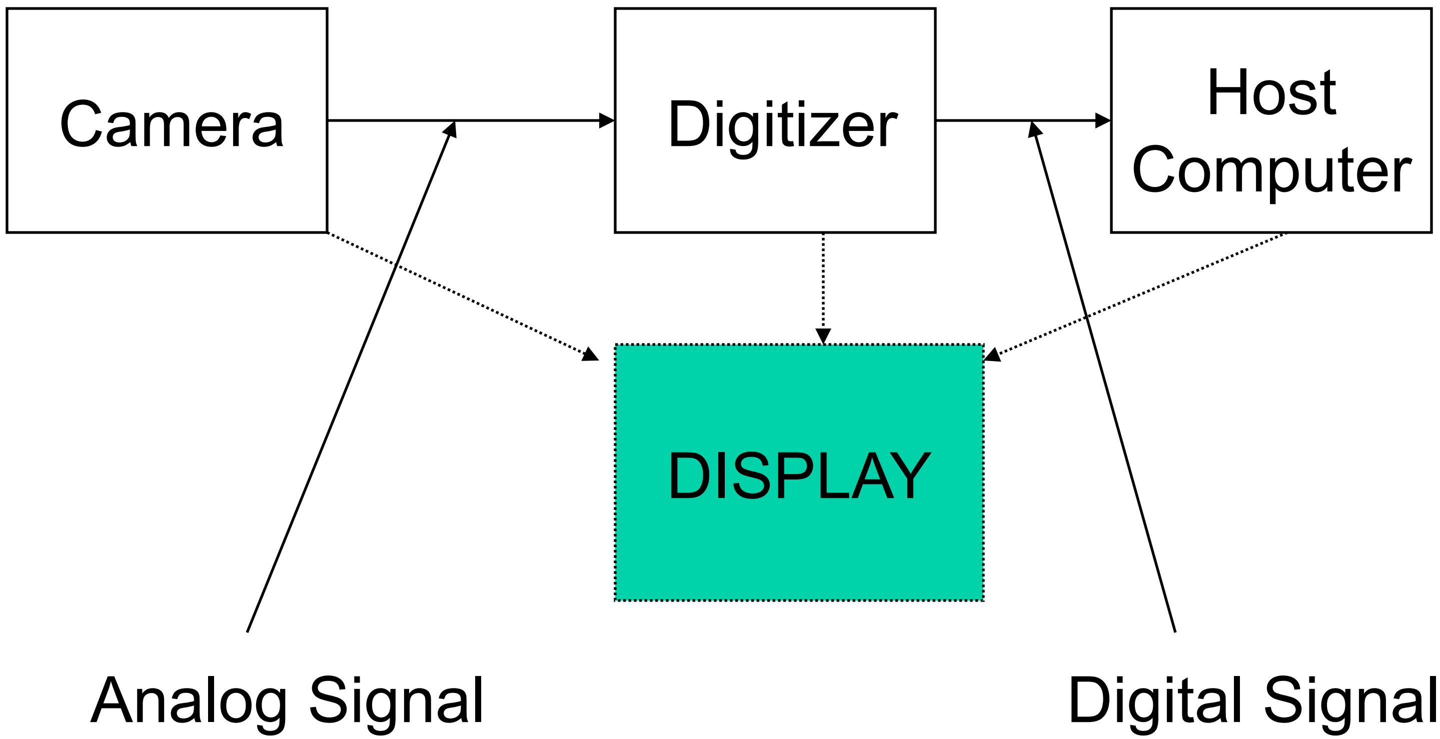
How Color Cameras Work

- 1 CCD cameras
  - A **Bayer** pattern is placed in front of the CCD
  - A **Demosaicing** process reads the pixels in a region and computes color and intensity
- 3 CCD camera use a beam splitter and 3 separate CCDs
  - higher color fidelity
  - needs lots of light
  - requires careful alignment of ccds

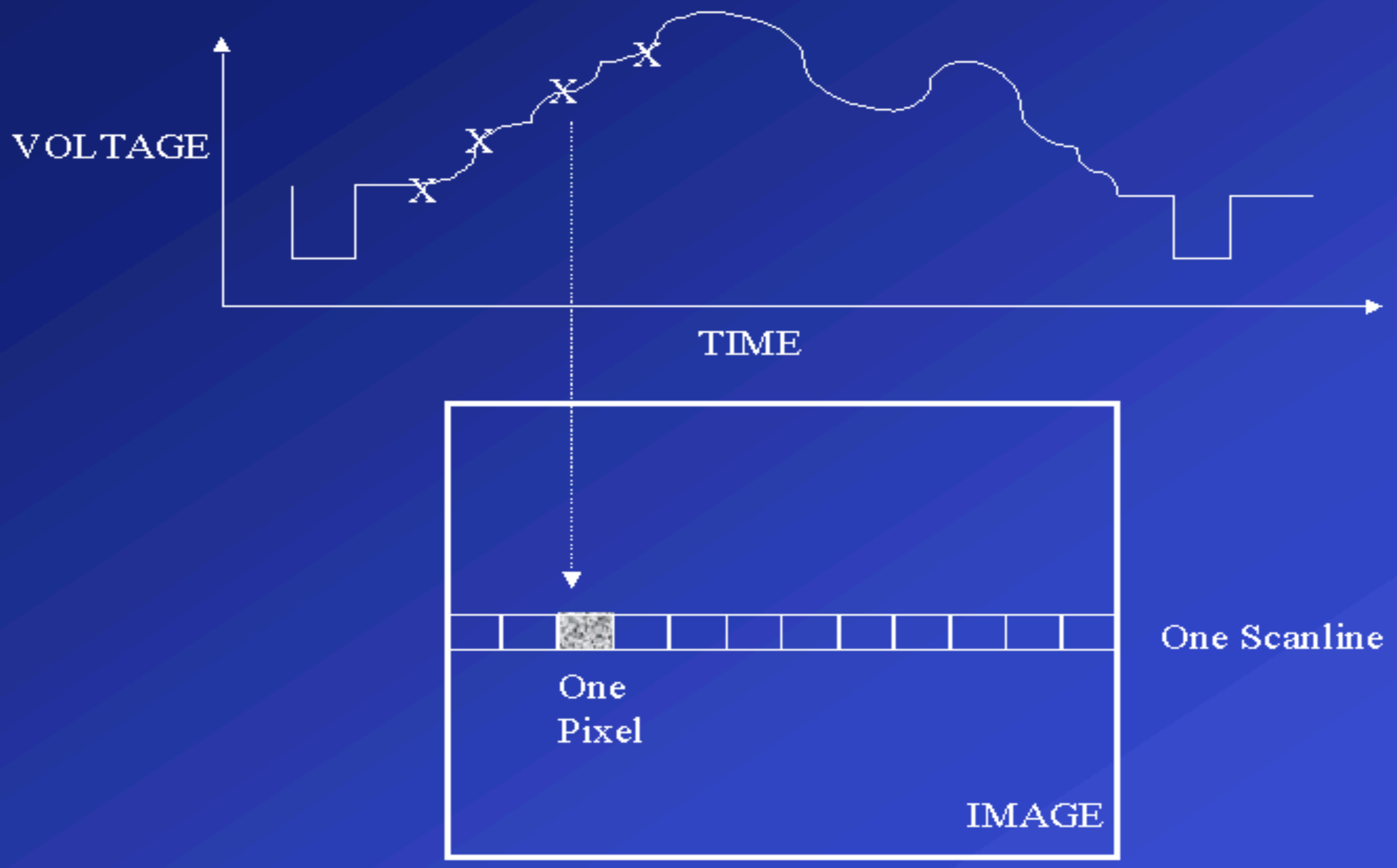




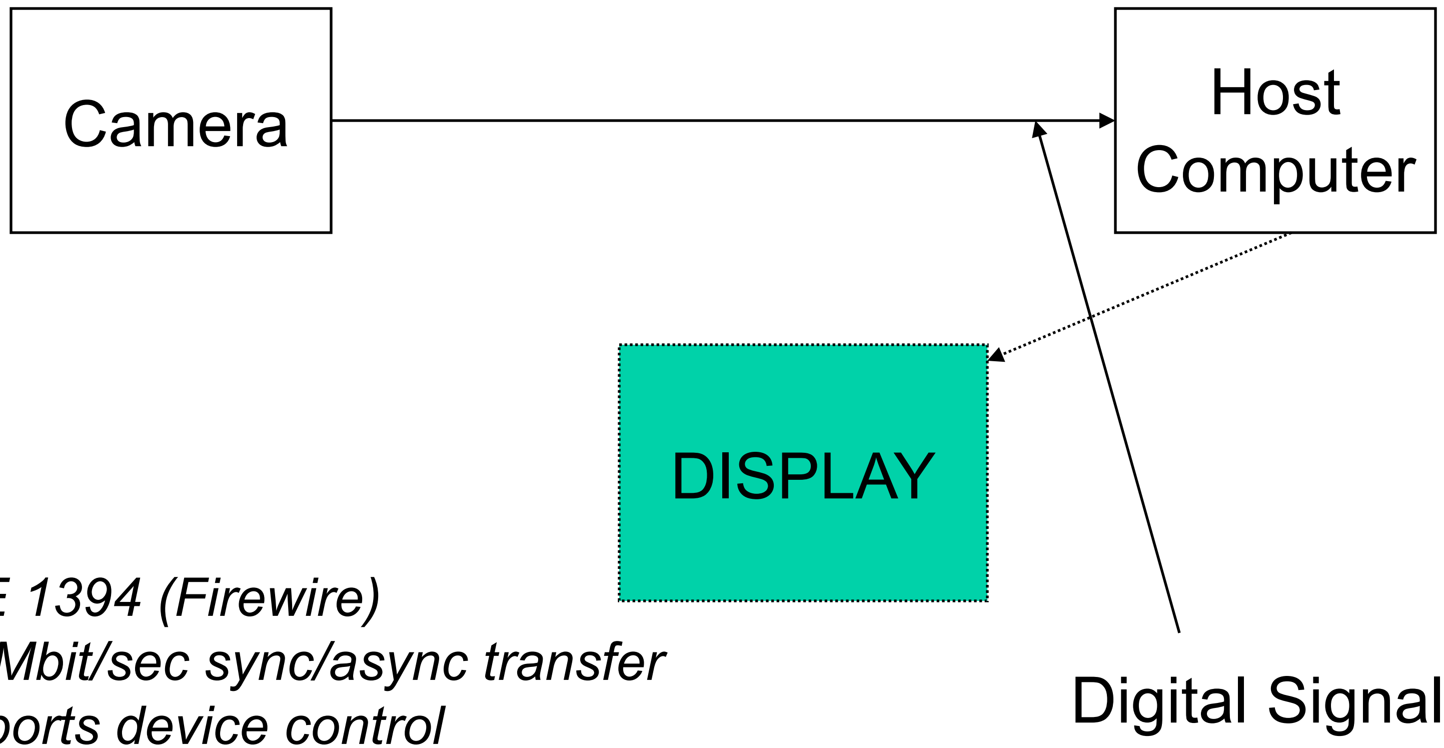
A “Traditional” Camera



What's under the Hood



A Modern Digital Camera



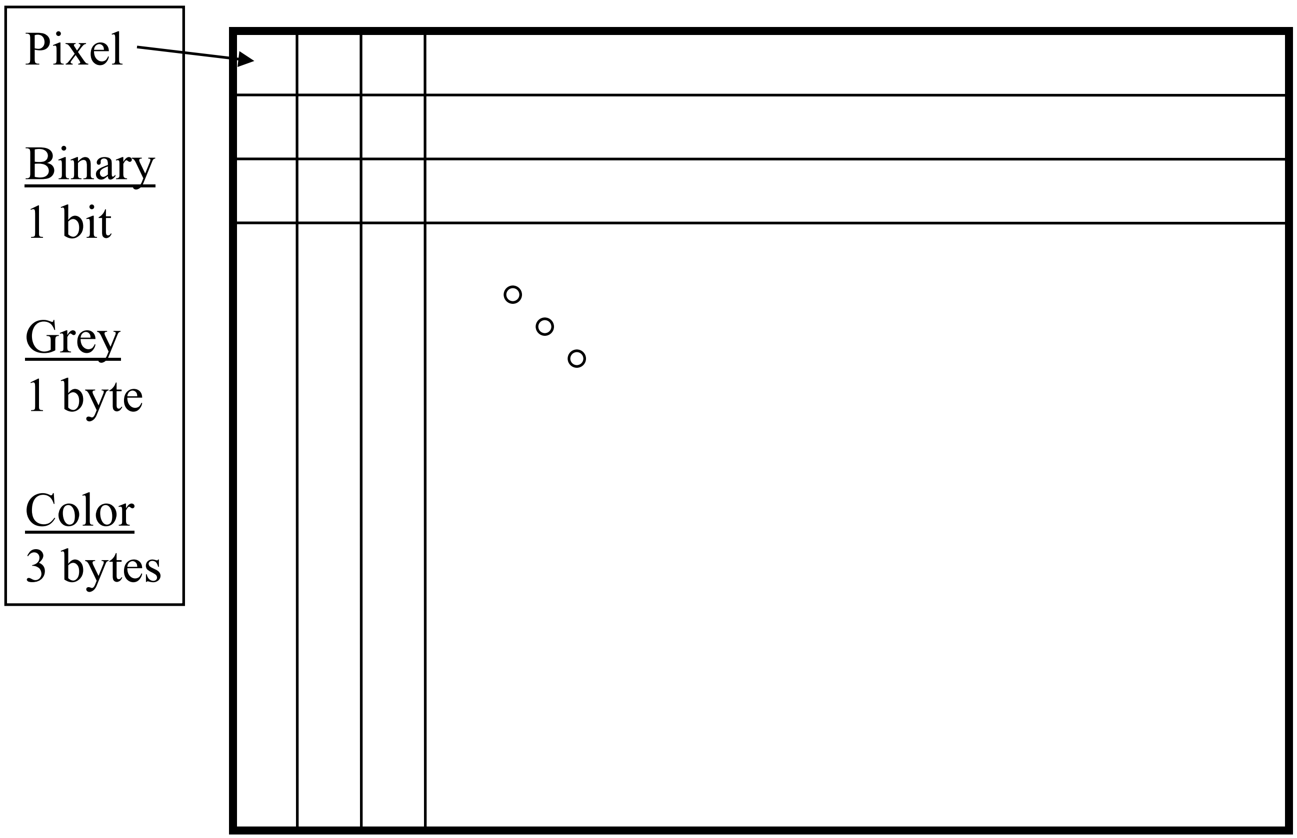
IEEE 1394 (Firewire)
400 Mbit/sec sync/async transfer
Supports device control

USB 2.0
480 Mbit/sec (~280Mbit/sec in practice)
Less flexible, but simpler to implement

Other Issues

- Automatic Gain Control (AGC): adjusting amplification and black level to get a “good fit” of the incident light power to the range of the image
- Shuttering: Electronic “switch” that controls how long the CCD is “exposed.”
- White balance: Adjustment of the mapping from measured spectral quantities to image RGB quantities (we’ll talk about this more when we get to color).

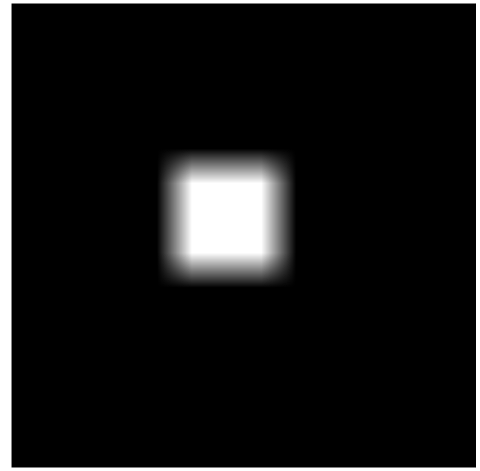
THE ORGANIZATION OF A 2D IMAGE

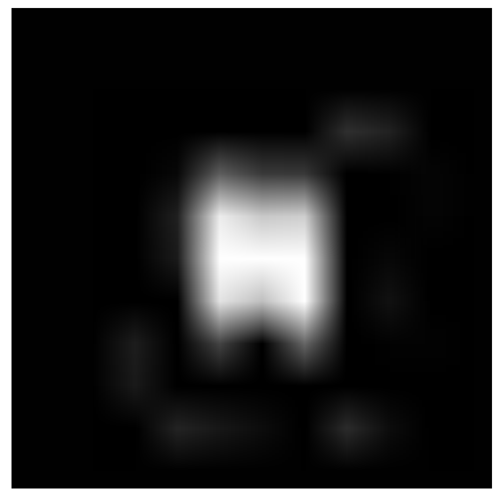


Storing Images

- Non-lossy schemes
  - pbm/pgm/ppm/pnm
    - code for file type, size, number of bands, and maximum brightness
  - tif (lossless and lossy versions)
  - bmp
  - gif (grayscale)

- Lossy schemes
  - gif (color)
  - jpg
    - uses Y Cb Cr color representation; subsamples the color
    - Uses DCT on result
    - Uses the fact the human system is less sensitive to color than spatial detail





GIF IMAGE FORMAT

- GIF (Graphics Interchange Format)
  - Limited to 8 bits/pixel for both color and gray-scale.

<u>8-bit index</u>	RED	GREEN	BLUE
0	R0	G0	B0
1	R1	G1	B1
2	R2	G2	B2
	⋮	⋮	⋮
254	R254	G254	B254
255	R255	G255	B255

TIFF IMAGE FORMAT

- TIFF (Tagged Image File Format)
  - More general than GIF
  - Allows 24 bits/pixel
  - Supports 5 types of image compression including:
    - RLE (Run length encoding)
    - LZW (Lempel-Ziv-Welch)
    - JPEG (Joint Photographic Experts Group)