

# Introduction

---

Selim Aksoy

Department of Computer Engineering

Bilkent University

[saksoy@cs.bilkent.edu.tr](mailto:saksoy@cs.bilkent.edu.tr)



# What is computer vision?

---

- Analysis of digital images by a computer.
- Stockman and Shapiro: *making useful decisions about real physical objects and scenes based on sensed images.*
- Trucco and Verri: *computing properties of the 3D world from one or more digital images.*
- Ballard and Brown: *construction of explicit, meaningful description of physical objects from images.*
- Forsyth and Ponce: *extracting descriptions of the world from pictures or sequences of pictures.*



# Why study computer vision?

---

- Possibility of building intelligent machines is fascinating.
- Capability of understanding the visual world is a prerequisite for such machines.
- Much of the human brain is dedicated to vision.
- Humans solve many visual problems effortlessly, yet we have little understanding of visual cognition.



# Why study computer vision?

---

- An image is worth 1000 words.
- Images and videos are everywhere.
- Fast growing collections and many useful applications.
- Goals of vision research:
  - Give machines the ability to understand scenes.
  - Aid understanding and modeling of human vision.
  - Automate visual operations.



# Challenge



- What do you see in the picture?
  - A hand holding a man
  - A hand holding a shiny sphere
  - An Escher drawing

Adapted from Octavia Camps, Penn State



# Perception and grouping

---

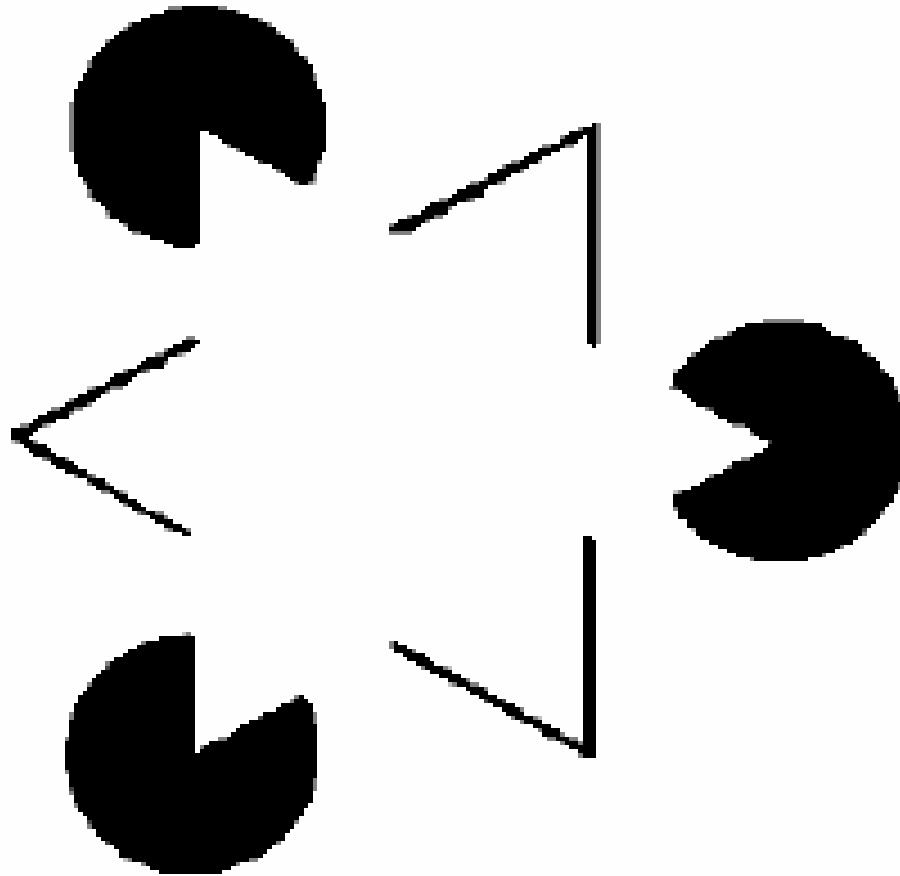


Subjective  
contours



# Perception and grouping

---



Subjective  
contours

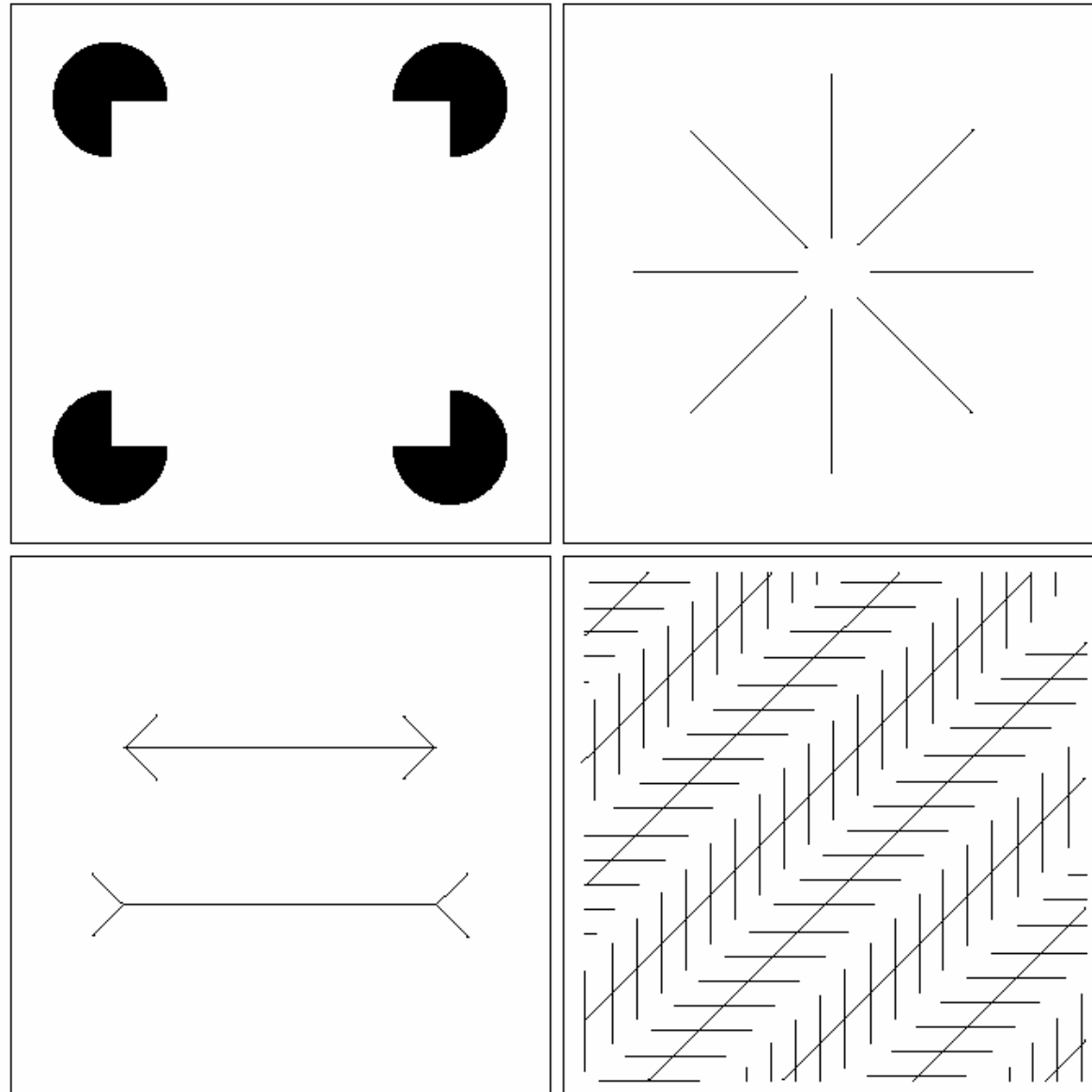
Adapted from Michael Black, Brown University



# Perception and grouping

a b  
c d

**FIGURE 2.9** Some well-known optical illusions.

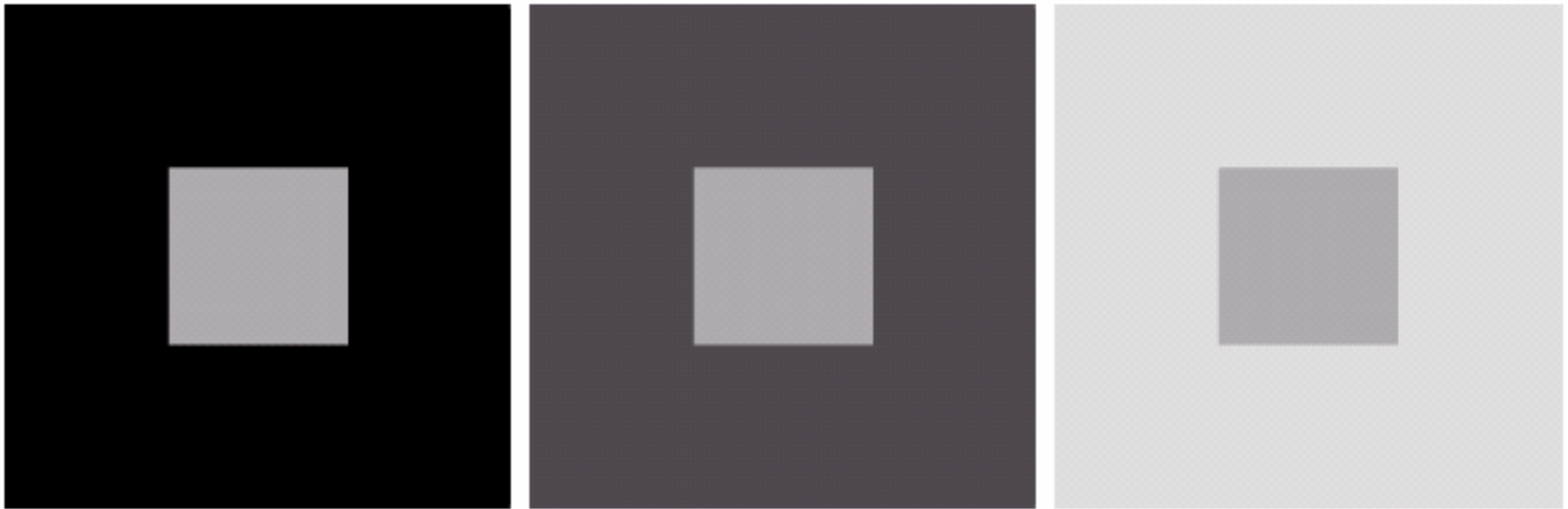


Adapted from Gonzales and Woods



# Perception and grouping

---



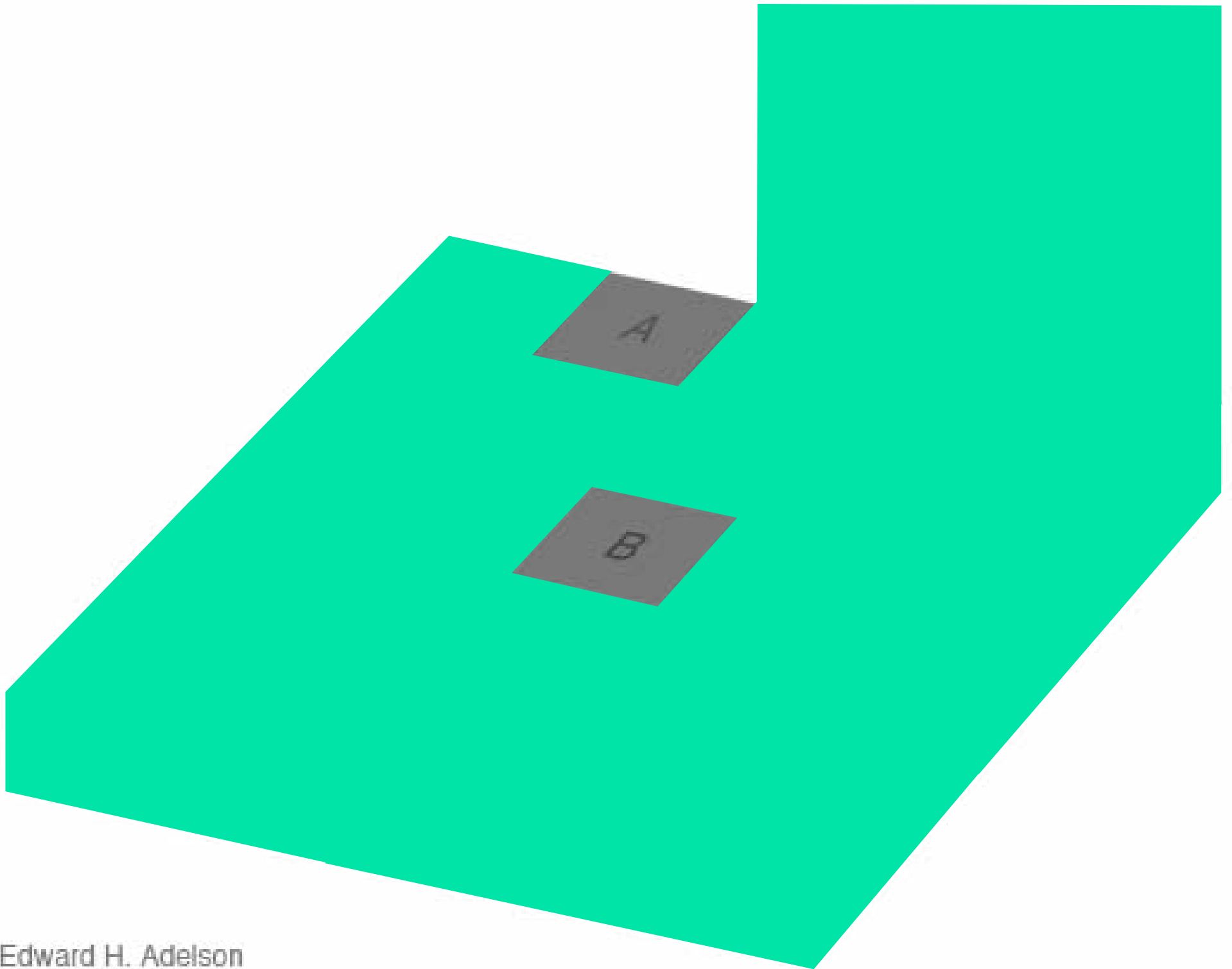
a b c

**FIGURE 2.8** Examples of simultaneous contrast. All the inner squares have the same intensity, but they appear progressively darker as the background becomes lighter.

---

Adapted from Gonzales and Woods







# Perception and grouping

---



Occlusion

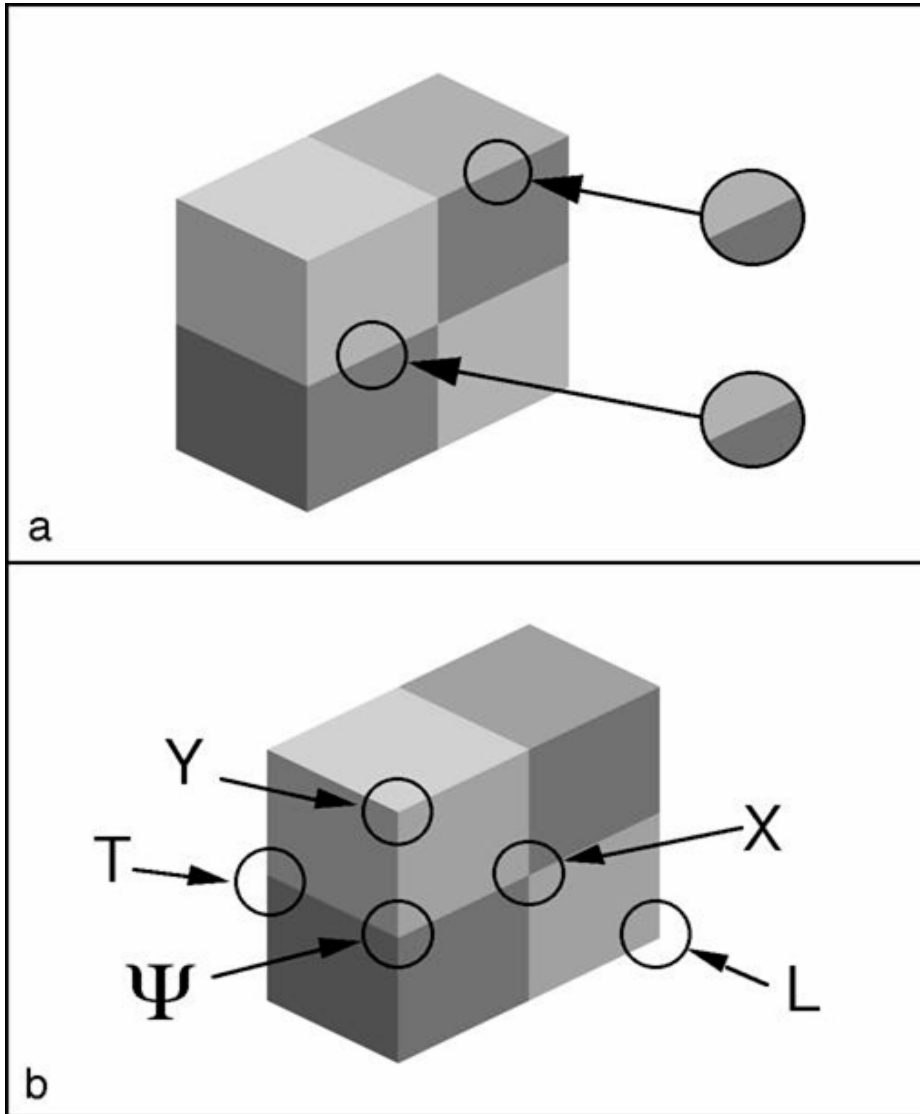
\* types of  
“junctions” give  
cues about  
surfaces,  
occlusion, and  
light.

I. Rock, The Logic of Perception, 1983.

Adapted from Michael Black, Brown University



# Perception and grouping



- The shape of junctions constrains the possible interpretations of the scene.
- Ambiguous: paint and surface boundaries can be confused.

Adapted from Michael Black, Brown University



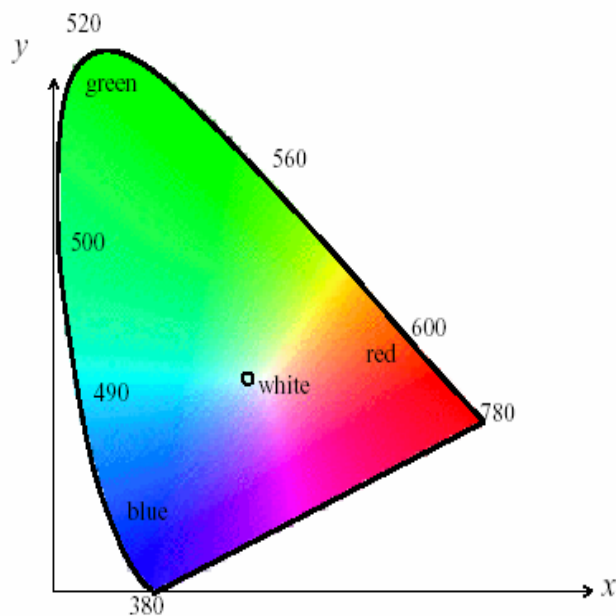
# Recognition

---

- How can different cues such as color, texture, shape, motion, etc., can be used for recognition?
  - Which parts of image should be recognized together?
  - How can objects be recognized without focusing on detail?
  - How can objects with many free parameters be recognized?
  - How do we structure very large model bases?



# Color



Adapted from Martial Hebert, CMU



# Texture

---





# Segmentation

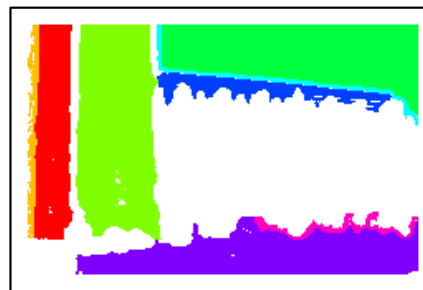
Original Images



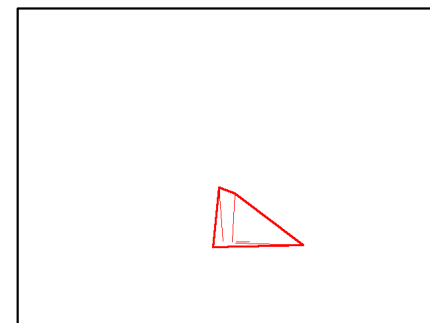
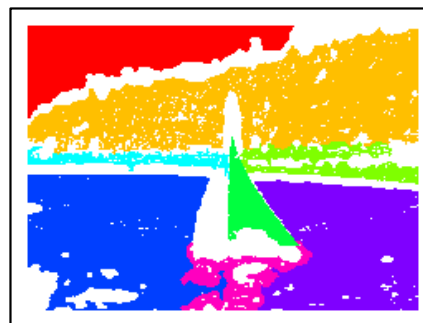
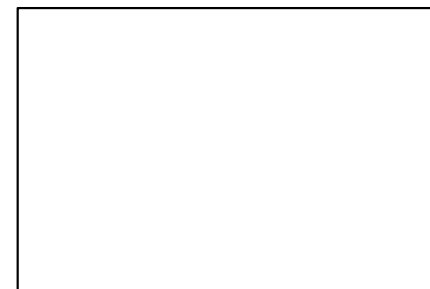
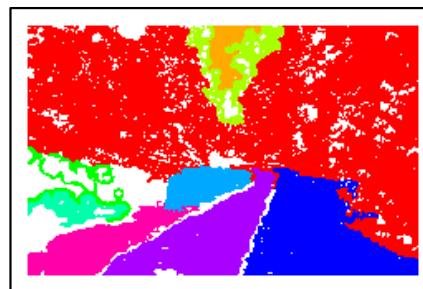
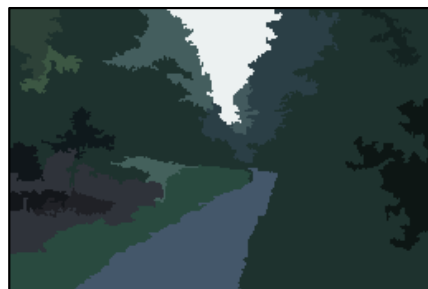
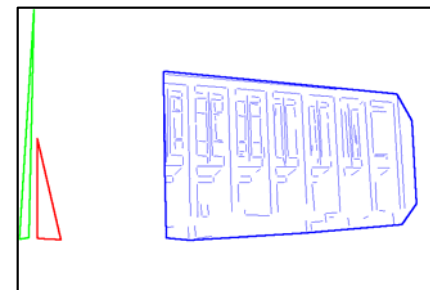
Color Regions



Texture Regions



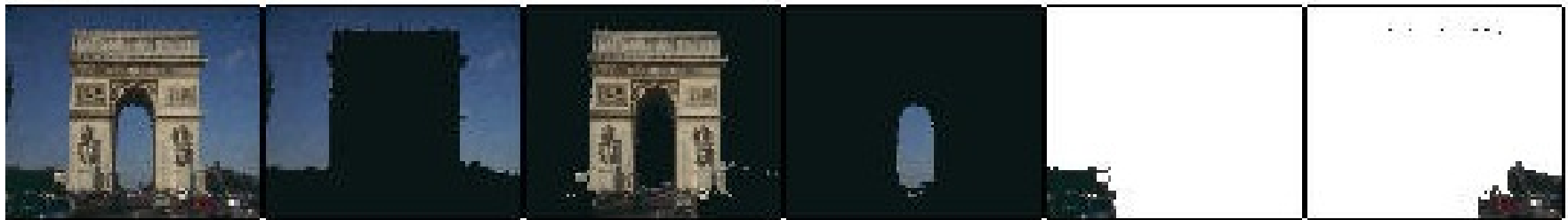
Line Clusters



Adapted from Linda Shapiro, U of Washington

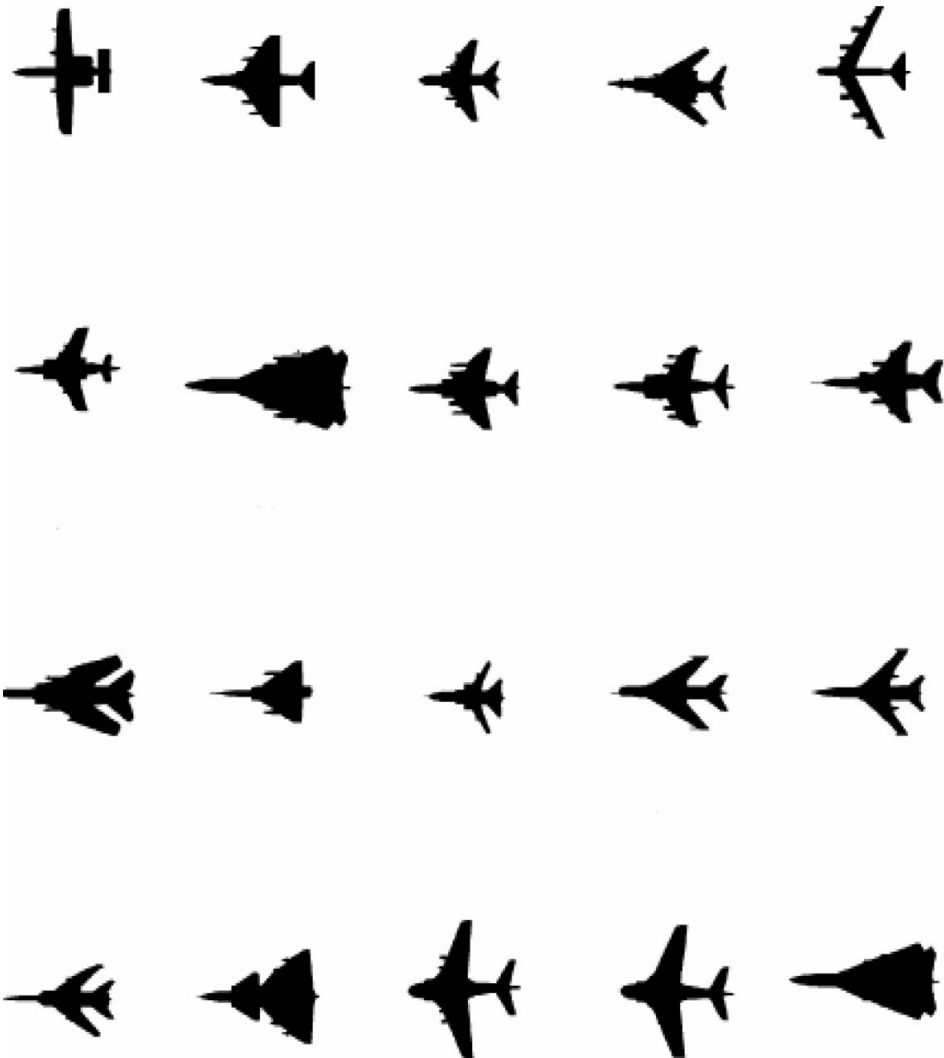


# Segmentation





# Shape



Model database



Recognized objects

Adapted from Enis Cetin, Bilkent University



# Motion

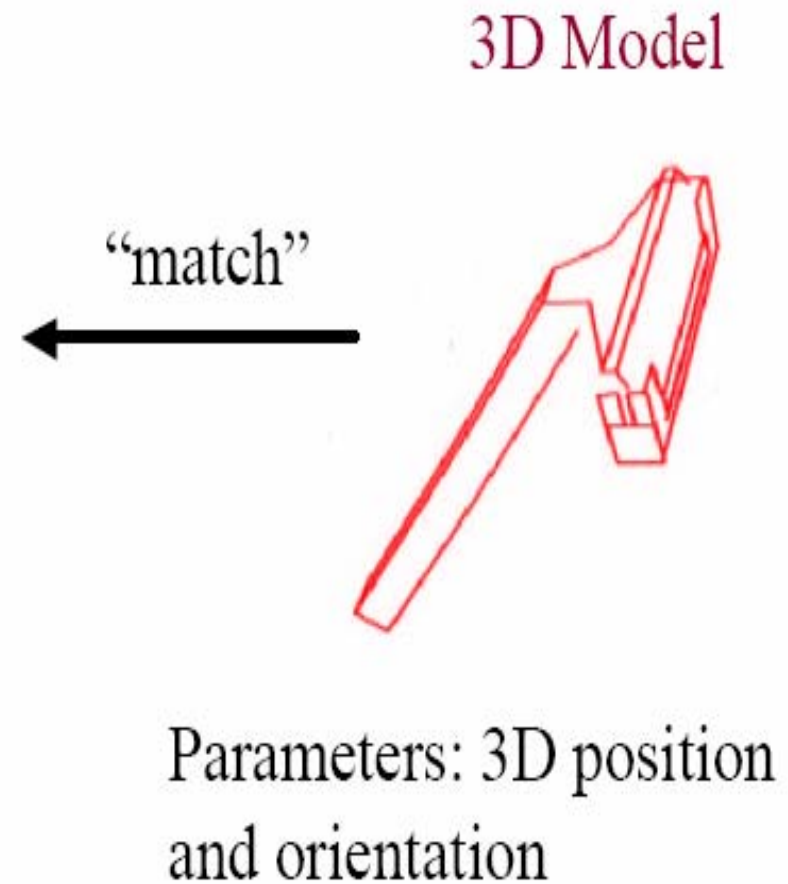
---



Adapted from Michael Black, Brown University



# Recognition

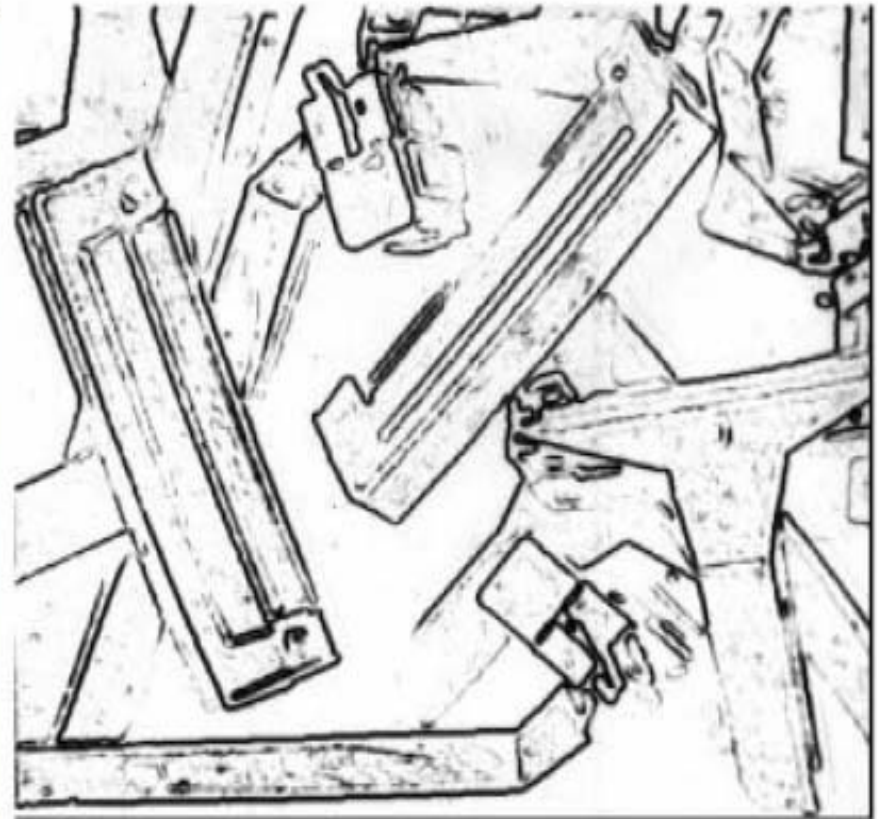


Adapted from Michael Black, Brown University



# Recognition

---



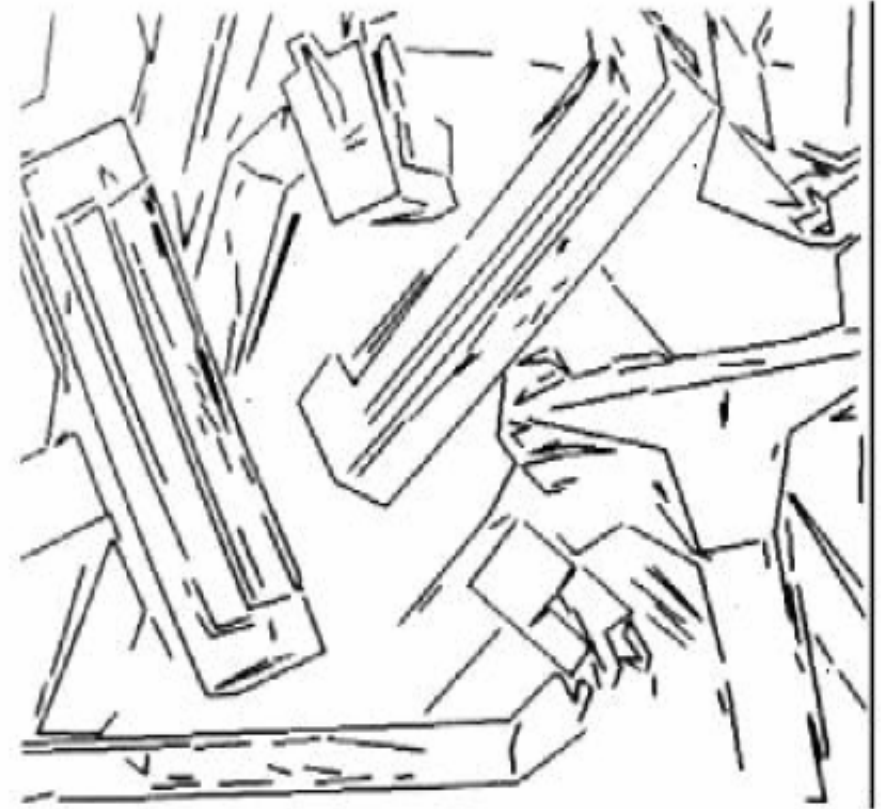
“Filter” image to find brightness changes.

Adapted from Michael Black, Brown University



# Recognition

---



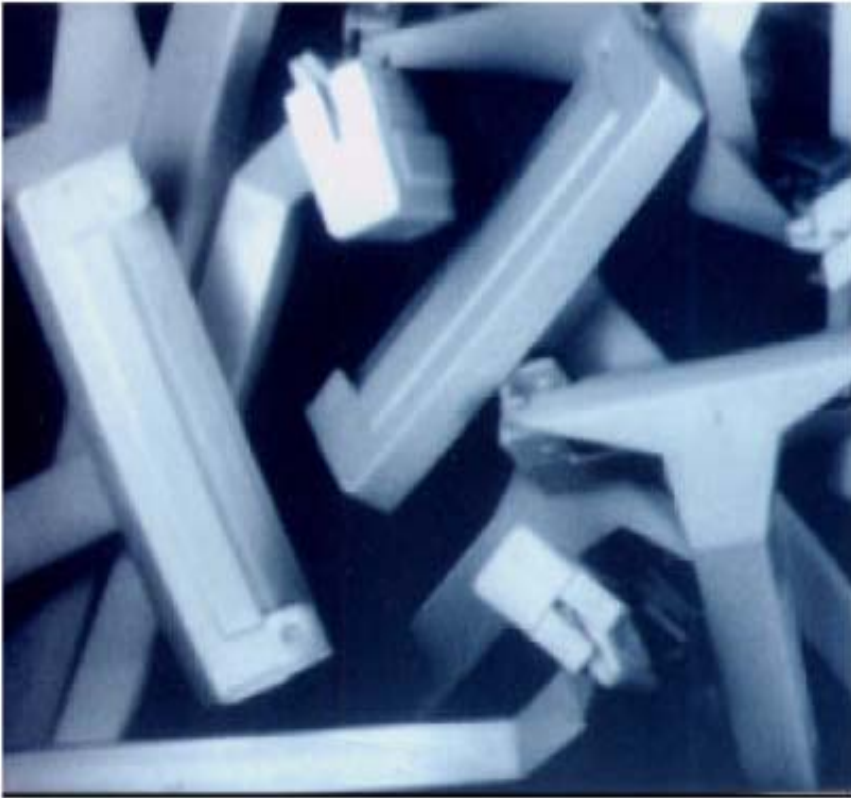
“Fit” lines to the raw measurements.

Adapted from Michael Black, Brown University



# Recognition

---



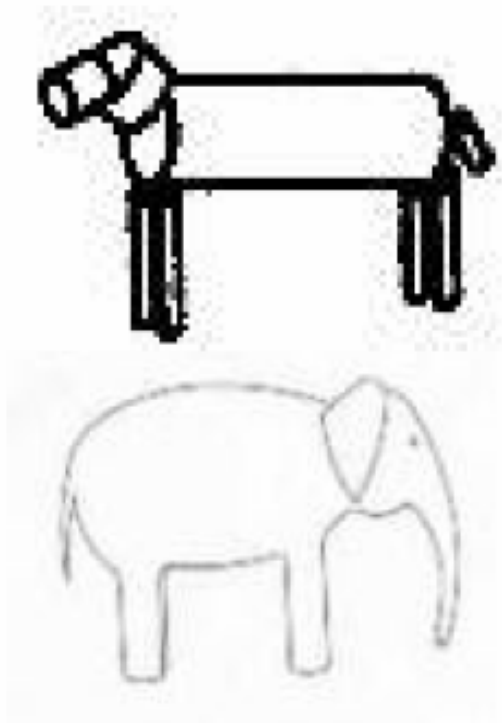
“Project” model into image and “match” to lines  
(solving for 3D pose).

Adapted from Michael Black, Brown University

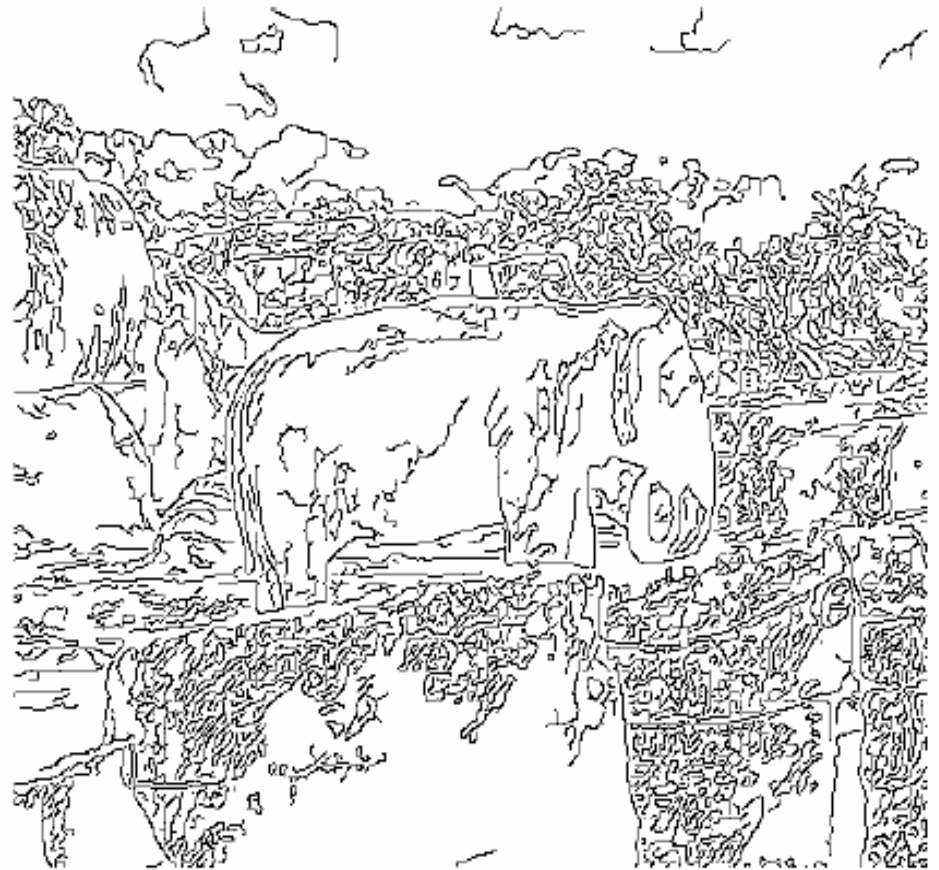


# Recognition

---



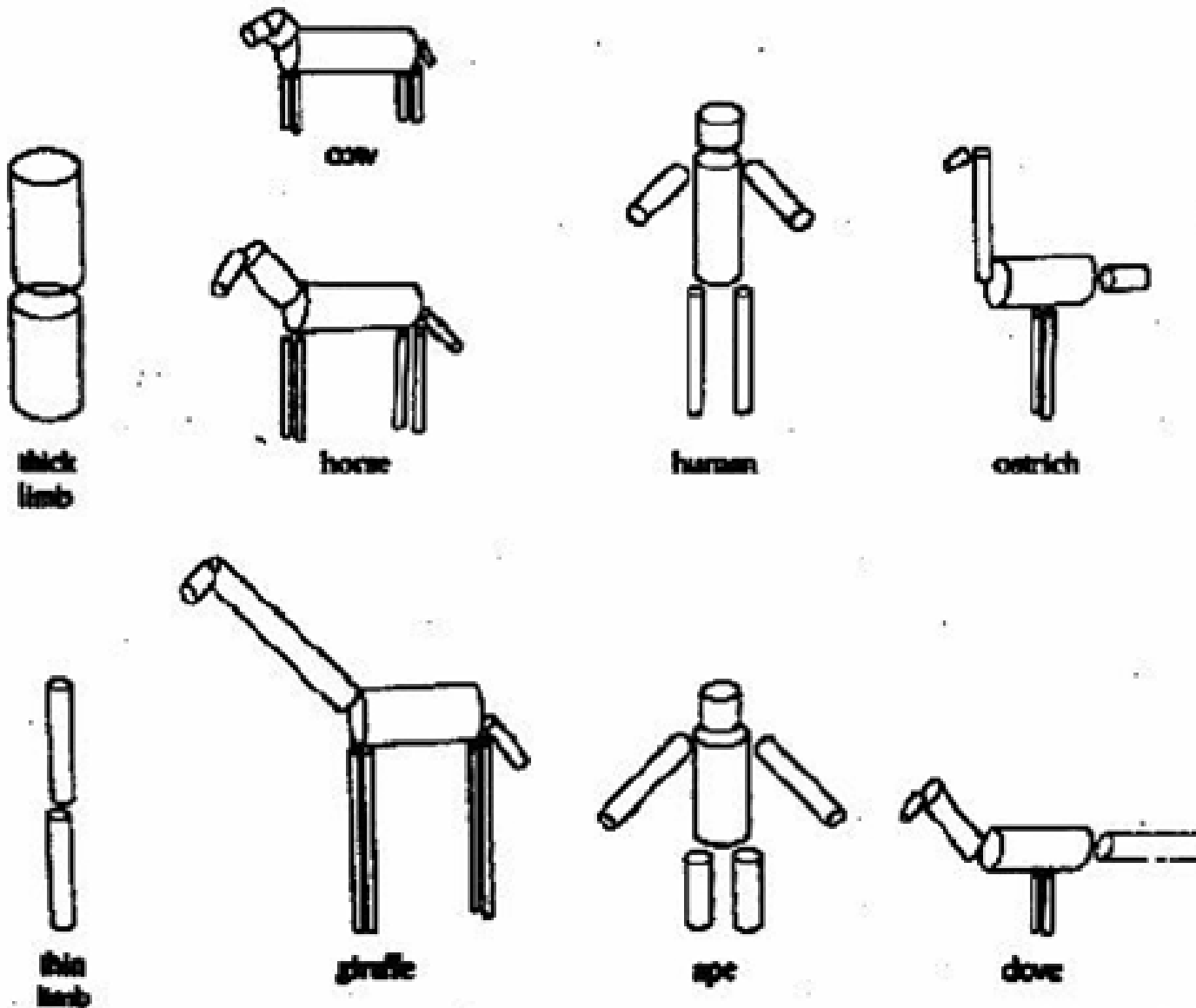
Match “model” to  
measurements?



Adapted from Michael Black, Brown University



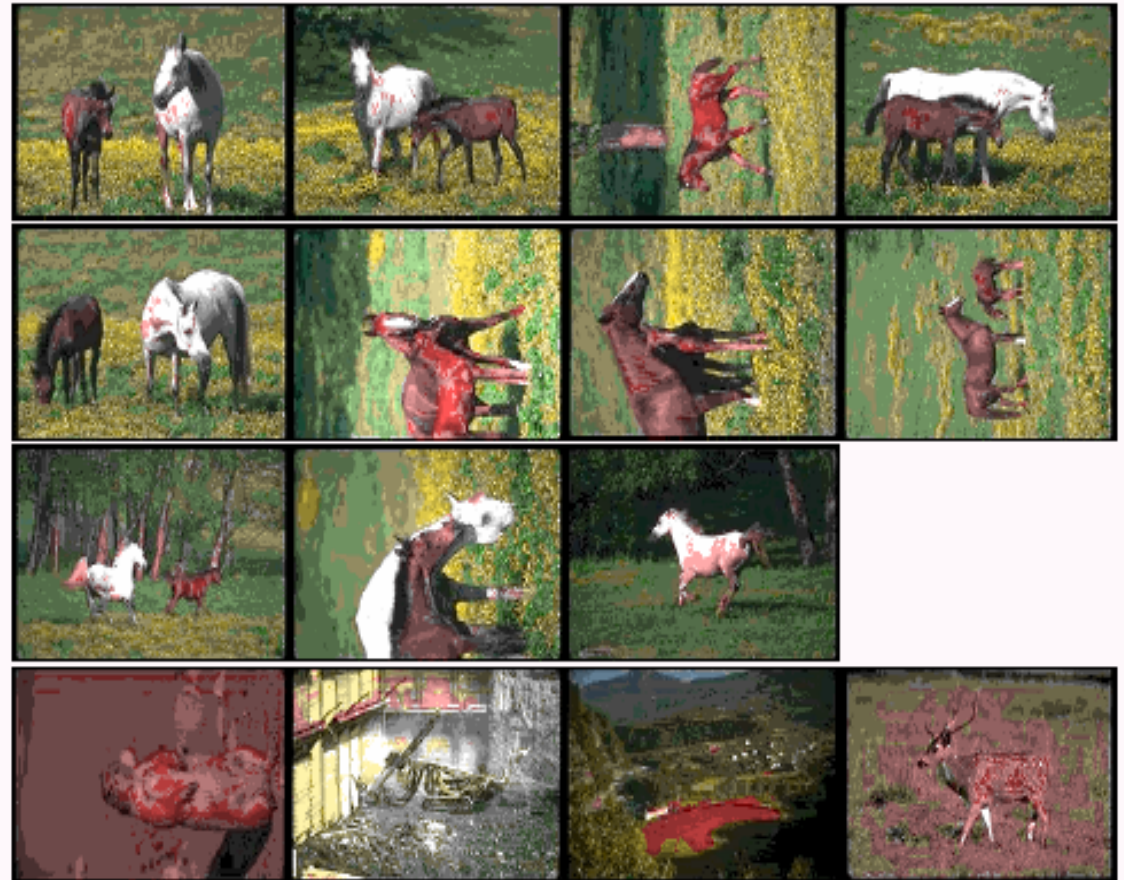
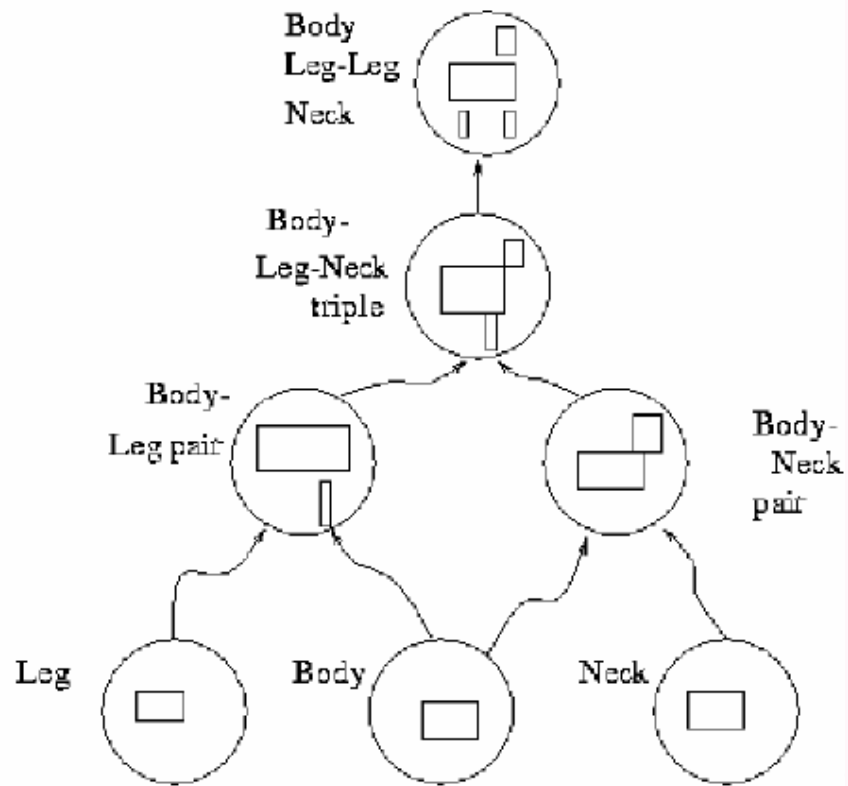
# Recognition



Marr & Nishihara



# Recognition

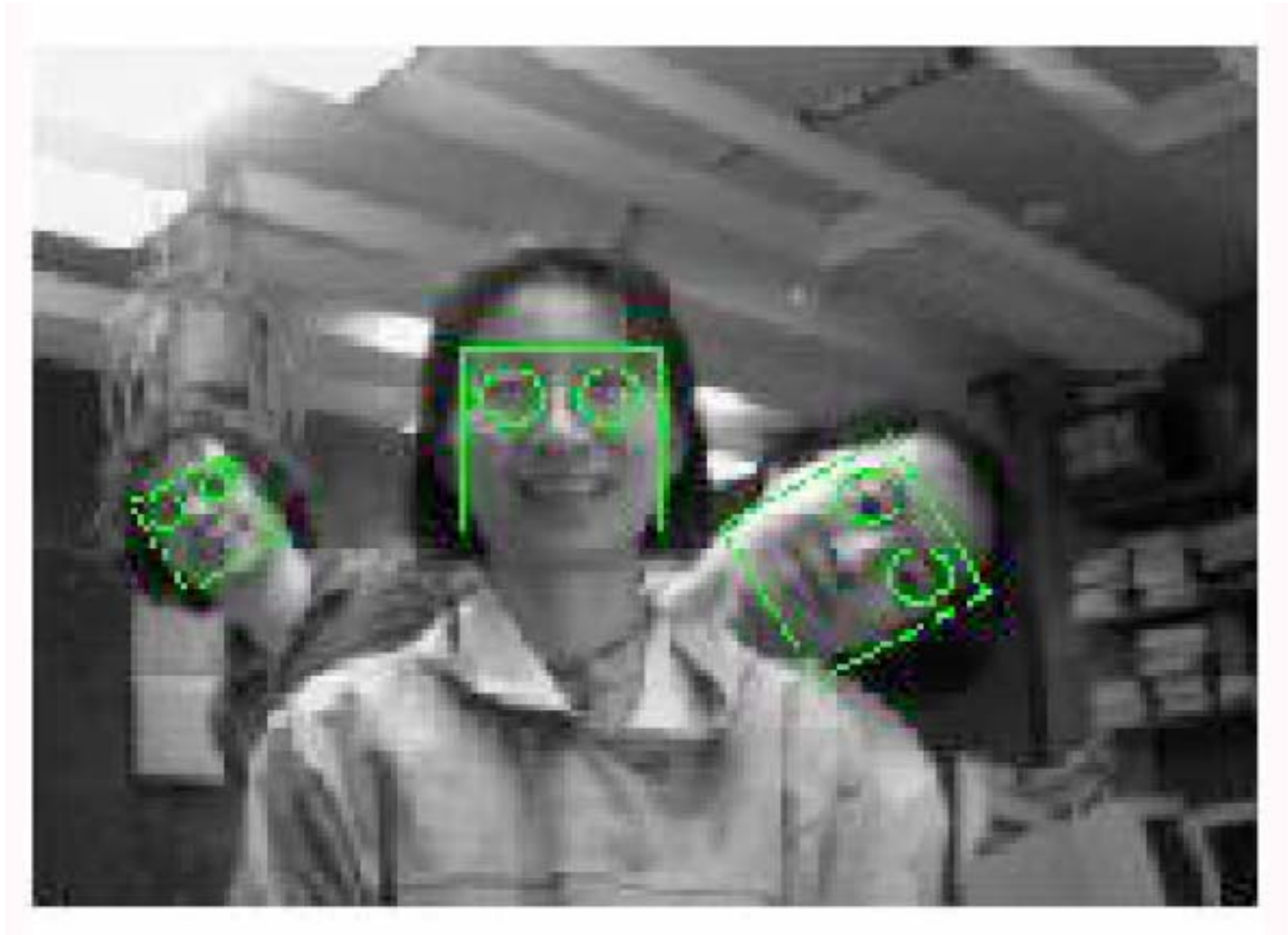


Adapted from David Forsyth, UC Berkeley



# Detection

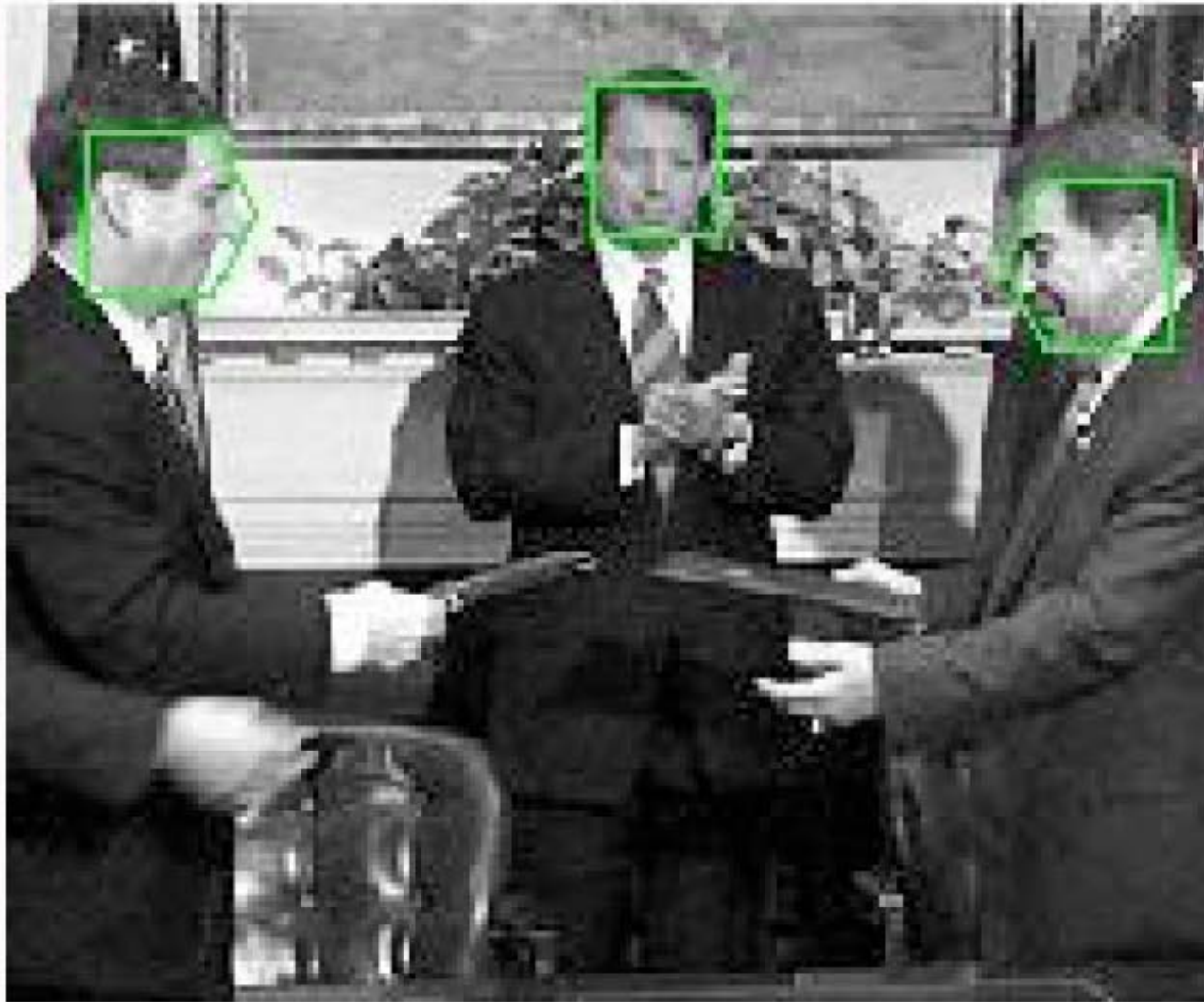
---



Adapted from David Forsyth, UC Berkeley



# Detection



Adapted from David Forsyth, UC Berkeley

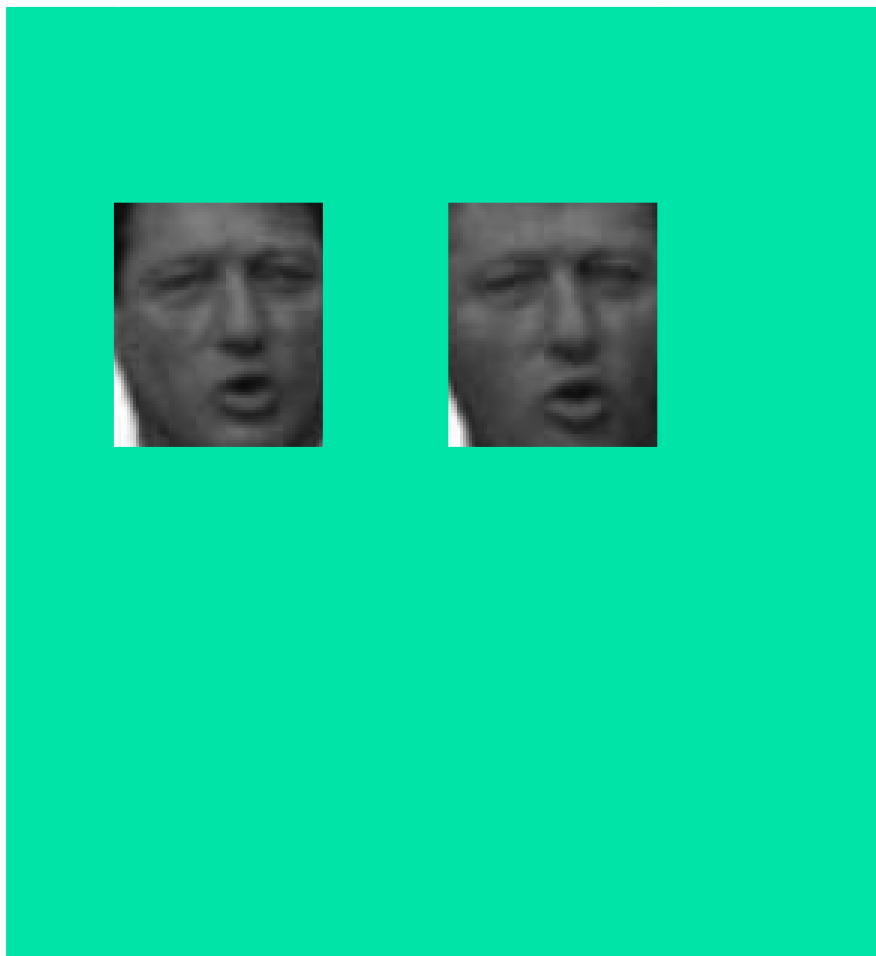


# Detection

---

What are our  
“models”?

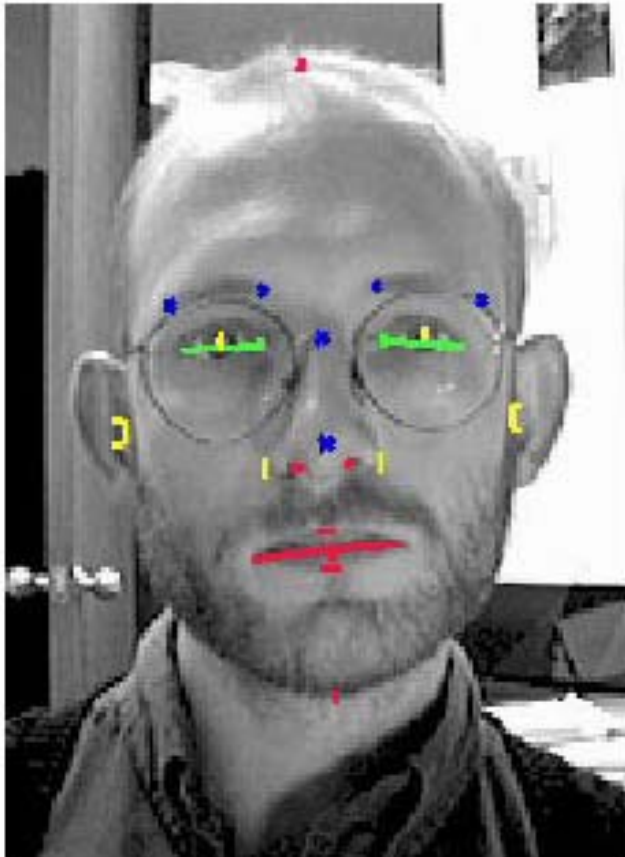
How good  
are they?



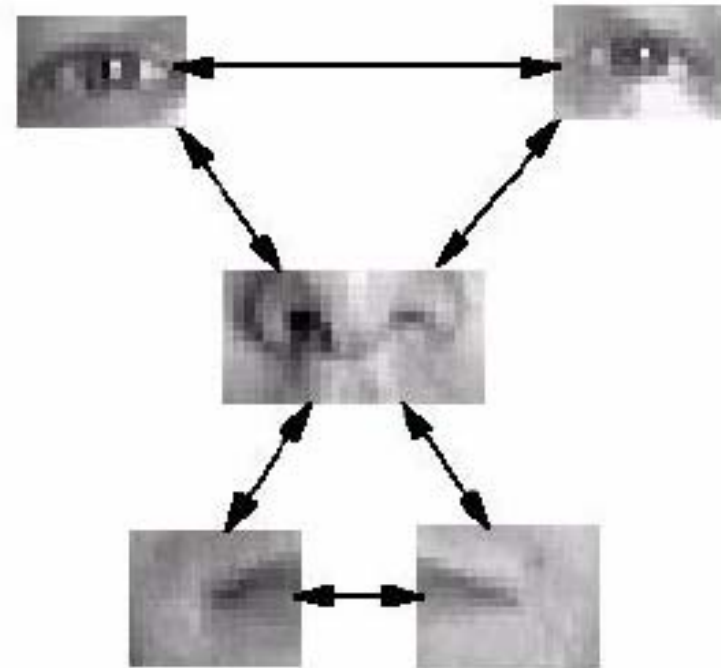
Adapted from Michael Black, Brown University



# Parts and relations



## Patch Model



<http://www.research.ibm.com/ecvg/biom/facereco.html>

Adapted from Michael Black, Brown University



# Parts and relations

---



How flexible are the spatial relations of the parts?

Adapted from Michael Black, Brown University



# Context

---



Adapted from Antonio Torralba, MIT



# Context

---



Adapted from Antonio Torralba, MIT



# Context

---



Adapted from Derek Hoiem, CMU



# Context



Adapted from  
Derek Hoiem, CMU



# Stages of computer vision

---

- Low-level

image  $\rightarrow$  image

- Mid-level

 image  $\rightarrow$  features / attributes

Image analysis / image understanding

- High-level

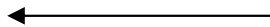
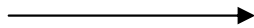
features  $\rightarrow$  “making sense”, recognition



# Low-level

---

sharpening



blurring

Adapted from Linda Shapiro, U of Washington



# Low-level



original image

Canny  
→



edge image

# Mid-level



edge image

ORT  
→

data  
structure



circular arcs and line segments

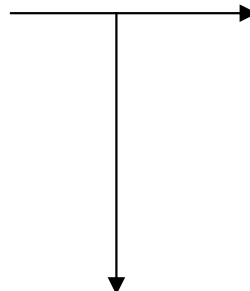


# Mid-level

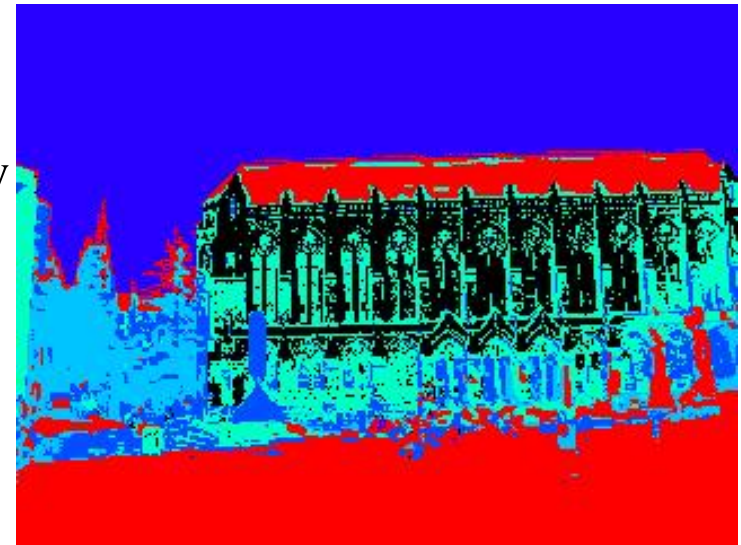


original color image

K-means  
clustering  
(followed by  
connected  
component  
analysis)



data  
structure

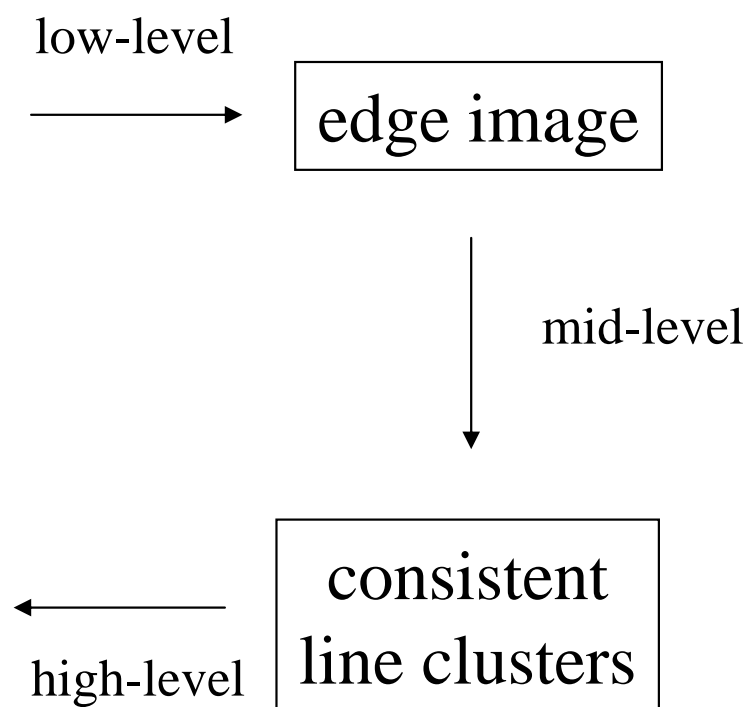
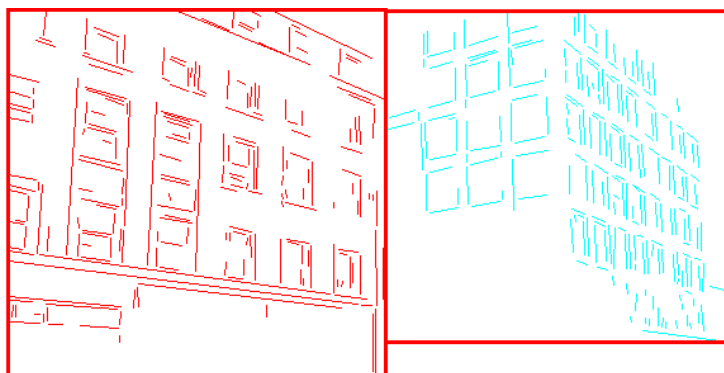


regions of homogeneous color

Adapted from Linda Shapiro, U of Washington



# Low-level to high-level



Adapted from Linda Shapiro, U of Washington



# Applications

---

- Industrial inspection, quality control
- Medical image analysis
- Remote sensing
- Security
  - Surveillance
  - Biometrics
  - Target recognition
  - Tracking
- Robotics
- Document analysis
- Multimedia
- Assisted living
- Human-computer interfaces



# Medical image analysis

---

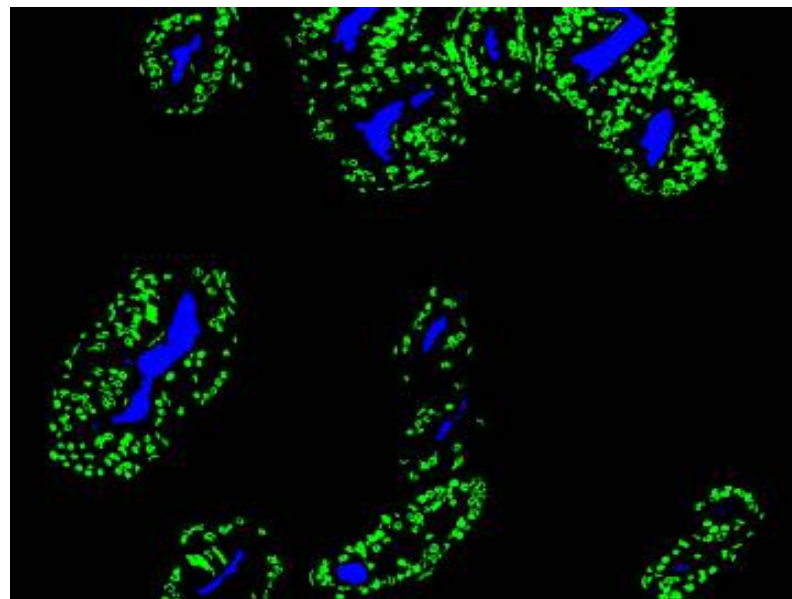
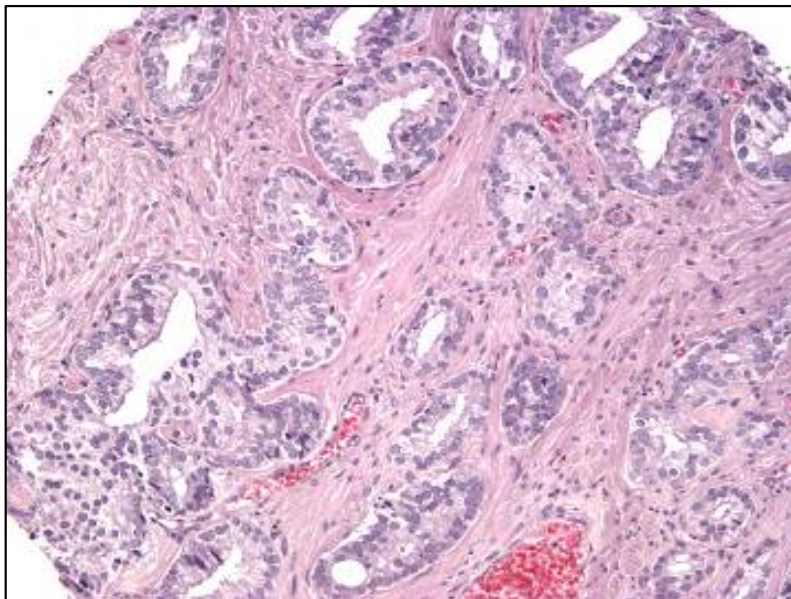
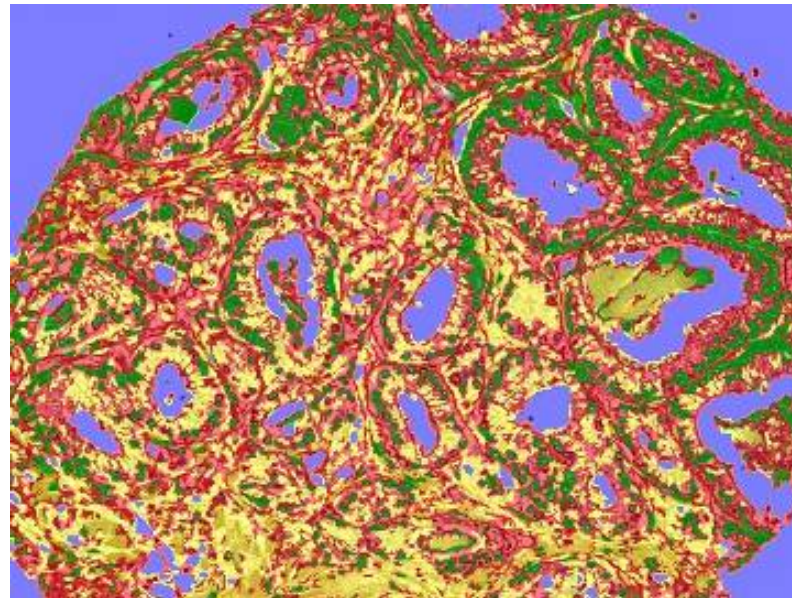
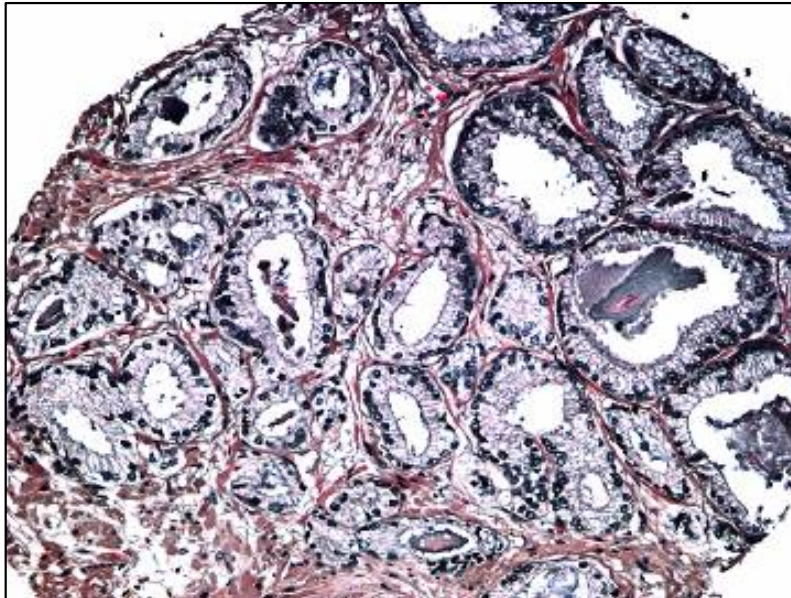


CT image of  
abdomen

Adapted from Linda Shapiro, U of Washington



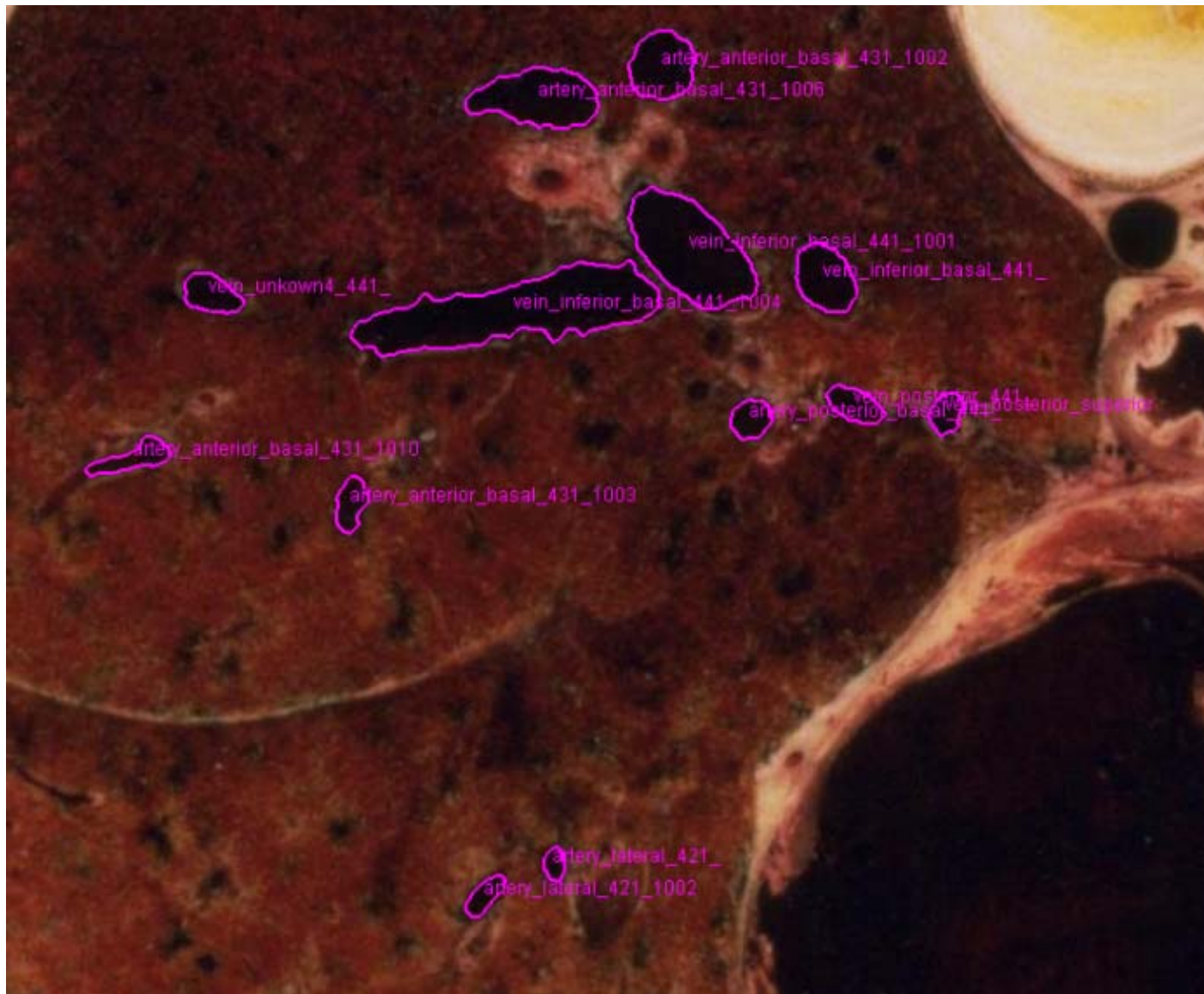
# Medical image analysis



Cancer  
detection  
and  
grading



# Medical image analysis



Slice of  
lung



# Land cover analysis





# Object recognition

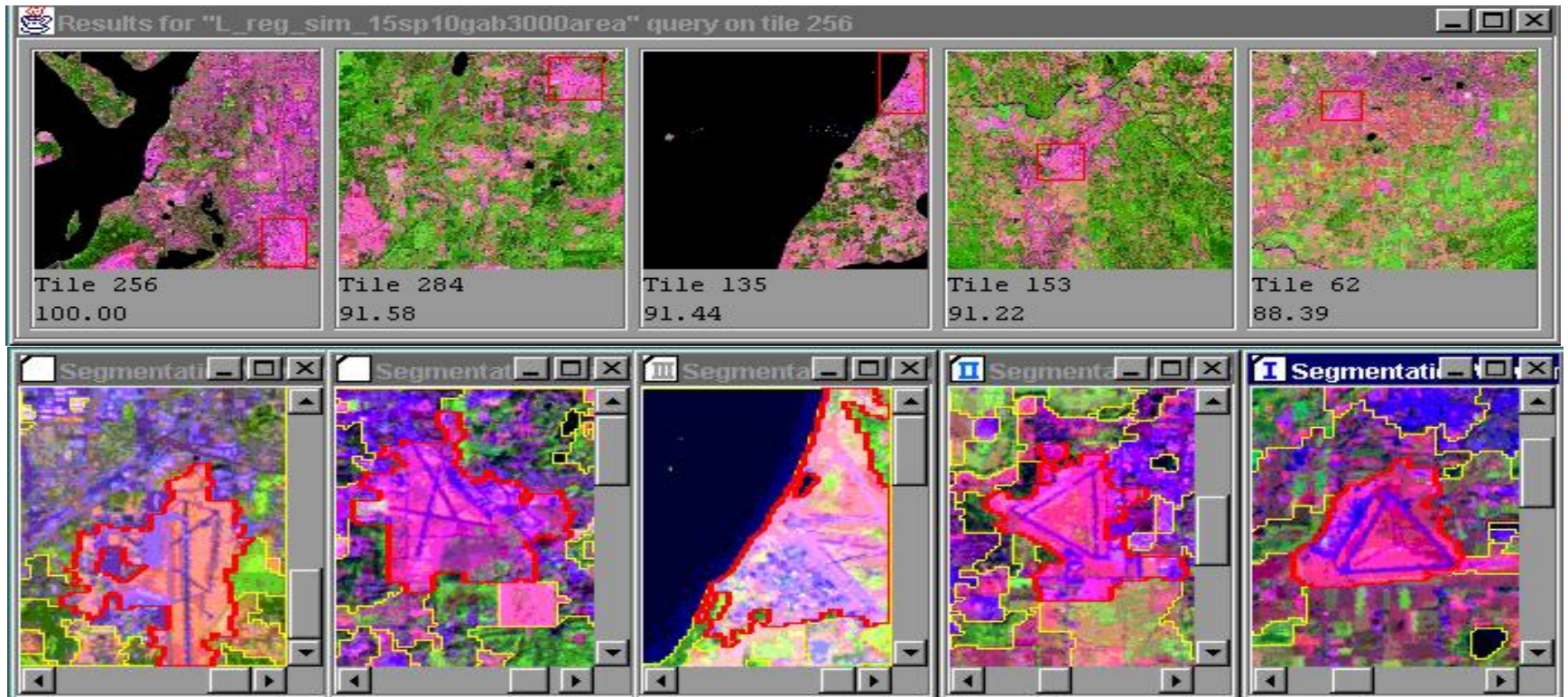
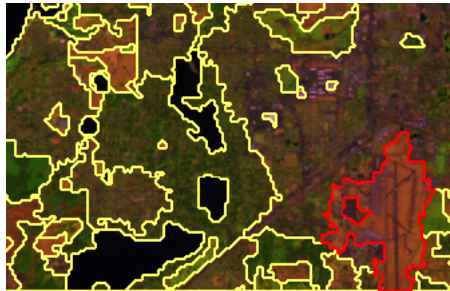
---



Recognition of buildings and building groups



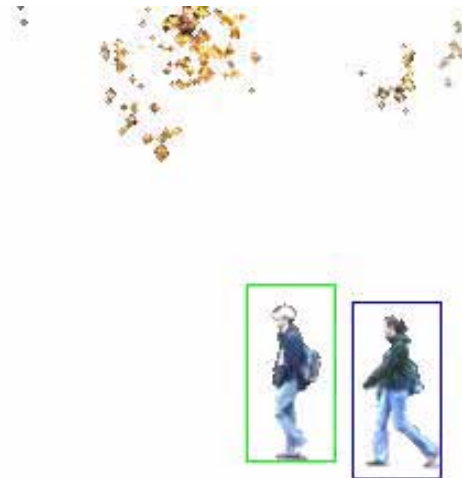
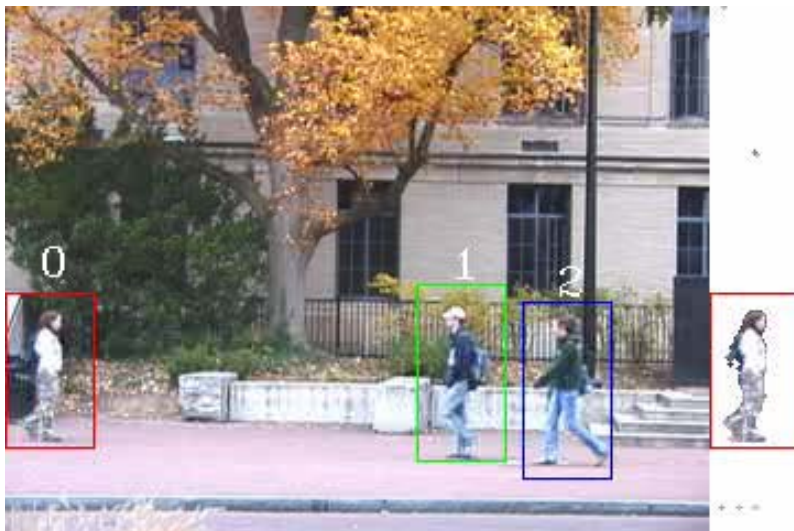
# Content-based retrieval



Finding similar regions: airports



# Tracking



Adapted from Octavia Camps, Penn State



# Tracking

---



Adapted from Martial Hebert, CMU



# Tracking

---



Adapted from Martial Hebert, CMU



# Biometrics



(a)



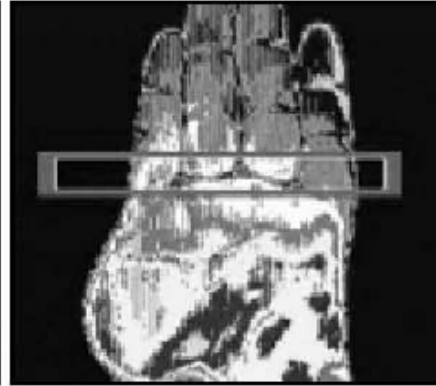
(b)



(c)



(d)



(e)



(f)



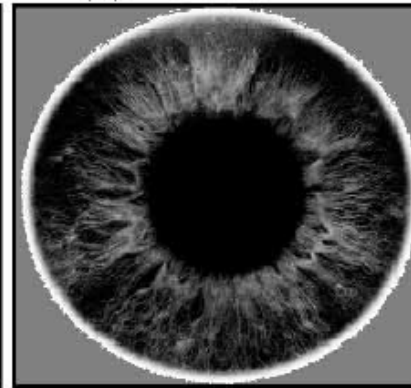
(g)



(h)



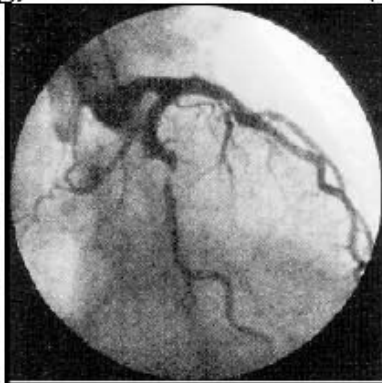
(i)



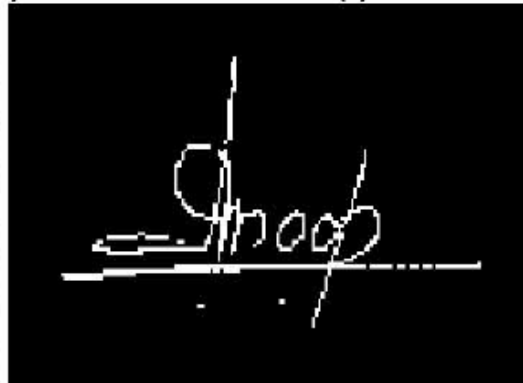
(j)



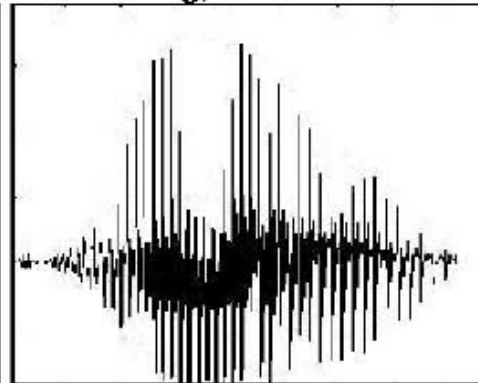
(k)



(l)



(m)



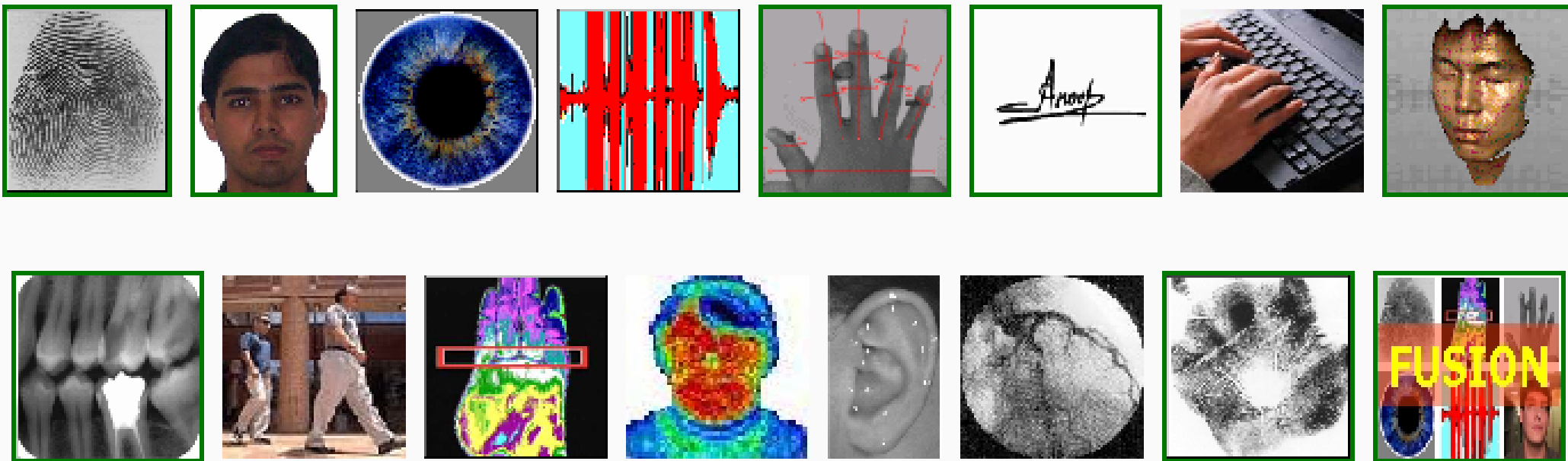
(n)

Adapted from  
Anil Jain,  
Michigan State



# Biometrics

---

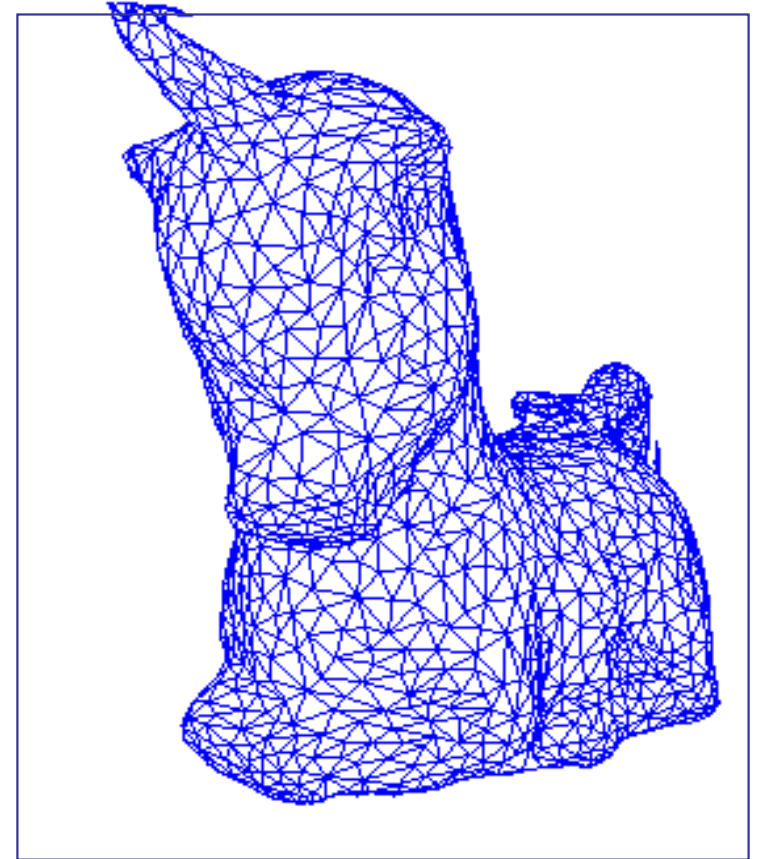


Adapted from Anil Jain, Michigan State



# Robotics

---

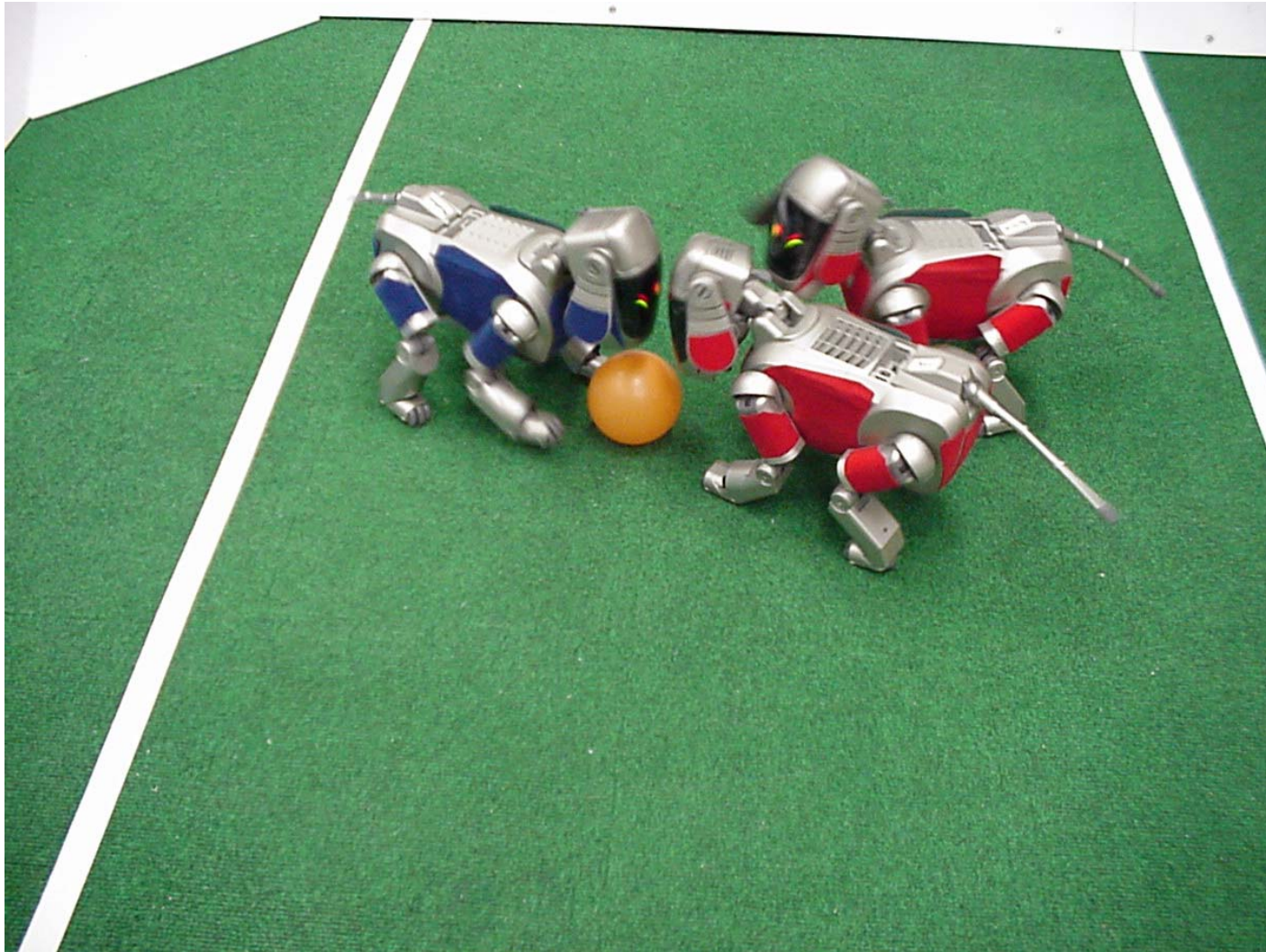


Adapted from Linda Shapiro, U of Washington



# Robotics

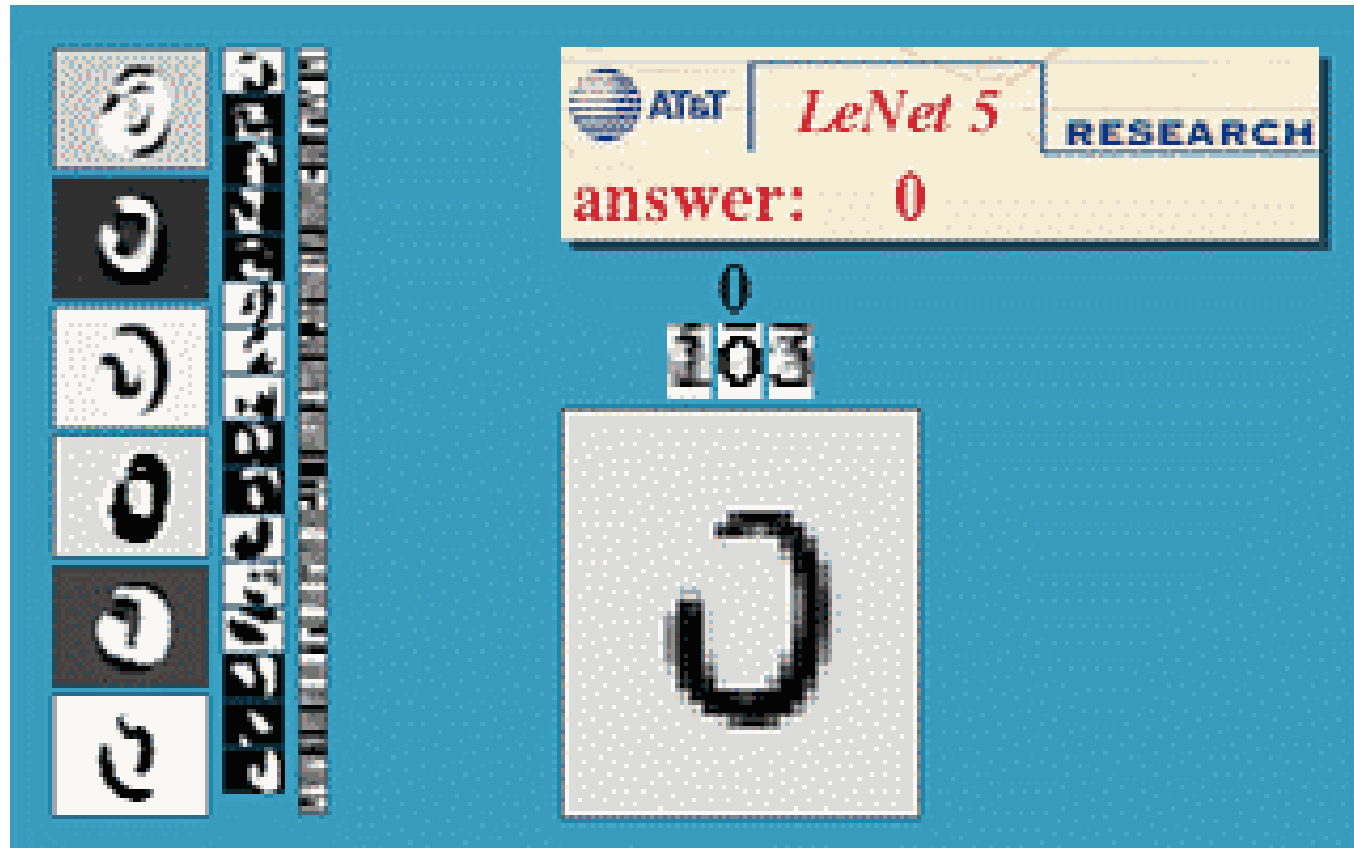
---



Adapted from Steven Seitz, U of Washington



# Document analysis



Digit recognition, AT&T labs  
<http://www.research.att.com/~yann>

Adapted from Steven Seitz, U of Washington



# Document analysis

---

儘眼望遠極，  
佰程無窮哩。  
壹物明域現，  
以迺吾後脊！

I looked as hard as I could see,  
beyond 100 plus infinity  
an object of bright intensity  
– it was the back of me!

Figure 1.5: (Left) Chinese characters and (right) English equivalent. Is it possible that a machine could automatically translate one into the other? Chinese characters and poem courtesy of John Weng.

Adapted from Shapiro and Stockman





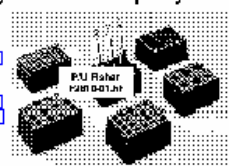
Peltier Cooler  
C1544-01.HI

## Model 147 IsoTemp® Dry Bath

[illegible]

Electrical Requirements	Cat. No.	Qty.
120V, 60Hz, 3000W (50 amp service)	11-215-100	1000
240V, 60Hz, 800W	11-215-1004	3000
Maximum distance from nearest 2" x 4" Joist: 16" Max.		

**Interchangeable Heating Blocks for IsoTemp® Dry Baths**

[illegible]

## Incu-Block<sup>®</sup> Partial Immersion Thermometers

For all standard bags: no blocks and water baths. Critical temperatures 125°, 10°, 37°, 58° C are marked with arrows. Available with shatter proof, contamination proof Teller® coating. Total length: 172 mm. In-mass flow: 35 mm.

Time Slot, min	MSB-DOS	Est. Hrs	Feasible
8	35	11-7:15-105	21-8
9	29	11-7:15-102	21-8
10	26, 15, 10, 10	11-7:15-120	21-10
11	12	11-7:15-100	21-8
12	12	11-7:15-121	21-8
13	12	11-7:15-111	21-8
14	12	11-7:15-113	21-8
15	12	11-7:15-128	21-10
16	12	11-7:15-115	21-8
17	11	11-7:15-117	21-8
21	5	11-7:15-119	21-5

### Alcohol Thermometers

For more thermometers, including digital types

see page 952

Ep: wa-rei-a-ma-e-fameyo, i-named-chi-ot-a-n-u-wa-lu

To Order Call: 1-800-6-60-335-876 Fax 1-505-251-0100

\* Transcription of the name of the Commission is based on the original of the letter.

Adapted from Linda Shapiro, U of Washington

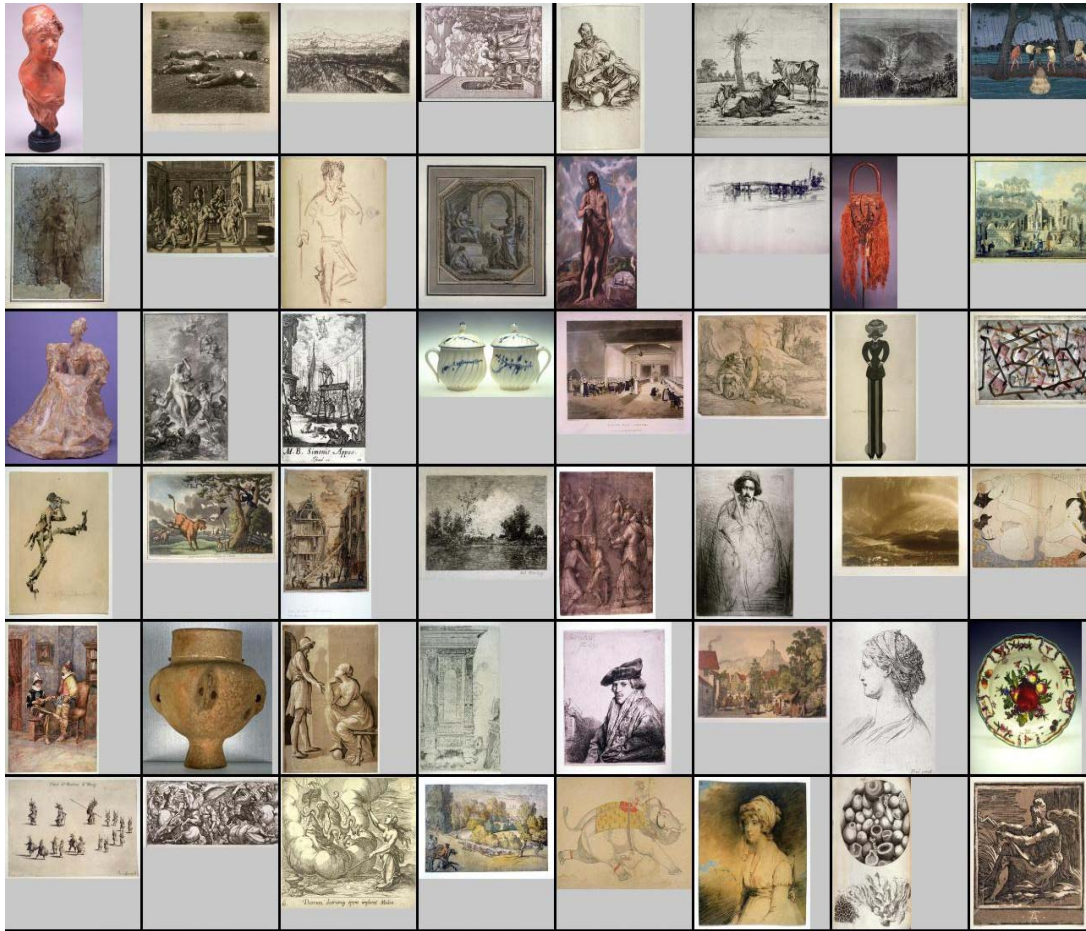


# Scene classification





# Organizing image archives



Adapted from Pinar Duygulu, Bilkent University



# Content-based retrieval

**UW ISL Image Database**

Query Image: image1108.ppm  
Load Random

Database: COREL Database  
Similarity Model: LAR + COOC + MVG  
LAR + COOC + FIT  
LAR + COOC + Lp

Graph Theoretic Clus  
Combined Classifiers  
Bayes Network  
MARS Model  
ETHZ Model  
Relevance Feedback

Change Working Dir.  
Num. Retrieved: (12)

<< Search >>

**Relevant Images:**

Image1106, d=-5.3067 Image1137, d=-5.2340 Image1108, d=-5.1543 Image1109, d=-5.0533

Image1134, d=-4.9668 Image1159, d=-4.8420 Image1115, d=-4.8137 Image1102, d=-4.7332

Image1123, d=-4.7003 Image1148, d=-4.6917 Image1129, d=-4.6206 Image1132, d=-4.5992

**Irrelevant Images:**

image752, d=62.6941 image1065, d=62.4209 image736, d=62.0796 image716, d=61.8161

Quit

**UW ISL Image Database**

Query Image: image1723.ppm  
Load Random

Database: COREL Database  
Similarity Model: LAR + COOC + MVG  
LAR + COOC + FIT  
LAR + COOC + Lp

Graph Theoretic Clus  
Combined Classifiers  
Bayes Network  
MARS Model  
ETHZ Model  
Relevance Feedback

Change Working Dir.  
Num. Retrieved: (12)

<< Search >>

**Relevant Images:**

Image1776, d=0.0194 Image1703, d=0.0228 Image1755, d=0.0282 Image1716, d=0.0313

Image1726, d=0.0324 Image1745, d=0.0332 Image1772, d=0.0352 Image1737, d=0.0358

Image1741, d=0.0361 Image1724, d=0.0378 Image1795, d=0.0395 Image1740, d=0.0415

**Irrelevant Images:**

image1465, d=73.3047 image1288, d=71.9889 image1491, d=65.6676 image1230, d=53.2475

Quit



# Content-based retrieval

**UW ISL Image Database**

Query Image: image245.ppm  
Load Random

Database: ISL Database  
Similarity Model: LAR + COOC + MVG  
LAR + COOC + FIT  
LAR + COOC + Lp

Graph Theoretic Clus  
Combined Classifiers  
Bayes Network  
MARS Model  
ETHZ Model  
Relevance Feedback

Change Working Dir.  
Num. Retrieved: (12)

<< Search >>

Relevant Images:

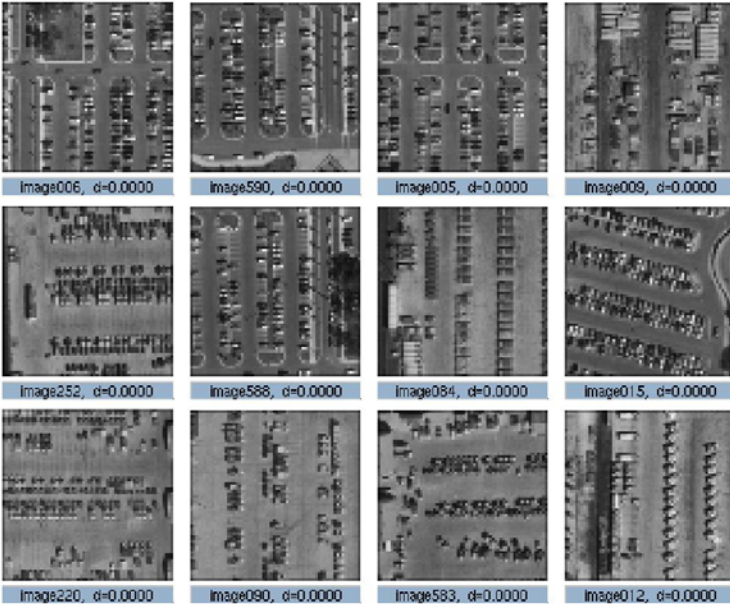


Image006, d=0.0000 Image590, d=0.0000 Image005, d=0.0000 Image009, d=0.0000  
Image252, d=0.0000 Image588, d=0.0000 Image084, d=0.0000 Image015, d=0.0000  
Image220, d=0.0000 Image090, d=0.0000 Image583, d=0.0000 Image012, d=0.0000

Irrelevant Images:




image208, image053, image539, image049,

**UW ISL Image Database**

Query Image: image292.ppm  
Load Random

Database: COREL Database  
Similarity Model: LAR + COOC + MVG  
LAR + COOC + FIT  
LAR + COOC + Lp

Graph Theoretic Clus  
Combined Classifiers  
Bayes Network  
MARS Model  
ETHZ Model  
Relevance Feedback

Change Working Dir.  
Num. Retrieved: (12)

<< Search >>

Relevant Images:

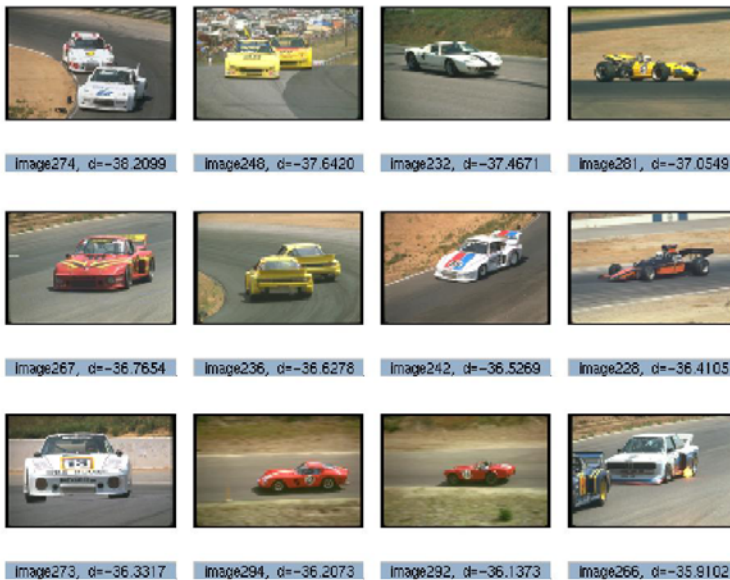


Image274, d=-38.2099 Image248, d=-37.6420 Image232, d=-37.4671 Image281, d=-37.0549  
Image267, d=-36.7654 Image236, d=-36.6278 Image242, d=-36.5269 Image228, d=-36.4105  
Image273, d=-36.3317 Image294, d=-36.2073 Image292, d=-36.1373 Image266, d=-35.9102

Irrelevant Images:






image1065, d=410.5549 image752, d=401.2940 image723, d=389.1098 image736, d=385.4378



# Content-based retrieval

Notepad: Bidworld Query Results: image +10019 (Patched)



File Edit View Go Communications Help







Query image: 10019      Query blobs



Querying from 35000 images (2000 returned by the filter).



	feature importance:				
	overall	color	texture	location	shape
blob	very	very	somewhat	not	not
background	somewhat	very	not	not	not

























# Face detection and recognition





# Object recognition

Part 1 – Det:5e-18



Part 2 – Det:8e-22



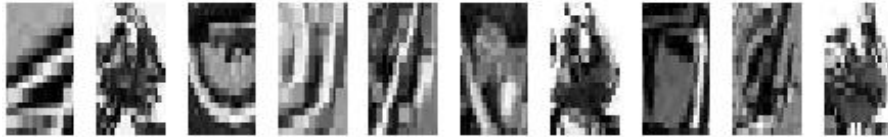
Part 3 – Det:6e-18



Part 4 – Det:1e-19



Part 5 – Det:3e-17



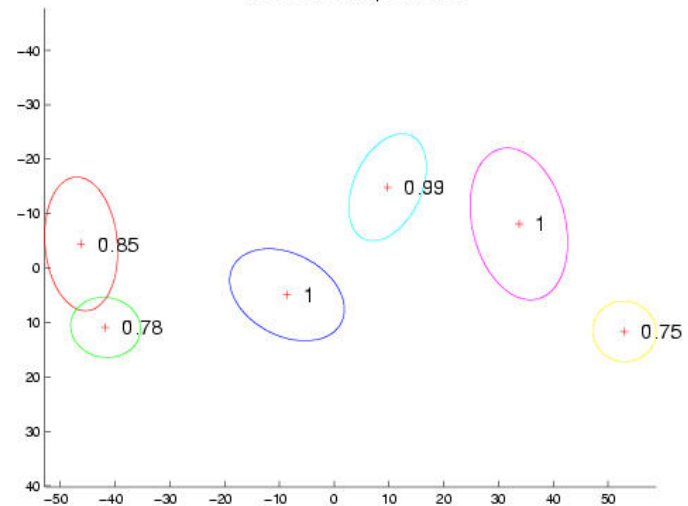
Part 6 – Det:4e-24



Background – Det:5e-19



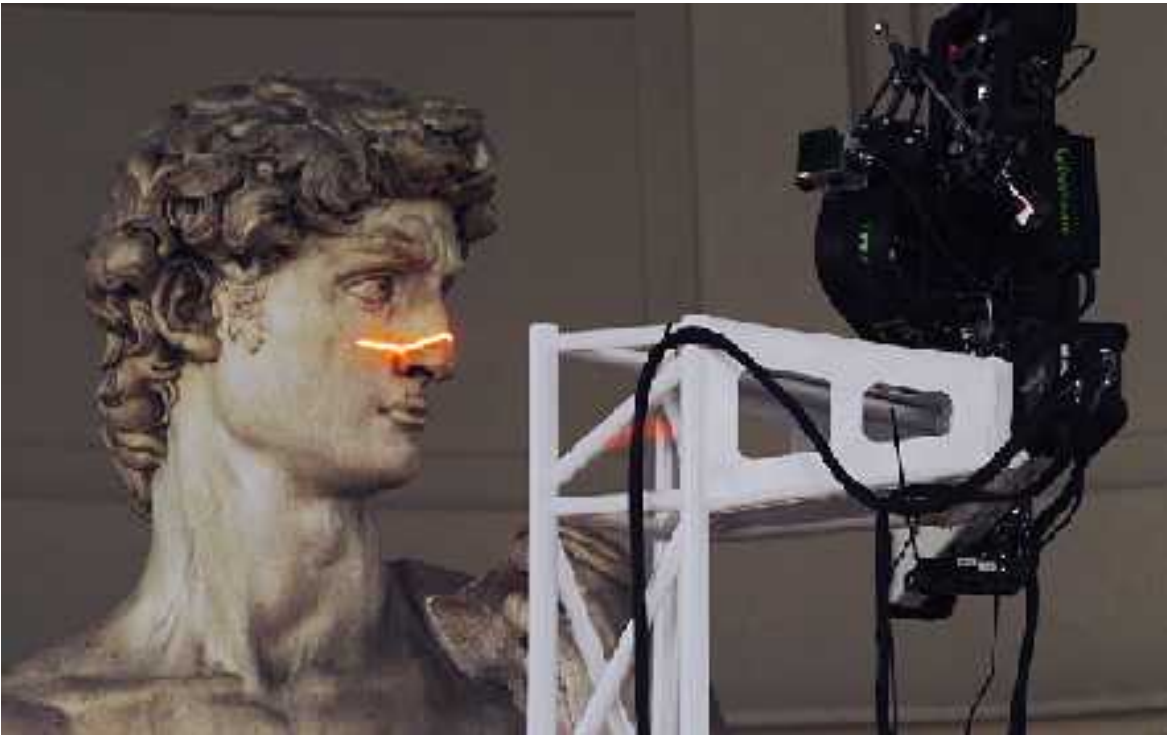
Motorbike shape model



Adapted from Rob Fergus, MIT



# 3D scanning



Adapted from Linda Shapiro, U of Washington



# 3D reconstruction

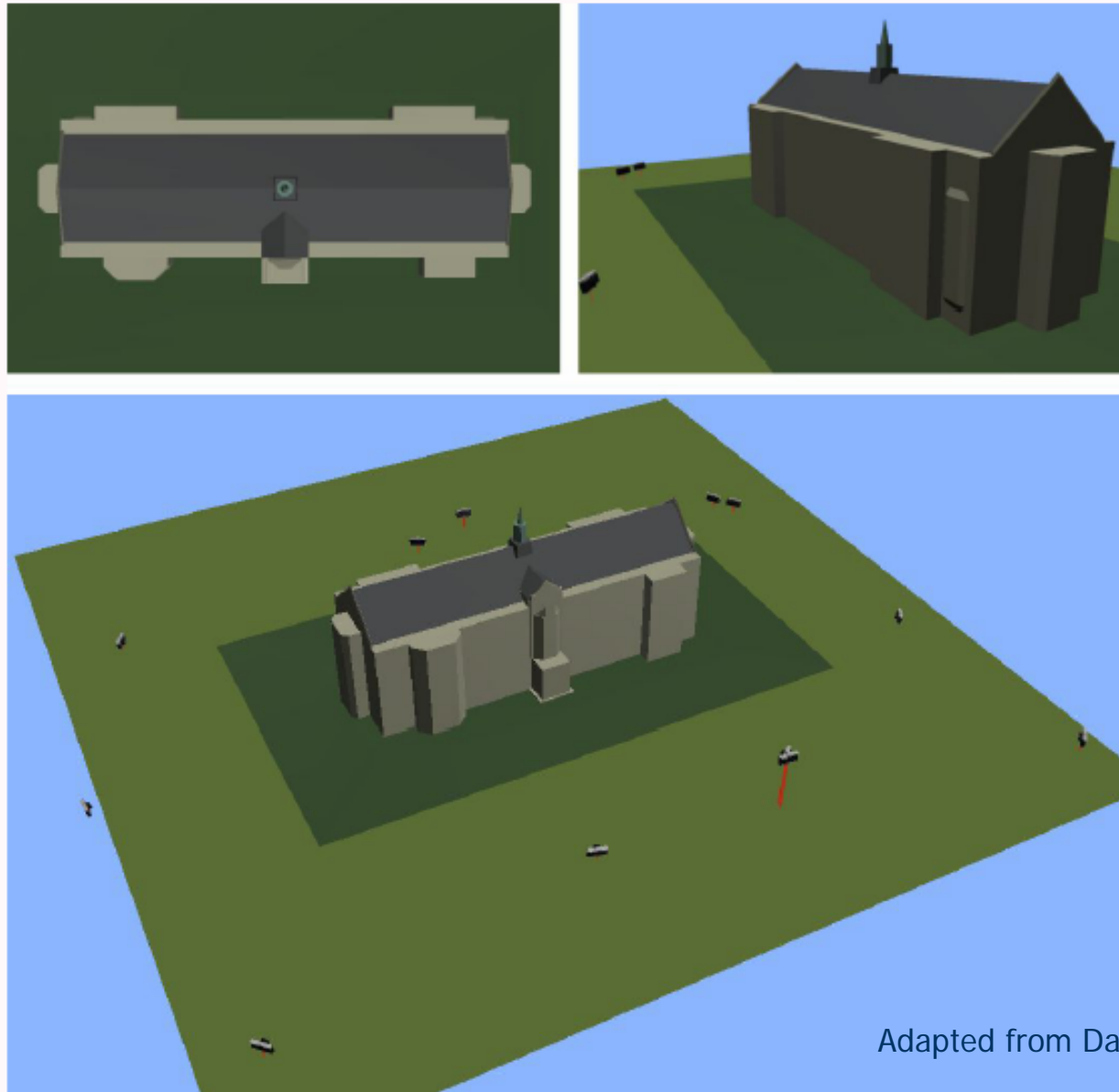
---



Adapted from David Forsyth, UC Berkeley



# 3D reconstruction



Adapted from David Forsyth, UC Berkeley



# Motion capture

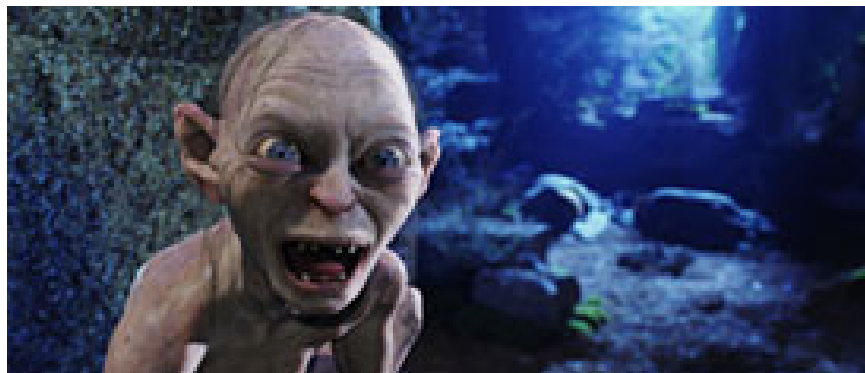
---





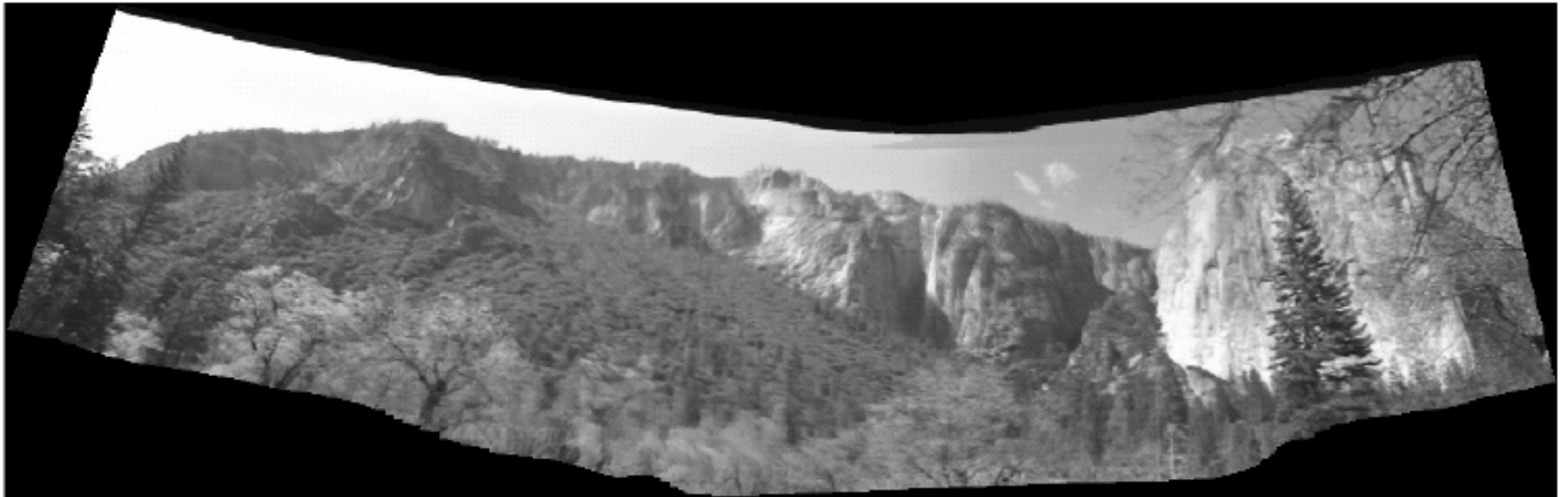
# Visual effects

---





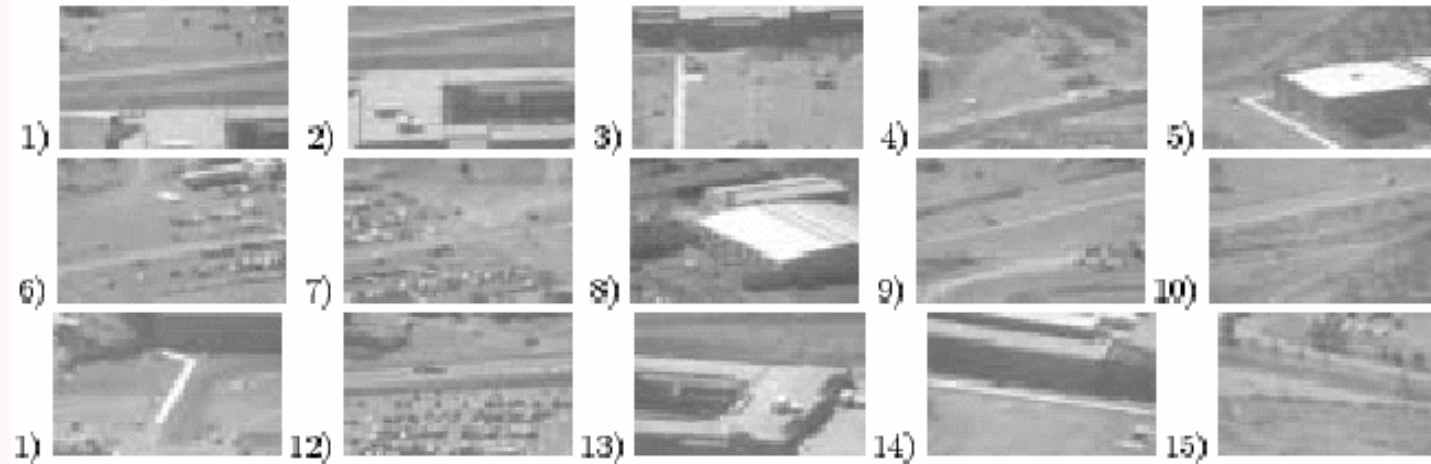
# Mozaic



Adapted from David Forsyth, UC Berkeley



# Mozaic



Adapted from David Forsyth, UC Berkeley



# Critical issues

---

- What information should be extracted?
- How can it be extracted?
- How should it be represented?
- How can it be used to aid analysis and understanding?