Introduction

Selim Aksoy
Department of Computer Engineering
Bilkent University
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?

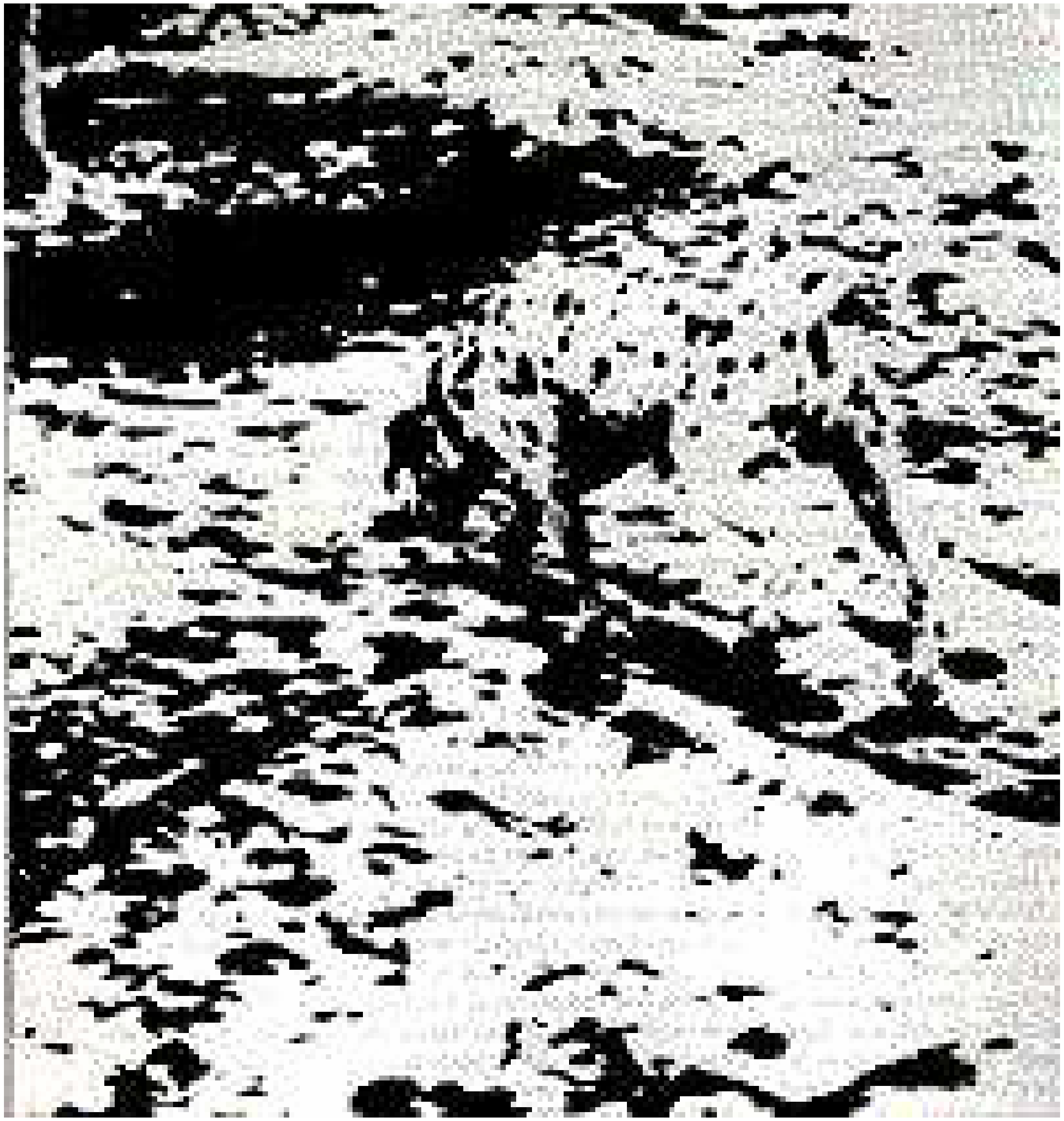
- An image is worth 1000 words.
- Images and videos are everywhere.
- Fast growing collections and many useful applications.
- Goals of vision research:
  - Give machines the ability to understand scenes.
  - Aid understanding and modeling of human vision.
  - Automate visual operations.

Challenge

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

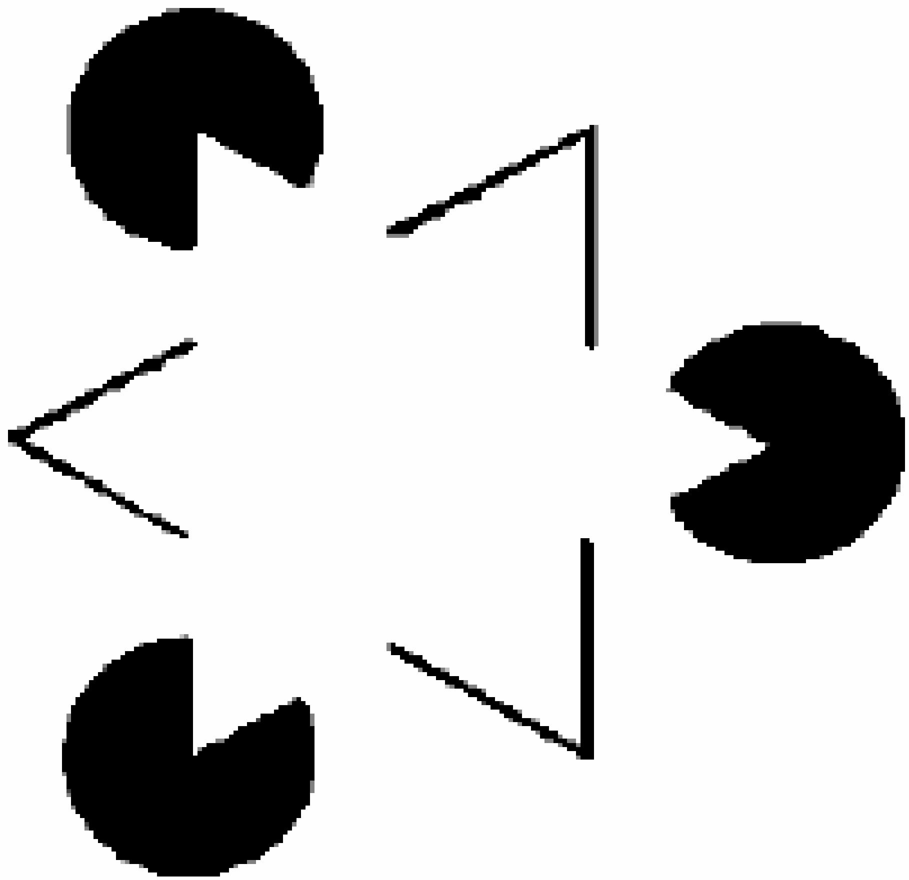
Adapted from Octavia Camps, Penn State

Perception and grouping



Subjective contours

Perception and grouping



Subjective contours

Adapted from Michael Black, Brown University

Perception and grouping

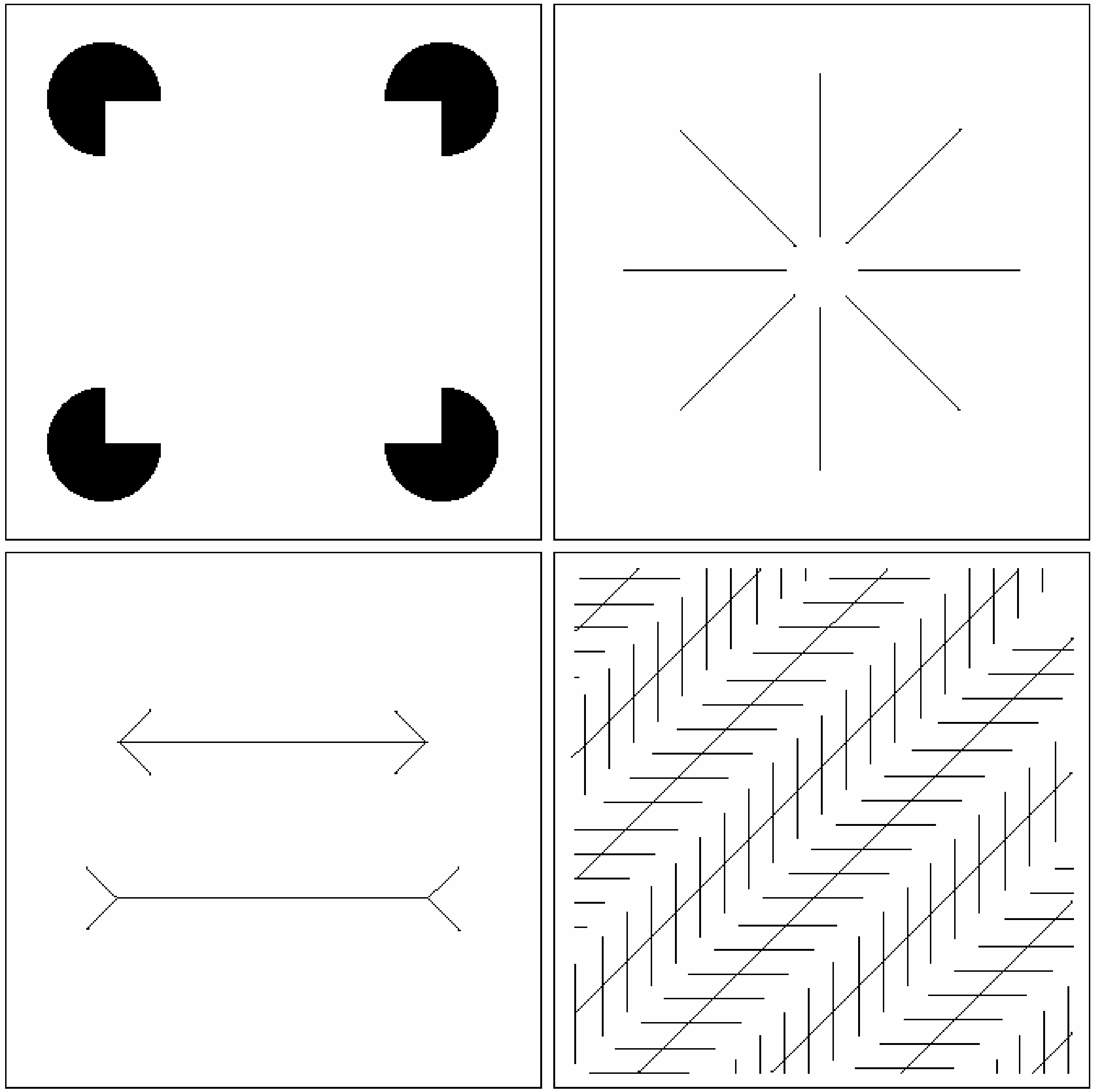


FIGURE 2.9 Some well-known optical illusions.

Adapted from Gonzales and Woods

Perception and grouping

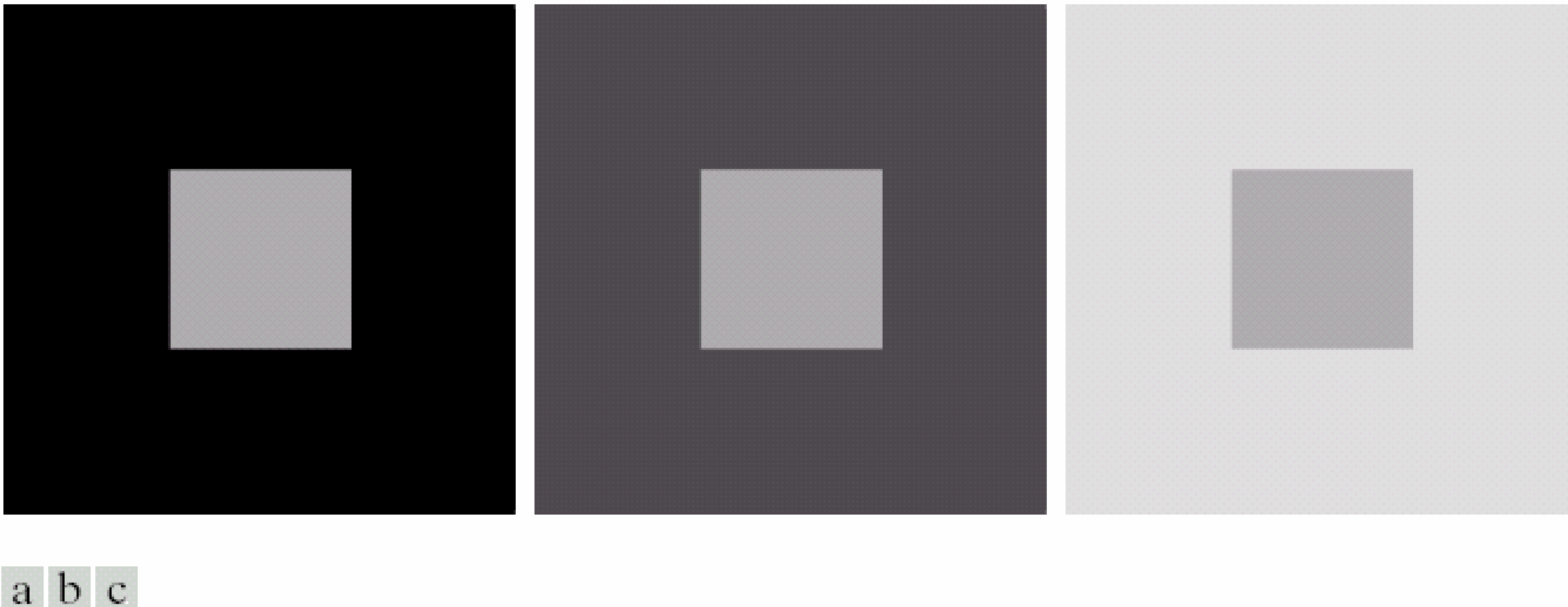
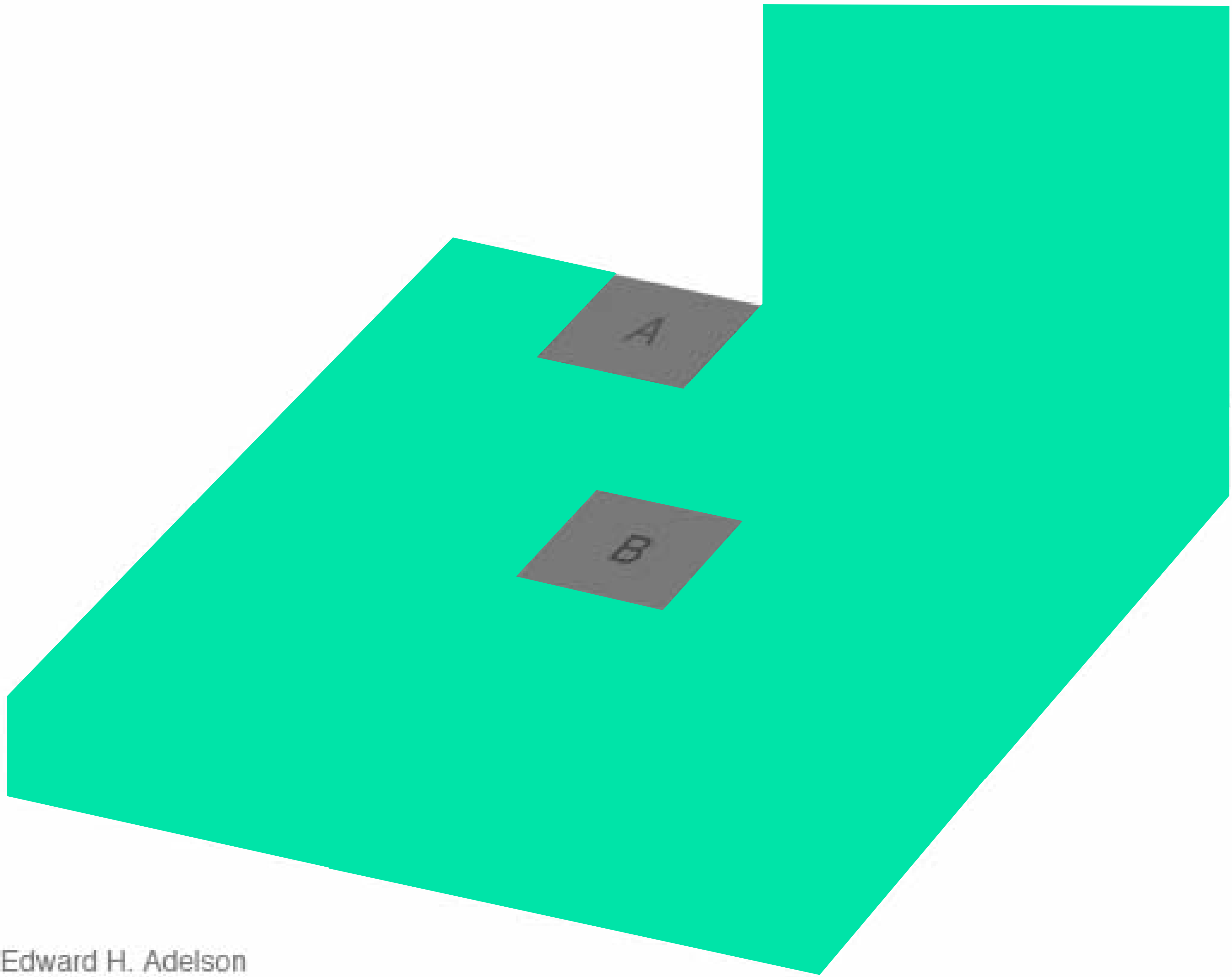


FIGURE 2.8 Examples of simultaneous contrast. All the inner squares have the same intensity, but they appear progressively darker as the background becomes lighter.

Adapted from Gonzales and Woods



Edward H. Adelson

Perception and grouping

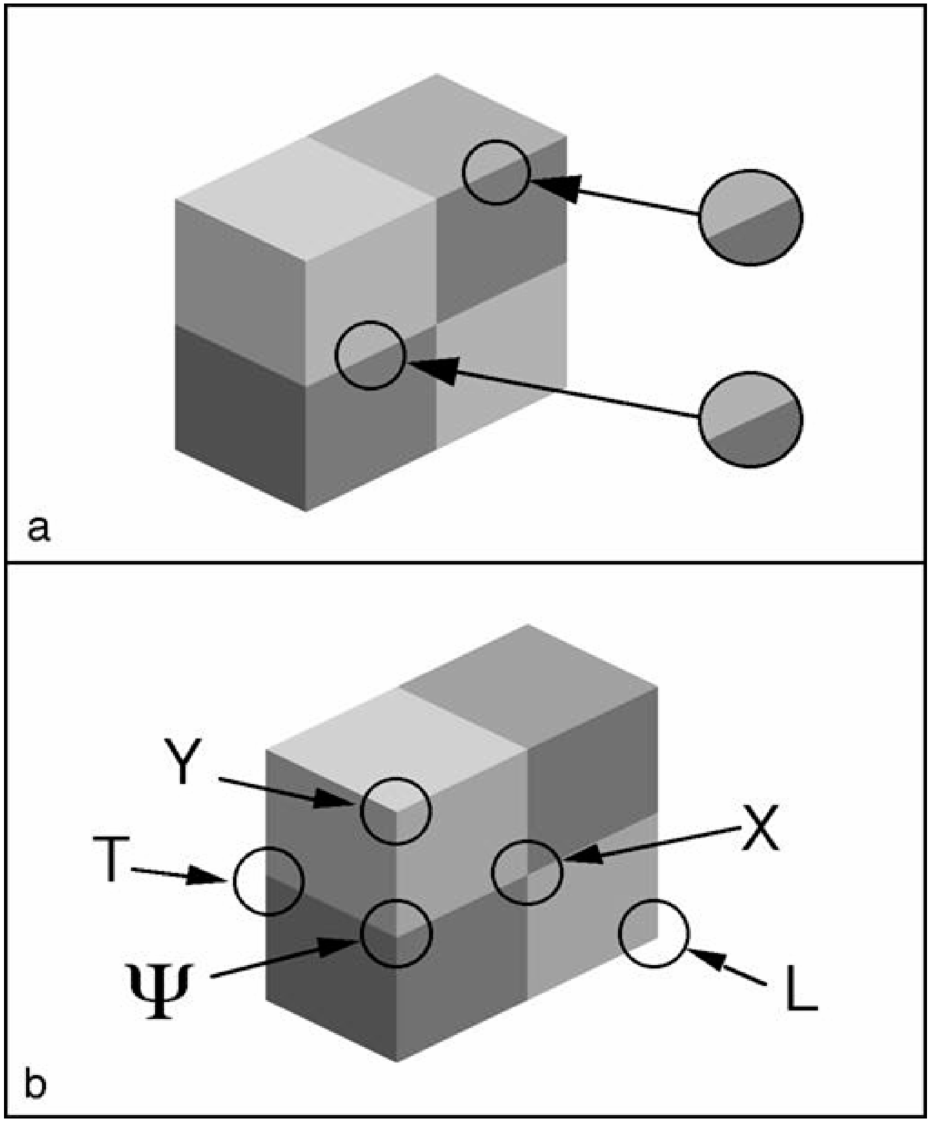
* types of "junctions" give cues about surfaces, occlusion, and light.

I. Rock, The Logic of Perception, 1983.

Occlusion

Adapted from Michael Black, Brown University

Perception and grouping



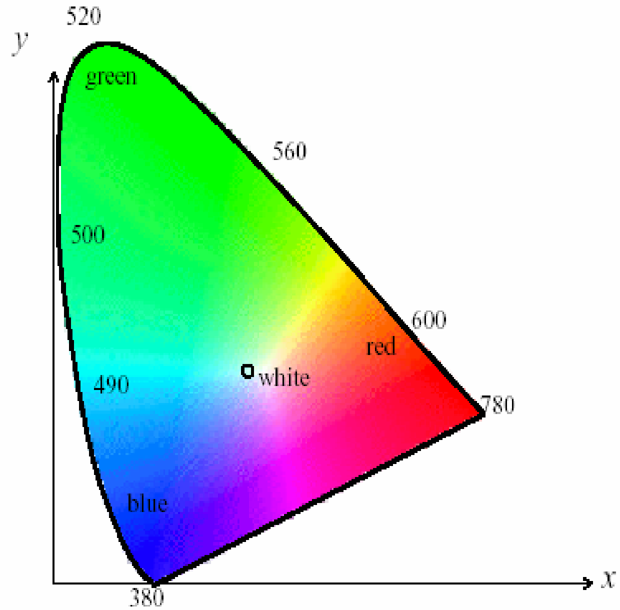
- The shape of junctions constrains the possible interpretations of the scene.
- Ambiguous: paint and surface boundaries can be confused.

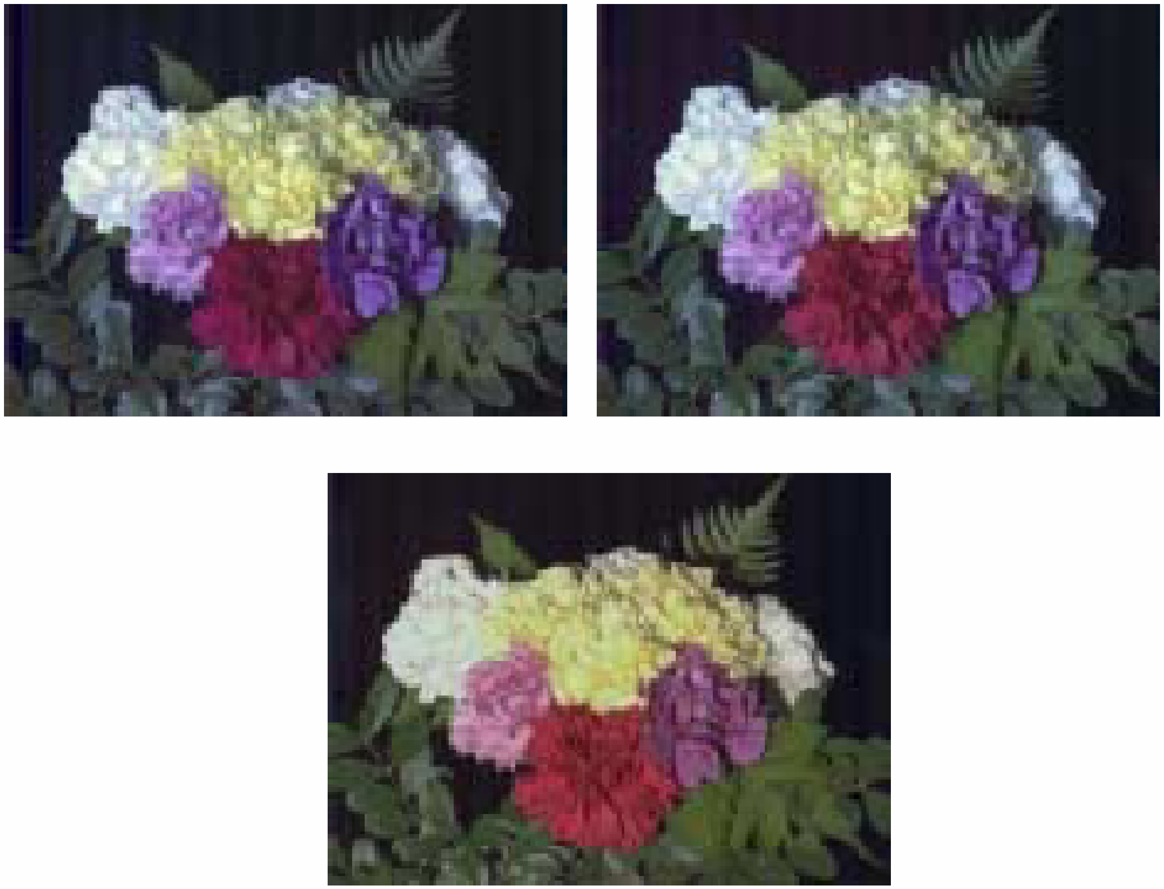
Adapted from Michael Black, Brown University

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





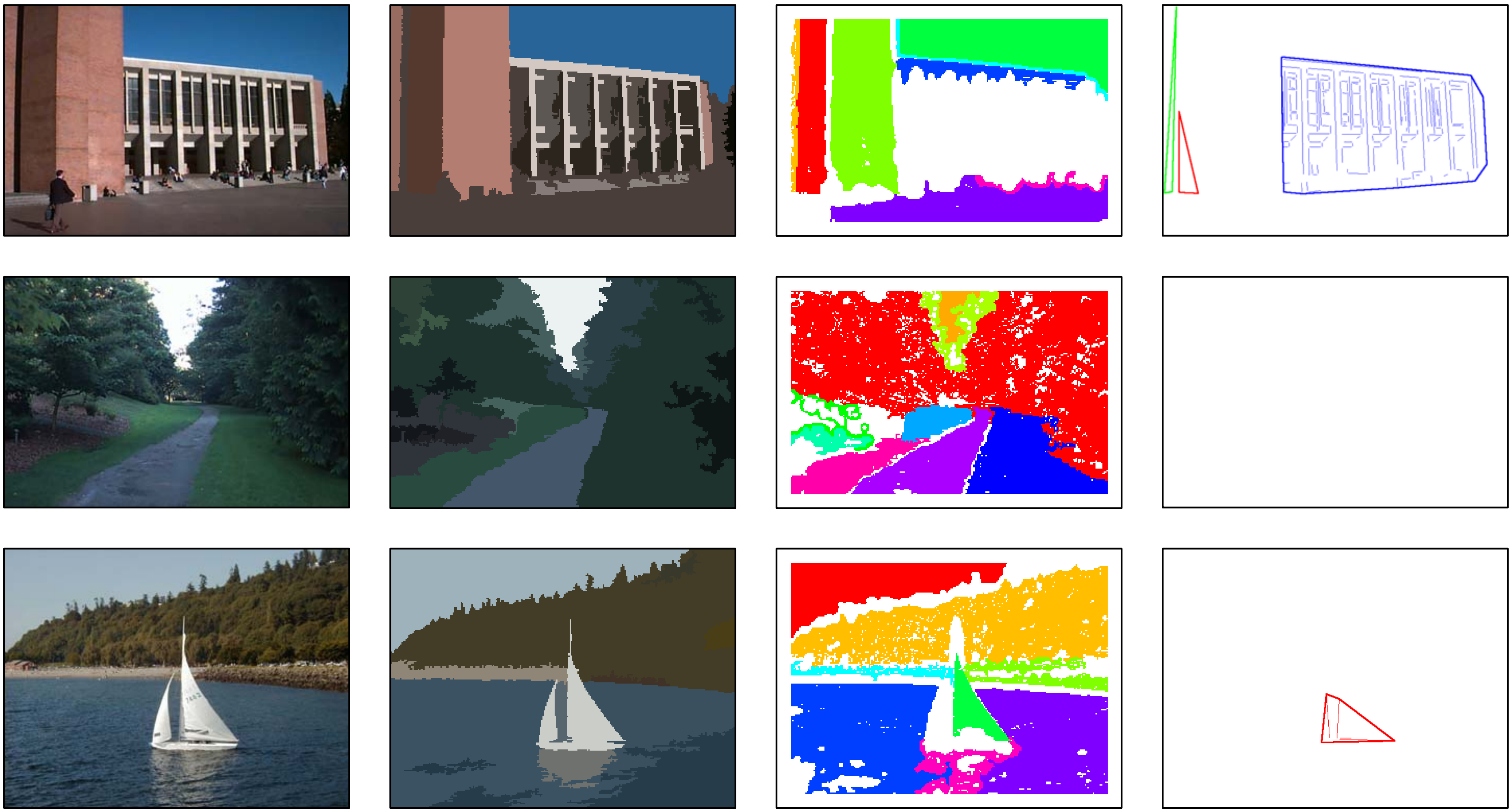
Adapted from Martial Hebert, CMU

Texture

Adapted from David Forsyth, UC Berkeley

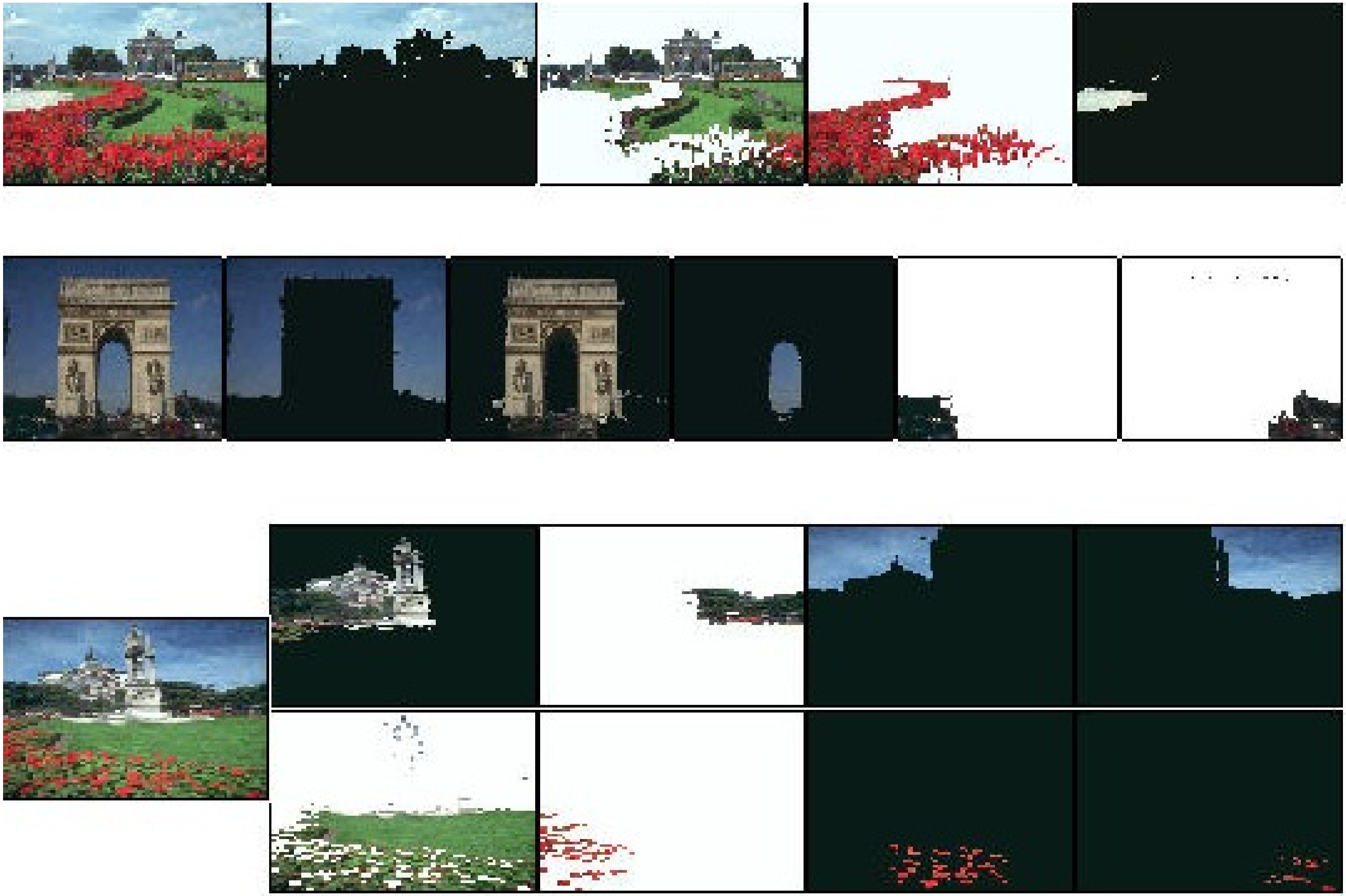
Segmentation

Original Images | Color Regions | Texture Regions | Line Clusters



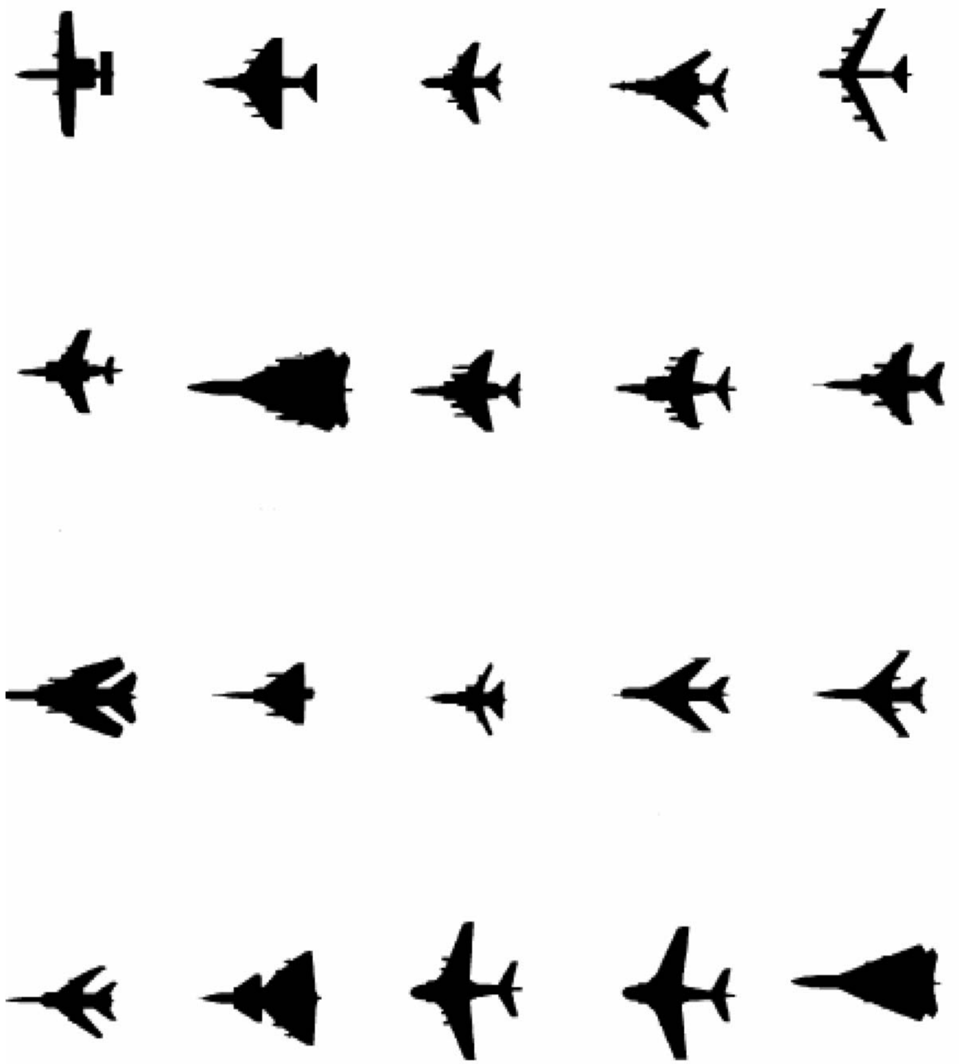
Adapted from Linda Shapiro, U of Washington

Segmentation



Adapted from Jianbo Shi, U Penn

Shape



Recognized objects

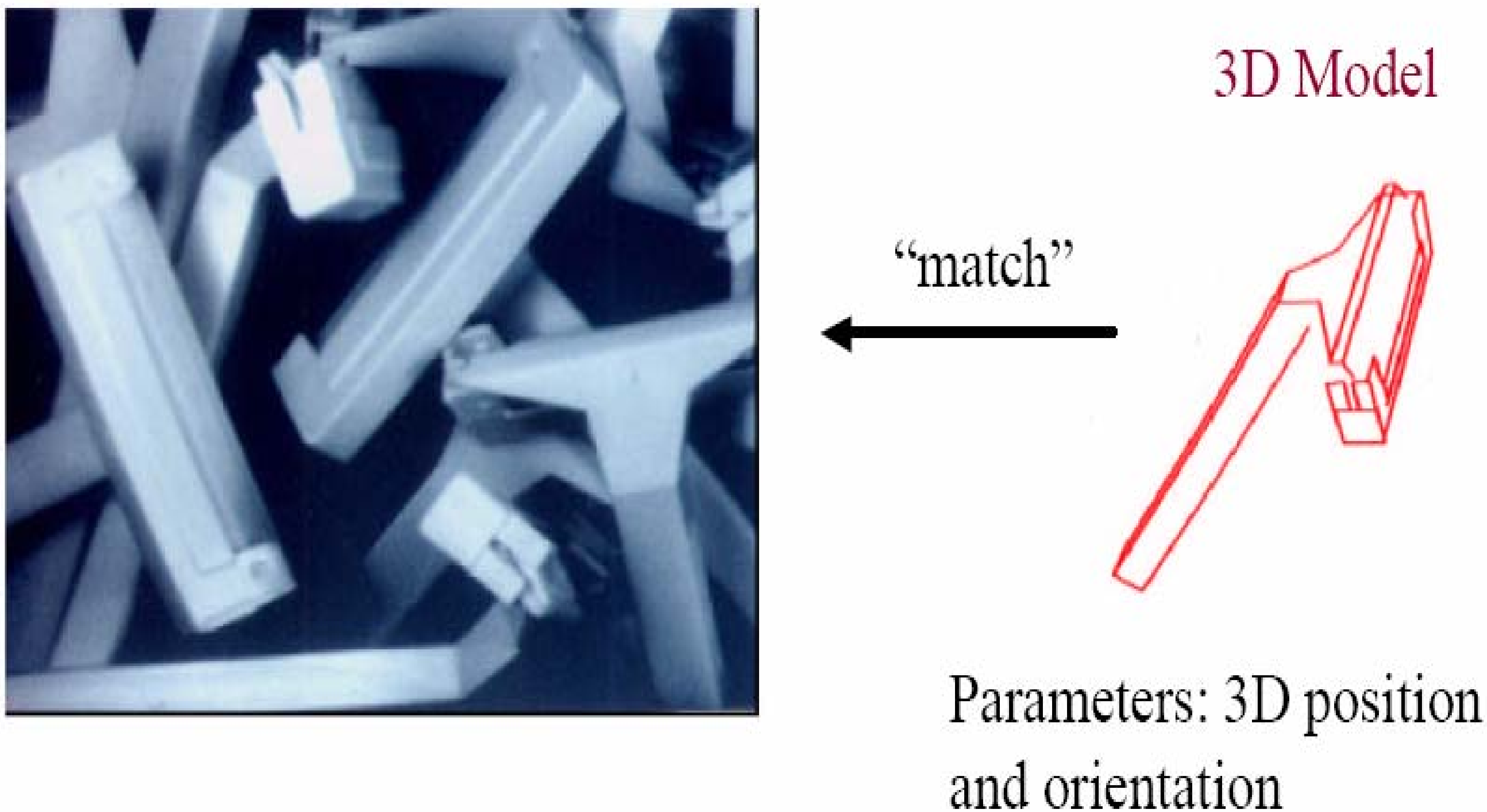
Model database

Adapted from Enis Cetin, Bilkent University

Motion

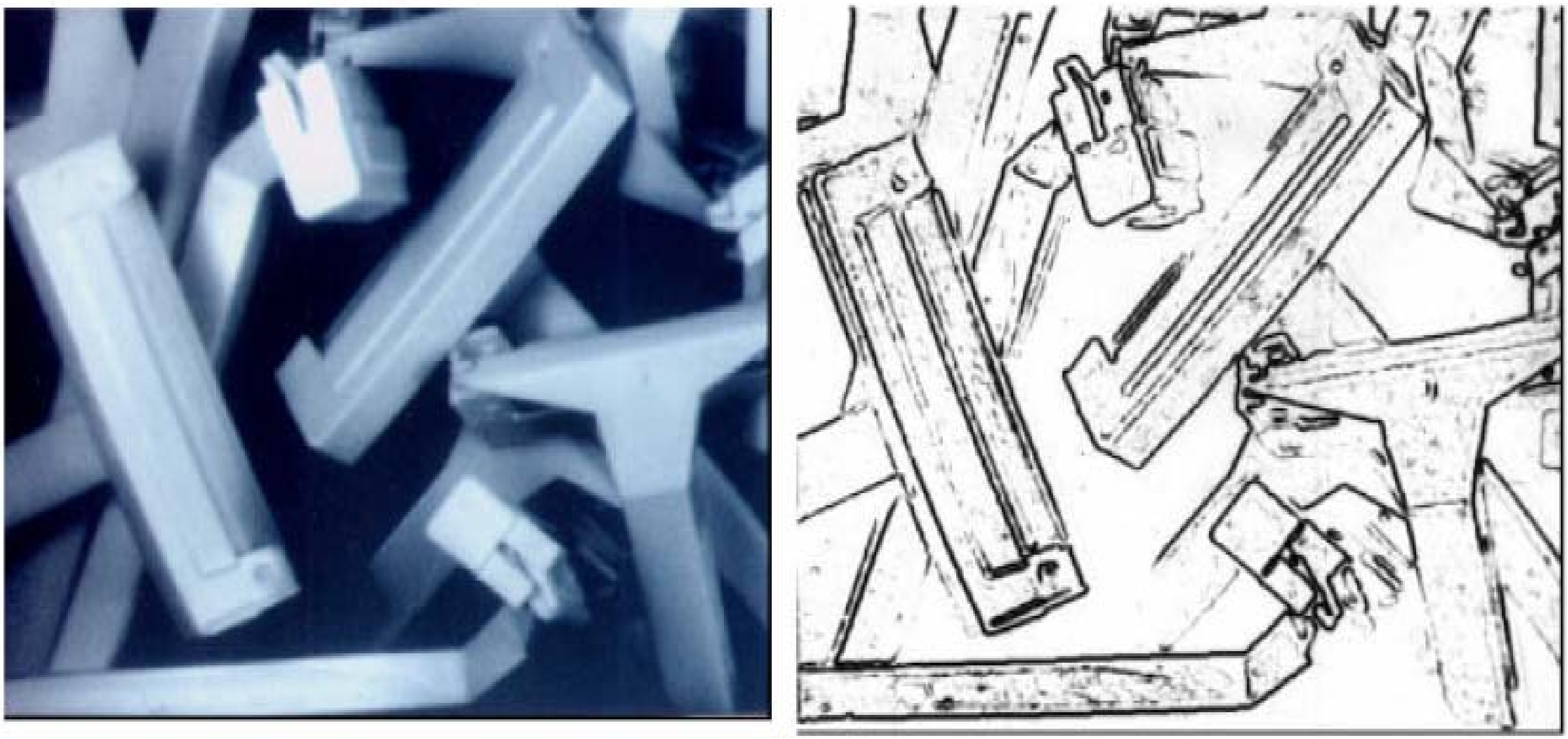
Adapted from Michael Black, Brown University

Recognition



Adapted from Michael Black, Brown University

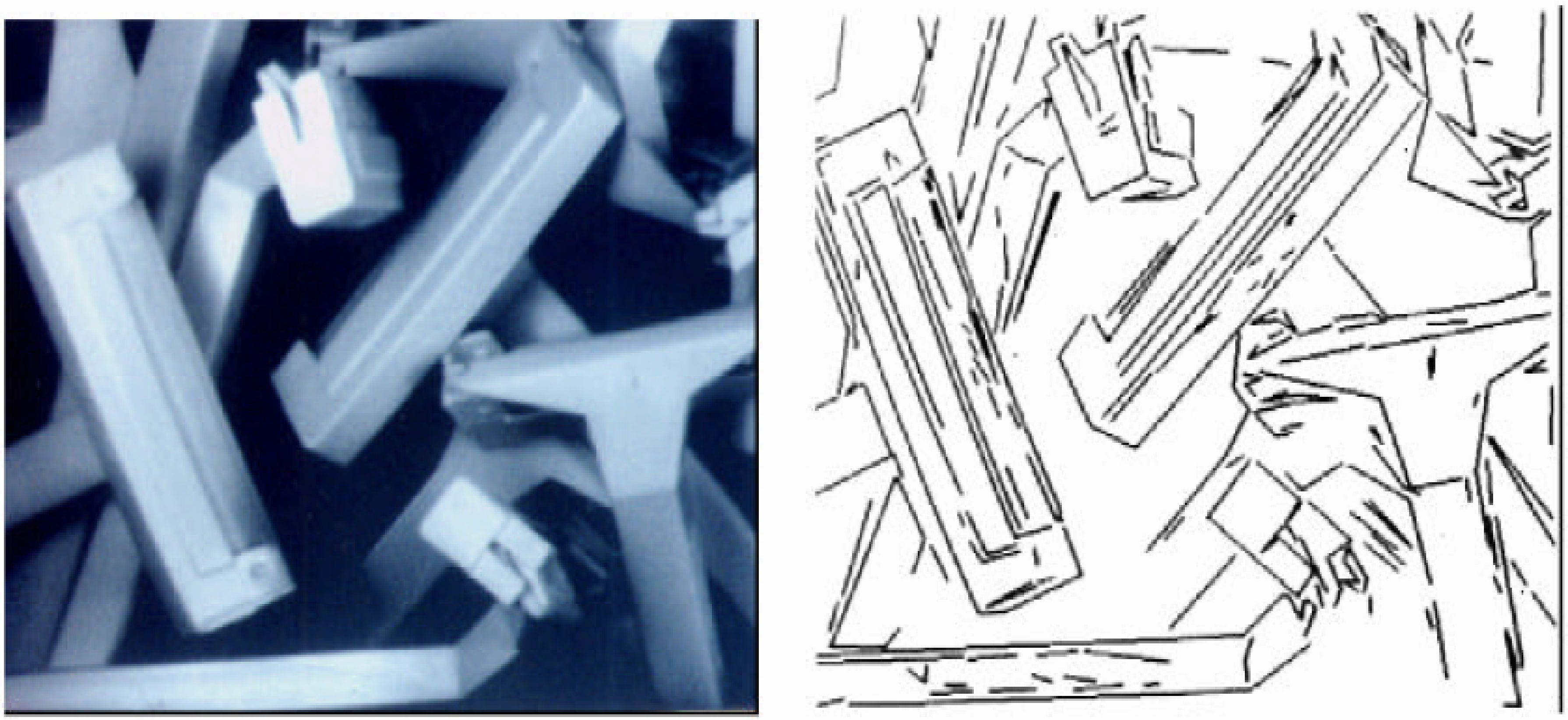
Recognition



"Filter" image to find brightness changes.

Adapted from Michael Black, Brown University

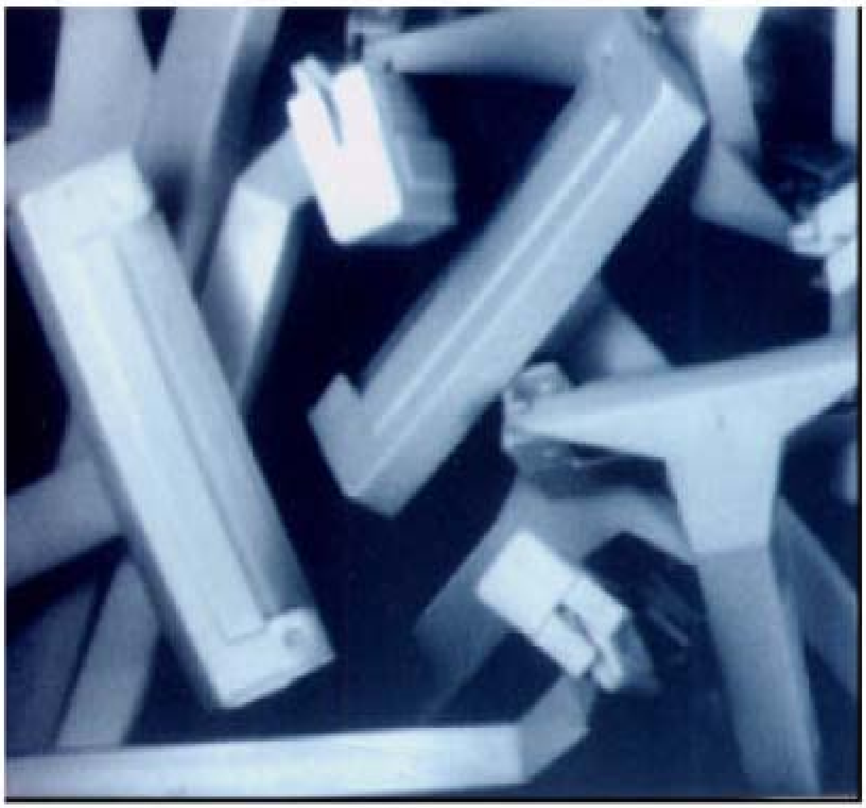
Recognition

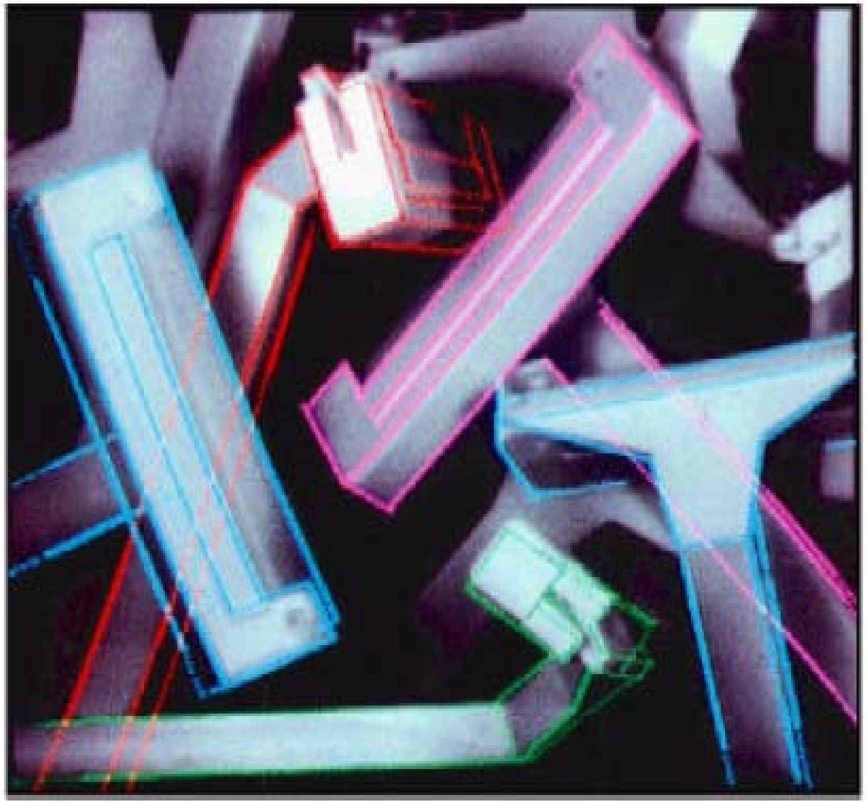


“Fit” lines to the raw measurements.

Adapted from Michael Black, Brown University

Recognition

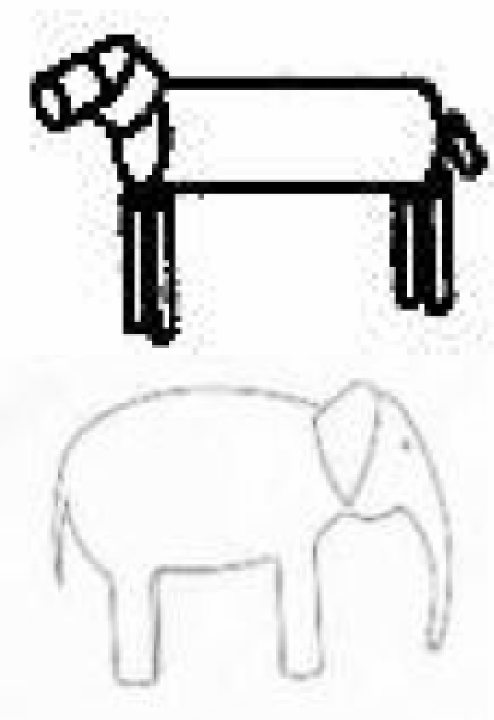




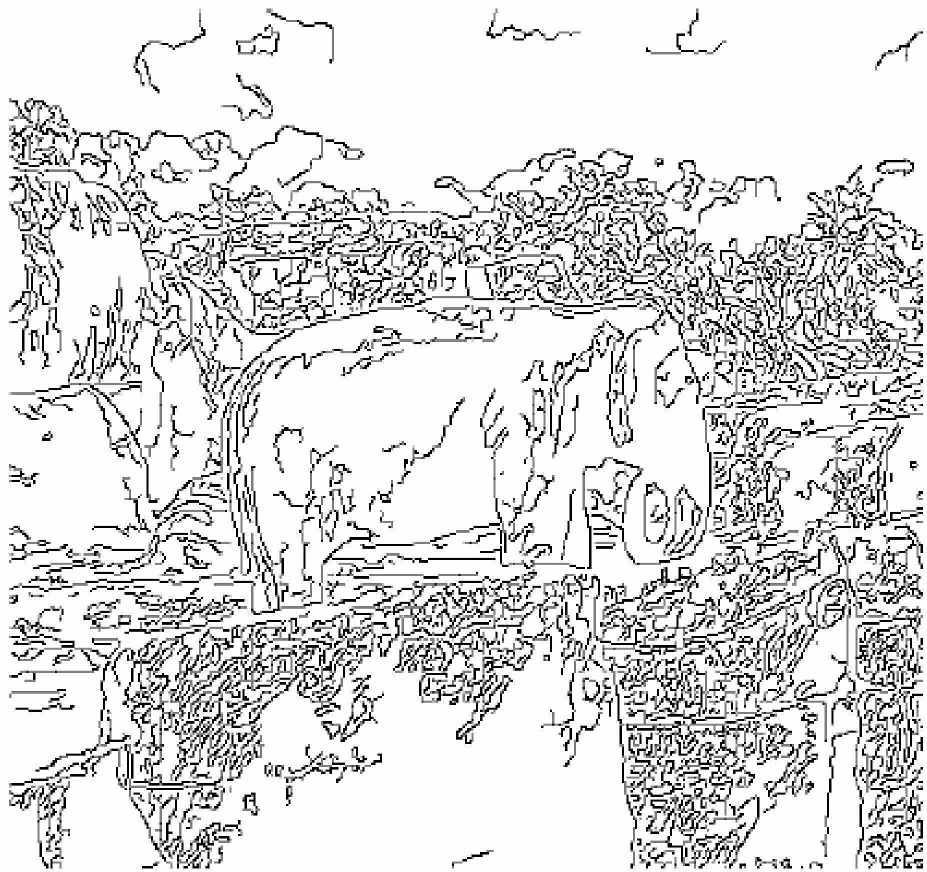
"Project" model into image and "match" to lines (solving for 3D pose).

Adapted from Michael Black, Brown University

Recognition

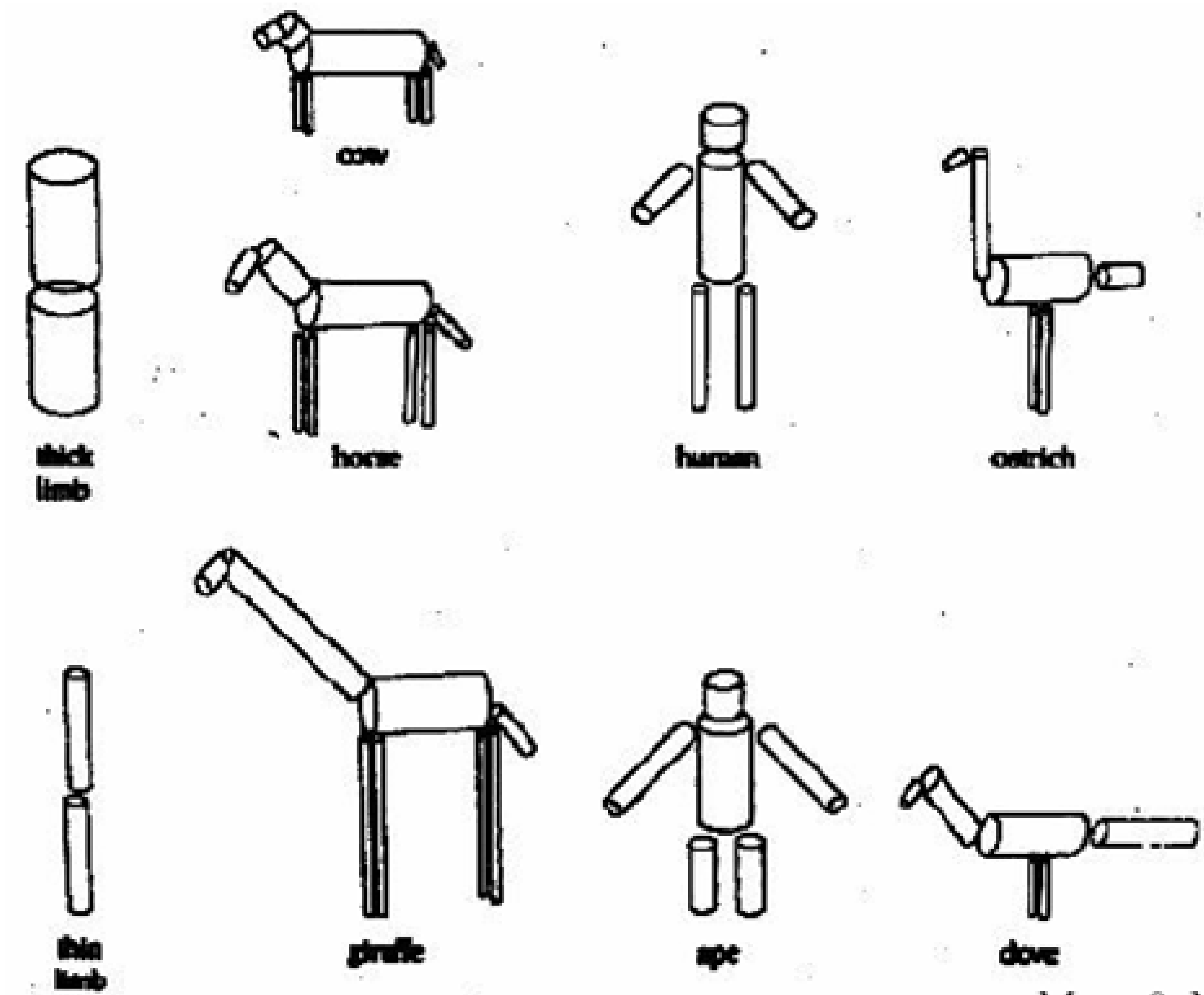


Match “model” to measurements?



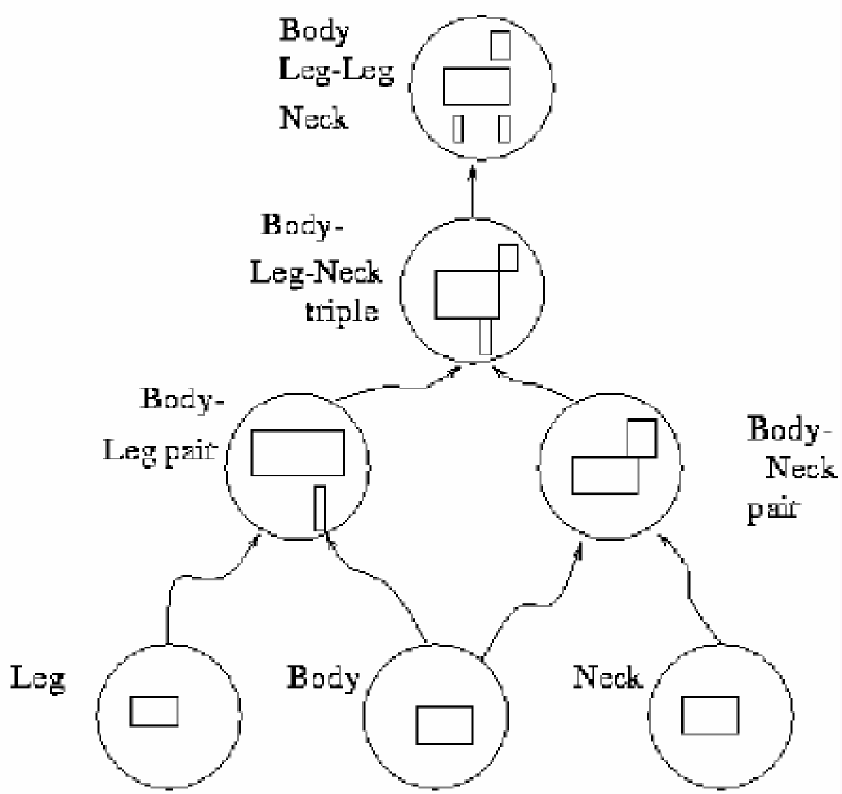
Adapted from Michael Black, Brown University

Recognition



Marr & Nishihara

Recognition

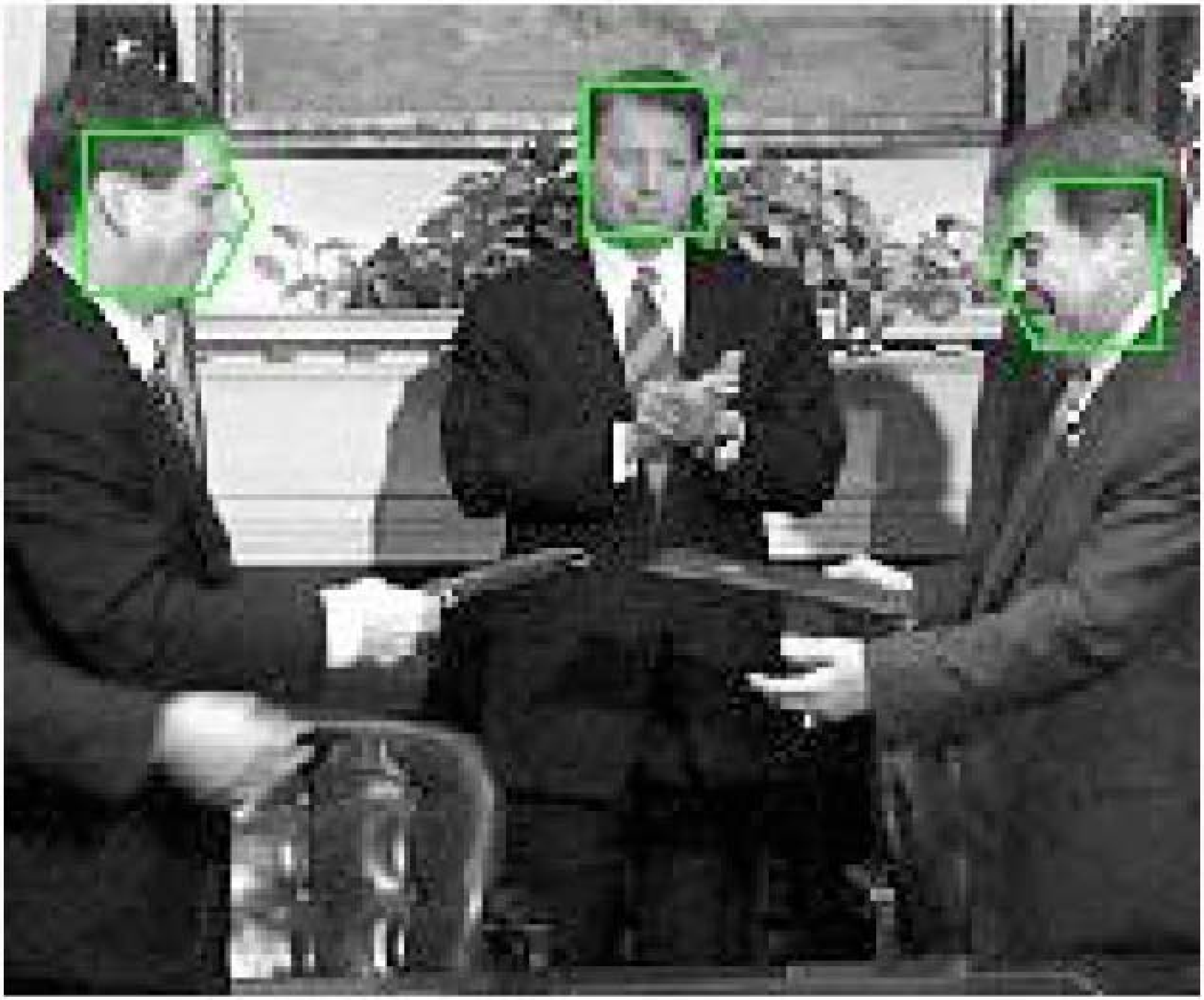


Adapted from David Forsyth, UC Berkeley

Detection

Adapted from David Forsyth, UC Berkeley

Detection

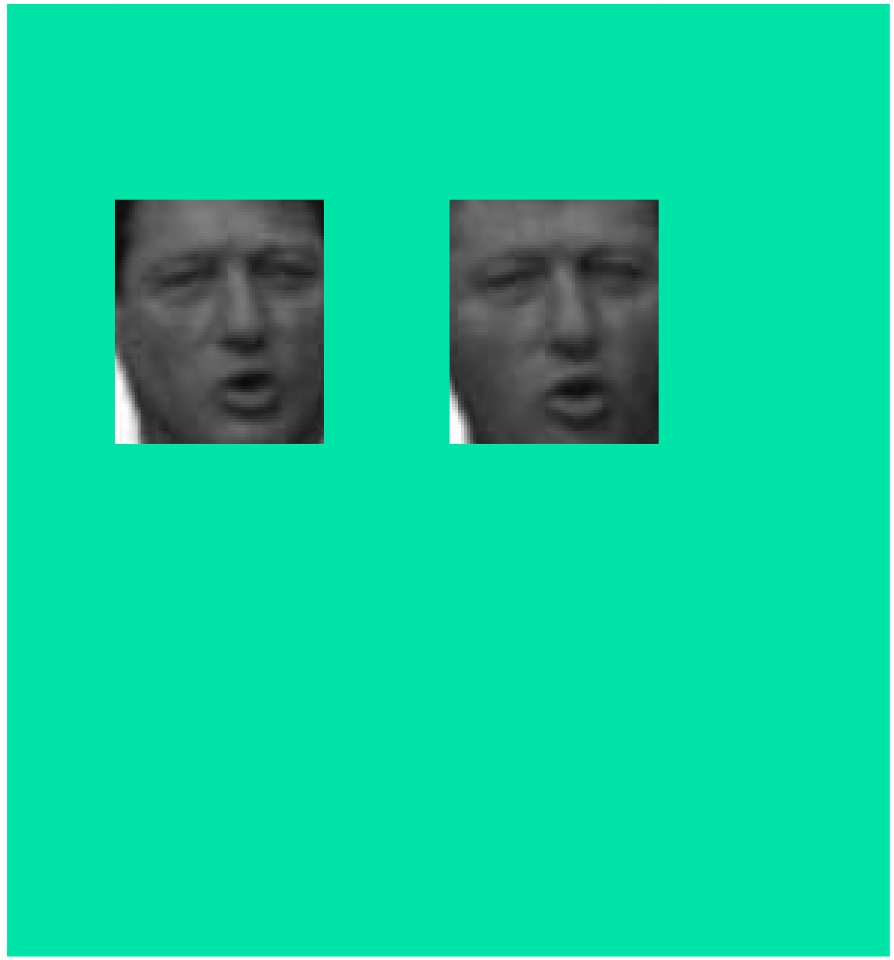


Adapted from David Forsyth, UC Berkeley

Detection

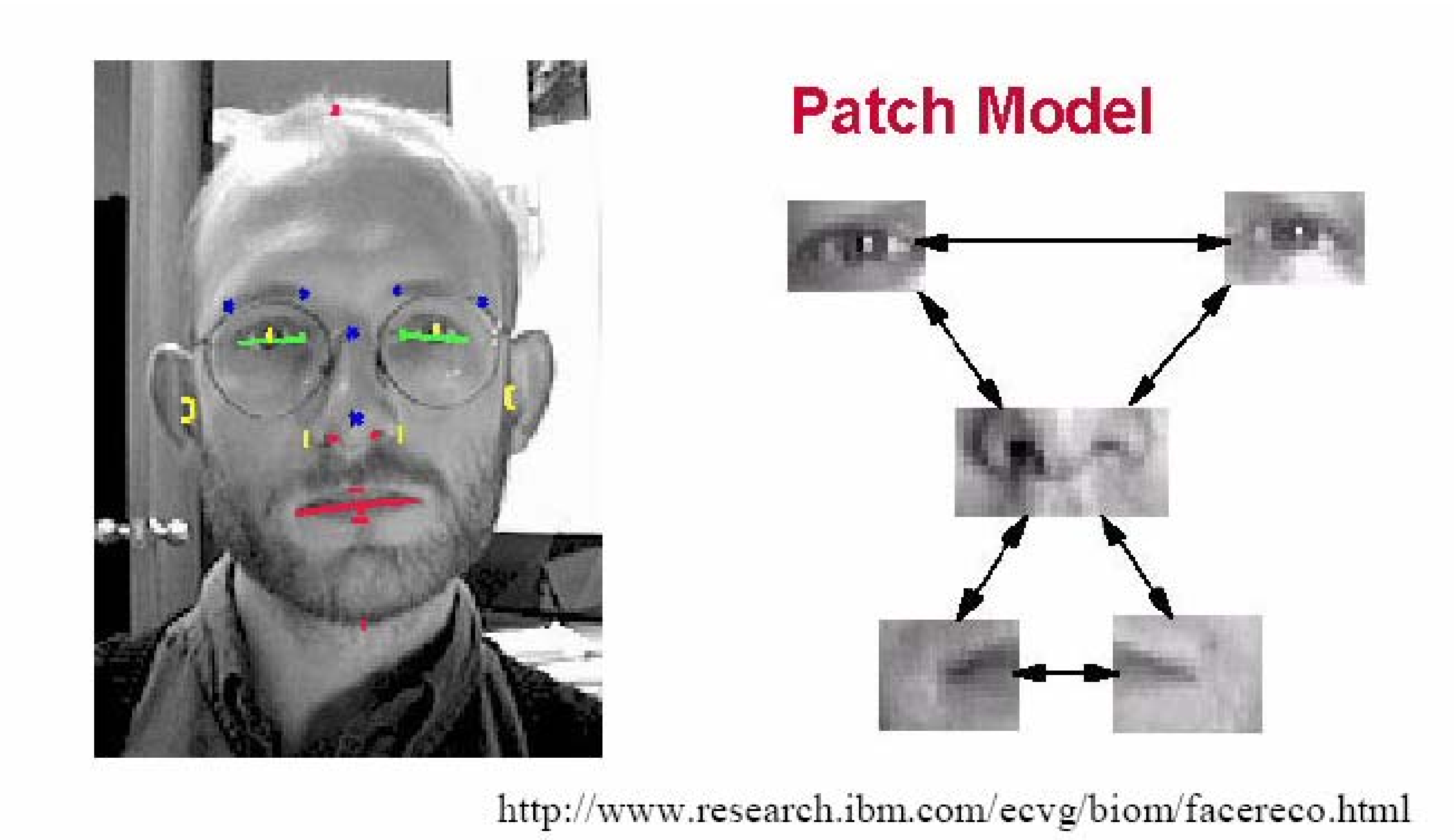
What are our "models"?

How good are they?



Adapted from Michael Black, Brown University

Parts and relations



Adapted from Michael Black, Brown University

Parts and relations

How flexible are the spatial relations of the parts?

Adapted from Michael Black, Brown University

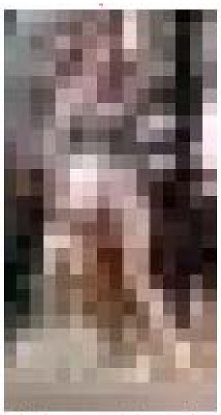
Context

Adapted from Antonio Torralba, MIT

Context

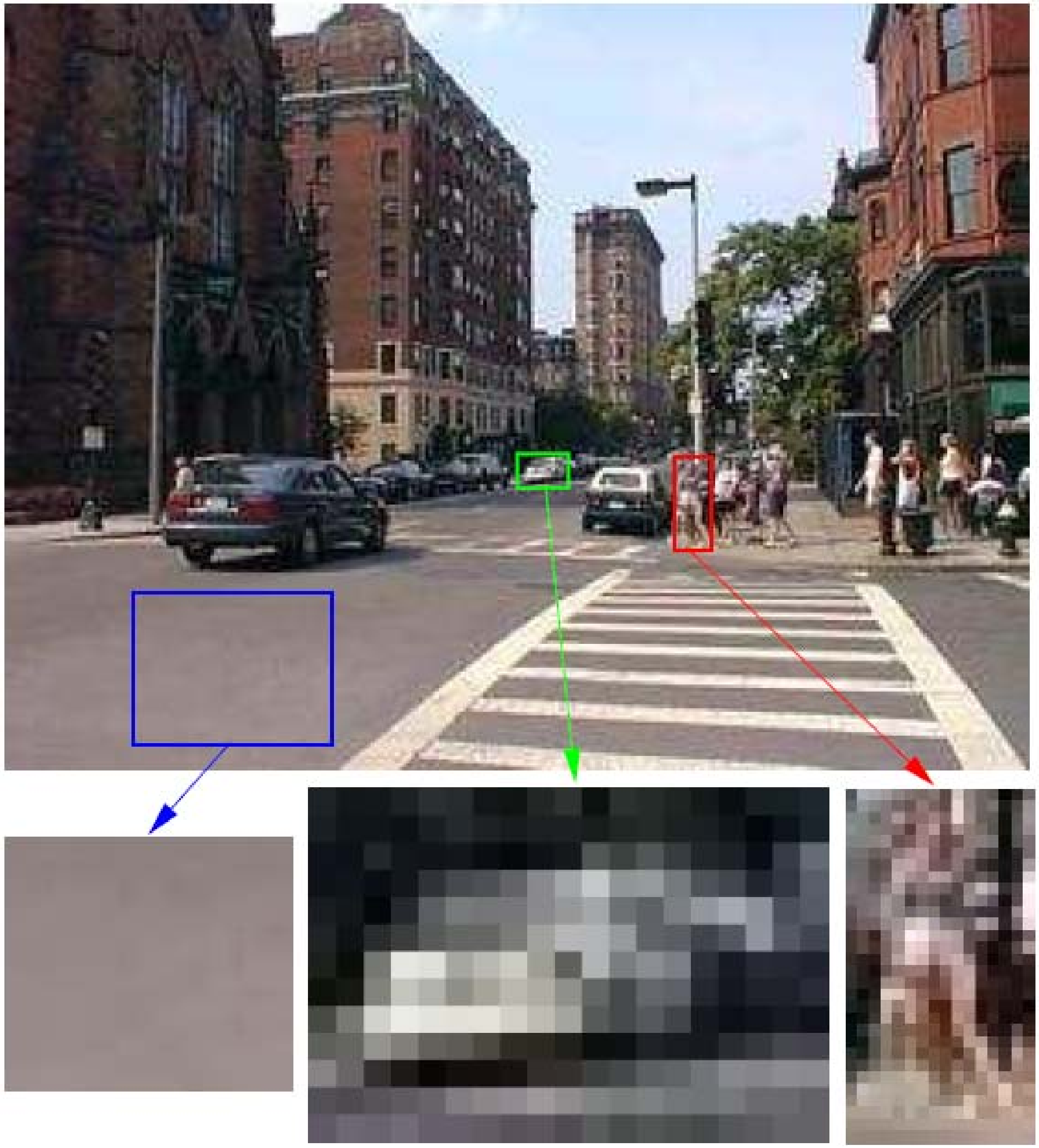
Adapted from Antonio Torralba, MIT

Context



Adapted from Derek Hoiem, CMU

Context



Adapted from
Derek Hoiem, CMU

Stages of computer vision

- Low-level

  image → image

- Mid-level

  image → features / attributes

  Image analysis / image understanding

- High-level

  features → “making sense”, recognition

Low-level

sharpening

blurring

Adapted from Linda Shapiro, U of Washington

Low-level

original image

Canny

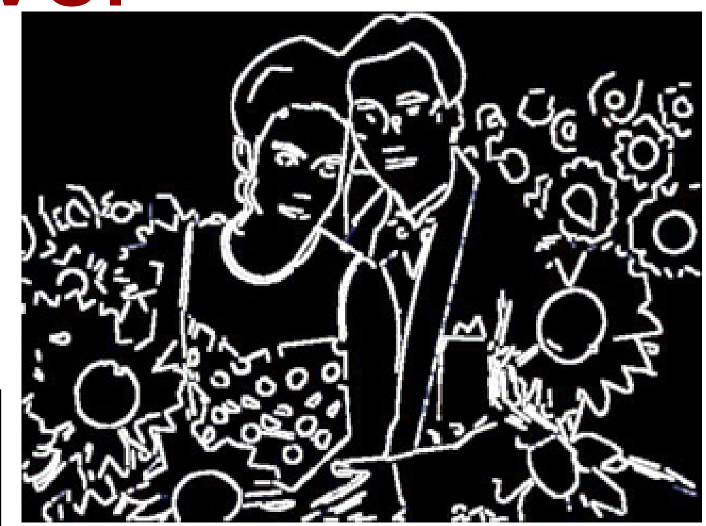
edge image

Mid-level

edge image

ORT

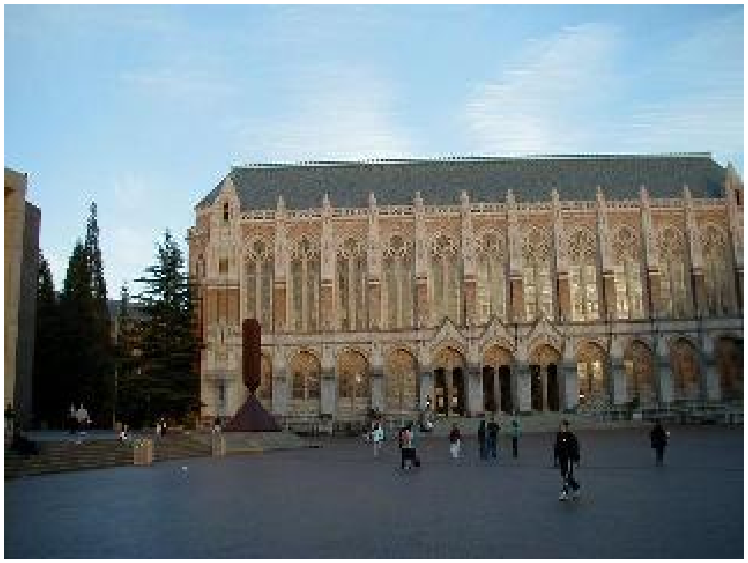
data structure



circular arcs and line segments

Adapted from Linda Shapiro, U of Washington

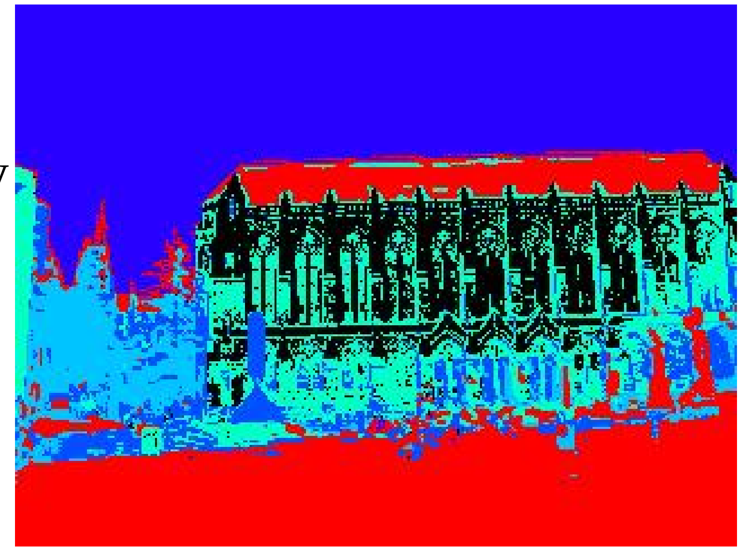
Mid-level



original color image

K-means clustering

(followed by connected component analysis)



regions of homogeneous color

data structure

Adapted from Linda Shapiro, U of Washington

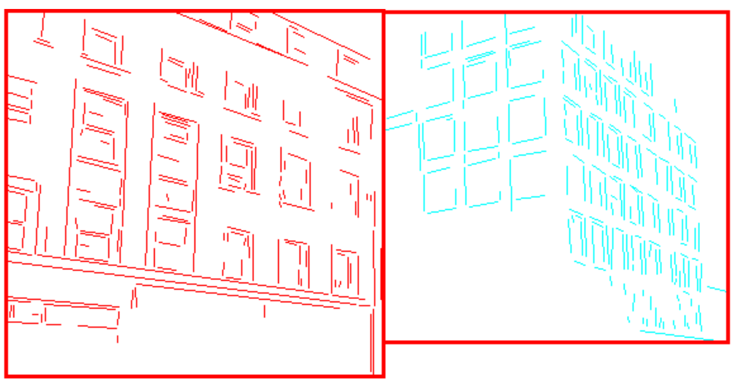
Low-level to high-level

low-level → edge image

mid-level ↓

consistent line clusters

high-level ←



Adapted from Linda Shapiro, U of Washington

Applications

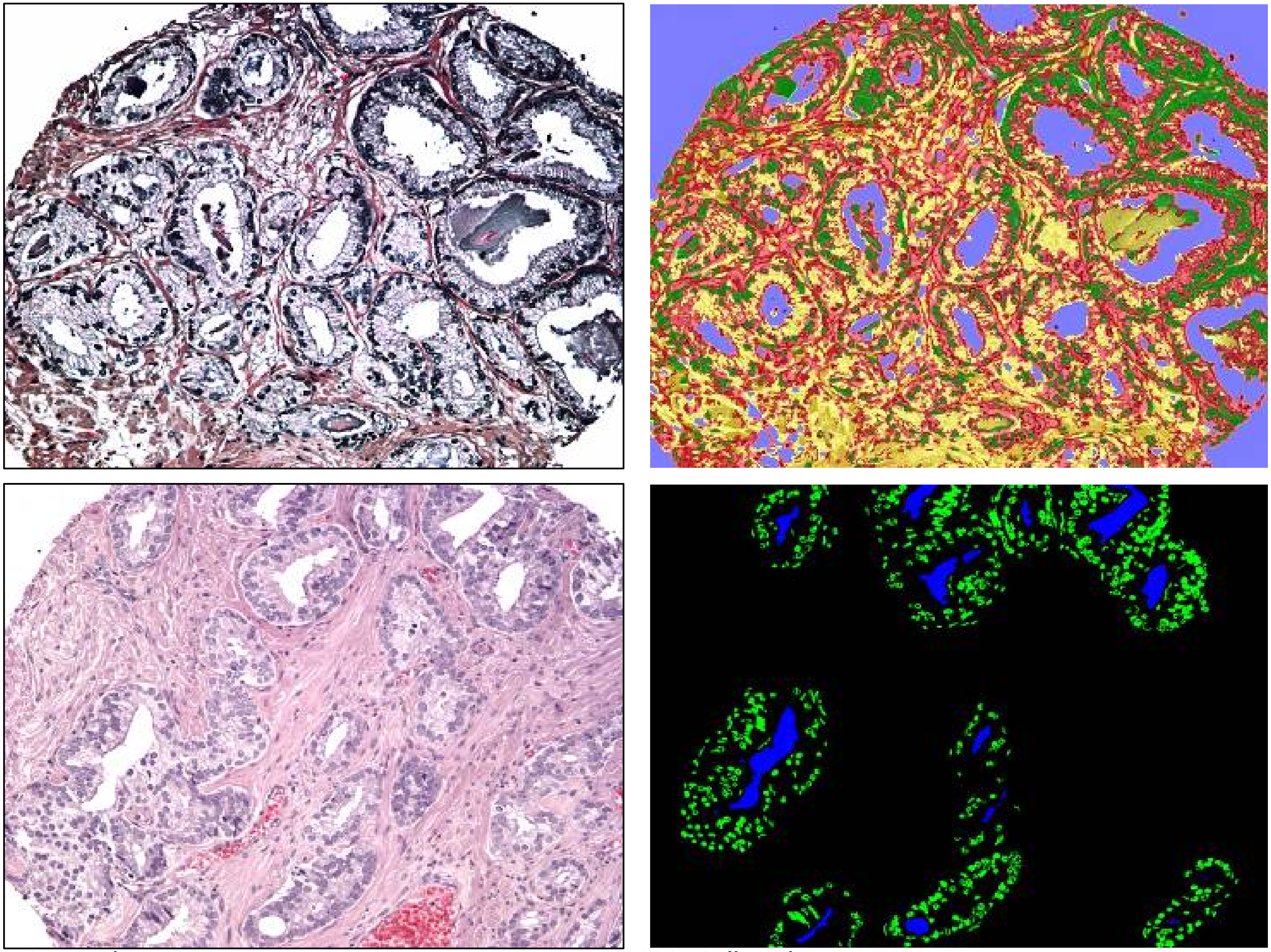
- Industrial inspection, quality control
- Medical image analysis
- Remote sensing
- Security
  - Surveillance
  - Biometrics
  - Target recognition
  - Tracking
- Robotics
- Document analysis
- Multimedia
- Assisted living
- Human-computer interfaces

Medical image analysis

CT image of abdomen

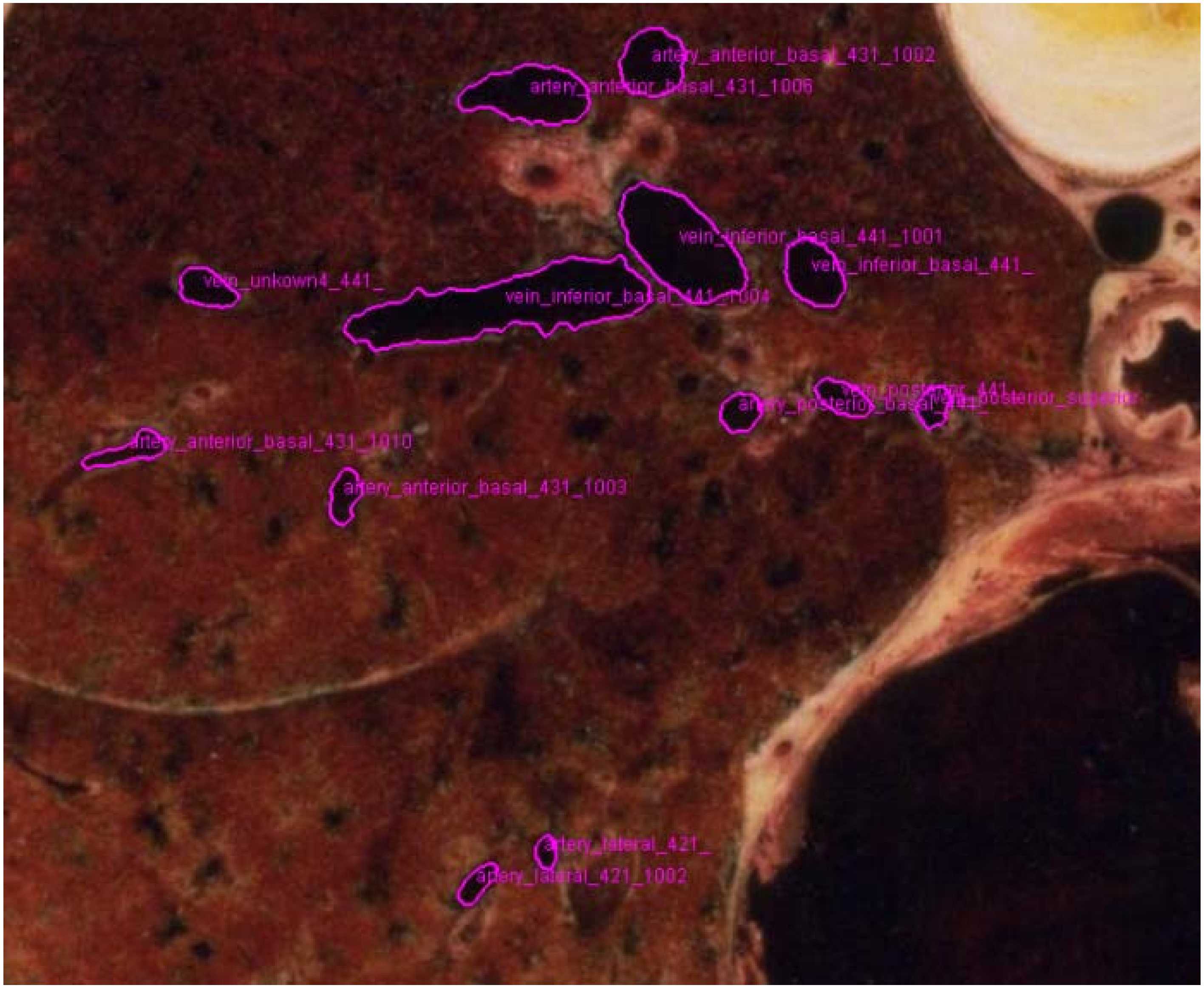
Adapted from Linda Shapiro, U of Washington

Medical image analysis



Cancer detection and grading

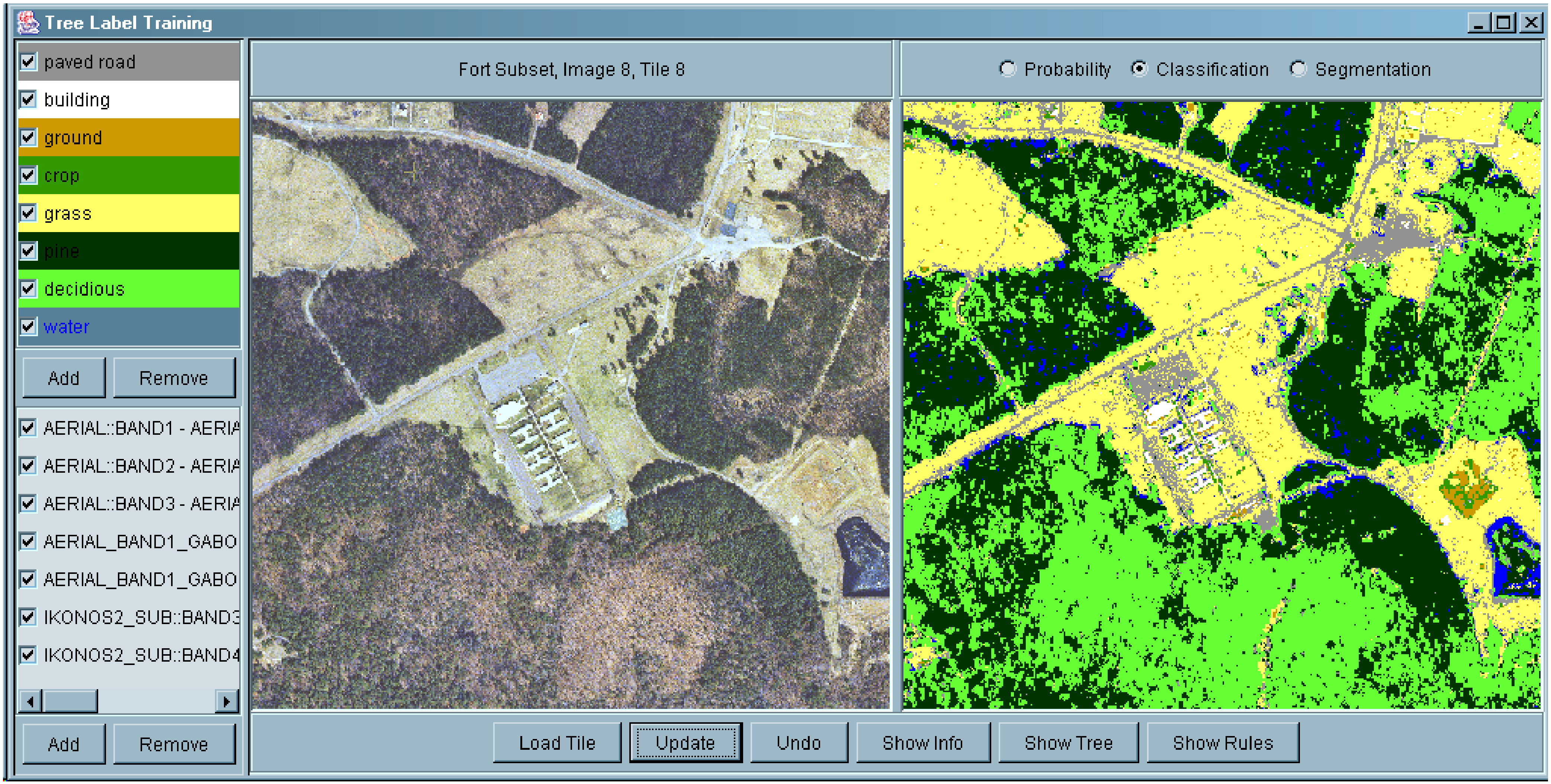
Medical image analysis



Slice of lung

Adapted from Linda Shapiro, U of Washington

Land cover analysis

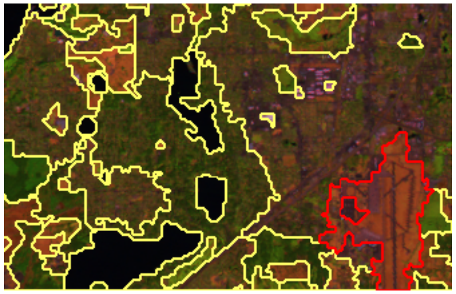


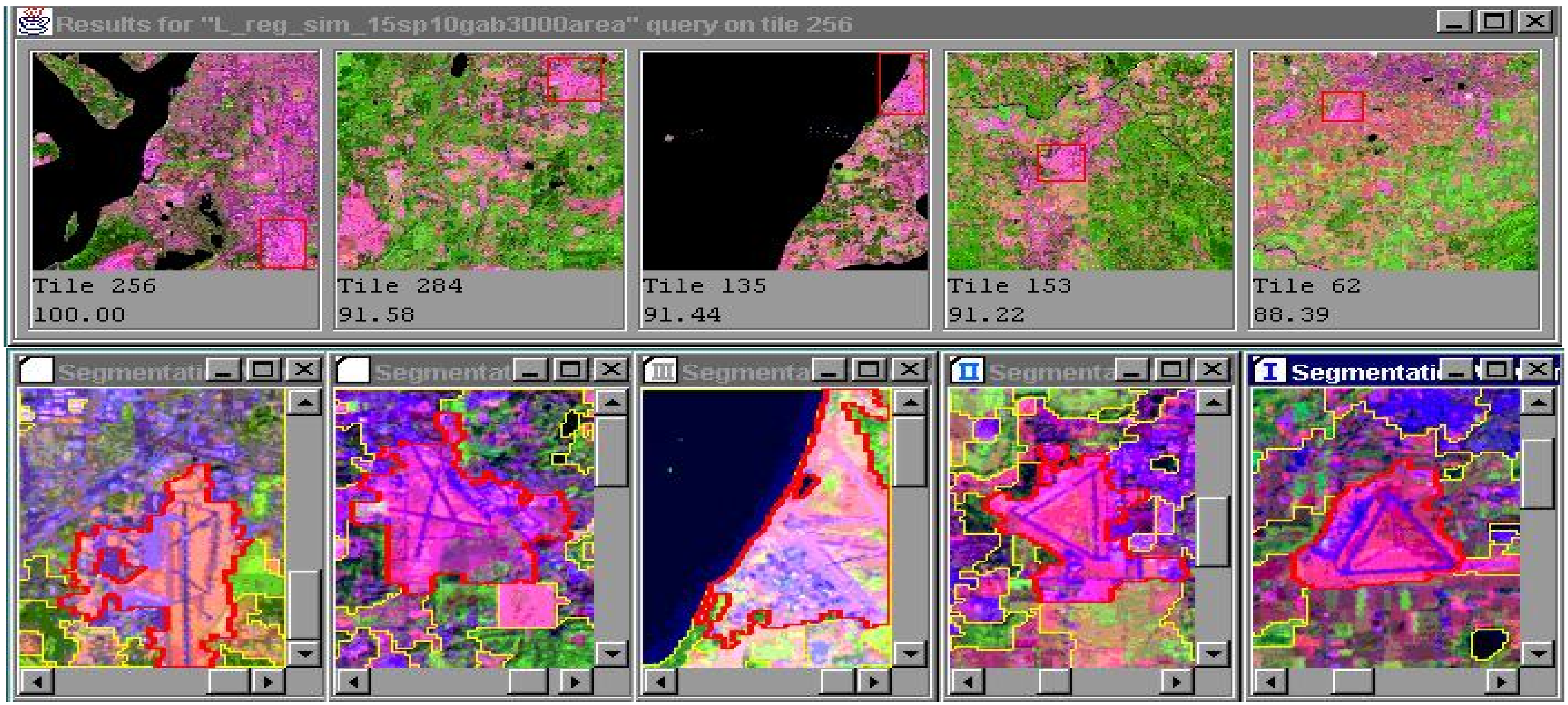
Object recognition

Recognition of buildings and building groups

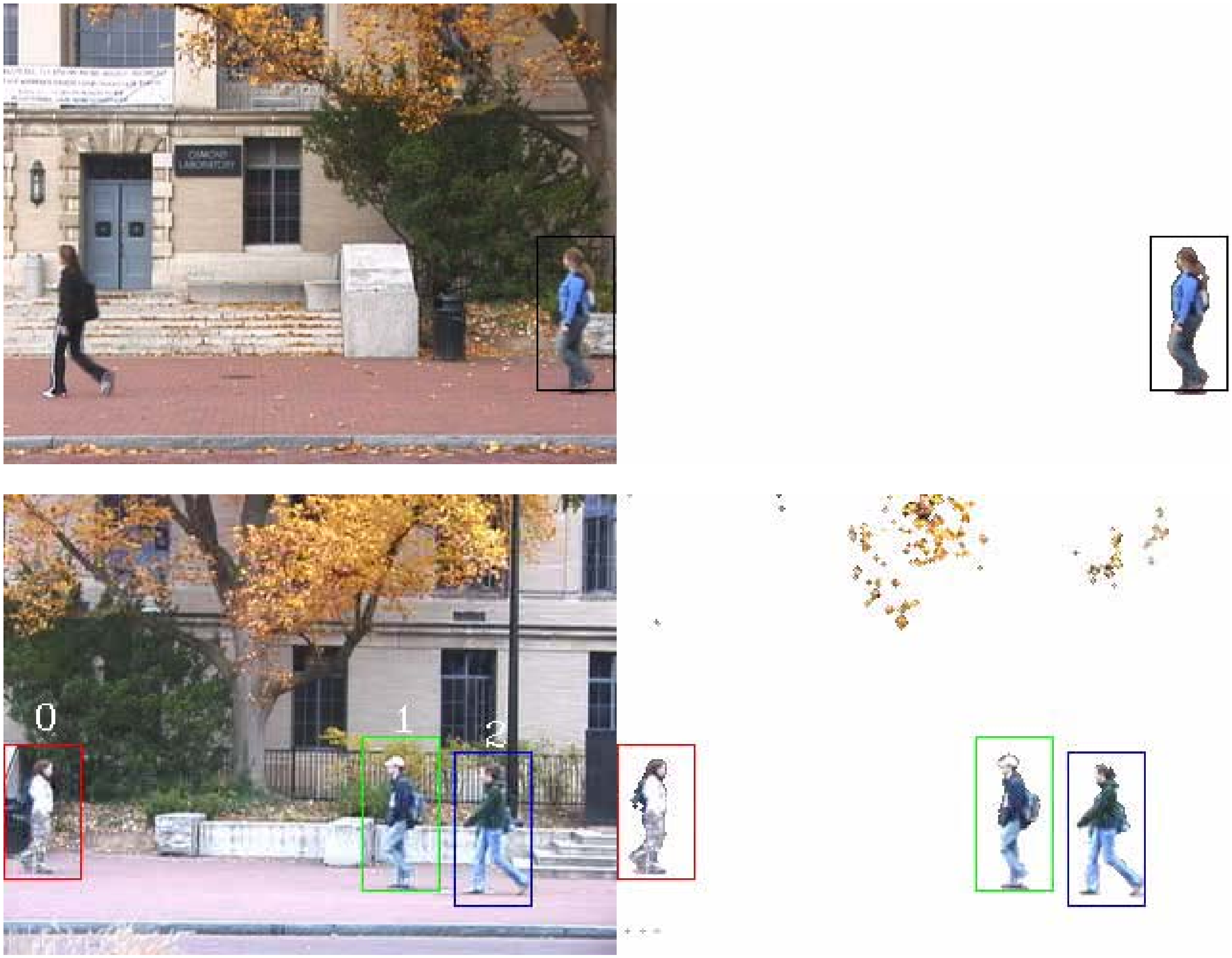
Content-based retrieval





Finding similar regions: airports

Tracking



Adapted from Octavia Camps, Penn State

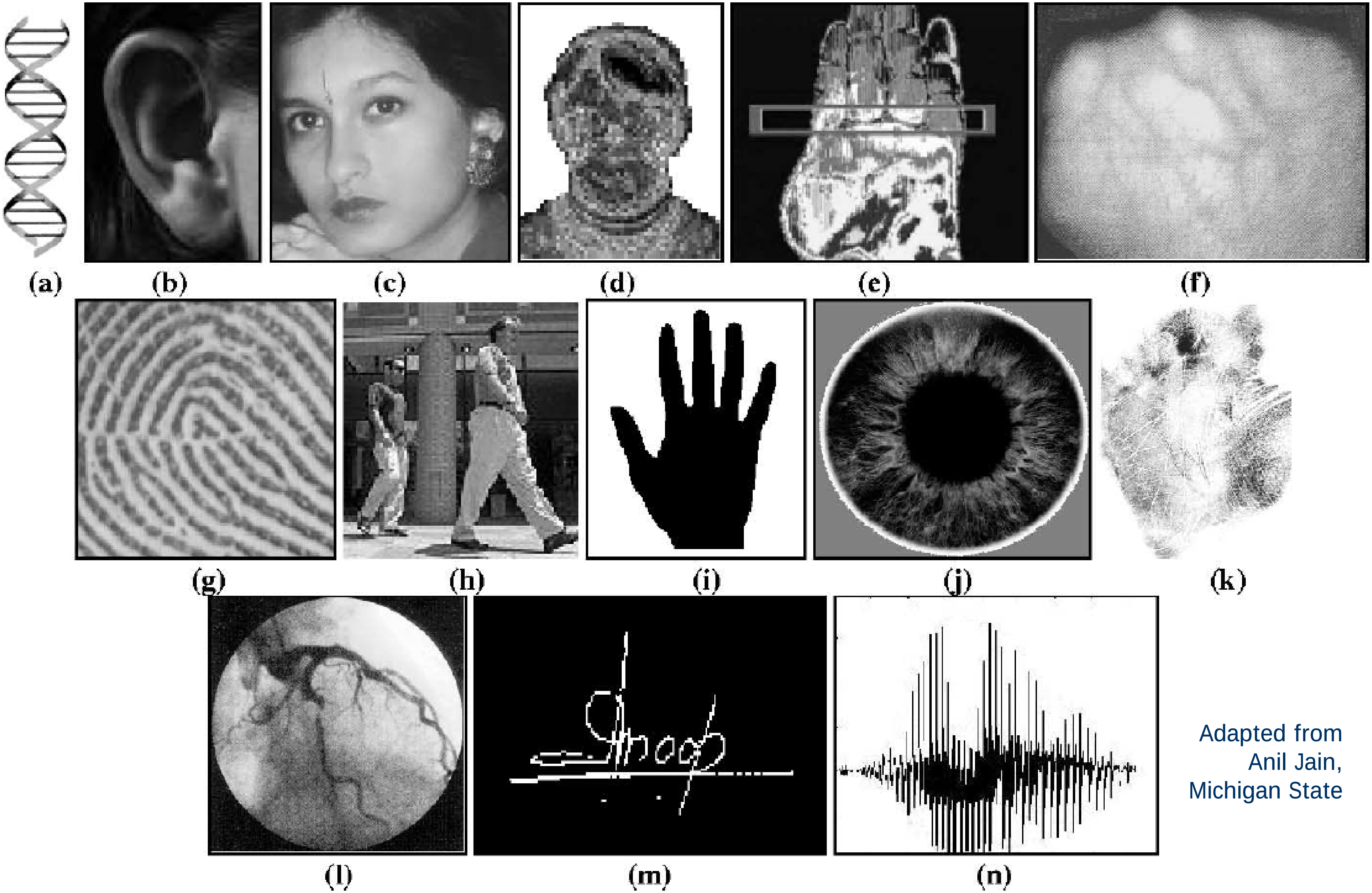
Tracking

Adapted from Martial Hebert, CMU

Tracking

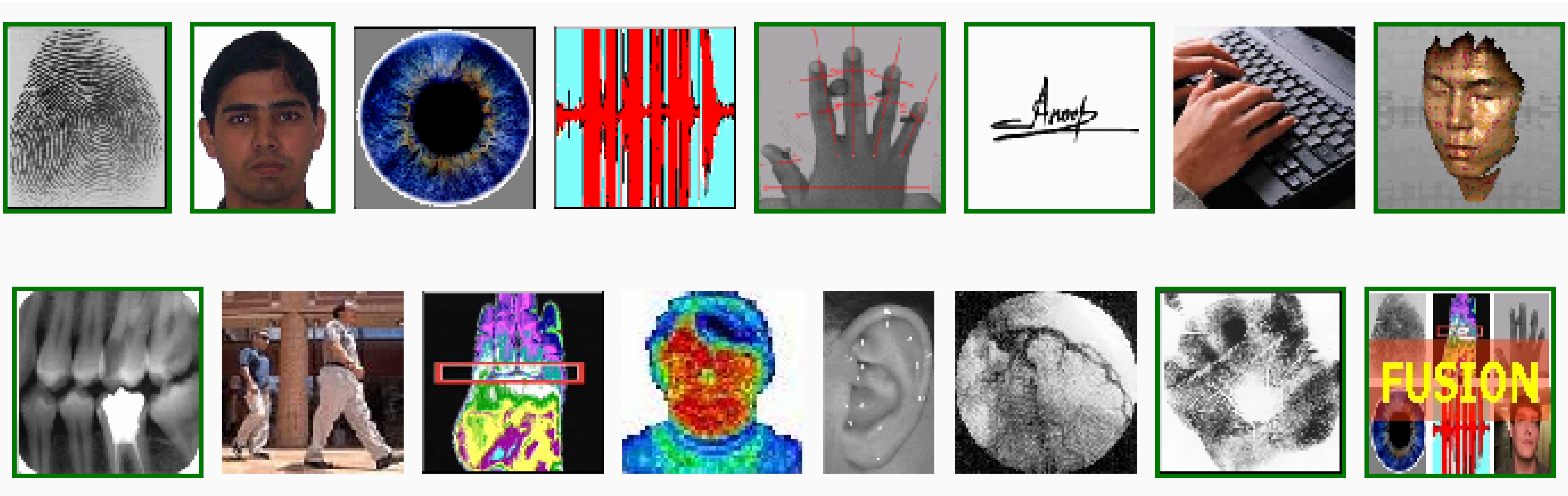
Adapted from Martial Hebert, CMU

Biometrics



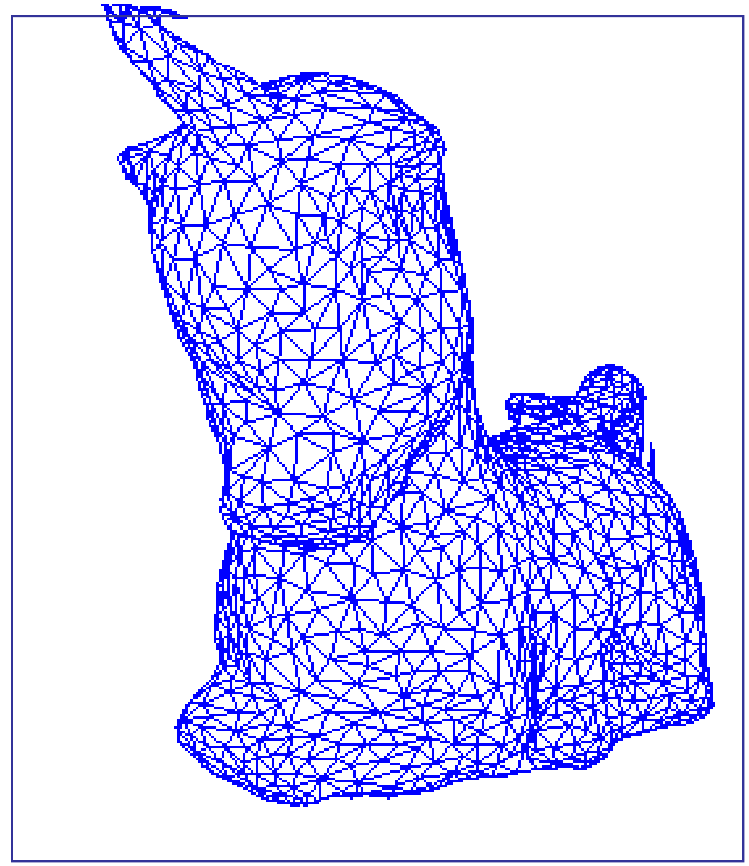
Adapted from Anil Jain, Michigan State

Biometrics



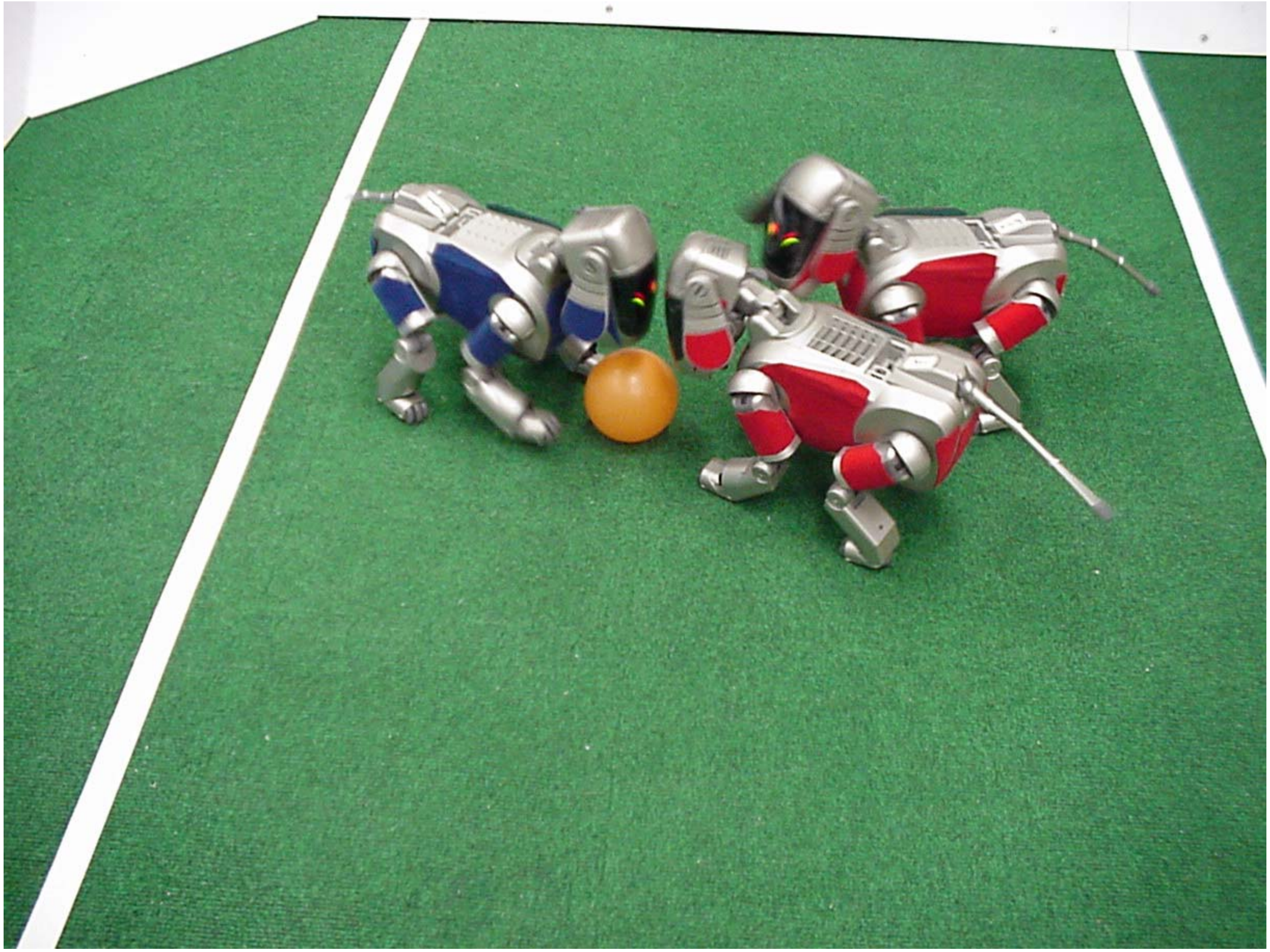
Adapted from Anil Jain, Michigan State

Robotics



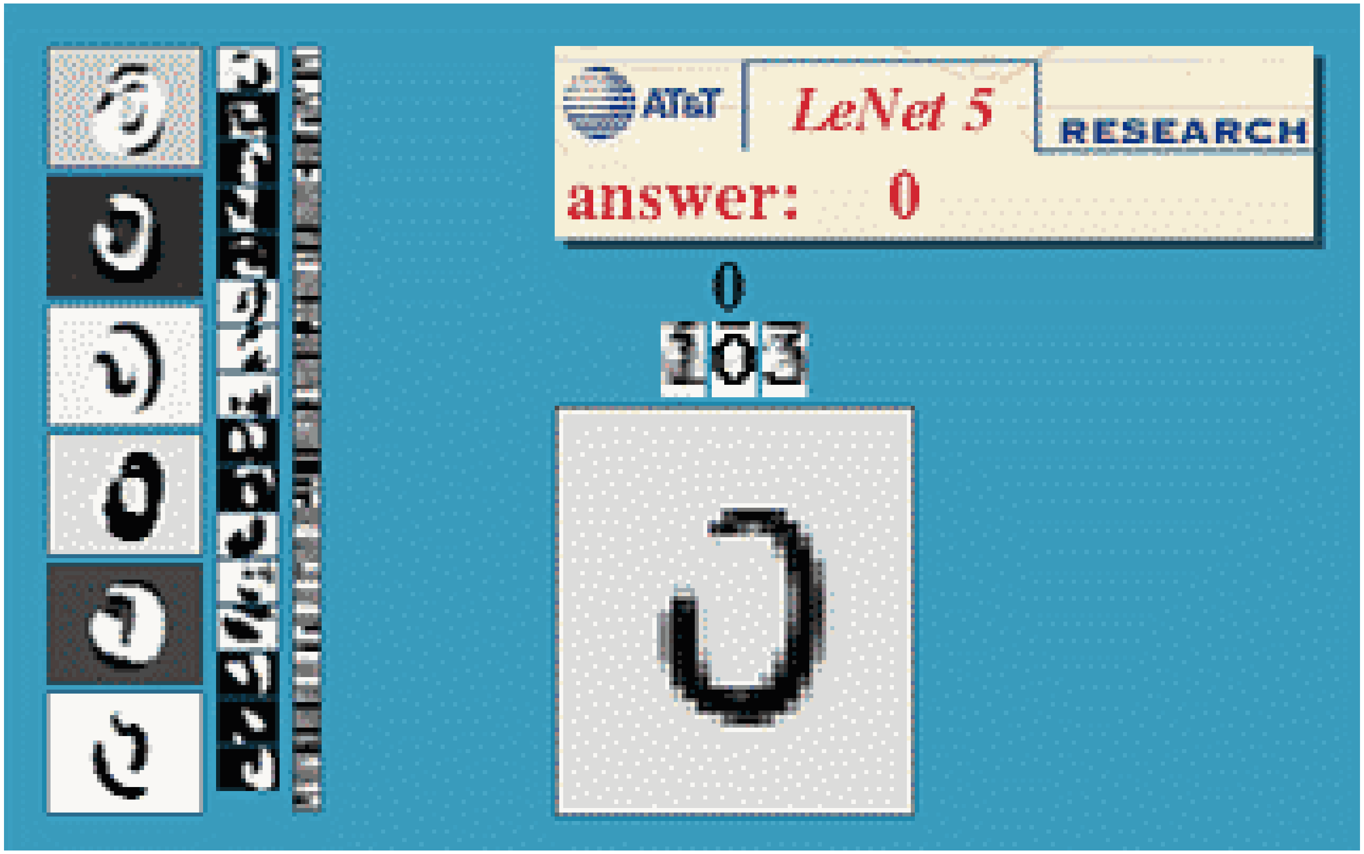
Adapted from Linda Shapiro, U of Washington

Robotics



Adapted from Steven Seitz, U of Washington

Document analysis

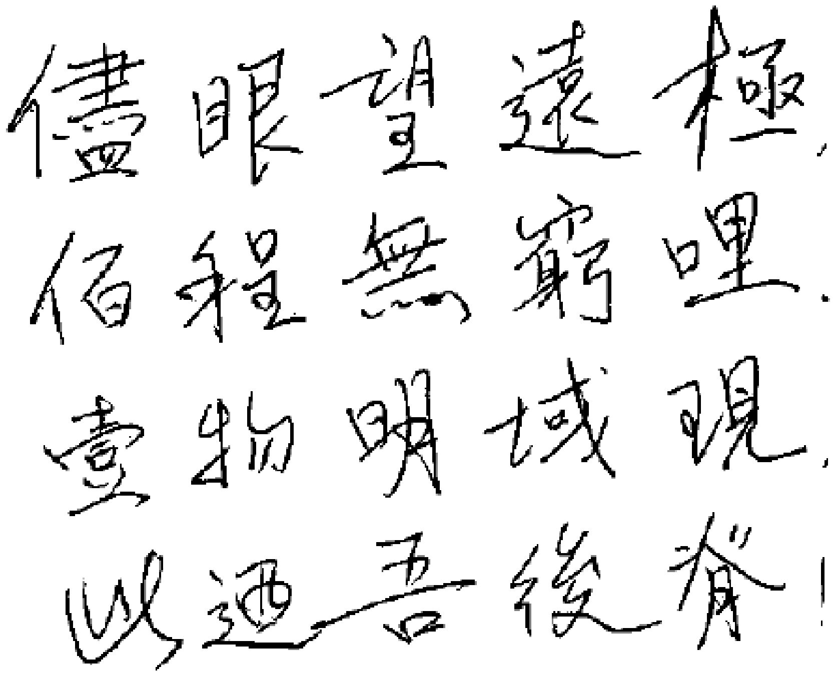


Digit recognition, AT&T labs

http://www.research.att.com/~yann

Adapted from Steven Seitz, U of Washington

Document analysis

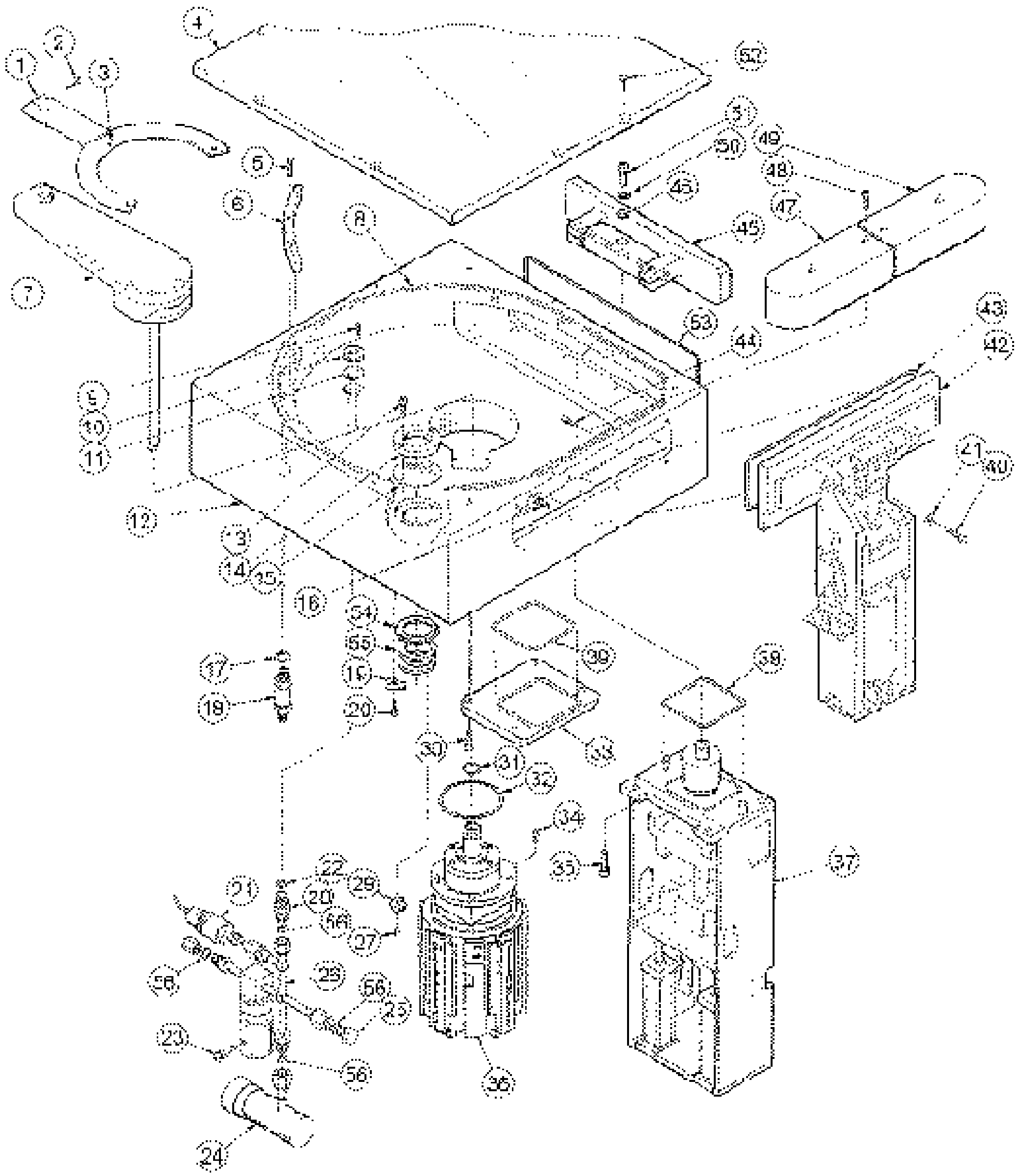


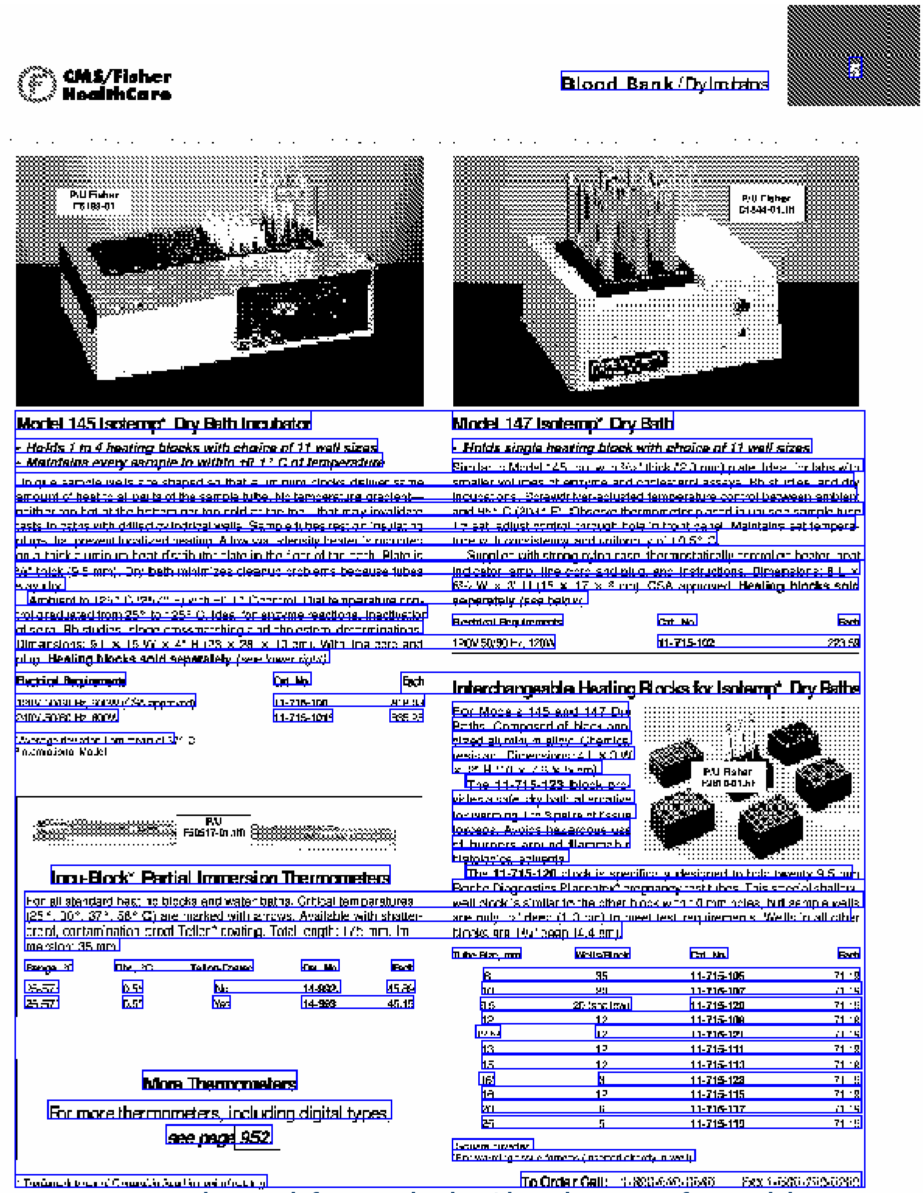
I looked as hard as I could see,
beyond 100 plus infinity
an object of bright intensity
– it was the back of me!

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

Document analysis




CMS/Fisher HealthCare

Blood Bank/Dry Incubators

Model 145 Isotemp Dry Bath Incubator

Model 147 Isotemp Dry Bath

Interchangeable Heating Blocks for Isotemp Dry Baths

Incu-Block Partial Immersion Thermometers

More Thermometers

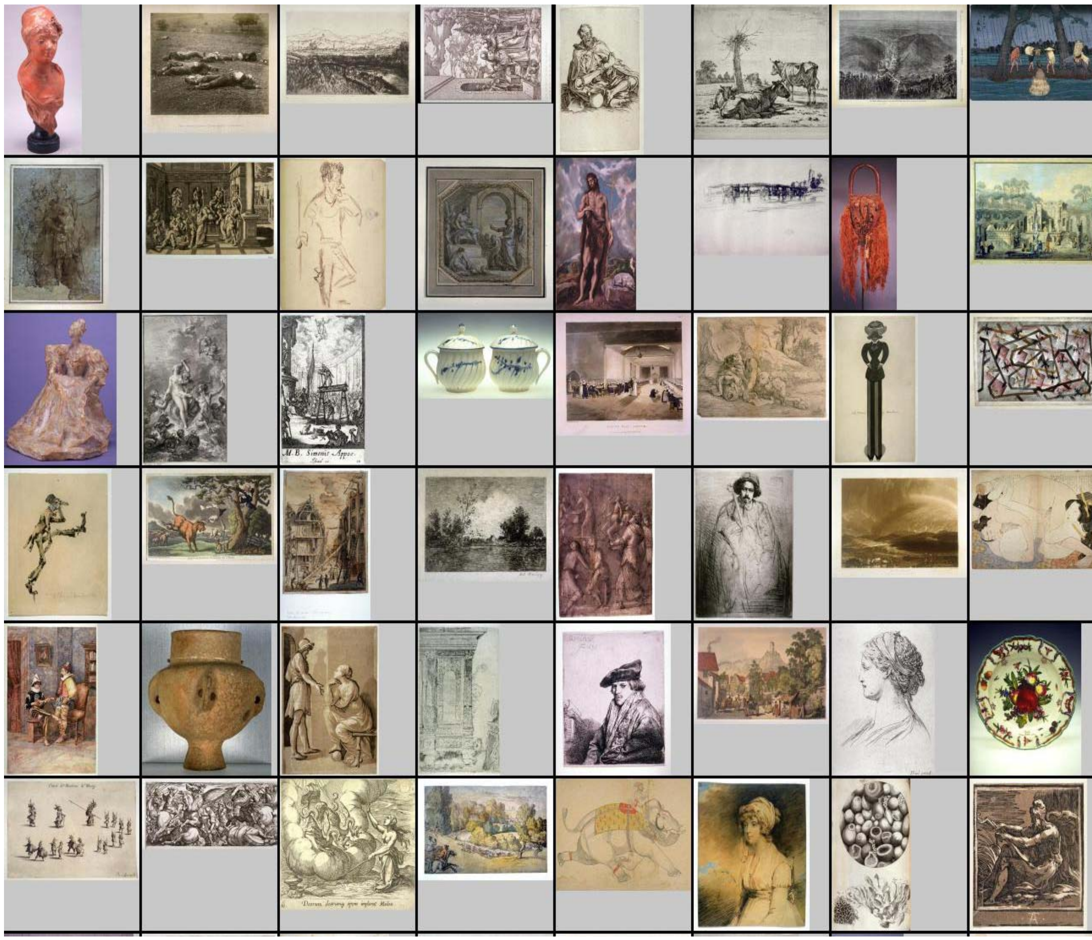
For more thermometers, including digital types, see page 952

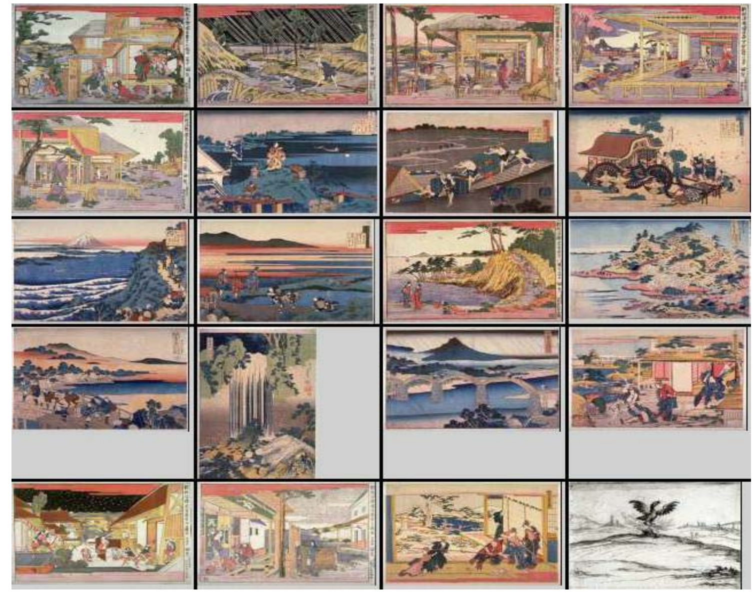
To Order Call:

Adapted from Linda Shapiro, U of Washington

Scene classification

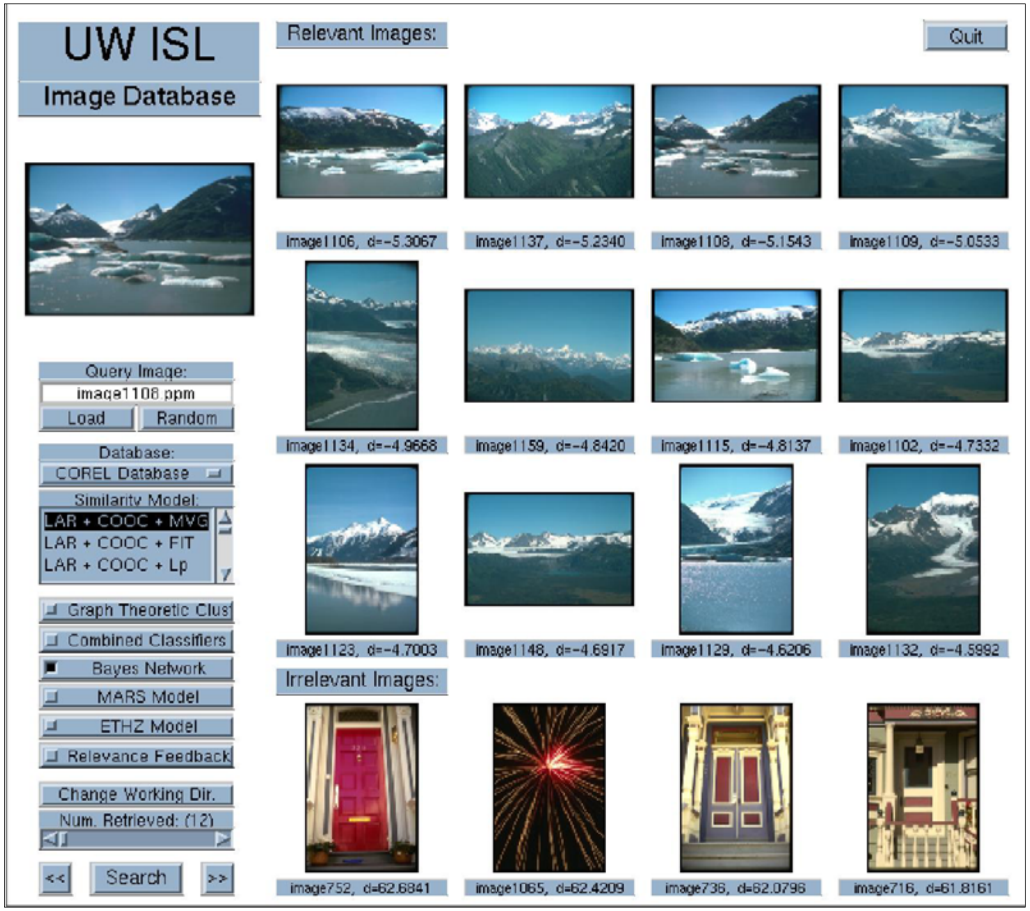
Organizing image archives

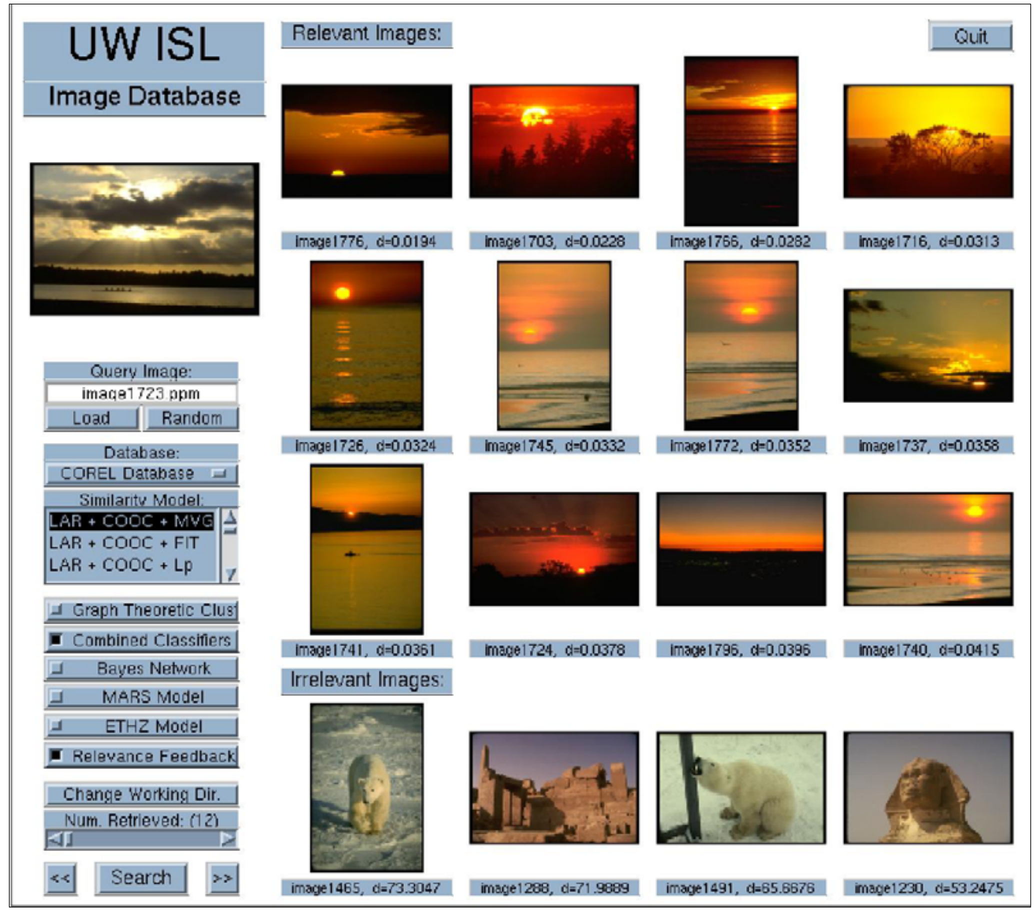




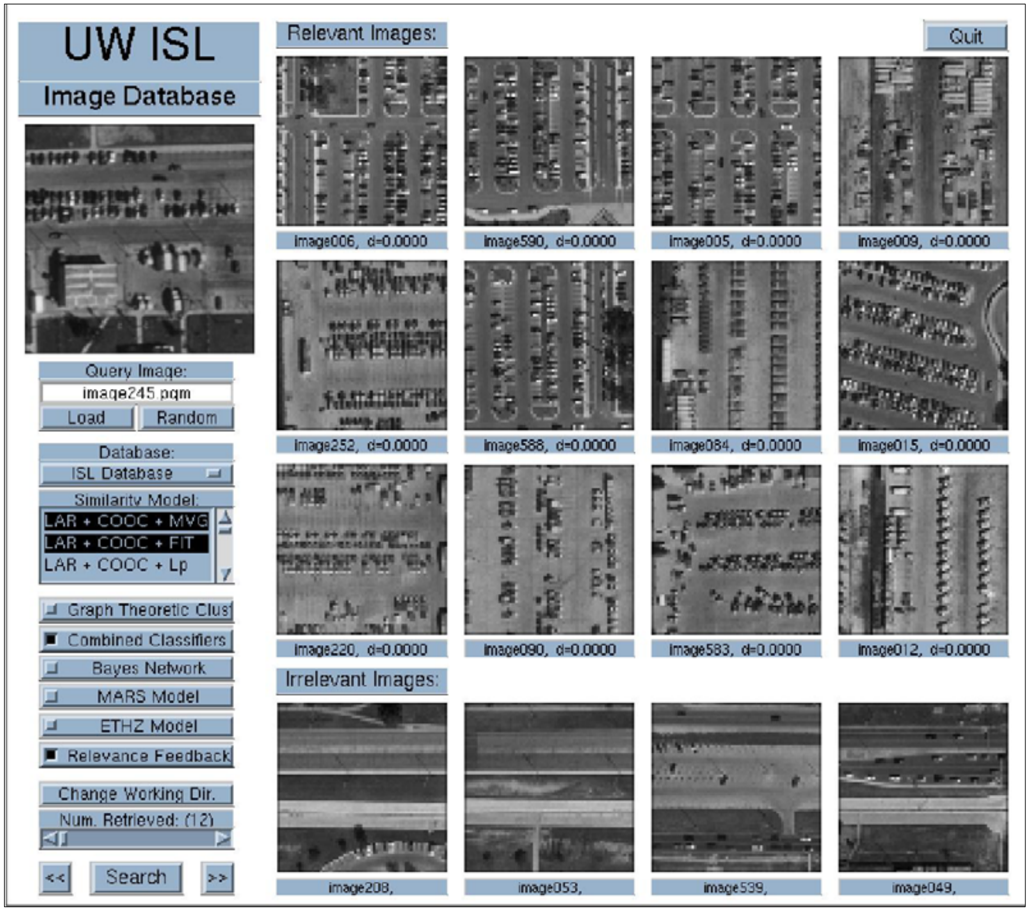
Adapted from Pinar Duygulu, Bilkent University

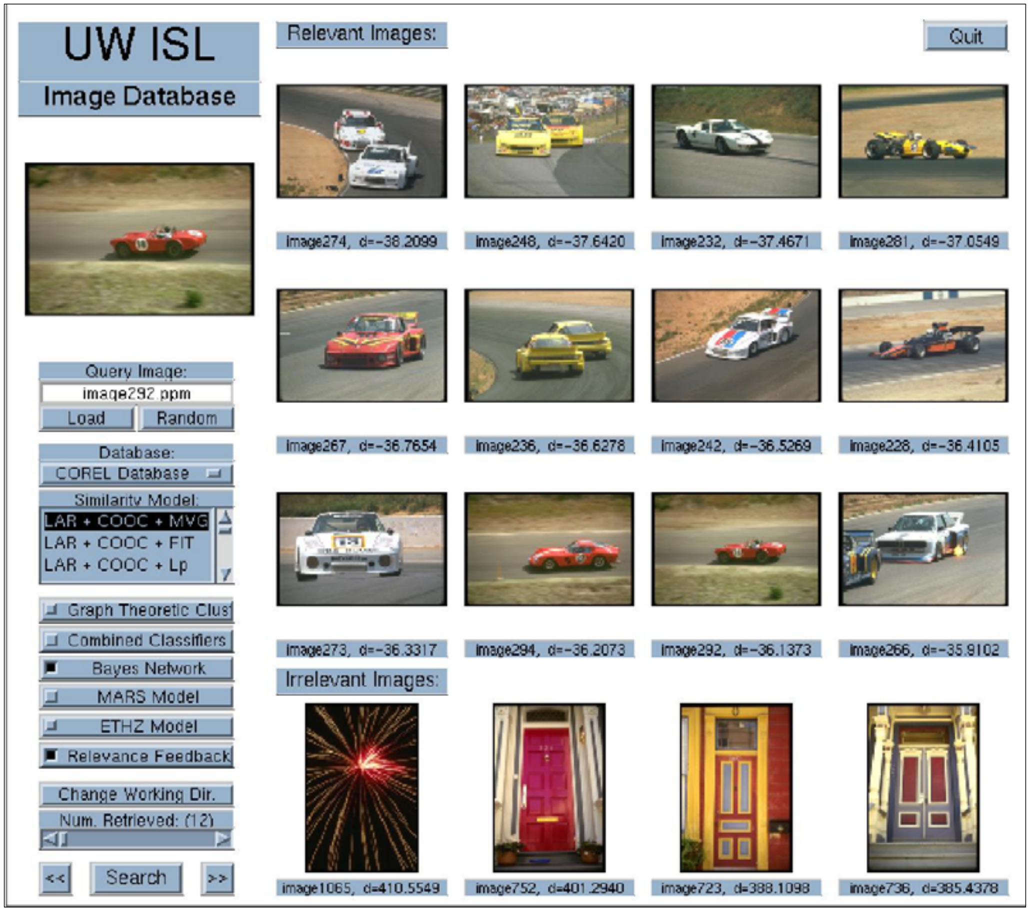
Content-based retrieval



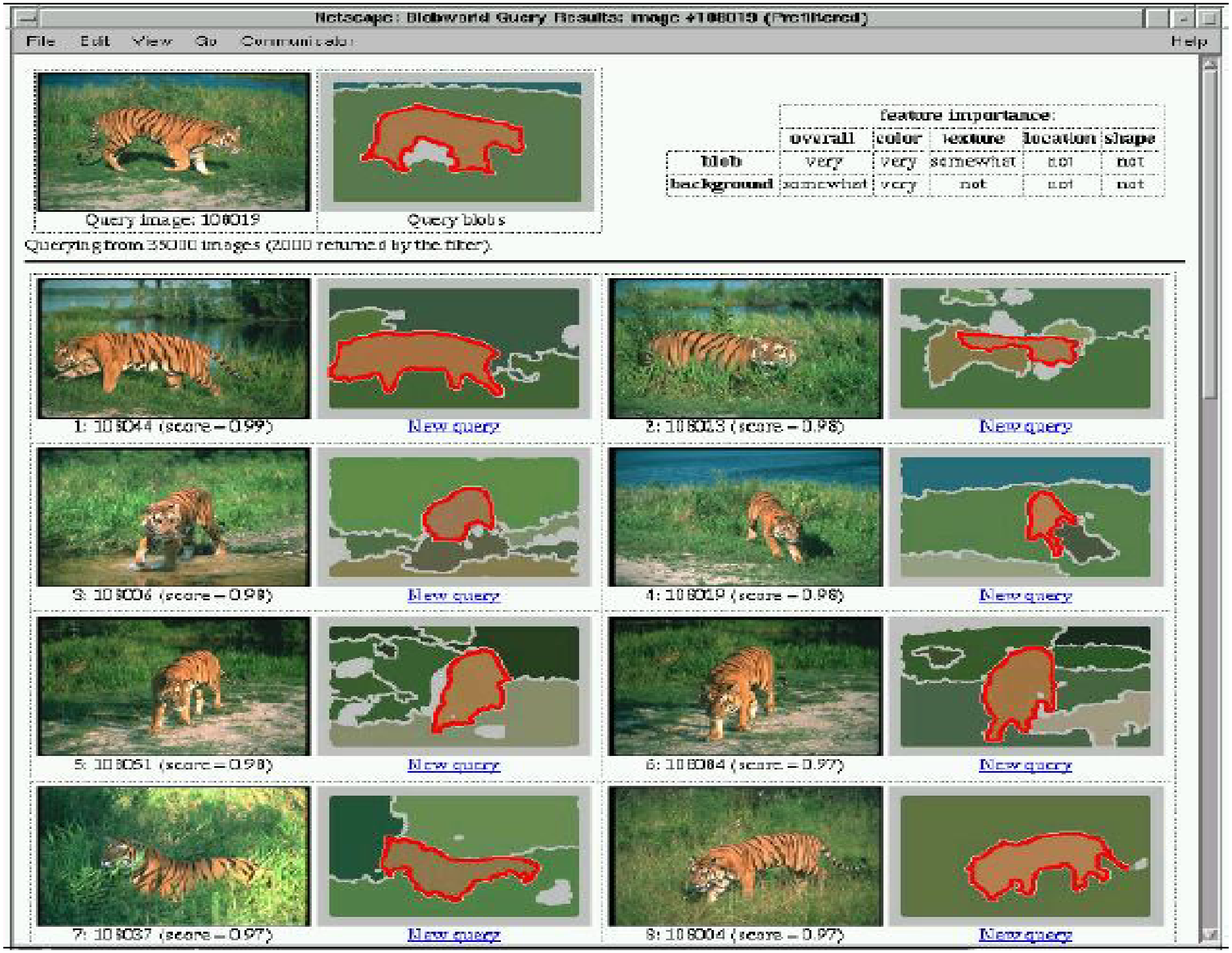


Content-based retrieval



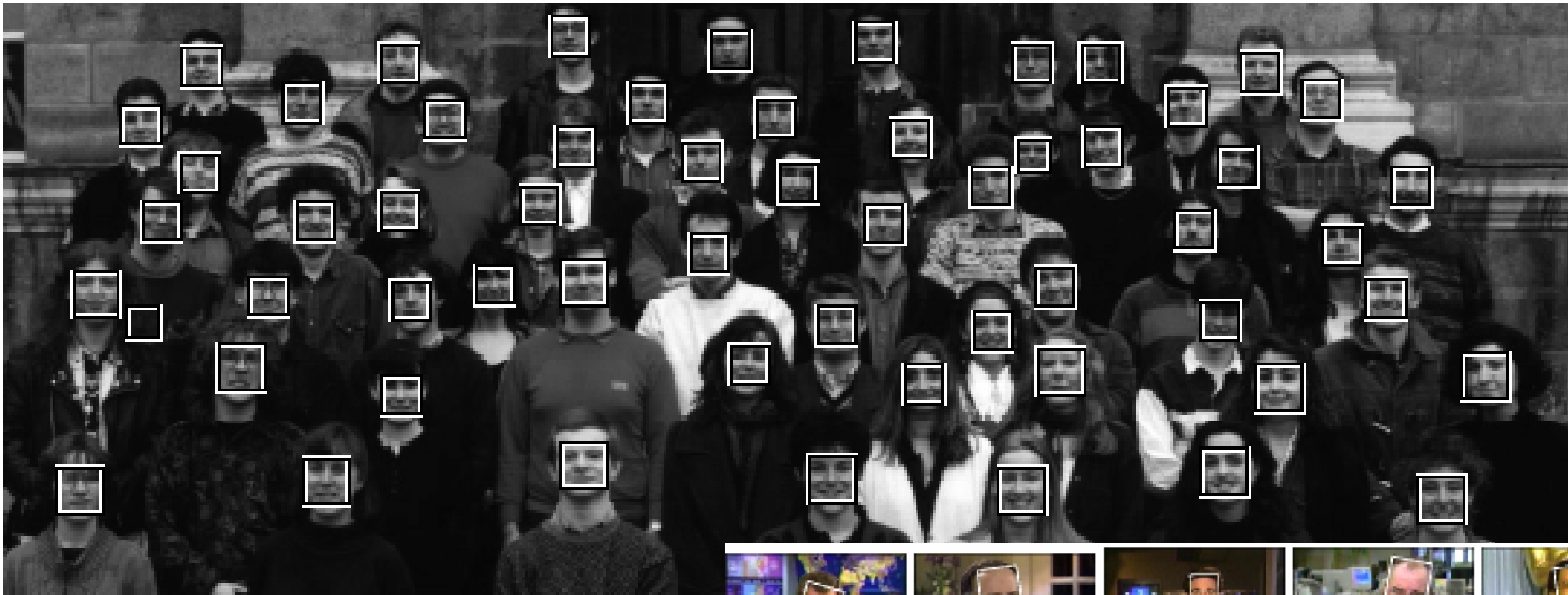


Content-based retrieval

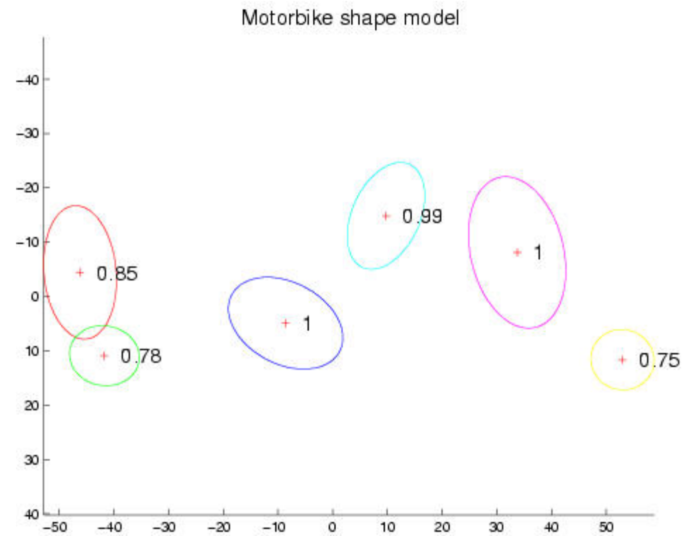


Adapted from David Forsyth, UC Berkeley

Face detection and recognition

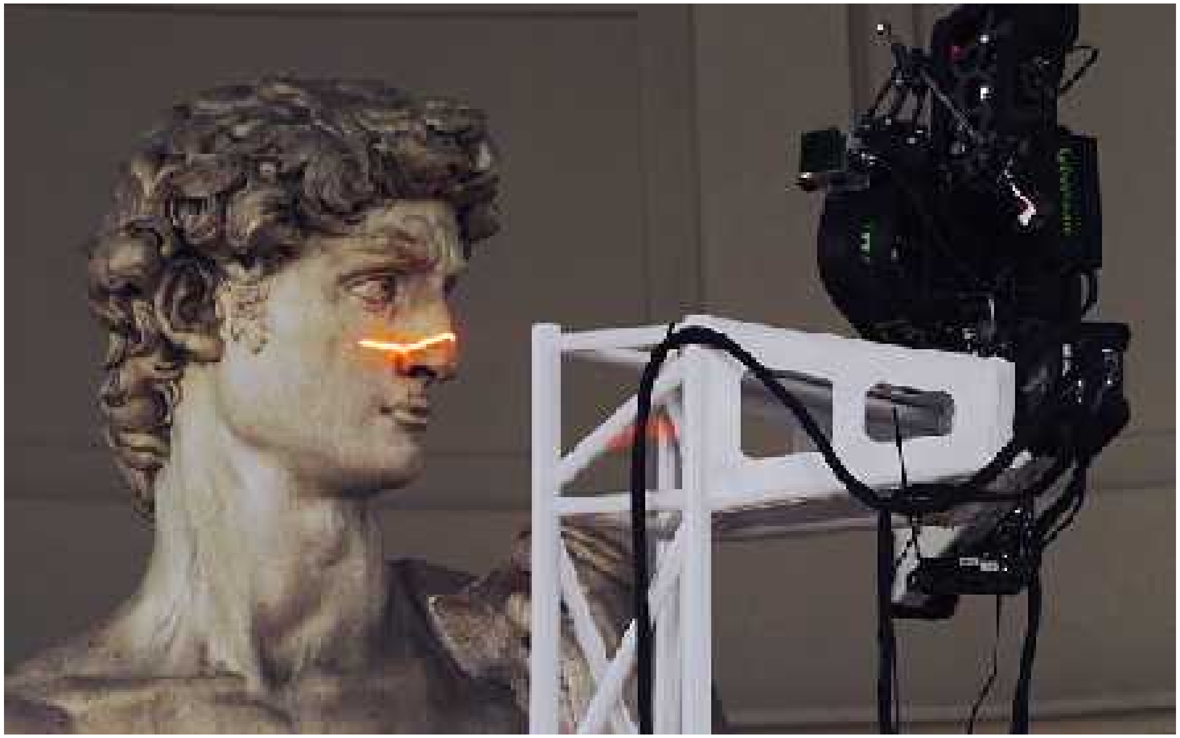


Object recognition



Adapted from Rob Fergus, MIT

3D scanning

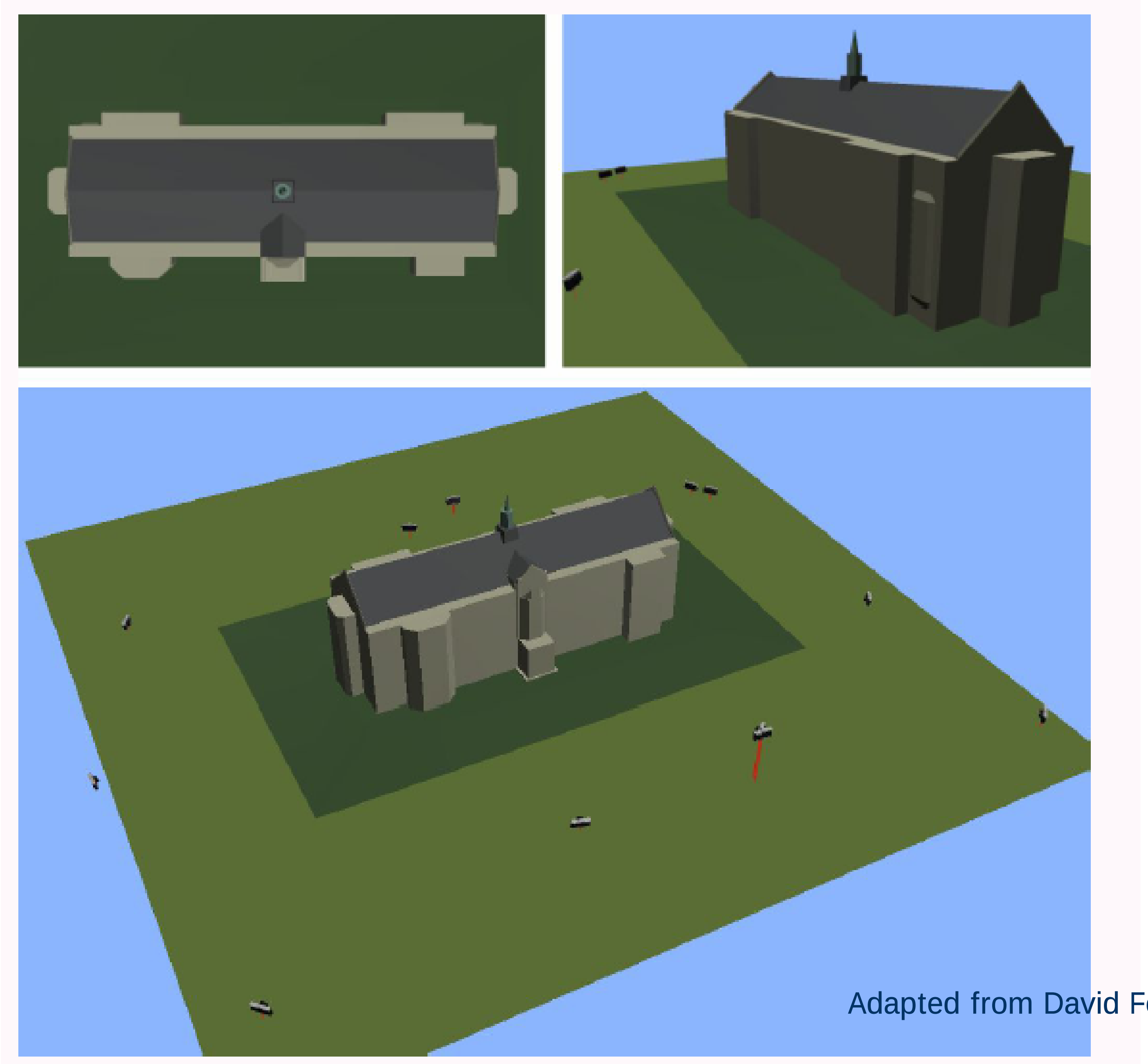


Adapted from Linda Shapiro, U of Washington

3D reconstruction

Adapted from David Forsyth, UC Berkeley

3D reconstruction



Adapted from David Forsyth, UC Berkeley

Motion capture

Adapted from Linda Shapiro, U of Washington

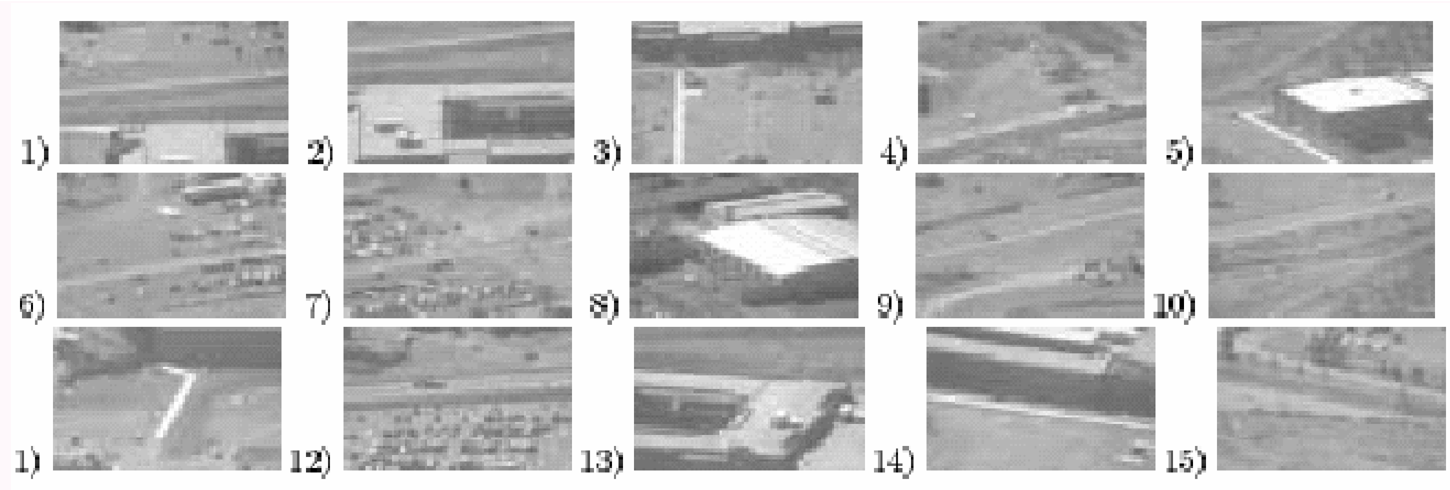
Visual effects

Adapted from Linda Shapiro, U of Washington

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?