Introduction

Introduction to Computer Vision

CSE 152

Lecture 1

What is Computer Vision?

- Trucco and Verri (Text): Computing properties of the 3-D world from one or more digital images
- Sockman 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"

Why is this hard?

What is in this image?

1. A hand holding a man?
2. A hand holding a mirrored sphere?
3. An Escher drawing?

- Interpretations are ambiguous
- The forward problem (graphics) is well-posed
- The "inverse problem" (vision) is not

We all make mistakes

"640K ought to be enough for anybody." – Bill Gates, 1981

"…" – Marvin Minsky

What do you see?

- Changing viewpoint
- Moving light source
- Deforming shape

What was happening

- Changing viewpoint
- Moving light source
- ✓ Deforming shape

Should Computer Vision follow from our understanding of Human Vision?

Yes & No

1. Who would ever be crazy enough to even try creating machine vision?
2. Human vision “works”, and copying is easier than creating.
3. Secondary benefit – in trying to mimic human vision, we learn about it.

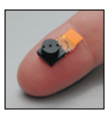
1. Why limit oneself to human vision when there is even greater diversity in biological vision
2. Why limit oneself to biological when there may be greater diversity in sensing mechanism?
3. Biological vision systems evolved to provide functions for “specific” tasks and “specific” environments. These may differ for machine systems
4. Implementation – hardware is different, and synthetic vision systems may use different techniques/methodologies that are more appropriate to computational mechanisms

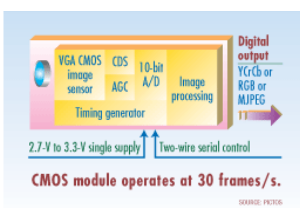
Why study Computer Vision?

- Images and movies are everywhere
- Fast-growing collection of useful applications
  - building representations of the 3D world from pictures
  - automated surveillance (who’s doing what)
  - Hollywood special effects
  - face recognition
- Various deep and attractive scientific mysteries
  - how does object recognition work?
  - Beautiful marriage of math, biology, physics, engineering
- Greater understanding of human vision

The Near Future: Ubiquitous Vision

- Five years from now, digital cameras will cost 1 cent (sensor cost).
- Digital video will be a widely available commodity component embedded in cell phones, PDA’s, doorbells, bridges, security systems, cars, etc.
- 99.9% of digitized video won’t be seen by a person.
- That doesn’t mean that only 0.1% is important!





Applications: touching your life

- Football
- Movies
- Surveillance
- HCI – hand gestures, American Sign Language
- Face recognition & Biometrics
- Road monitoring
- Industrial inspection
- Robotic control
- Autonomous driving
- Space: planetary exploration, docking
- Medicine – pathology, surgery, diagnosis
- Microscopy
- Military
- Remote Sensing

Related Fields

- Image Processing
- Computer Graphics
- Pattern Recognition
- Perception
- Robotics
- AI

Image Interpretation - Cues

- Variation in appearance in multiple views
  - stereo
  - motion
- Shading & highlights
- Shadows
- Contours
- Texture
- Blur
- Geometric constraints
- Prior knowledge

Shading and lighting

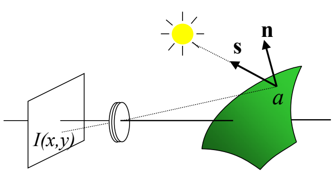
Shading as a result of differences in lighting is

1. A source of information
2. An annoyance

Illumination Variability

"The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity."
-- Moses, Adini, Ullman, ECCV '94

Image Formation



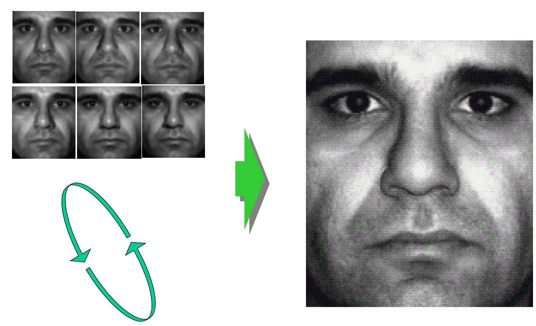
At image location *(x,y)* the intensity of a pixel *I(x,y)* is

$$I(x,y) = a(x,y)\, \mathbf{n}(x,y) \bullet \mathbf{s}$$

where

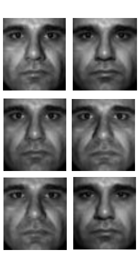
- *a(x,y)* is the albedo of the surface projecting to *(x,y)*.
- $\mathbf{n}(x,y)$ is the unit surface normal.
- $\mathbf{s}$ is the direction and strength of the light source.

Lighting variation



Single Light Source

Shading reveals shape



Basic idea: 3 or more images under slightly different lighting

The course

- Part 1: The Physics of Imaging
- Part 2: Early Vision (Segmentation)
- Part 3: Reconstruction (Shape-from-X)
- Part 4: Recognition

Part I of Course: The Physics of Imaging

- How images are formed
  - Cameras
    - What a camera does
    - How to tell where the camera was located
  - Light
    - How to measure light
    - What light does at surfaces
    - How the brightness values we see in cameras are determined
  - Color
    - The underlying mechanisms of color
    - How to describe it and measure it

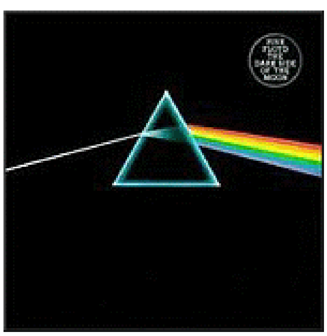
Cameras, lenses, and sensors

Figure 1.16 The first photograph on record, *la table servie*, obtained by Nicéphore Niepce in 1822. *Collection Harlinge–Viollet.*

- •Pinhole cameras
- •Lenses
- •Projection models
- •Geometric camera parameters

From Computer Vision, Forsyth and Ponce, Prentice-Hall, 2002.

Color



Part II: Early Vision in One Image

- Representing small patches of image
  - For three reasons
    - Sharp changes are important in practice --- known as "edges"
    - Representing texture by giving some statistics of the different kinds of small patch present in the texture.
      - Tigers have lots of bars, few spots
      - Leopards are the other way
    - We wish to establish correspondence between (say) points in different images, so we need to describe the neighborhood of the points

Segmentation

- Which image components "belong together"?
- Belong together ≅ lie on the same object
- Cues
  - similar color
  - similar texture
  - not separated by contour
  - form a suggestive shape when assembled

Texture Patterns
[Leung, Malik]

- Regular texture pattern, repeated texture elements
- Segment image based on texture
- Surface shape from texture pattern

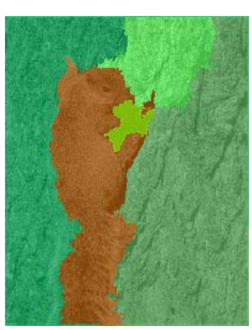
Boundary Detection

http://www.robots.ox.ac.uk/~vdg/dynamics.html

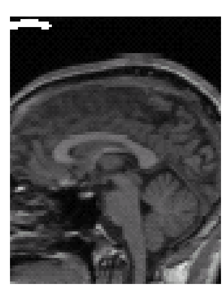
Boundary Detection: Local cues

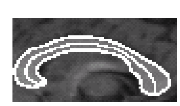
Gradients

(Sharon, Balun, Brandt, Basri)



Boundary Detection





Finding the Corpus Callosum

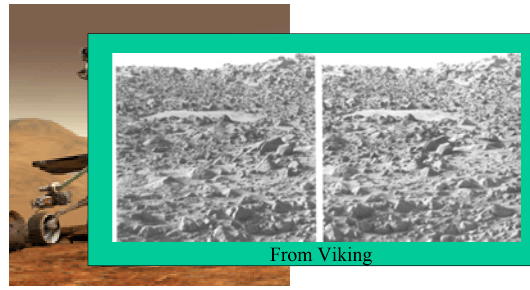
(G. Hamarneh, T. McInerney, D. Terzopoulos)

Part 3: Reconstruction from Multiple Images

- Photometric Stereo
  - What we know about the world from lighting changes.
- The geometry of multiple views
- Stereopsis
  - What we know about the world from having 2 eyes
- Structure from motion
  - What we know about the world from having many eyes
    - or, more commonly, our eyes moving.

Mars Rover

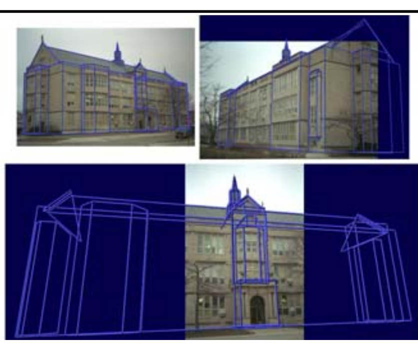
Spirit



Façade (Debevec, Taylor and Malik, 1996)
Reconstruction from multiple views, constraints, rendering

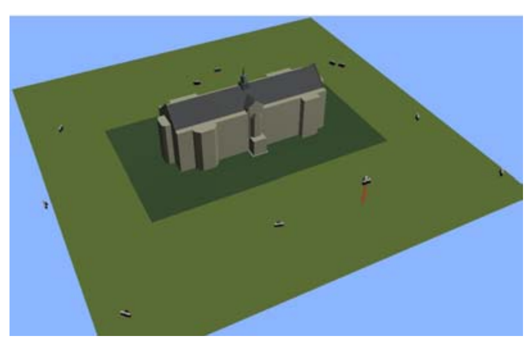
Architectural modeling:
- photogrammetry;
- view-dependent texture mapping;
- model-based stereopsis.

Images with marked features



Recovered model edges reprojected through recovered camera positions into the three original images

Resulting model & Camera Positions



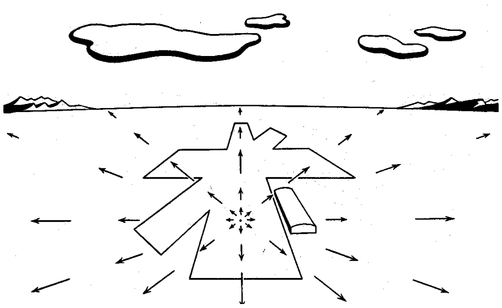
Façade

- The Camponile Movie

Video-Motion Analysis

- Where "things" are moving in image – segmentation.
- Determining observer motion (egomotion)
- Determining scene structure
- Tracking objects
- Understanding activities & actions

Forward Translation & Focus of Expansion [Gibson, 1950]



Tracking

- Use a model to predict next position and refine using next image
- Model:
  - simple dynamic models (second order dynamics)
  - kinematic models
  - etc.
- Face tracking and eye tracking now work rather well

Visual Tracking

Main Challenges

1. 3-D Pose Variation
2. Occlusion of the target
3. Illumination variation
4. Camera jitter
5. Expression variation

etc.

[Ho, Lee, Kriegman]

Part 4: Recognition

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

Recognition Challenges

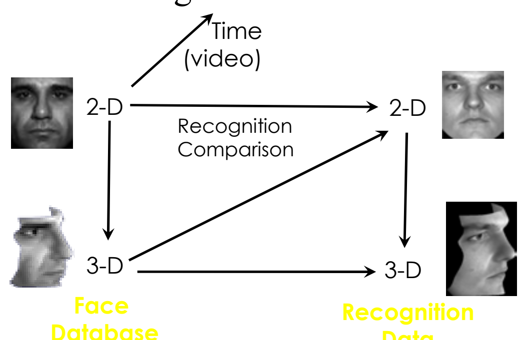
- Within-class variability
  - Different objects within the class have different shapes or different material characteristics
  - Deformable
  - Articulated
  - Compositional
- Pose variability:
  - 2-D Image transformation (translation, rotation, scale)
  - 3-D Pose Variability (perspective, orthographic projection)
- Lighting
  - Direction (multiple sources & type)
  - Color
  - Shadows
- Occlusion – partial
- Clutter in background -> false positives

Face Detection: First Step

Why is Face Recognition Hard?

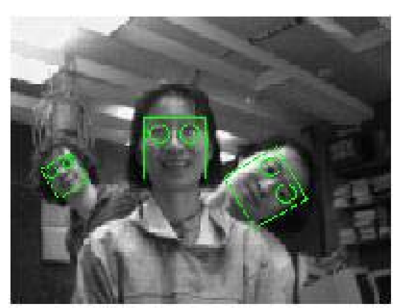
Many faces of Madona

Face Recognition: 2-D and 3-D



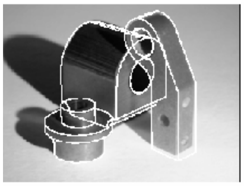
Object Recognition: 2-D Image-based

- Some objects are 2D patterns
  - e.g. faces
- Build an explicit pattern matcher
  - discount changes in illumination by using a parametric model
  - changes in background are hard
  - changes in pose are hard



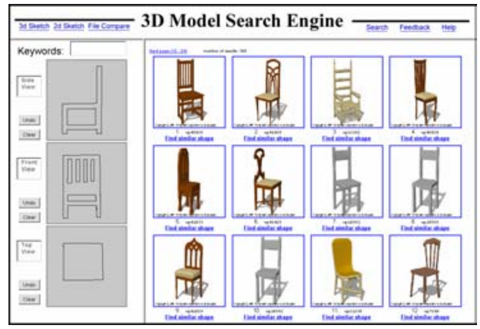
http://www.ri.cmu.edu/projects/project_271.html

Model-Based Vision



- Given 3-D models of each object
- Detect image features (often edges, line segments, conic sections)
- Establish correspondence between model &image features
- Estimate pose
- Consistency of projected model with image.

Object Classes: Chairs



(Funkhauser, Min, Kazhdan, Chen, Halderman, Dobkin, Jacobs)

Announcements

- Class Web Page is up:
  – http://www.cs.ucsd.edu/classes/sp05/cse152/
- Assignment 0: "**Getting Started with Matlab**" (to be posted soon), due 4/7/05
- Read Chapters 1 Trucco & Verri

The Syllabus