Introduction

Computer Vision I

CSE 291A00

Lecture 1

What is Computer Vision?

- Trucco and Verri: computing properties of the 3D 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

What do you see?

- Changing viewpoint
- Moving light source
- Deforming shape

What was happening

- Changing viewpoint
- Moving light source
- ✓Deforming shape

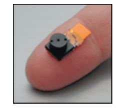
Why study Computer Vision?

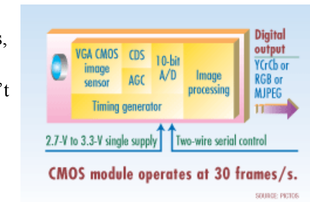
- Images and movies are everywhere
- Fast-growing collection of useful applications
  - building representations of the 3D world from pictures
  - automated surveillance (who's doing what)
  - movie post-processing
  - face recognition
- Various deep and attractive scientific mysteries
  - how does object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

The real reason?

The Near Future: Ubiquitous Vision

- Five years from now, digital cameras will cost 1 cent.
- Digital video will be a widely available commodity component embedded in cell phones, doorbells, PDA's, 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!





Some Objectives

- Segmentation
  - Breaking images and video into meaningful pieces
- Reconstructing the 3D world
  - from multiple views
  - from shading
  - from structural models
- Recognition
  - What are the objects in a scene?
  - What is happening in a video?
- Control
  - Obstacle avoidance
  - Robots, machines, etc.

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

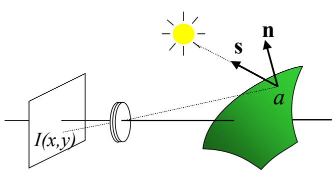
Shading as a result of differences in lighting is

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



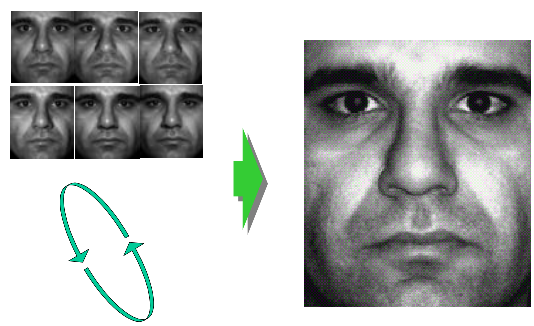
At image location *(x,y)* the intensity of a pixel *I(x,y)* is

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

where

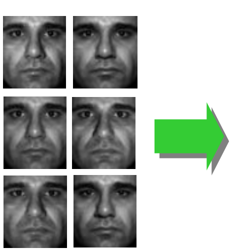
- *a(x,y)* is the albedo of the surface projecting to *(x,y)*.
- **n***(x,y)* is the unit surface normal.
- **s** is the direction and strength of the light source.

Lighting variation

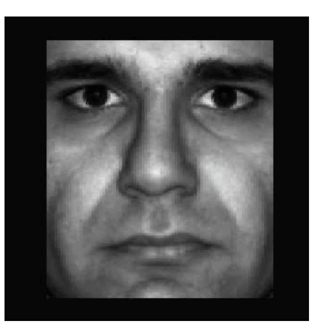


Single Light Source

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
- Part 3: Reconstruction
- 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

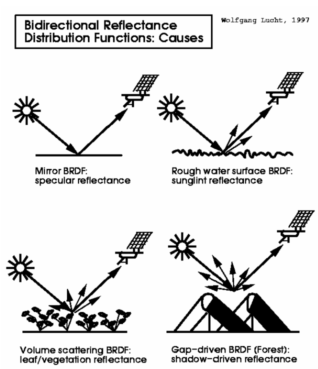
Cameras, lenses, and sensors

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.

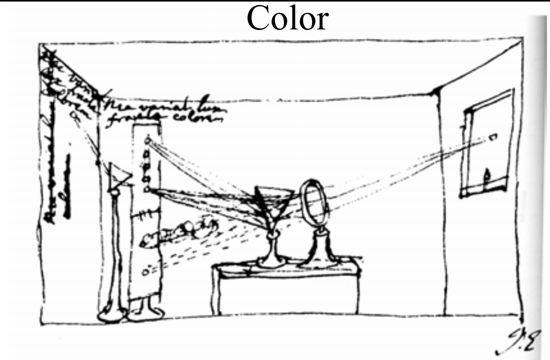
Radiometry



Wolfgang Lucht

brdf/brdfexpl.html

Color



4.1 NEWTON'S SUMMARY DRAWING of his experiments with light. Using a point source of light and a prism, Newton separated sunlight into its fundamental components. By reconverging the rays, he also showed that the decomposition is reversible.

From Foundations of Vision, by Brian Wandell, Sinauer Assoc., 1995

Part II: Early Vision in One Image

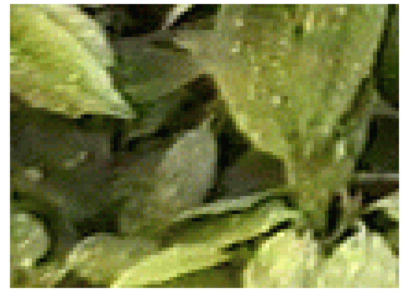
- Representing small patches of image
  - For three reasons
    - We wish to establish correspondence between (say) points in different images, so we need to describe the neighborhood of the points
    - Sharp changes are important in practice --- known as "edges"
    - Representing texture by giving 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

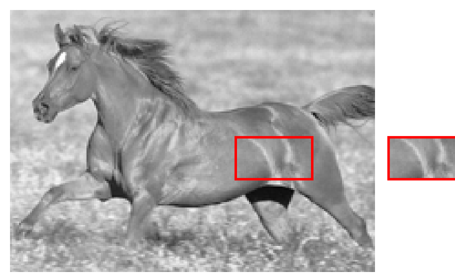
Segmentation

- 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

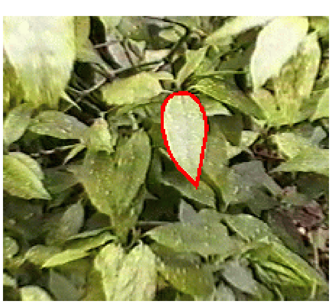
Boundary Detection: Local cues

Boundary Detection: Local cues



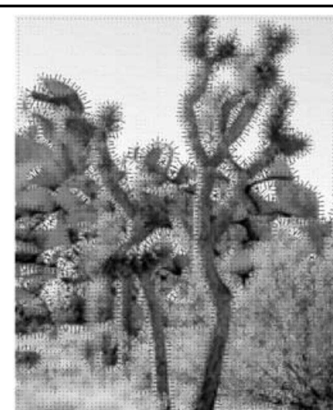


Boundary Detection

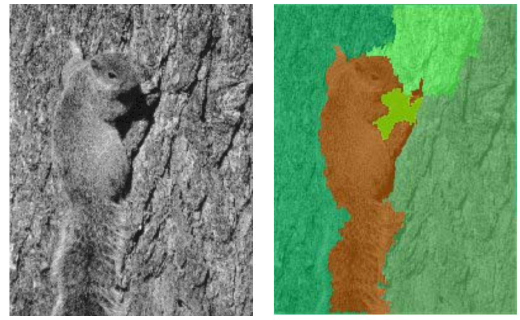


http://www.robots.ox.ac.uk/~vdg/dynamics.html

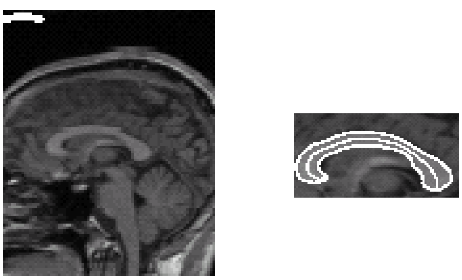
Gradients



(Sharon, Balun, Brandt, Basri)



Boundary Detection



Finding the Corpus Callosum

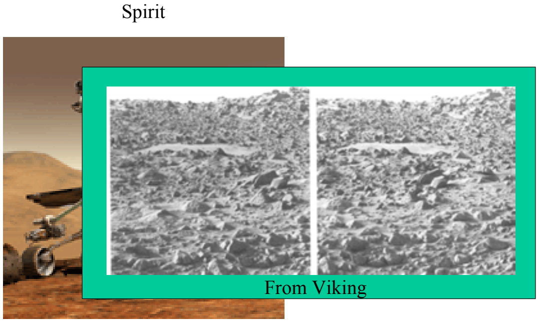
(G. Hamarneh, T. McInerney, D. Terzopoulos)

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

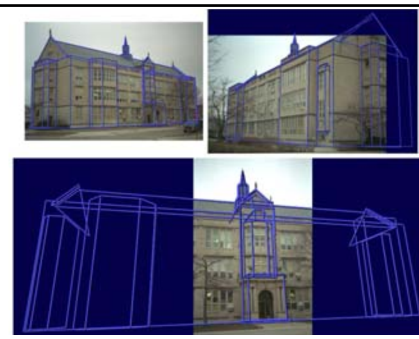
Spirit



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

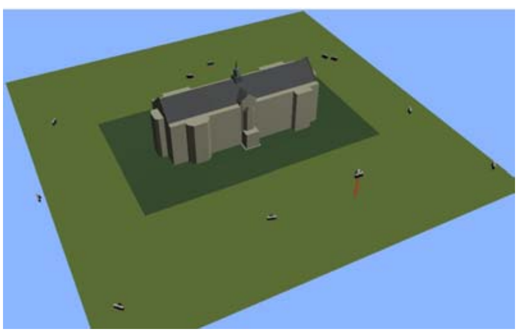
Architectural modeling:
- photogrammetry;
- view-dependent texture mapping;
- model-based stereopsis.

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

Part 4:Recognition: Two approaches

- Detection
  - Find locations in images where class of objects occurs
- Recognition
  - Classify neighborhood of location
- Most useful for specific class of objects (e.g., faces, cars, planes)

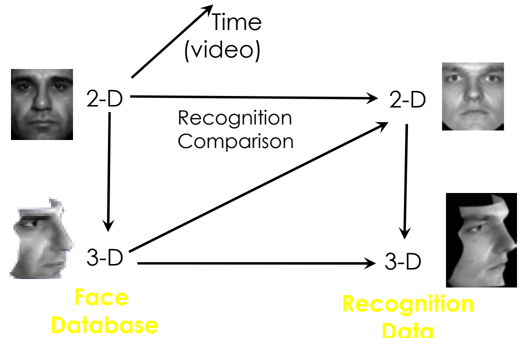
- Segmentation:
  - Which bits of image should be grouped together?
- Recognition:
  - What labels should be attached to each image region.
- Most useful for interpreting entire scene.

Face Detection: First Step

Why is Face Recognition Hard?

Many faces of Madona

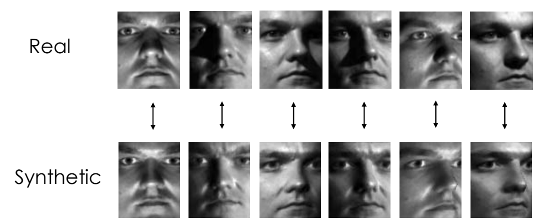
Face Recognition: 2-D and 3-D



Yale Face Database B

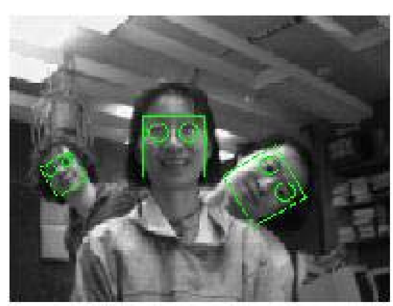
64 Lighting Conditions
9 Poses
=> 576 Images per Person

Real vs. Synthetic



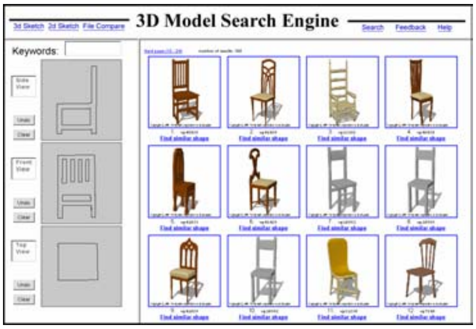
Object Recognition: 2-D Image-based

- 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

Object Classes: Chairs



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

Object Recognition: 3-D Model-based

- Have a 3-D model of the object
- Have representations of classes of objects
- Parts/Whole
- Function

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

Tracking in IR images

Visual Tracking

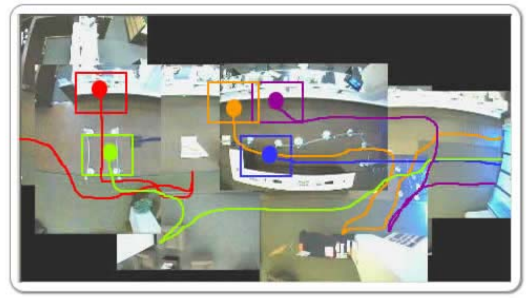
- Estimate location in image of object of interest - color and geometry

Tracking

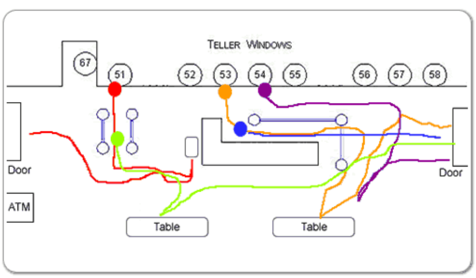
(www.brickstream.com)

Tracking

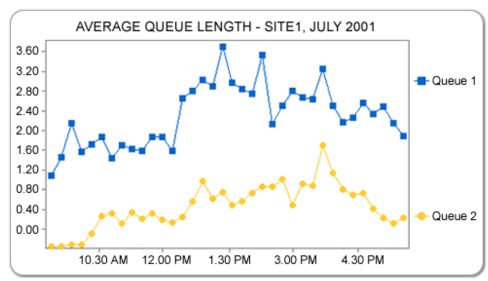
Tracking



Tracking

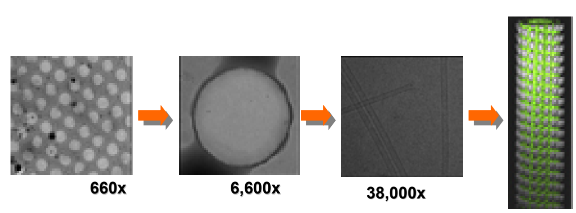


Tracking



A couple applications

Intelligent Microscope for Transmission Electron Microscopy



http://www.itg.uiuc.edu/technology/automated_microscopy/

Visually guided surgery

Compositing Real Objects in Video

The Syllabus

cse252a: early vision and recognition

- Cameras
- Human Vision
- Photometry (radiance, irradiance, BRDF)
- Illumination cones
- Shape from Shading, Photometric Stereo
- Curves & Surfaces
- Color
- Filtering
- Edges & Features
- Stereo Matching
- Optical Flow and Motion
- Tracking
- Statistical pattern recognition (Bayes, SVM, Kernel methods)
- Object Recognition
- Behavior Recognition (HMM's)

cse252b: Multiview Geometry & Segmentation

- Multiview Geometry
- Affine Structure from Motion
- Projective Structure from Motion
- Robust F-matrix estimation
- Image Segmentation
- Texture: Synthesis, Recognition, Shape-from
- Motion Segmentation
- Object Detection
- Image Registration
- Image Based Rendering