INTRODUCTION TO COMPUTER VISION

The slides are from several sources through James Hays (Brown); Srinivasa Narasimhan (CMU); Silvio Savarese (U. of Michigan); Bill Freeman and Antonio Torralba (MIT), including their own slides.
Computer vision

- Automatic control
- Robotics
- Robot vision
- Data mining
- Image retrieval
- Machine learning
- Artificial intelligence
- Signal processing
- Compression
- Non linear SP
- Multi-variate SP
- Biological vision
- Visual Psychophysics
- Neurobiology
- Physics
- Imaging
- Optics
- Smart cameras
- Acquisition methods
- Computer graphics
- Applied math
- Geometry
- Optimization
- Statistics
- Visual pattern recognition
- Computer graphics
- Acquisition methods
Computer vision

Computer vision studies the **tools and theories** that enable the design of machines that can **extract useful information from imagery data** (images and videos) toward the goal of **interpreting the world**

- Scene
- Objects
- People
- Actions

**Information:** visual cues, 3D structure, motion flows, etc…

**Interpretation:** recognize objects, scenes, actions, events
Have we reached humans?

... not yet

- computer vision is still no match for human perception
- but catching up, particularly in certain areas
Applications of computer vision

- Factory inspection
- Assistive technologies
- Surveillance
- Autonomous driving, robot navigation
- Driver assistance (collision warning, lane departure warning, rear object detection)
- Security

Sources: K. Grauman, L. Fei-Fei, S. Laznebick
Human Vision

- Can do amazing things like:
  - Recognize people and objects
  - Navigate through obstacles
  - Understand mood in the scene
  - Imagine stories

- But still is not perfect: After millions of years are you sure?
  - Suffers from Illusions Illusions may not be errors after all....
  - Ignores many details
  - Ambiguous description of the world
  - Doesn’t care about accuracy of world
Computer Vision

What we see

What a computer sees
Ridiculously brief history of computer vision

in fact was Seymour Papert

• 1966: Minsky assigns computer vision as an undergrad summer project
• 1960’s: interpretation of synthetic worlds
• 1970’s: some progress on interpreting selected images
• 1980’s: ANNs come and go; shift toward geometry and increased mathematical rigor
• 1990’s: face recognition; statistical analysis in vogue
• 2000’s: broader recognition; large annotated datasets available; video processing starts
A Simple World

MACHINE PERCEPTION OF THREE-DIMENSIONAL SOLIDS

by

LAWRENCE GILMAN ROBERTS

Submitted to the Department of Electrical Engineering on May 10, 1963, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition system. However, the multi-tudinous attempts at character recognition, including my own, have not led very far. The reason, I feel, is that the study of abstract, two-dimensional forms leads us away from, not toward, the techniques necessary for the recognition of three-dimensional objects. The per-

Complete Convex Polygons. The polygon selection procedure would select the numbered polygons as complete and convex. The number indicates the probable number of sides. A polygon is incomplete if one of its points is a collinear joint of another polygon.

Why is this hard?

Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

Sinha & Adelson 93
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".
Non-accidental properties

**Perceptual Organization and Visual Recognition**

David G. Lowe

---

**Principle of Non-Accidentalness:** Critical information is unlikely to be a consequence of an accident of viewpoint.

**Three Space Inference from Image Features**

<table>
<thead>
<tr>
<th>2-D Relation</th>
<th>3-D Inference</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Collinearity of points or lines</td>
<td>Collinearity in 3-Space</td>
<td></td>
</tr>
<tr>
<td>2. Curvilinearity of points of arcs</td>
<td>Curvilinearity in 3-Space</td>
<td></td>
</tr>
<tr>
<td>3. Symmetry (Skew Symmetry ?)</td>
<td>Symmetry in 3-Space</td>
<td></td>
</tr>
<tr>
<td>4. Parallel Curves (Over Small Visual Angles)</td>
<td>Curves are parallel in 3-Space</td>
<td></td>
</tr>
<tr>
<td>5. Vertices—two or more terminations at a common point</td>
<td>Curves terminate at a common point in 3-Space</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4. Five nonaccidental relations. (From Figure 5.2, Perceptual organization and visual recognition [p. 77] by David Lowe. Unpublished doctoral dissertation, Stanford University. Adapted by permission.)*
An other way to divide computer vision is in three parts:

1. Geometry
2. Low & Mid-level vision
3. High level vision

More correct: low, middle and high-level vision.
1. Geometry
2. Low & Mid-level vision
3. High level vision

Geometry:
- How to extract 3D information?
- Which cues are useful?
- What are the mathematical tools?
Visual cues: texture • shading • contours • shadows • reflections
Visual cues: texture • shading • contours • shadows • reflections
Visual cues:  texture • shading • contours • shadows • reflections
Visual cues:  texture • shading • contours • **shadows** • reflections
Visual cues:  texture • shading • contours • shadows • reflections
Visual cues:  texture • shading • contours • shadows • reflections

Number of observers: **monocular** • multiple views
Visual cues: texture • shading • contours • shadows • reflections

Number of observers: monocular • multiple views
Projective structure from motion: Here be dragons!

Stereo

Epipolar geometry

Tomasi & Kanade (1993)

Structure from motion

Image sources: S. Laznebick
Visual cues:  texture • shading • contours • shadows • reflections

Number of observers:  monocular • multiple views

Active lighting:  laser stripes • structured lighting patterns
3D Scanning

Scanning Michelangelo’s “The David”

- The Digital Michelangelo Project
- 2 BILLION polygons, accuracy to .29mm

Courtesy of Stanford computer graphics lab
1. Geometry
2. Low & Mid-level vision
3. High level vision

Mid-level vision:
- Extract useful building blocks
- Region segmentation
- Motion flows
Extract useful building blocks
Extract useful building blocks

Alignment

Extract planar regions

Object segmentation
Image enhancement

Image Inpainting, M. Bertalmio et al.
http://www.iua.upf.es/~mbertalmio/restoration.html
Automatic Panorama Stitching

Sources: M. Brown
Automatic Panorama Stitching
Camera tracking and V.R. insertions

Courtesy of Exford Visual Geometry Group
1. Geometry
2. Low & Mid-level vision
3. High level vision

High level operations
- Recognition of objects and people
- Places
- Actions & events
Object recognition and categorization

- Building
- Downtown Chicago
- Clock
- Person
- Car
- Pedestrians crossing street
Detecting and tracking people
Face recognition
Recognizing scenes
Sorting out millions of images/videos
Challenges: viewpoint variation
Challenges: illumination

image credit: J. Koenderink
Challenges: scale
Challenges: deformation
Challenges: occlusion

Magritte, 1957
Challenges: background clutter

Kilmeny Niland. 1995
Challenges: local ambiguity

Source: Rob Fergus and Antonio Torralba
Challenges: local ambiguity

Source: Rob Fergus and Antonio Torralba
Challenges: Motion
Challenges: object intra-class variation
Some, mostly computer vision success stories.
Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software

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

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition
Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, ...
Vision-based biometrics

“How the Afghan Girl was Identified by Her Iris Patterns”  Read the story wikipedia
Login without a password...

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/
Object recognition (in mobile phones)

Point & Find, Nokia
Google Goggles
Sports

Sportvision first down line

Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html
Mobileye

Vision systems in high-end BMW, GM, Volvo models

• Pedestrian collision warning
• Forward collision warning
• Lane departure warning
• Headway monitoring and warning
Google cars

Vision-based interaction: Xbox Kinect


http://electronics.howstuffworks.com/microsoft-kinect.htm

Industrial robots

Vision-guided robots
Medical imaging

3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT
Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.