Instructors Antonio Torralba
TA Jianxiong Xiao

Lecture TR 1:00PM - 2:30PM (3-333)
Assignments

• Problem sets (2/3)
  – Almost weekly
  – Graded as 1-3
  – Late policy
  – Collaboration policy

• Final project (1/3)
  – Project proposal
  – 5min class presentation
  – Report

• No exams or quizzes
Readings

http://groups.csail.mit.edu/vision/courses/6.869/materials.html
Matlab Tutorial
Friday Sep 7

• If you are a Matlab expert, no need to attend.
Lecture 1
A Simple Vision System
What is vision?

- What does it mean, to see? “to know what is where by looking”.
- How to discover from images what is present in the world, where things are, what actions are taking place.

from Marr, 1982
The importance of images

Some images are more important than others

“Dora Maar au Chat”
Pablo Picasso, 1941

100 million $
Finally, Widespread Public Success of Computer Vision Applications

- Mobile phones (Google Goggles, Amazon SnapTell, Panoramas, Augmented Reality, MoneyReader, Word Lens, many more...)
- User interfaces (Microsoft Kinect, Sony EyeToy, ...)
- Automotive driving (MobilEye, Iteris, Google, many auto companies, ...)
- Film and Video (Sports Video analysis, Hawkeye, match-move software, ...)
- Medical vision (registration, diagnosis, augmented reality...)
- Industrial vision, defense, surveillance, biometrics, safety, marketing, agriculture, biology, resource management, ...

* But, the term “computer vision” is still not widely recognized.
Why is vision hard?
The structure of ambient light
The structure of ambient light
The intensity $P$ can be parameterized as:

$$P(\theta, \phi, \lambda, t, X, Y, Z)$$

"The complete set of all convergence points constitutes the permanent possibilities of vision." Gibson
Why is vision hard?
Measuring light vs. measuring scene properties

We perceive two squares, one on top of each other.
Measuring light vs. measuring scene properties

by Roger Shepard ("Turning the Tables")

Depth processing is automatic, and we can not shut it down…
Measuring light vs. measuring scene properties
Measuring light vs. measuring scene properties
Measuring light vs. measuring scene properties

(c) 2006 Walt Anthony
Assumptions can be wrong

Ames room
Why is vision hard?
Some things have strong variations in appearance
Some things know that you have eyes

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".
Problem set 1
The “one week” vision project

The goal of the first problem set is to solve vision
A Simple Visual System

• A simple world
• A simple image formation model
• A simple goal
A Simple World
A Simple World

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 multitudinous 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.
A Simple World
A simple image formation model

Simple world rules:
• Surfaces can be horizontal or vertical.
• Objects will be resting on a white horizontal ground plane
A simple image formation model

Perspective projection

Parallel (orthographic) projection
A simple image formation model

(right-handed reference system)
A simple image formation model

World coordinates

image coordinates

World coordinates

image coordinates

\[ x = X + x_0 \]
\[ y = \cos(\theta) Y - \sin(\theta) Z + y_0 \]
A simple goal

To recover the 3D structure of the world

We want to recover $X(x,y)$, $Y(x,y)$, $Z(x,y)$ using as input $I(x,y)$
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
A simple visual system
The input image

I(x,y)

x

y

0

255
• Proposition 1. The primary task of early vision is to deliver a small set of useful measurements about each observable location in the plenoptic function.

• Proposition 2. The elemental operations of early vision involve the measurement of local change along various directions within the plenoptic function.

  Adelson, Bergen. 91

• Goal: to transform the image into other representations (rather than pixel values) that makes scene information more explicit

Cavanagh, Perception 95
Edges

- Occlusion
- Change of Surface orientation
- Contact edge
- Vertical 3D edge
- Horizontal 3D edge
- Shadow boundary
Finding edges in the image

Image gradient:

\[ \nabla I = \left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \]

Approximation image derivative:

\[ \frac{\partial I}{\partial x} \approx I(x, y) - I(x - 1, y) \]

Edge strength:

\[ E(x, y) = |\nabla I(x, y)| \]

Edge orientation:

\[ \theta(x, y) = \angle \nabla I = \arctan \frac{\partial I/\partial y}{\partial I/\partial x} \]

Edge normal:

\[ n = \frac{\nabla I}{|\nabla I|} \]
Finding edges in the image

\[ \nabla I = \left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \quad \text{n} = \frac{\nabla I}{|\nabla I|} \]
Edge classification

- Figure/ground segmentation

- Occlusion edges
  - Occlusion edges are owned by the foreground

- Contact edges
From edges to surface constraints

• **Ground**
  
  \[ Y(x,y) = 0 \quad \text{if (x,y) belongs to a ground pixel} \]

• **Contact edge**
  
  \[ Y(x,y) = 0 \quad \text{if (x,y) belongs to foreground and is a contact edge} \]

Next, things get a bit more complicated.
Generic view assumption: the observer should not assume that he has a special position in the world… The most generic interpretation is to see a vertical line as a vertical line in 3D.

Freeman, 93
Non-accidental properties

Perceptual Organization and Visual Recognition

David G. Lowe

D. Lowe, 1985

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.)
Non-accidental properties in the simple world

Using $\theta(x,y)$

Using $E(x,y)$

Using $\theta(x,y)$
From edges to surface constraints

• Vertical edges

\[
\begin{align*}
x &= X + x_0 \\
y &= \cos(\theta) Y - \sin(\theta) Z + y_0
\end{align*}
\]

\[
Z = \text{constant along the edge}
\]

\[
\frac{\partial Y}{\partial y} = \frac{1}{\cos(\theta)}
\]
From edges to surface constraints

• Horizontal edges

\[ x = X + x_0 \]
\[ y = \cos(\theta) Y - \sin(\theta) Z + y_0 \]

\[ Y = \text{constant along the edge} \]
\[ \frac{\partial Y}{\partial \mathbf{t}} = 0 \]

Where \( \mathbf{t} \) is the vector parallel to the edge

\[ \mathbf{t} = (-n_y, n_x) \]

\[ \frac{\partial Y}{\partial \mathbf{t}} = -n_y \frac{\partial Y}{\partial x} + n_x \frac{\partial Y}{\partial y} \]
From edges to surface constraints

• What happens where there are no edges?

Assumption of planar faces:

\[
\begin{align*}
\frac{\partial^2 Y}{\partial x^2} &= 0 \\
\frac{\partial^2 Y}{\partial y^2} &= 0 \\
\frac{\partial^2 Y}{\partial y \partial x} &= 0
\end{align*}
\]

Information has to be propagated from the edges
A simple inference scheme

All the constraints are linear

\[ Y(x,y) = 0 \]

\[
\frac{\partial Y}{\partial y} = \frac{1}{\cos(\theta)} \\
\frac{\partial Y}{\partial t} = 0
\]

if \((x,y)\) belongs to a ground pixel

if \((x,y)\) belongs to a vertical edge

if \((x,y)\) belongs to an horizontal edge

if \((x,y)\) is not an edge

\[
\frac{\partial^2 Y}{\partial x^2} = 0 \\
\frac{\partial^2 Y}{\partial y^2} = 0 \\
\frac{\partial^2 Y}{\partial y \partial x} = 0
\]
Discrete approximation

We can transform every differential constrain into a discrete linear constraint on \( Y(x,y) \)

\[
\frac{dY}{dx} \approx Y(x,y) - Y(x-1,y)
\]

A slightly better approximation
(it is symmetric, and it averages horizontal derivatives over 3 vertical locations)
Discrete approximation

Transform the “image” $Y(x,y)$ into a column vector:

$$\frac{dY}{dx} \approx Y(x,y) - Y(x-1,y) = Y(2,2) - Y(1,2) =$$

$$x=2, y=2$$

$Y(x,y)$

$x=0$

$y=0$
A simple inference scheme

\[ A Y = b \]
\[ Y = (A^T A)^{-1} A^T b \]
Results

- Edge strengths
- 3D orientation
- Contact edges
- Depth discontinuities

Edge normals
Changing view point

Input

New view points:
Violations of simple world assumptions

Shading is due to painted stripes
Violations of simple world assumptions

Shading is due to illumination
Violations of simple world assumptions

Figure 24.9 The impossible steps. On the left, the horizontal stripes appear to be due to paint; on the right, they appear to be due to shading.

Impossible steps
Impossible steps
What is missing?

• Recognition