Computer Vision

Micha Lindenbaum
CS dept. Technion

Slides from: Ullman, Seitz, Torralba, Moses, Avidan

Context

The world

Sensing Imaging

Graphics Rendering

Image(s)

Image Processing

A description of the world

Computer Vision
Vision
Image understanding
Why an image?

Other sensors:
• Sound
• Tactile - touch -
• Electromagnetic field
• Chemical ...

Advantage of images:
• Intuitive, and therefore convenient
• Versatile
• Practical and powerful (evolution)
• Sensors are available

Images

An image = an object that describes visual perception

• Can be obtained by an imaging process (photography)
• (Also by painting, comp. rendering)
Camera history

- Optics principles, Mo-ti (400BC), Aristolles (350 BC) - light travels on straight lines, pinhole imaging

- Eye mechanism - Al-Haytham (1000AC)

- Camera obscura - Leonardo (1600),

- Camera obscura were common during the Renaissance by artists, and for looking at sun eclipse (Kepler 1650)

- The first picture by Niepce at 1822
Now – almost all cameras are digital, color, and video

Billions of cameras, iphone6 most popular (flickr, 2016)

To shoot or not to shoot - not the question anymore.

Automatically shooting cameras in scientific, industrial, security, entertainment, cars.

Images
Images

Representing a digital image –
• The sensors are matrices of individual sensors
  – picture elements (pixels)
• Every pixel has a value
  – Intensity of light, denoted grey level,
    usually quantized to 8-12 bits
  – Light – usually a mixture of wavelength (colors),
    images are usually represented by intensity in 3 bands (e.g.
    RGB), gives and illusion of natural color.
• For analysis – An image is a 2D function \( I(x,y), I(i,j) \)
• Input to a vision process
  – Single image, or a time sequence of image (video)
  – Several images from different cameras
Computer Vision Tasks

- The ultimate goal: Scene interpretation:
  Describing the observed world (scene) by
  - Where - Where are objects, surfaces?
  - What - which familiar objects are present?

- Partial tasks: where are the objects' boundaries? Are there straight lines?

Tasks Examples

Source: Torralba
Object categorization & recognition

Scene understanding:

A VIEW OF A PARK ON A NICE SPRING DAY

Source: Torralba
Action recognition

Do not feed the ducks sign

PEOPLE WALKING IN THE PARK

PERSON FEEDING DUCKS IN THE PARK

DUCKS LOOKING FOR FOOD

Source: Torralba

Where

The yellow object is farther than the red object

Far from the camera

Close to the camera

The “DUCKS” objects are on top of the “GRASS” plane

Source: Torralba
Motivations

- **Application related**
  How to get reliable and cheap vision capabilities for a particular task (e.g., car driving). Can use a variety of tools, e.g., many cameras, range finders, interaction with user, bar codes.

- **Natural science**
  How is vision carried out in humans and in other animals? Considers also neural knowledge. E.g., Where, in the brain, recognition happens?

- **Scientific, curiosity, algorithmic**
  Vision is challenging yet possible
Applications - already classic

- Optical character recognition

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

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

Source: Seitz

Applications - inspection

- Industrial (and agriculture) inspection

Source: AMAT - wafer inspection

Source: Orbot - PCB inspection

Source: CogniTens - automotive module
Applications - Biometric

- Login w.o. a password (fingerprint, face)

Source: Seitz

Applications - media

- Blue screen
- Special effects for movies

• The Matrix movies

Source: Seitz
Applications - media

• Large scale 3D modeling

Image from Microsoft’s Virtual Earth
(see also: Google Earth)

Source: Seitz

Applications - Computer Interaction

Primesense (Kinect)

Source: Moses
Applications - Medical

• Many applications

Source: Seitz

3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Applications - Transportation

Mobileye

Source: mobileye
Vision seems easy

Minsky (to an undergraduate student 1966):

“Spend the summer linking a camera to a computer and getting the computer to describe what it saw”

Why is Vision difficult ?

- The image depends on many factors
  For natural images
  - Objects' identity, position, color, reflection properties, occlusions, class/object variation
  - Illumination sources' location, intensity, light distribution, color, light propagation
  - Camera location, sensor and lens properties, deformations

\[ I_{\text{given}} = I(\text{Shape, position, camera param}...) \]
Why is Vision difficult? (cont.)

- The image does not provide direct/explicit information on the (object in the) scene.
- Photometric uncertainty: e.g. dark or light object?
- Geometric uncertainty: e.g. far or close object?

\[ I_{\text{given}} = I(\text{Shape}, \text{position}, \text{camera param}...) \]

Vision by Synthesis

Synthesis based approach for vision: minimize the difference between given and synthesized

\[
\text{arg min}_{\text{Shape, position, other...}} \| I_{\text{given}} - I(\text{Shape, position},...) \|
\]

The synthesis-based/inverse task is difficult:
- Realistic synthesis is difficult and complex in itself
- Optimization is difficult: non-convex, non-unique, in high dimension.
- Wasteful: Often, we are interested in only few unknowns
Vision by image Comparison

Goal: find an object

Preprocessing: Make an image of the object (model)

Detection: compare the model to all subimages.

Source: Torralba

Find the chair
How to make vision easier?

- Focus on image information that changes less (invariants)
- Solve easier, partial problems,
- Make the problem easier by controlled illumination, multiple cameras, video, range, limited context, human interaction, ...
- Use knowledge & learning.

Current status: Many successful applications in limited contexts, no general vision,

Amazing progress

The human visual system
one vision system that works, but how?
The human visual system

- Useful,
- Effective,
- Fast (e.g. 20-80ms recognition).
- Occupies about 50% of cortex,
- We have partial knowledge about function
- Hard to investigate:
  - Anatomy & Physiology - limited conclusions
  - Psychophysics - brain as black box
  - Single cell recording - not so meaningful
  - fMRI - density limited - 10,000 neuron/mm^3

Anatomical findings: e.g. eye
Anatomy & Physiology:
Retina, Photoreceptors
Cons and Rods

- Retina: nonuniform array of photoreceptors, computation units, blood vessels
- light → neural activity: logarithmic function

Computation in Retina

Which computation?
Psychophysics: What do you see?

Brightness illusion
Ganglion cells give “Center-surround” response
- Sensitive to spatial intensity change
- Weak response to constant intensity

A slice through the system
Single neuron recording in the primary visual cortex (V1)

Hubel & Wiesel

V1 performs some kind of Orientation analysis

Physiology & fMRI: Visual areas

• Some areas have known roles. Other do not.
• Preprocessing in V1++
• The “dorsal” part is responsible to motion, object locations, and eye control - “Where?”
• The “ventral” part, is responsible for recognition - “What?”
### General plan

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction, HVS</td>
</tr>
<tr>
<td>2-3</td>
<td>Imaging processes, Geometry &amp; Photometry</td>
</tr>
<tr>
<td>3-4</td>
<td>Edge detection, Scale space, pyramids</td>
</tr>
<tr>
<td>5</td>
<td>Interest operators</td>
</tr>
<tr>
<td>6-7</td>
<td>Grouping, clustering, segmentation</td>
</tr>
<tr>
<td>7-8</td>
<td>Reconstruction, Shape from X</td>
</tr>
<tr>
<td>9-10</td>
<td>Correspondence and motion</td>
</tr>
<tr>
<td>11</td>
<td>Learning for vision tasks</td>
</tr>
<tr>
<td>12-13</td>
<td>Object Recognition and Categorization</td>
</tr>
</tbody>
</table>