Introduction to Artificial Intelligence

V22.0472-001 Fall 2009

Lecture 26: Computer Vision

Rob Fergus – Dept of Computer Science, Courant Institute, NYU
Slides from Andrew Zisserman

What is Computer Vision?

- Vision is about discovering from images what is present in the scene and where it is.
- In Computer Vision a camera (or several cameras) is linked to a computer. The computer interprets images of a real scene to obtain information useful for tasks such as navigation, manipulation and recognition.

Applications

- Intelligent machines (AI)
- Industrial inspection  
  e.g. light bulbs, electronic circuits
- Automotive  
  e.g. Ford, GM, DARPA Grand Challenge
- Security  
  e.g. facial recognition in airports
- Toys (Aibo dog)
- Image/video retrieval
- Digital cameras are everywhere now….

Application: Assisted driving

Pedestrian and car detection

Lane detection

- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,

Application: Computational photography

Application: Improving online search

Query: STREET

Organizing photo collections
The problem

- When we “see” something, what’s involved?
- Take a picture with a digital camera, but computer doesn’t understand the image – it’s just a load of coloured dots (pixels)
- Want to make a computer understand images

Real world scene - Sensing device - Interpreting device - Interpretation

A person/
A person with folded arms/
Prof. Pietro Perona etc.

The Human Eye

- Retina measures about 5 × 5 cm and contains 10^6 sampling elements (rods and cones).
- The eye’s spatial resolution is about 0.01° over a 150° field of view (not evenly spaced, there is a fovea and a peripheral region).
- Intensity resolution is about 11 bits/element, spectral range is 400–700nm.
- Temporal resolution is about 100 ms (10 Hz).
- Two eyes give a data rate of about 3 GBytes/s!

The Human Eye

- Vision is the most powerful of our own senses.
- Around 1/3 of our brain is devoted to processing the signals from our eyes.
- The visual cortex has around \(O(10^{11})\) neurons.

Computer Vision: A whole series of problems

- What is in the image?
  - Object recognition problem
  - Where is it?
    - 3D spatial layout
    - Shape
  - How is the camera moving?
  - What is the action?

Brief Look at Two Problems

1. Stereo Reconstruction
   Measuring the 3D structure of the scene
2. Object Recognition
   What objects are present in the scene

Brief Look at Two Problems

1. Stereo Reconstruction
   Measuring the 3D structure of the scene
2. Object Recognition
   What objects are present in the scene
Image is a projection of world

Stereo Reconstruction

Shape (3D) from two (or more) images

Scenarios

The two images can arise from
• A stereo rig consisting of two cameras
  • the two images are acquired simultaneously
or
• A single moving camera (static scene)
  • the two images are acquired sequentially

The two scenarios are geometrically equivalent
**Imaging Geometry**

- central projection
- camera centre, image point and scene point are collinear
- an image point back projects to a ray in 3-space
- depth of the scene point is unknown

**The objective**

Given two images of a scene acquired by known cameras compute the 3D position of the scene (structure recovery)

Basic principle: triangulate from corresponding image points
- Determine 3D point at intersection of two back-projected rays

**An algorithm for Stereo Reconstruction**

1. For each point in the first image determine the corresponding point in the second image
   (this is a search problem)
2. For each pair of matched points determine the 3D point by triangulation
   (this is an estimation problem)

**The correspondence problem**

Given a point \( x \) in one image find the corresponding point in the other image

This appears to be a 2D search problem, but it is reduced to a 1D search by the epipolar constraint

**Notation**

The two cameras are \( P \) and \( P' \), and a 3D point \( x \) is imaged as

\[
x = Px \quad x' = P'x
\]

- \( P \): 3 x 4 matrix
- \( x \): 4-vector
- \( x' \): 3-vector

Warning

for equations involving homogeneous quantities ‘\( = \)’ means ‘equal up to scale’
Epipolar geometry
Given an image point in one view, where is the corresponding point in the other view?

- A point in one view "generates" an epipolar line in the other view
- The corresponding point lies on this line

Epipolar line

Epipolar constraint
- Reduces correspondence problem to 1D search along an epipolar line

Geometry:
3D models of planar objects

[Fitzgibbon et. al.]
[Zisserman et. al.]

Brief Look at Two Problems

1. Stereo Reconstruction
   Measuring the 3D structure of the scene

2. Object Recognition
   What objects are present in the scene

Detection: localize the street-lights in the image

Classification: is there a street-light in the image?
Challenges 1: view point variation

Michelangelo 1475-1564

Challenges 2: illumination

slide credit: S. Ullman

Challenges 3: occlusion

Magritte, 1957

Challenges 4: scale

Challenges 5: deformation

Xu, Beihtong 1943

Challenges 6: background clutter

Klimt, 1913
Within-class variations

Bag-of-features models

Object → Bag of ‘words’

Objects as texture

• All of these are treated as being the same

• Similar to Bag-of-Words models for text document analysis, e.g. Spam/Ham problem

Representing categories: Parts and Structure


Parts-and-shape representation

• Model:
  – Object as a set of parts
  – Relative locations between parts
  – Appearance of part

Motorbikes

Figure from [Fischler & Elschlager 73]
Current Performance

- Trains
  - Class images: Highest ranked
  - Class images: Lowest ranked
  - Non-class images: Highest ranked
  - Viewpoint bias

Current Performance

- Person
  - Class images: Highest ranked
  - Class images: Lowest ranked
  - Non-class images: Highest ranked
  - cf. 2007 motorbikes