6.869 Advances in Computer Vision: Learning and Interfaces
Spring 2005

Announcements

Course Information
- Syllabus
- Problem Sets and Exams
- Grading and Requirements
- Internet Resources

Contacts
http://courses.csail.mit.edu/6.869

Administration

- Syllabus
- Grading
- Collaboration Policy
- Project

Syllabus

The topics studied in this course will include:
- Image statistics, image representations, and texture models
- Color Vision
- Graphical models, Bayesian methods
- Motion Random Fields, applications to low-level vision
- Appearance models
- Statistical methods
- Contouring & Segmentation
- Object recognition
- Tracking and Spatial Propagation
- Visual Surveillance and Activity Monitoring
Course requirements

- Two take-home exams
- Five problem sets with lab exercises in Matlab
- No final exam
- Final project

Grading

- Problem sets are graded check, check-plus, check-minus
- Contribution to grade:
  - 5 problem sets: 30%
  - 2 take-home exams: 40%
  - final project: 30%

Collaboration Policy

Problem sets may be discussed, but all written work and coding must be done individually. Take-home exams may not be discussed. Individuals found submitting duplicate or substantially similar materials due to inappropriate collaboration may get an F in this class and other sanctions.

Project

The final project may be
- An original implementation of a new or published idea
- A detailed empirical evaluation of an existing implementation of one or more methods
- A paper comparing three or more papers not covered in class, or surveying recent literature in a particular area

A project proposal not longer than two pages must be submitted and approved by April 1st. I can provide ideas or suggestions for projects.

Problem Set 0

- Out today, due 2/12
- Matlab image exercises
  - load, display images
  - pixel manipulation
  - RGB color interpolation
  - image warping / morphing with interp2
  - simple background subtraction
- All psets graded loosely: check, check-, 0.
- (Outstanding solutions get extra credit.)
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

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

Why study Computer Vision?

• One can “predict the future” (and avoid bad things…)! 
• 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 finding
• Greater understanding of human vision
• Various scientific questions
  – how does object recognition work?

The course, in broad categories

• Images and image formation
• Low-level vision
• High-level vision
• Implementations and applications

What is object recognition?

• People draw distinctions between what is seen
  – This could mean “is this a fish or a bicycle?”
  – It could mean “is this George Washington?”
  – It could mean “is this poisonous or not?”
  – It could mean “is this slippery or not?”
  – It could mean “will this support my weight?”
• Area of research:
  • How to build programs that can draw useful distinctions based on image properties.
Computer vision class, fast-forward

Images and image formation

Cameras, lenses, and sensors

- Pinhole cameras
- Lenses
- Projection models
- Geometric camera parameters

Radiometry…not covered (see 6.801)

Wolfgang Lucht

http://geography.bu.edu/brdf/brdfexpl.html


Color

Low-level vision

Figure 5.16: The first photograph on record, in table service, obtained by Niépce in 1825. Collectie Hartwig-Meertens.


6.1: 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 recombining the rays, he also showed that the decomposition is reversible.

Image filtering

- Review of linear systems, convolution
- Bandpass filter-based image representations
- Probabilistic models for images

SIFT (scale invariant feature transforms)

David Lowe, IJCV 2004

Non-linear filtering, and applications

Models of texture

Learning and vision

Bayesian framework for vision
Bayesian framework for vision

Coincidental appearance of face profile in rock?

http://www.cs.dartmouth.edu/whites/old_man.html

Bayesian framework for vision

Coincidental appearance of faces in rock?

http://bensguide.gpo.gov/3-5/symbols/print/mountrushmore.html

Eigenfaces: linear bases for faces

http://www.cs.dartmouth.edu/whites/old_man.html

Statistical classifiers

– MIT Media Lab face localization results.
– Applications: database search, human machine interaction, video conferencing.

Support vector machines and boosting

Support vector machines and boosting

Large-margin classifier

http://www.support-vector.net/nello.html

“The kernel trick”

http://www.support-vector.net/nello.html
Recent, now classic, paper on face detection:

*Rapid Object Detection Using a Boosted Cascade of Simple Features*

Paul Viola  Michael J. Jones  
Mitsubishi Electric Research Laboratories (MERL)  
Cambridge, MA

Most of this work was done at Compaq CRL before the authors moved to MERL

Face Detection Goal

Many Uses  
- User Interfaces  
- Interactive Agents  
- Security Systems  
- Video Compression  
- Image Database Analysis

Use of context for object detection

Images by Antonio Torralba

The world, to a face detector

Structure from Motion

What is the shape of the scene?

Segmentation (perceptual grouping)

How many ways can you segment six points?  
(or curves)
Segmentation

• Which image components “belong together”?
• Belong together=lie on the same object
• Cues
  – similar colour
  – similar texture
  – not separated by contour
  – form a suggestive shape when assembled
Applications

Tracking

Follow objects and estimate location..
- radar / planes
- pedestrians
- cars
- face features / expressions
Many ad-hoc approaches...
General probabilistic formulation: model density over time.

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
Articulated Models

Articulated tracking

• Constrained optimization
• Coarse-to-fine part iteration
• Propagate joint constraints through each limb
• Real-time on Ghz pentium...

Find most likely model consistent with observations…. (and previous configuration)

Computer vision applications as ocean-going vessels
Game: Decathlete

Optical-flow-based Decathlete figure motion analysis

Decathlete 100m hurdles

Decathlete javelin throw

Companies and applications

- Cognex
- Reactrix
- Poseidon
- Mobileye
- Eyetoy
- Identix
- Roomba
Motion magnification

And…

• Visual Category Learning
• Image Databases
• Image-based Rendering
• Medical Imaging

Skills learned from this class

• Goal: You’ll be able to go to a computer vision conference and understand what’s going on in most of the presentations.
• You’ll have the skills and awareness of the literature to start building the vision systems you want.
Cameras, lenses, and calibration

Today:
• Camera models
• Projection equations
• Calibration methods

Images are projections of the 3-D world onto a 2-D plane…

7-year old’s question

Why is there no image on a white piece of paper?

Pinhole cameras

• Geometry

Distant objects are smaller

Virtual image, perspective projection

• Abstract camera model - box with a small hole in it

Parallel lines meet

Common to draw film plane in front of the focal point. Moving the film plane merely scales the image.
The equation of projection

- Cartesian coordinates:
  - We have, by similar triangles, that
    \[(x, y, z) \rightarrow (f \frac{x}{z}, f \frac{y}{z}, -f)\]
  - Ignore the third coordinate, and get
    \[(x, y, z) \rightarrow (f \frac{x}{z}, f \frac{y}{z})\]

Vanishing points

- Each set of parallel lines (=direction) meets at a different point
  - The vanishing point for this direction
- Sets of parallel lines on the same plane lead to collinear vanishing points.
  - The line is called the horizon for that plane

Geometric properties of projection

- Points go to points
- Lines go to lines
- Planes go to the whole image or a half-plane
- Polygons go to polygons
- Degenerate cases
  - line through focal point to point
  - plane through focal point to line

What if you photograph a brick wall head-on?

Wandell, Foundations of Vision, Sinauer, 1995
Pinhole camera demonstrations

• Film camera, box, demo. Apertures, lens.

• The image is the convolution of the aperture with the scene.

Weak perspective

• Issue
  – perspective effects, but not over the scale of individual objects
  – collect points into a group at about the same depth, then divide each point by the depth of its group
  – Adv: easy
  – Disadv: wrong

Orthographic projection

Example use of orthographic projection: inferring human body motion in 3-d

Advantage of orthographic projection

Our simplified rendering conditions are as follows: the body is transparent, and each marker is rendered to the image plane orthographically. The figural motion described by human motion basis coefficients \( \hat{\theta} \), the rendered image sequence \( \hat{y} \), is:

\[
\hat{y} = P U \hat{\theta},
\]

(1)

where \( P \) is the projection operator which collapses the \( y \) dimension of the image sequence \( U \hat{\theta} \).

Leventon and Freeman, Bayesian Estimation of Human Motion, MERL TR08-06
Orthography can lead to analytic solutions

have our multi-dimensional gaussian,

Prior probability

\[ P(\theta) = k_3 e^{-\|\theta\|^2/\sigma^2} \]  

where \( k_3 \) is another normalization constant. If we model the observation noise as i.i.d. gaussian with variance \( \sigma \), we have, for the likelihood term of Bayes theorem,

Likelihood function

\[ P(y|\theta) = k_4 e^{-\|y-f(\theta)\|^2/(2\sigma^2)} \]  

with normalization constant \( k_4 \).

The posterior distribution is the product of these two gaussians. That yields another gaussian, with mean and covariance found by a matrix generalization of “completing the square” [7]. The squared error optimal estimate for \( \alpha \) is then

\[ \alpha = \text{SVD}(P^T SP + \sigma I)^{-1}(\hat{y} - (P\bar{\theta})) \]  

Analytic solution for inferred 3-d motion

Leventon and Freeman, Bayesian Estimation of Human Motion, MERL TR98-06

But, alas

“The results for the simplified problem appear promising. However serious questions arise because of the simplifying assumptions, which trivialize a number of the hard issues of the problem in the real world. Eg. scaling effects that arise from perspective projection are ignored, by assuming orthographic projection. …”

Reviewer’s comments

Water glass refraction

http://data.pg2k.hd.org/_e


Snell’s law

\[ n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2) \]
Spherical lens

First order optics

\[ \sin(\theta) \approx \theta \]

\[ \theta \approx \frac{D/2}{f} \]

Paraxial refraction equation

\[ \alpha_1 = \gamma + \beta_1 \approx h \left( \frac{1}{R} + \frac{1}{d_1} \right) \]

\[ \alpha_2 = \gamma - \beta_2 \approx h \left( \frac{1}{R} - \frac{1}{d_2} \right) \]

\[ n_1 \alpha_1 \approx n_2 \alpha_2 \Leftrightarrow \frac{n_1}{d_1} + \frac{n_2}{d_2} = \frac{n_2 - n_1}{R} \]

The thin lens, first order optics

\[ \frac{1}{z} - \frac{1}{z'} = \frac{1}{f} \]

\[ f = \frac{R}{2(n-1)} \]
What camera projection model applies for a thin lens?

Candle and laser pointer demo

More accurate models of real lenses
- Finite lens thickness
- Higher order approximation to $\sin(\theta)$
- Chromatic aberration
- Vignetting

Thick lens

Third order optics

$$\sin(\theta) \approx \theta - \frac{\theta^3}{6}$$

Paraxial refraction equation, 3rd order optics
Spherical aberration (from 3rd order optics)

Other 3rd order effects
- Coma, astigmatism, field curvature, distortion.

Astigmatic distortion

Lens systems
Lens systems can be designed to correct for aberrations described by 3rd order optics

Vignetting

Chromatic aberration
(great for prisms, bad for lenses)
Other (possibly annoying) phenomena

- Chromatic aberration
  - Light at different wavelengths follows different paths; hence, some wavelengths are defocussed
  - Machines: coat the lens
  - Humans: live with it
- Scattering at the lens surface
  - Some light entering the lens system is reflected off each surface it encounters (Fresnel’s law gives details)
  - Machines: coat the lens, interior
  - Humans: live with it (various scattering phenomena are visible in the human eye)

Summary

- Want to make images
- Pinhole camera models the geometry of perspective projection
- Lenses make it work in practice
- Models for lenses
  - Thin lens, spherical surfaces, first order optics
  - Thick lens, higher-order optics, vignetting.

Next

- how positions in the image relate to 3-d positions in the world.

Translation

\[
^B P = ^A P + ^B O_A
\]

Find the rotation matrix

Project \( \hat{O}\hat{P} = \begin{pmatrix} \hat{i}_d & \hat{j}_d & \hat{k}_d \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \)

onto the B frame’s coordinate axes.

\[
\begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} = \begin{pmatrix} \hat{i}_b \cdot \hat{i}_d A_x & \hat{i}_b \cdot \hat{j}_d A_y & \hat{i}_b \cdot \hat{k}_d A_z \\ \hat{j}_b \cdot \hat{i}_d A_x & \hat{j}_b \cdot \hat{j}_d A_y & \hat{j}_b \cdot \hat{k}_d A_z \\ \hat{k}_b \cdot \hat{i}_d A_x & \hat{k}_b \cdot \hat{j}_d A_y & \hat{k}_b \cdot \hat{k}_d A_z \end{pmatrix} \begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix}
\]
Rotation matrix

\[ \begin{pmatrix}
B_x \\
B_y \\
B_z
\end{pmatrix} = \begin{pmatrix}
i_B \cdot i_A A_x & i_B \cdot j_A A_y & i_B \cdot k_A A_z \\
i_B \cdot i_A A_x & i_B \cdot j_A A_y & i_B \cdot k_A A_z \\
k_B \cdot i_A A_x & k_B \cdot j_A A_y & k_B \cdot k_A A_z
\end{pmatrix} \]

implies

\[ B P = {}^A R \begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix} \]

where

\[ {}^A R = \begin{pmatrix}
i_B \cdot i_A & i_B \cdot j_A & i_B \cdot k_A \\
i_B \cdot i_A & i_B \cdot j_A & i_B \cdot k_A \\
k_B \cdot i_A & k_B \cdot j_A & k_B \cdot k_A
\end{pmatrix} \]

Translation and rotation

Let's write

\[ B \begin{pmatrix}
P \\
A_x \\
A_y \\
A_z
\end{pmatrix} = {}^B R \begin{pmatrix}
A_x \\
A_y \\
A_z
\end{pmatrix} + {}^B O_A \]

as a single matrix equation:

\[ \begin{pmatrix}
B_x \\
B_y \\
B_z
\end{pmatrix} = \begin{pmatrix}
- & - & - \\
- & {}^B R & - \\
0 & 0 & 0
\end{pmatrix} \begin{pmatrix}
{}^B O_A \\
A_x \\
A_y \\
A_z
\end{pmatrix} \begin{pmatrix}
1
\end{pmatrix} \]