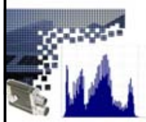


# 6.869 Advances in Computer Vision: Learning and Interfaces

Spring 2005

Tuesday and Thursday; 2:30 to 4:00pm in 36-153

## Announcements

## Course Information

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

## Contacts

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

## Contacts

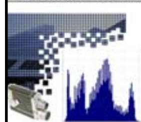
| | |
|---|---|
| **Instructor** | Professor William T. Freeman<br>billf at mit dot edu<br>32-D476<br>(617) 253- 8828 |
| Office Hours | By Appointment |
| **Teaching Assistant** | Xiaoxu Ma<br>xiaoxuma at mit dot edu<br>32-D542<br>(617) 258-5485 |
| Office Hours | Monday, Wed. 4-5pm in 32-D451 |

All offices are located on the fourth and fifth floor of the Dreyfoos building (Stata Center).

If you cannot attend our normally scheduled office hours, please send e-mail to schedule an alternate appointment.

# Administration

- Syllabus
- Grading
- Collaboration Policy
- Project



# 6.869 Advances in Computer Vision: Learning and Interfaces

Spring 2005

## Syllabus

The topics studied in this course will include:

- Image statistics, image representations, and texture models
- Color Vision
- Graphical models, Bayesian methods
- Markov Random Fields, applications to low-level vision
- Approximate inference methods
- Statistical classifiers
- Clustering & Segmentation
- Object recognition
- Tracking and Density Propagation
- Visual Surveillance and Activity Monitoring

## Course Calendar

| Lecture | Date | Description | Readings | Assignments | Materials |
|---|---|---|---|---|---|
| 1 | 2/1 | Course Introduction Cameras and Lenses | Req: FP 1.1, 2.1, 2.2, 2.3, 3.1, 3.2 | PS0 out | |
| 2 | 2/3 | Image Filtering | Req: FP 7.1 - 7.6 | | |
| 3 | 2/8 | Image Representations: Pyramids | Req: FP 7.7, 9.2 | | |
| 4 | 2/10 | Image Statistics | | PS0 due | |
| 5 | 2/15 | Texture | Req: FP 9.1, 9.3, 9.4 | PS1 out | |
| 6 | 2/17 | Color | Req: FP 6.1-6.4 | | |
| 7 | 2/22 | Guest Lecture: Context in vision | | | |
| 8 | 2/24 | Guest Lecture: Medical Imaging | | PS1 due | |
| 9 | 3/1 | Multiview Geometry | Req: Mikolajczyk and Schmid; FP 10 | PS2 out | |
| 10 | 3/3 | Local Features | Req: Shi and Tomasi; Lowe | | |

| | | | | | |
|---|---|---|---|---|---|
| 11 | 3/8 | Bayesian Analysis | | | |
| 12 | 3/10 | Markov Random Fields Belief Propagation | | PS2 due | |
| 13 | 3/15 | Model Based Recognition | Req: FP 18.1-18.5, Lowe | EX1 out | |
| 14 | 3/17 | Discriminative Models | | EX1 due | |
| | 3/22-3/24 | Spring Break (NO LECTURE) | | | |
| 15 | 3/29 | Face Detection and Recognition I | Req: FP 22 | | |
| 16 | 3/31 | Face Detection and Recognition II | | Project proposal due | |
| 17 | 4/5 | Segmentation and Clustering | Req: FP 14, 15.1-15.2, Comaniciu and Meer | PS3 out | |
| 18 | 4/7 | Segmentation and Fitting | Req: FP 15.3-15.5, 16 | | |
| 19 | 4/12 | Tracking I | Req: FP 17 | | |
| 20 | 4/14 | Articulated Tracking and Shape Inference | Req: FP Extra Chapter | PS3 due | |
| | 4/19 | No class (Patriot's Day Holiday) | | | |
| 21 | 4/21 | Approximate Inference Methods | | PS4 out | |

## 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

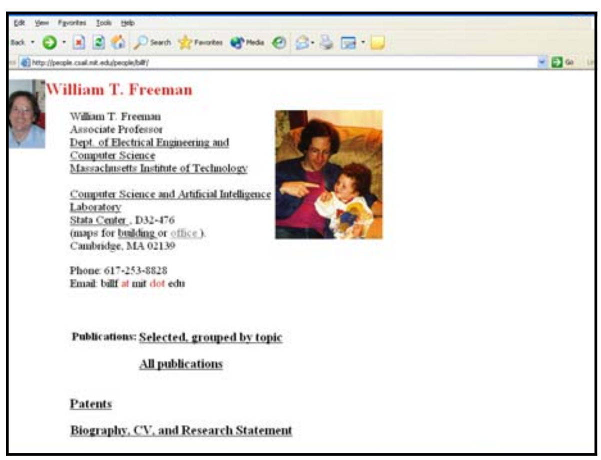
The final project may be

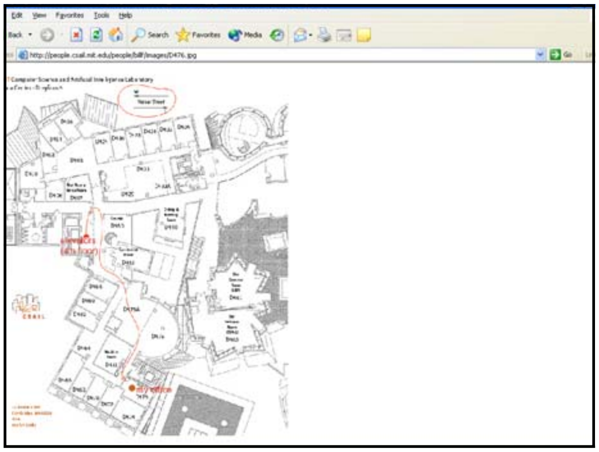
- An original implementation of a new or published idea
- A detailed empirical evaluation of an existing implementation of one or more methods
- A paper comparing three or more papers not covered in class, or surveying recent literature in a particular area

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?

# 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.

# The course, in broad categories

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

Computer vision class,
fast-forward

F»F

Images and image formation

Cameras, lenses, and sensors

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

**Figure 1.16** The first photograph on record, *la table servie*, obtained by Nicéphore Niepce in 1822. *Collection Harlinge–Viollet.*

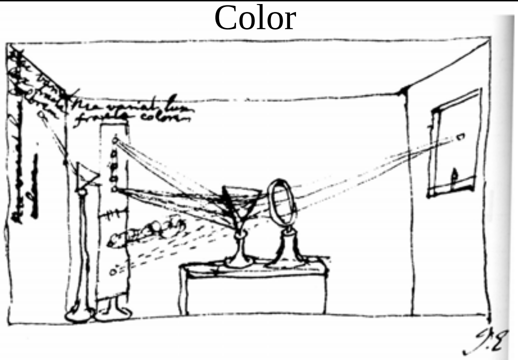
From Computer Vision, Forsyth and Ponce, Prentice-Hall, 2002.

Radiometry…not covered (see 6.801)

Wolfgang Lucht


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

Color



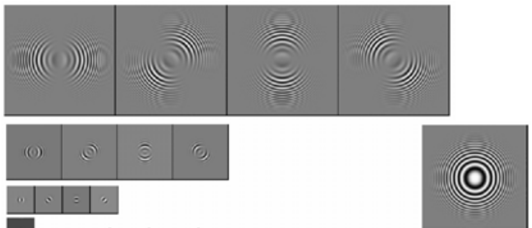
**4.1 NEWTON'S SUMMARY DRAWING** of his experiments with light. Using a point source of light and a prism, Newton separated sunlight into its fundamental components. By reconverging the rays, he also showed that the decomposition is reversible.

From Foundations of Vision, by Brian Wandell, Sinauer Assoc., 1995

Low-level vision

## Image filtering

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



*Oriented, multi-scale representation* *Image*

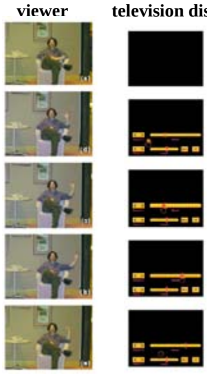
## SIFT (scale invariant feature transforms)

David Lowe, IJCV 2004

Figure 13: This example shows location recognition within a complex scene. The training images for locations are shown at the upper left and the 640x315 pixel test image taken from a different viewpoint is on the upper right. The recognized regions are shown on the lower image, with keypoints shown as squares and an outer parallelogram showing the boundaries of the training images under the affi ne transform used for recognition.

## Non-linear filtering, and applications

**viewer** **television display**



**template**

**image** **Normalized correlation**

12 Sample session of television viewing. (a) Television is off, but searching for the trigger gesture. (b) Viewer shows trigger gesture (open hand). Television set turns on and hand icon and graphics overlays appear. (c) The hand icon tracks the user's hand movement. User changes controls as with a mouse. (d) User has moved hand icon to change channel. (e) User closes hand to leave control mode. After one second, the hand icon and controls then disappear.

IEEE Computer Graphics and Applications, 18, no. 3, 1998

## Models of texture


Parametric model

Non-parametric model

**A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients**
***J. Portilla and E. Simoncelli,*** *International Journal of Computer Vision 40(1): 49-71, October 2000.*
*© Kluwer Academic Publishers.*

A. Efros and W. T Freeman, **Image quilting for texture synthesis and transfer**, SIGGRAPH 2001

## Learning and vision

## Bayesian framework for vision


"Good lord, Holmes! How did you come to know I'd seafood for lunch?"

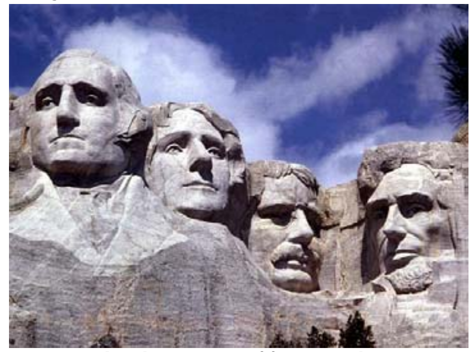
Gahan Wilson's Still Weird, Forge, 1994

## 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

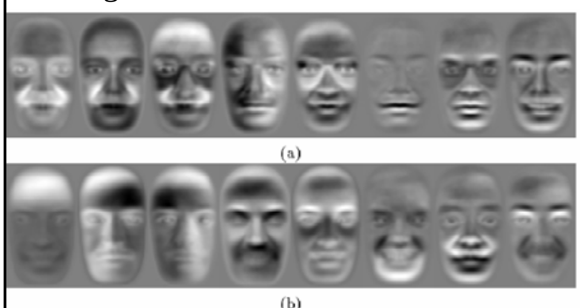


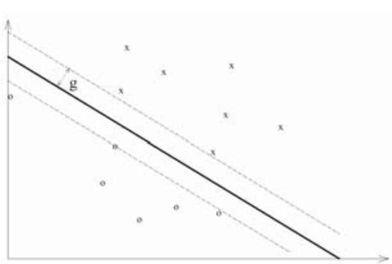

Figure 6: "Dual" Eigenfaces: (a) Intrapersonal, (b) Extrapersonal

Moghaddam, B.; Jebara, T.; Pentland, A., "Bayesian Face Recognition", *Pattern Recognition*, Vol 33, Issue 11, pps 1771-1782, November 2000

## Statistical classifiers

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

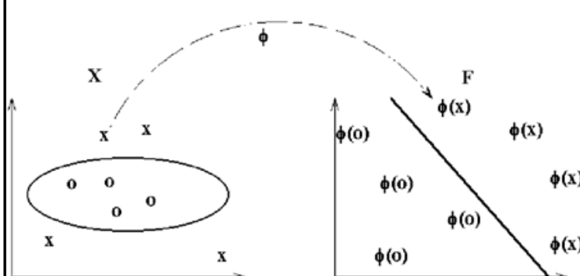
## Support vector machines and boosting




Large-margin classifier

**www.support-vector.net/nello.html**

## Support vector machines and boosting




"The kernel trick"

**www.support-vector.net/nello.html**

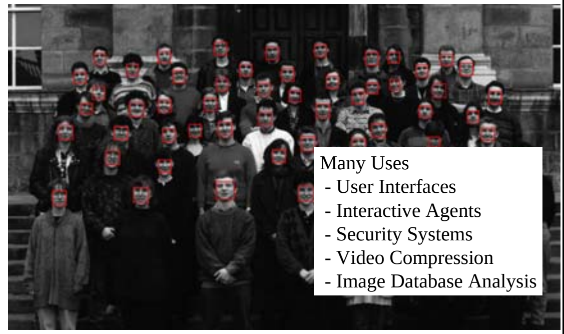
## Recent, now classic, paper on face detection:

*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




Viola and Jones, Robust object detection using a boosted cascade of simple features, CVPR 2001

## Use of context for object detection

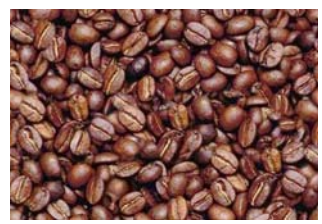

car


pedestrian

**Identical local image features!**

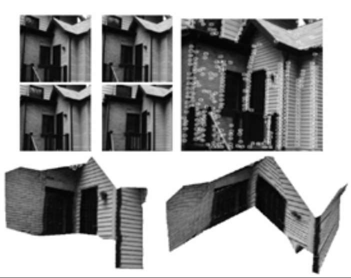
Images by Antonio Torralba

## The world, to a face detector



## Structure from Motion

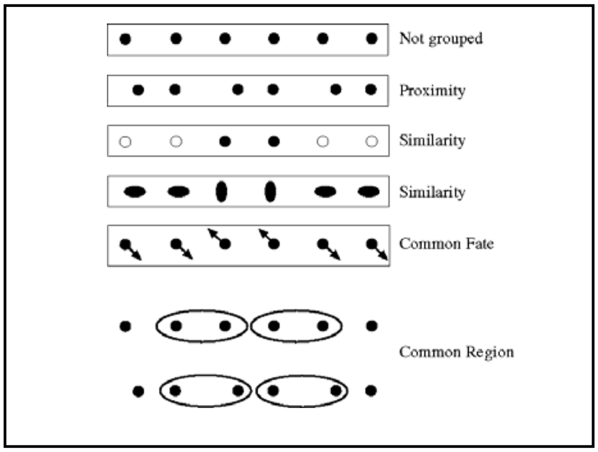
What is the shape of the scene?

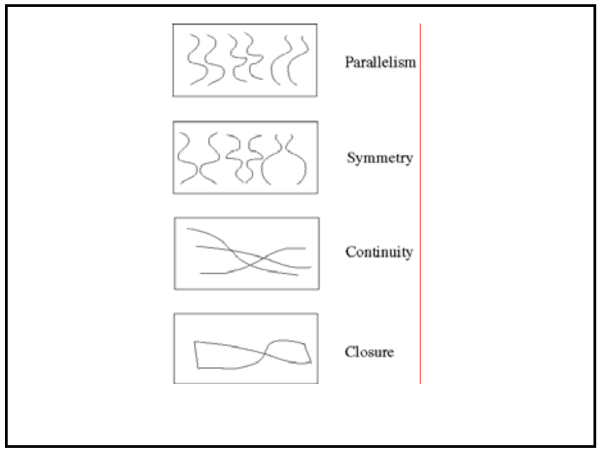


## Segmentation (perceptual grouping)

How many ways can you segment six points?

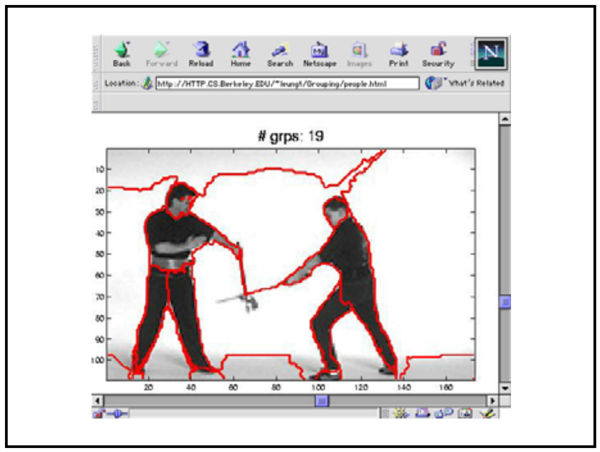
(or curves)

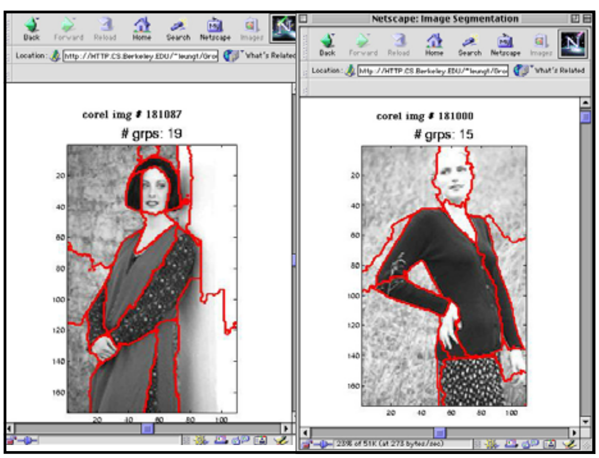


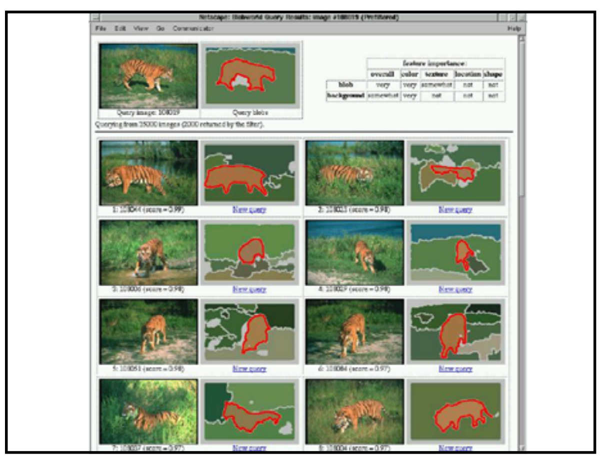




## Segmentation

- 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

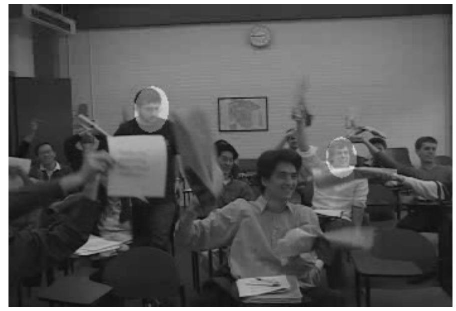
Follow objects and estimate location..
- radar / planes
- pedestrians
- cars
- face features / expressions

Many ad-hoc approaches…

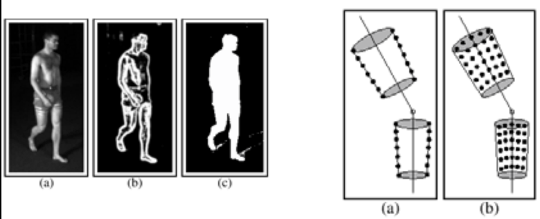
General probabilistic formulation: model density over time.

## 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

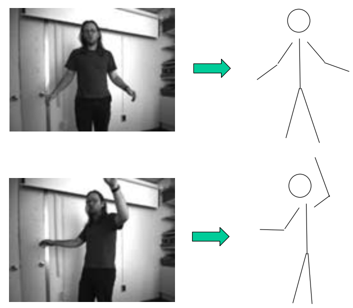


## Articulated Models




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

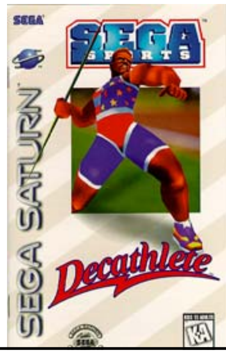
## Articulated tracking



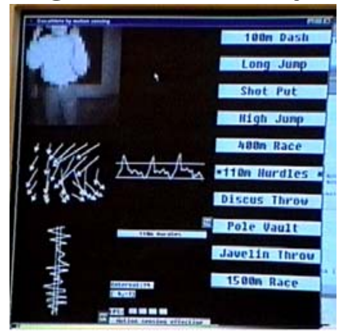
- Constrained optimization
- Coarse-to-fine part iteration
- Propagate joint constraints through each limb
- Real-time on Ghz pentium…


## Computer vision applications as ocean-going vessels

## Game: Decathlete




## Optical-flow-based Decathlete figure motion analysis

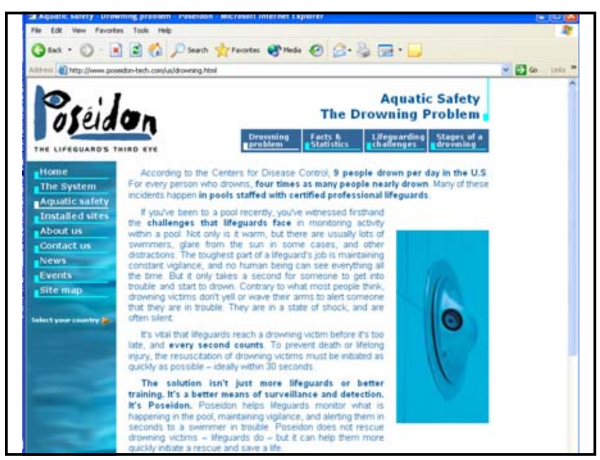


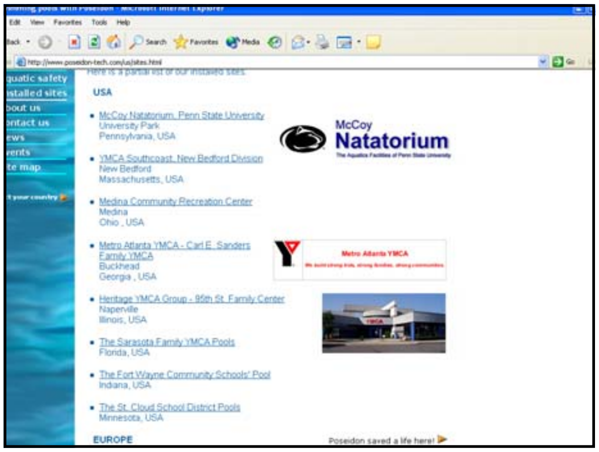

## Decathlete 100m hurdles

## 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

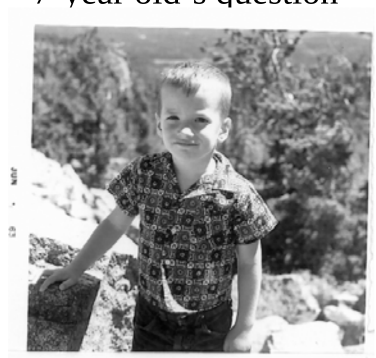
- Goal: You'll be able to go to a computer vision conference and understand what's going on in most of the presentations.
- You'll have the skills and awareness of the literature to start building the vision systems you want.

## 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

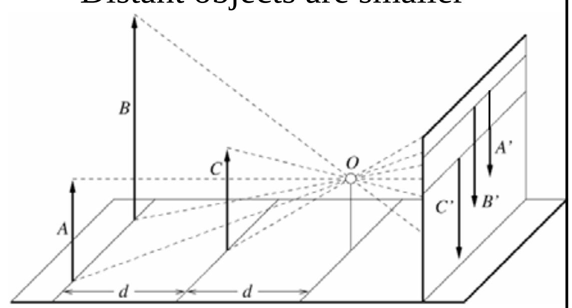


Why is there no image on a white piece of paper?

## Pinhole cameras

- Geometry

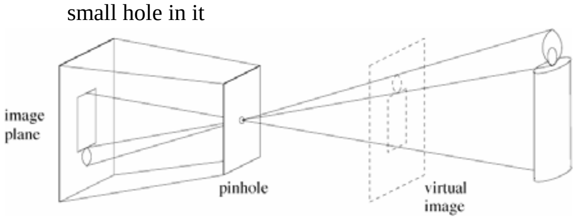
## Distant objects are smaller




Forsyth&Ponce

## Virtual image, perspective projection

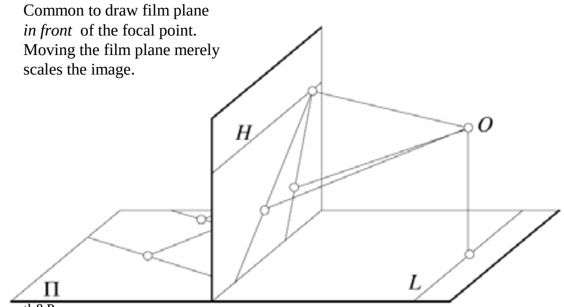
- Abstract camera model - box with a small hole in it




Forsyth&Ponce

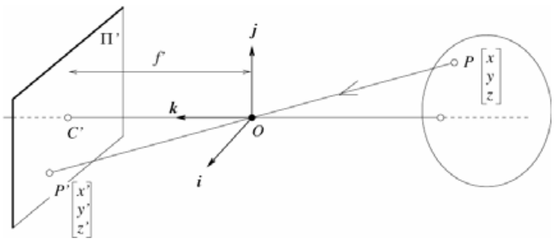
## Parallel lines meet

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




Forsyth&Ponce

## The equation of projection




## The equation of projection

- Cartesian coordinates:
  - We have, by similar triangles, that (x, y, z) -> (f x/z, f y/z, -f)
  - Ignore the third coordinate, and get

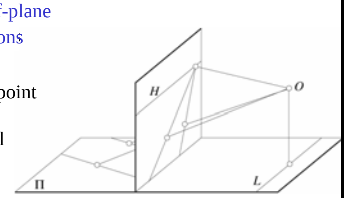
$$(x, y, z) \rightarrow (f\frac{x}{z}, f\frac{y}{z})$$

## Vanishing points

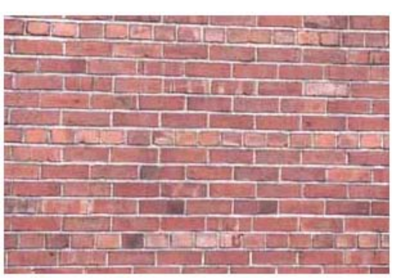
- Each set of parallel lines (=direction) meets at a different point
  - The *vanishing point* for this direction
- Sets of parallel lines on the same plane lead to *collinear* vanishing points.
  - The line is called the *horizon* for that plane
- We show this on the board…

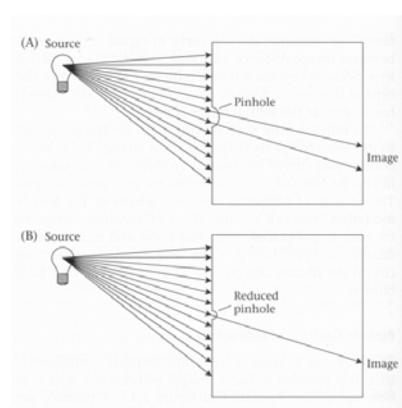
## Geometric properties of projection

- 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?

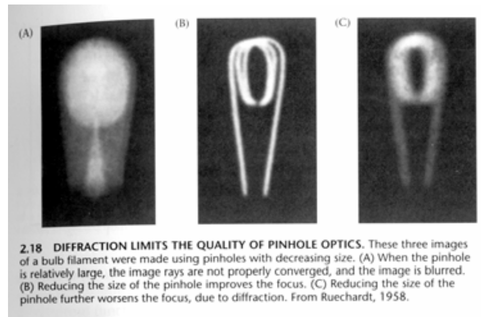





Wandell, Foundations of Vision, Sinauer, 1995

## Pinhole camera demonstrations

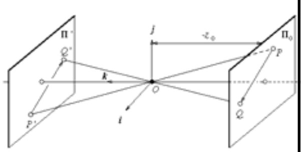
- Film camera, box, demo. Apertures, lens.
- The image is the convolution of the aperture with the scene.



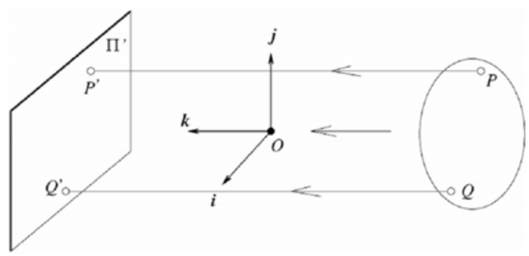

Wandell, Foundations of Vision, Sinauer, 1995

## Weak perspective

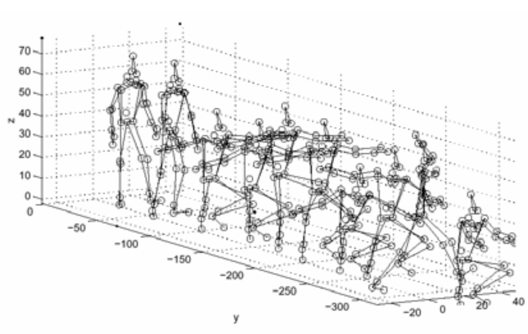
- Issue
  - perspective effects, but not over the scale of individual objects
  - collect points into a group at about the same depth, then divide each point by the depth of its group
  - Adv: easy
  - Disadv: wrong



## 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. For figural motion described by human motion basis coefficients $\vec{\alpha}$, the rendered image sequence, $\vec{y}$, is:

$$\vec{y} = PU\vec{\alpha}, \qquad (1)$$

where $P$ is the projection operator which collapses the $y$ dimension of the image sequence $U\vec{\alpha}$.

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

## Orthography can lead to analytic solutions

have our multi-dimensional gaussian,

Prior probability
$$P(\vec{\alpha}) = k_2 e^{-\vec{\alpha}'\Lambda^{-1}-\vec{\alpha}}, \quad (3)$$

where $k_2$ 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
$$P(\vec{y}|\vec{\alpha}) = k_3 e^{-|\vec{y}-PU\vec{\alpha}|^2/(2\sigma^2)}, \quad (4)$$

with normalization constant $k_3$.

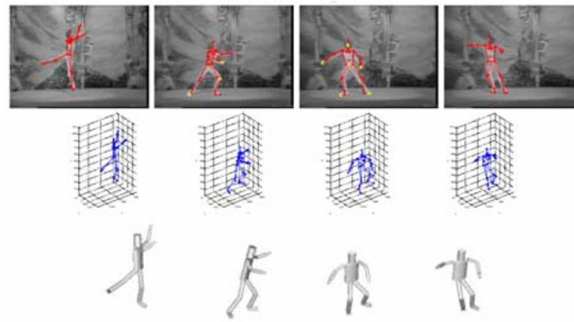
The posterior distribution is the product of these two gaussians. That yields another gaussian, with mean and covariance found by a matrix generalization of "completing the square" [7]. The squared error optimal estimate for $\alpha$ is then

$$\alpha = SU'P'(PUSU'P' + \sigma I)^{-1}(\vec{y} - (P\vec{m})) \quad (5)$$

Analytic solution for inferred 3-d motion

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

## Results



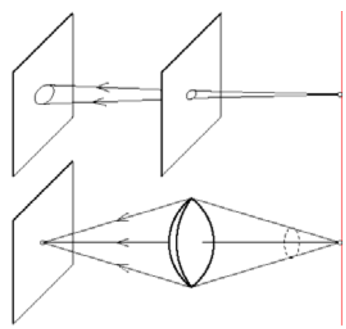
Leventon and Freeman, Bayesian Estimation of Human Motion, MERL TR98-06

## 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 perpective projection are ignored, by assuming orthographic projection. …"

**Reviewer's comments**

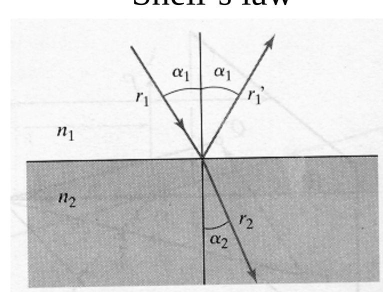
## The reason for lenses



## Water glass refraction

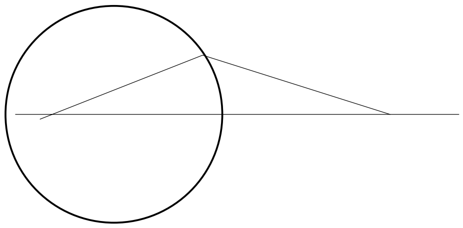
**http://data.pg2k.hd.org/_exhibits/natural-science/cat-black-and-white-domestic-short-hair-DSH-with-nose-in-glass-of-water-on-bedside-table-tweaked-mono-1-AJHD.jpg**

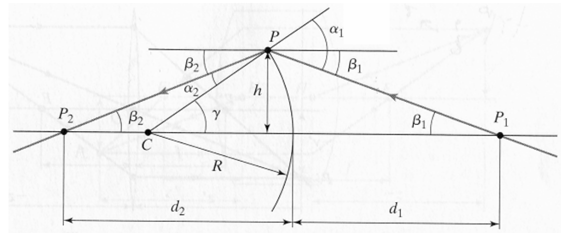
## Snell's law




$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2)$$

## Spherical lens

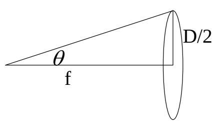





Forsyth and Ponce

## First order optics

$$\sin(\theta) \approx \theta$$




$$\theta \approx \frac{D/2}{f}$$

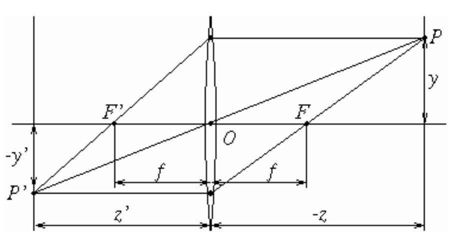
## Paraxial refraction equation

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

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

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

## The thin lens, first order optics




$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$

$$f = \frac{R}{2(n-1)}$$

Forsyth&Ponce

US Navy Manual

## 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

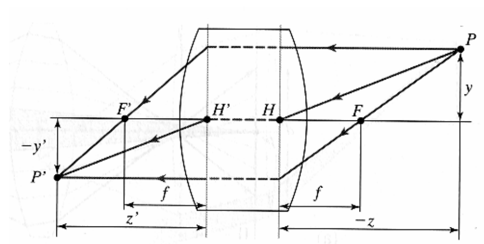
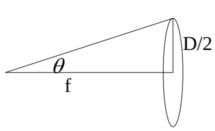



**Figure 1.11** A simple thick lens with two spherical surfaces.

Forsyth&Ponce

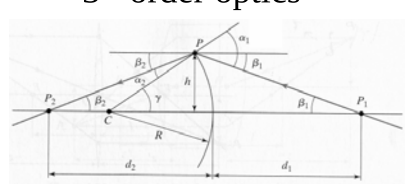
## Third order optics

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




$$\theta \approx \frac{D/2}{f} - \frac{\left(\frac{D/2}{f}\right)^3}{6}$$

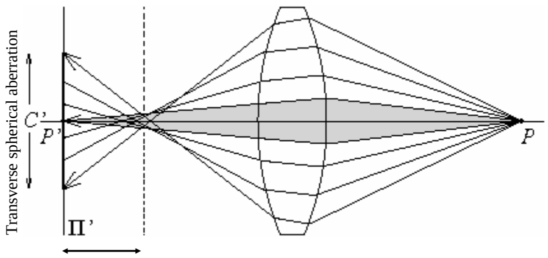
## Paraxial refraction equation, 3rd order optics



$$\frac{n_1}{d_1}+\frac{n_2}{d_2}=\frac{n_2-n_1}{R}+h^2\left[\frac{n_1}{2d_1}\left(\frac{1}{R}+\frac{1}{d_1}\right)^2+\frac{n_2}{2d_2}\left(\frac{1}{R}-\frac{1}{d_2}\right)^2\right]$$

Forsyth&Ponce

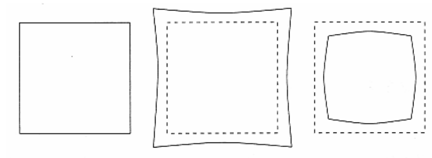
## Spherical aberration (from 3rd order optics




Forsyth&Ponce

## Other 3rd order effects

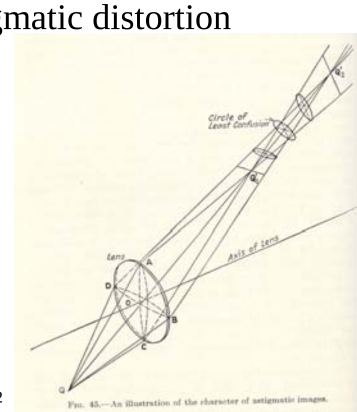
- Coma, astigmatism, field curvature, distortion.



no distortion | pincushion distortion | barrel distortion

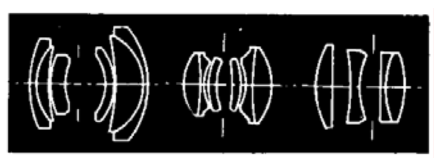
Forsyth&Ponce

## Astigmatic distortion




Hardy & Perrin,
The Principles of Optics, 1932

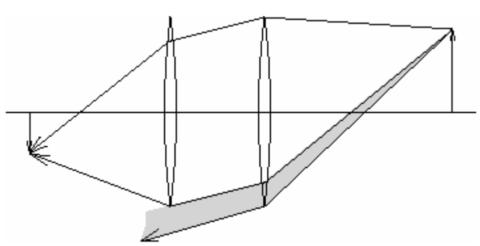
## Lens systems



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

Forsyth&Ponce

## Vignetting



Forsyth&Ponce

## 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

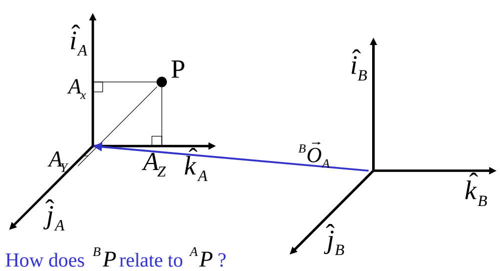
- Want to make images
- Pinhole camera models the geometry of perspective projection
- Lenses make it work in practice
- Models for lenses
  - Thin lens, spherical surfaces, first order optics
  - Thick lens, higher-order optics, vignetting.

## Next

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

## Translation

$^{A}P = \begin{pmatrix} A_x \\ A_Y \\ A_Z \end{pmatrix} \quad ^{B}P = \begin{pmatrix} B_x \\ B_Y \\ B_Z \end{pmatrix}$

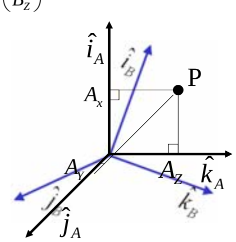



How does $^{B}P$ relate to $^{A}P$ ?

$$^{B}P = {}^{A}P + {}^{B}O_A$$

## Rotation

$^{A}P = \begin{pmatrix} A_x \\ A_Y \\ A_Z \end{pmatrix} \quad ^{B}P = \begin{pmatrix} B_x \\ B_Y \\ B_Z \end{pmatrix}$

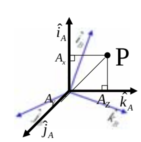



How does $^{B}P$ relate to $^{A}P$ ?

$$^{B}P = {}^{B}_{A}R \; {}^{A}P$$

## Find the rotation matrix

Project $\overrightarrow{OP} = \begin{pmatrix} \hat{i}_A & \hat{j}_A & \hat{k}_A \end{pmatrix} \begin{pmatrix} A_X \\ A_Y \\ A_Z \end{pmatrix}$




onto the B frame's coordinate axes.

$$\begin{pmatrix} B_X \\ B_Y \\ B_Z \end{pmatrix} = \begin{pmatrix} \hat{i}_B \bullet \hat{i}_A A_X & \hat{i}_B \bullet \hat{j}_A A_Y & \hat{i}_B \bullet \hat{k}_A A_Z \\ \hat{j}_B \bullet \hat{i}_A A_X & \hat{j}_B \bullet \hat{j}_A A_Y & \hat{j}_B \bullet \hat{k}_A A_Z \\ \hat{k}_B \bullet \hat{i}_A A_X & \hat{k}_B \bullet \hat{j}_A A_Y & \hat{k}_B \bullet \hat{k}_A A_Z \end{pmatrix}$$

## Rotation matrix

this

$$\begin{pmatrix} B_X \\ B_Y \\ B_Z \end{pmatrix} = \begin{pmatrix} \hat{i}_B \bullet \hat{i}_A A_X & \hat{i}_B \bullet \hat{j}_A A_Y & \hat{i}_B \bullet \hat{k}_A A_Z \\ \hat{j}_B \bullet \hat{i}_A A_X & \hat{j}_B \bullet \hat{j}_A A_Y & \hat{j}_B \bullet \hat{k}_A A_Z \\ \hat{k}_B \bullet \hat{i}_A A_X & \hat{k}_B \bullet \hat{j}_A A_Y & \hat{k}_B \bullet \hat{k}_A A_Z \end{pmatrix}$$

implies $^{B}P = {}^{B}_{A}R \; {}^{A}P$

where

$${}^{B}_{A}R = \begin{pmatrix} \hat{i}_B \bullet \hat{i}_A & \hat{i}_B \bullet \hat{j}_A & \hat{i}_B \bullet \hat{k}_A \\ \hat{j}_B \bullet \hat{i}_A & \hat{j}_B \bullet \hat{j}_A & \hat{j}_B \bullet \hat{k}_A \\ \hat{k}_B \bullet \hat{i}_A & \hat{k}_B \bullet \hat{j}_A & \hat{k}_B \bullet \hat{k}_A \end{pmatrix}$$

## Translation and rotation

Let's write $^{B}P = {}^{B}_{A}R \; {}^{A}P + {}^{B}O_A$

as a single matrix equation:

$$\begin{pmatrix} B_X \\ B_Y \\ B_Z \\ 1 \end{pmatrix} = \begin{pmatrix} - & - & - & | \\ - & {}^{B}_{A}R & - & {}^{B}O_A \\ - & - & - & | \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_X \\ A_Y \\ A_Z \\ 1 \end{pmatrix}$$