CMSC 426: Image Processing (Computer Vision)

David Jacobs

Today's class

- What is vision
- What is computer vision
- Layout of the class

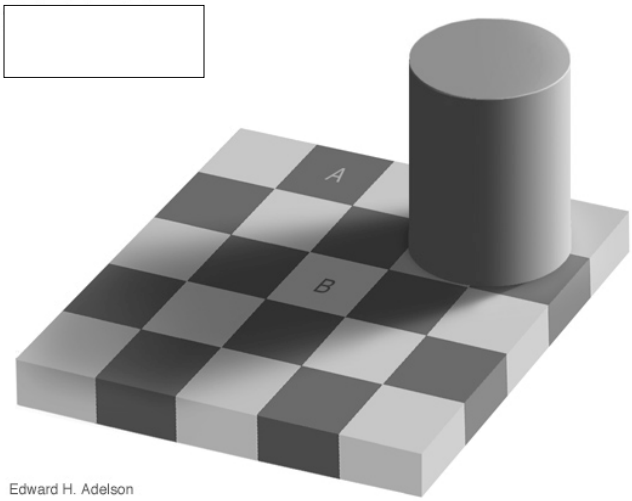
Vision

- ``to know what is where, by looking.'' (Marr).
- Where
- What

Why is Vision Interesting?

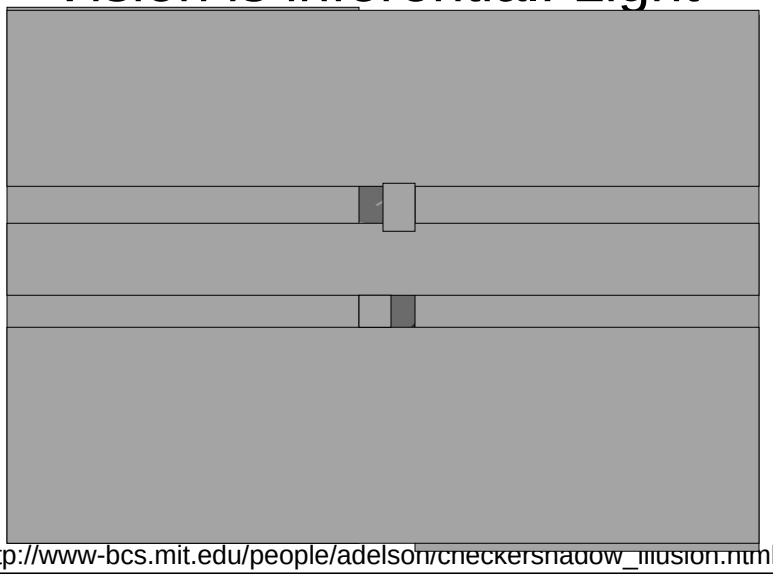
- Psychology
  - ~ 50% of cerebral cortex is for vision.
  - Vision is how we experience the world.
- Engineering
  - Want machines to interact with world.
  - Digital images are everywhere.

Vision is inferential: Light



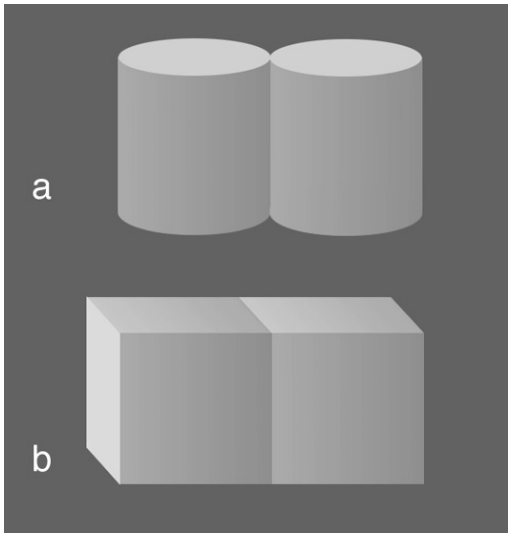
(http://www-bcs.mit.edu/people/adelson/checkershadow_illusion.html)

Vision is inferential: Light



(http://www-bcs.mit.edu/people/adelson/checkershadow_illusion.html)

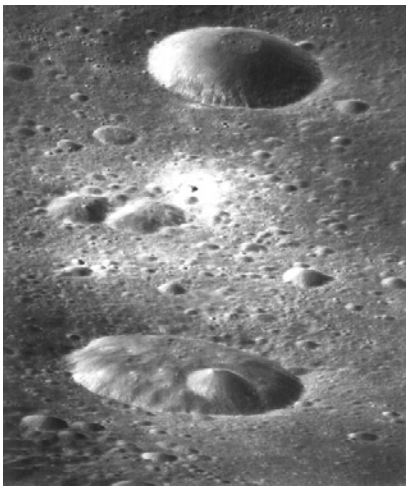
Vision is Inferential: Geometry



Vision is Inferential: Geometry

movie

Vision is Inferential: Prior Knowledge



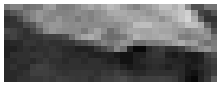
Vision is Inferential: Prior Knowledge

Computer Vision

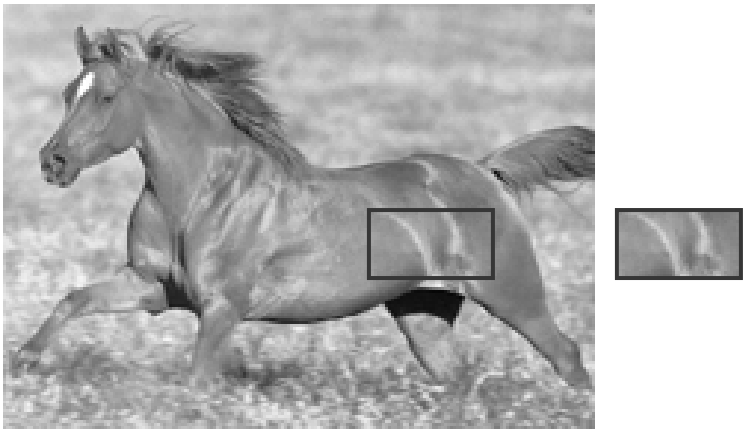
- Inference → Computation
- Building machines that see
- Applying computation to images
- Modeling biological perception

A Quick Tour of Computer Vision

Boundary Detection: Local cues



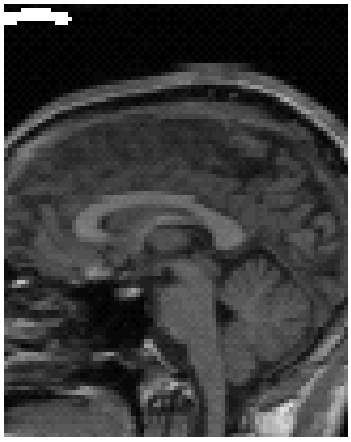
Boundary Detection: Local cues

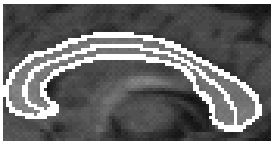


Boundary Detection

GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts

Boundary Detection





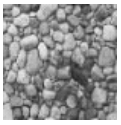
Finding the Corpus Callosum

(G. Hamarneh, T. McInerney, D. Terzopoulos)

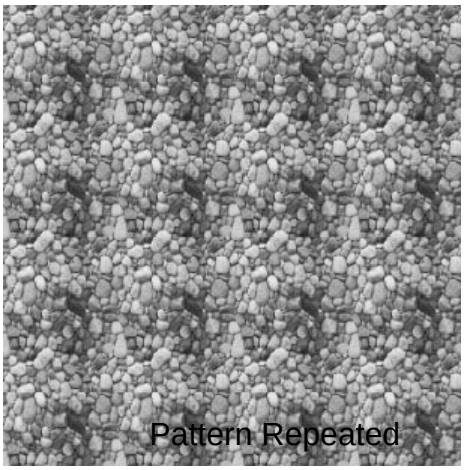
Texture in Boundary Detection

(Sharon, Balun, Brandt, Basri)

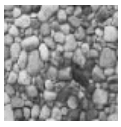
Texture Synthesis



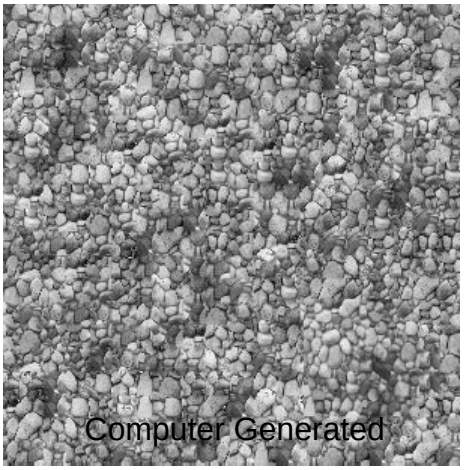
Photo



Texture

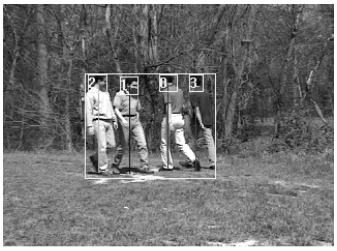


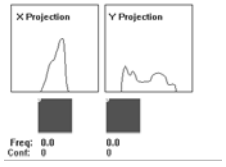
Photo



Background Subtraction

Tracking and Activity Classification





Tracking

(Comaniciu and Meer)

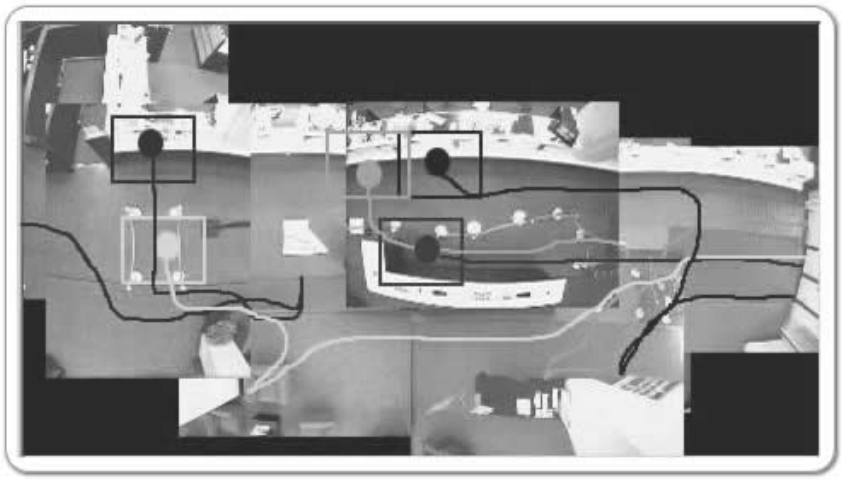
Understanding Action

Tracking and Understanding

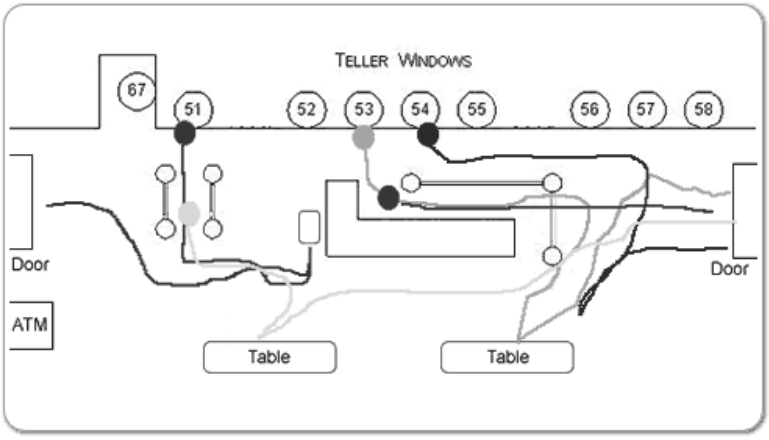
(www.brickstream.com)

Tracking

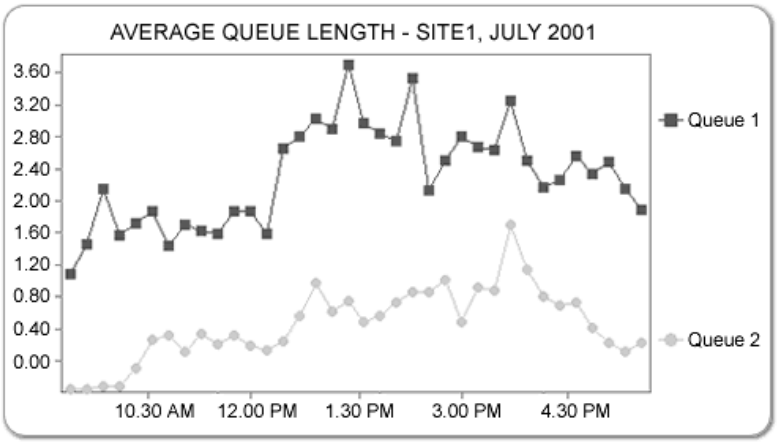
Tracking



Tracking



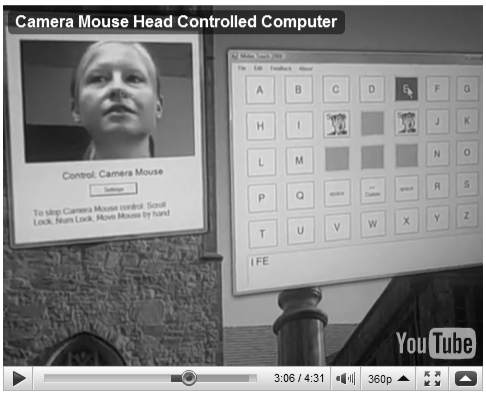
Tracking



Tracking: Application

About Camera Mouse

Camera Mouse is a program that allows you to control the mouse pointer on a Windows computer just by moving your head.

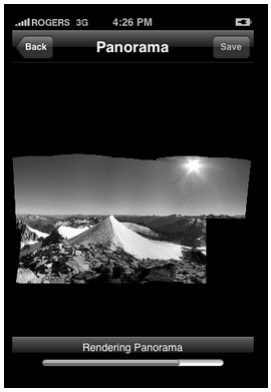


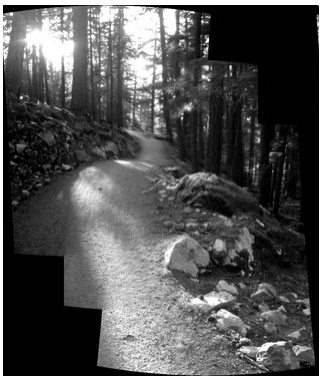
Testimonials

"One of our staff, who has cerebral palsy, has already had a try at the programme. We are very keen to try it on other young people at our institute and we are sure it is going to make life easier for many of them... Many many thanks."
-- India

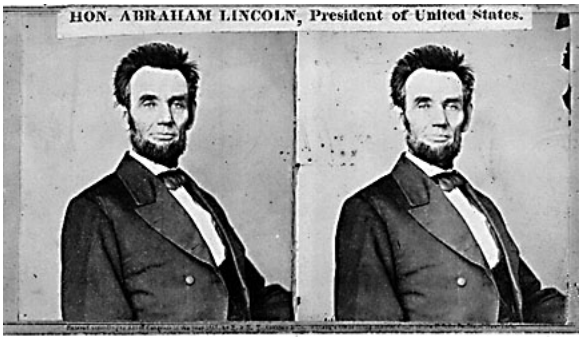
"thank you for the invention of the camera mouse."
-- Malaysia

Image Alignment



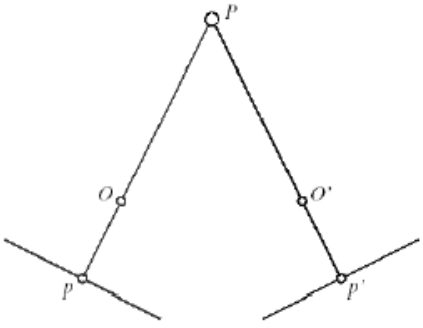


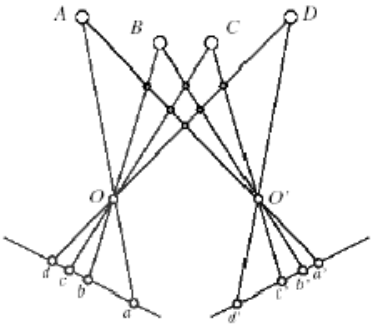
Stereo



Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923 (Slide courtesy Steve Seitz)

Stereo



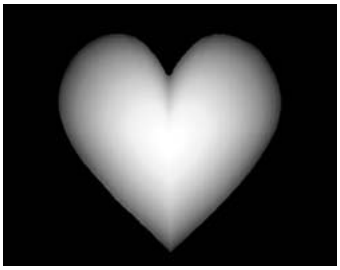


http://www.ai.mit.edu/courses/6.801/lect/lect01_darrell.pdf

Stereo

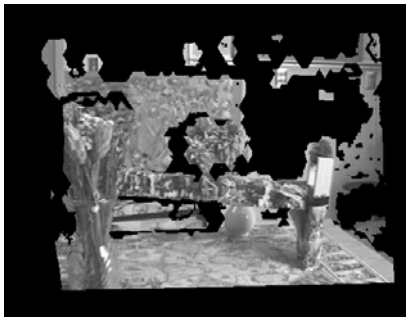
http://www.magiceye.com/

Stereo



http://www.magiceye.com/

Motion



Courtesy Yiannis Aloimonos

Photosynth

- http://photosynth.net/view.aspx?cid=368eeded-d574-4430-867b-94e003a354a7

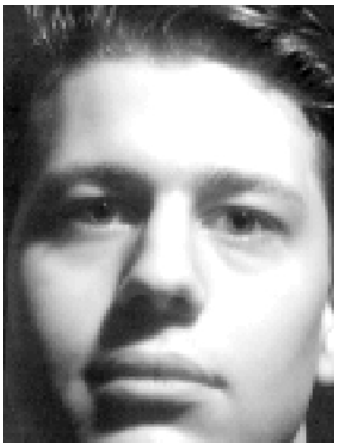
Motion - Application

(www.realviz.com)

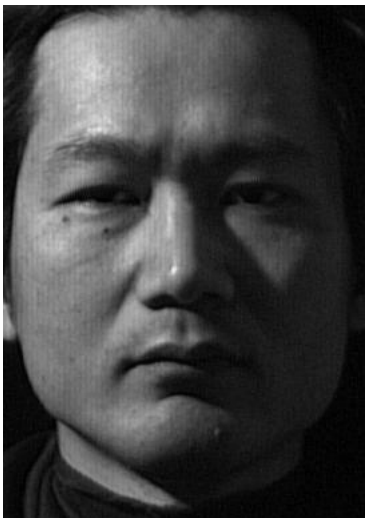
Pose Determination

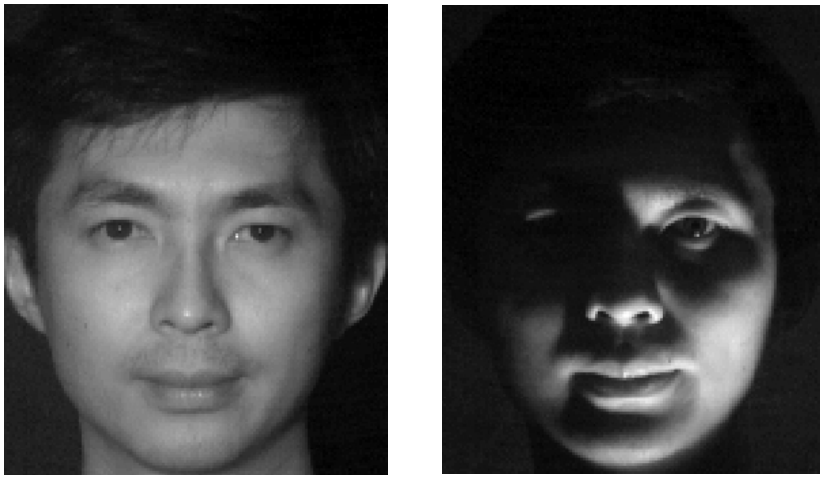
Visually guided surgery

Recognition - Shading

Lighting affects appearance



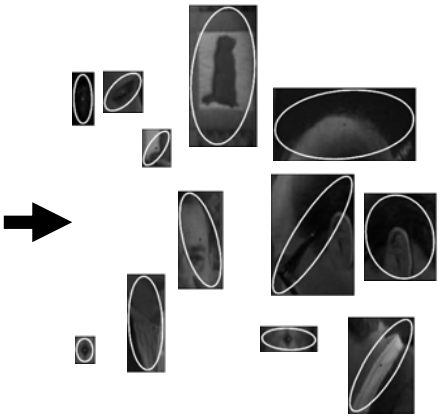
Object search in video: why is it hard?

- an object's imaged appearance varies …
  - scale changes
  - lighting changes
  - viewpoint changes
  - partial occlusion

- sheer amount of data
  - feature length movie ~ 100,000 -150,000 frames

Slide by Josef Sivic

Visual description – visual words

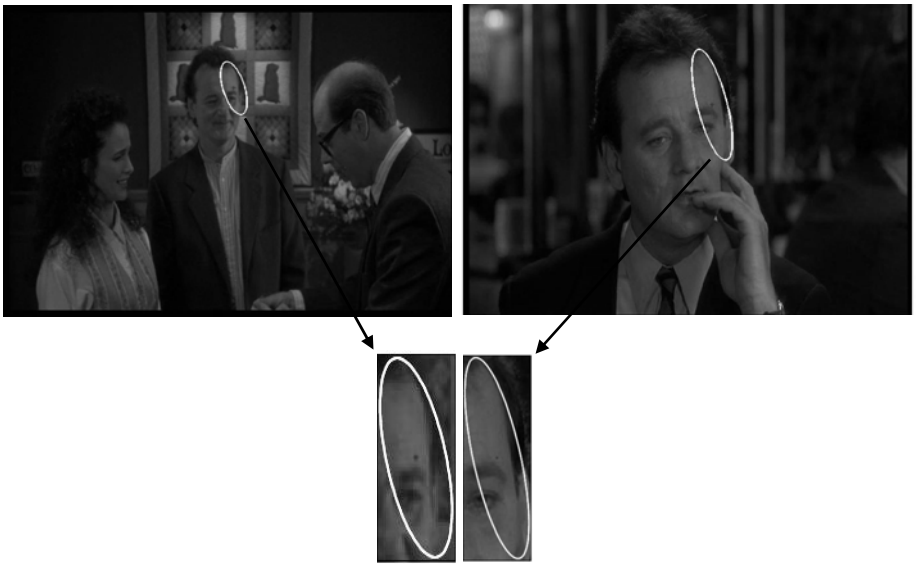


Image

visual nouns

Slide by Josef Sivic

Visual vocabulary unaffected by scale and viewpoint



The same visual word

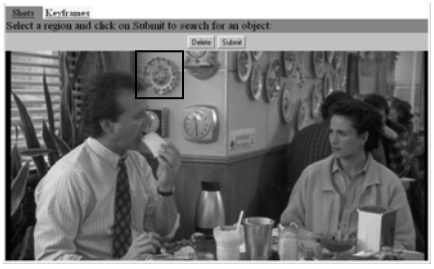
Slide by Josef Sivic

Image representation using visual words

Use efficient google like search on visual words

Slide by Josef Sivic

Example : Groundhog Day



Video Google, Sivic & Zisserman, ICCV 2003 (Slide from Antonio Torralba)

retrieved shots

Slide by Josef Sivic

Example: Casablanca

retrieved shots

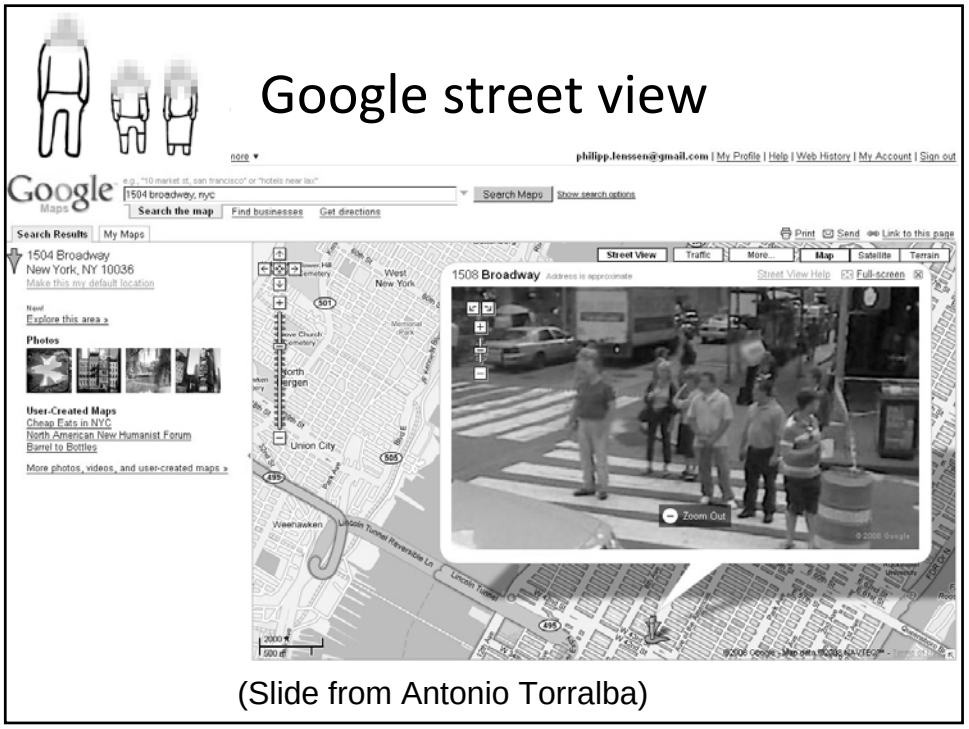
(Slide from Antonio Torralba)

Slide by Josef Sivic

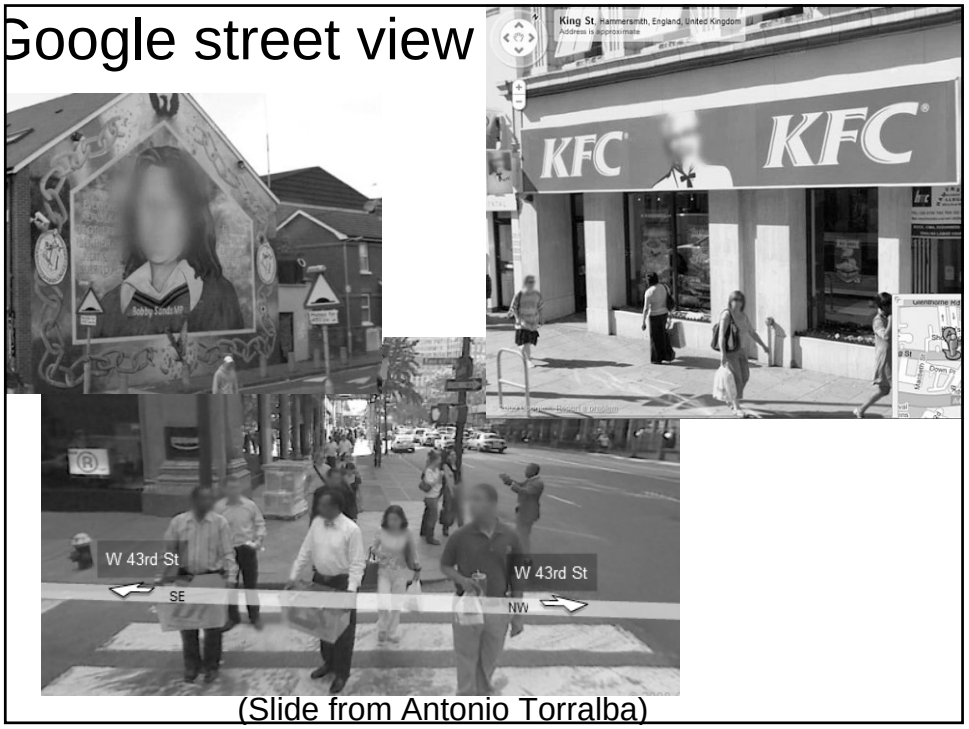
Face detection

[Face priority AE] When a bright part of the face is too bright

(Slide from Antonio Torralba)



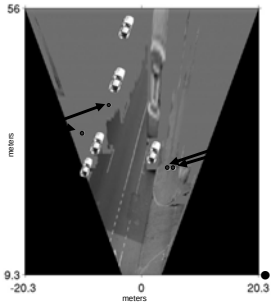
(Slide from Antonio Torralba)

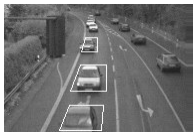


(Slide from Antonio Torralba)

Assisted driving

Pedestrian and car detection





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

Lane detection

(Slide from Antonio Torralba)

What we will cover

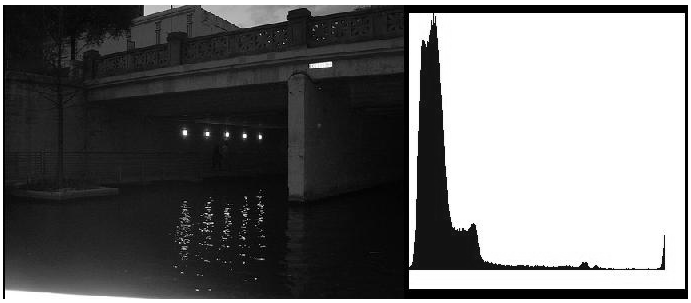
- Vision topics
  - Edge/Boundary Detection
  - Texture matching
  - Segmentation
  - Background Subtraction
  - Image Registration and Matching
  - Feature Detection
  - Optical Flow and Motion Fields
  - Stereo Reconstruction
  - Motion Understanding
  - Object Classification

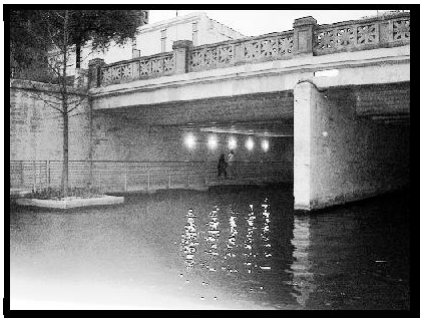
What we will cover

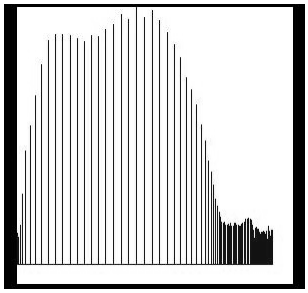
- Core Concepts
  - Histograms/Probability Distributions
  - Correlation and convolution
  - Interpolation
  - Combinatorial Optimization
  - Gradients
  - 3D Geometry

What we will cover

- Projects







Efros and Leung

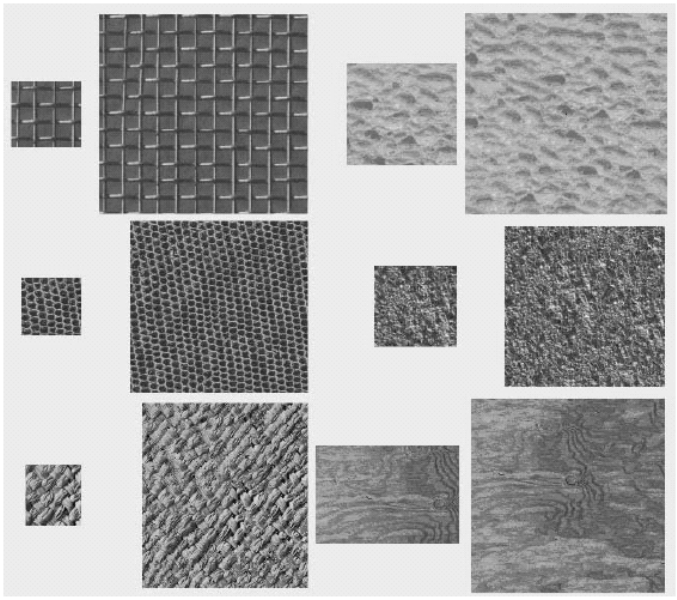
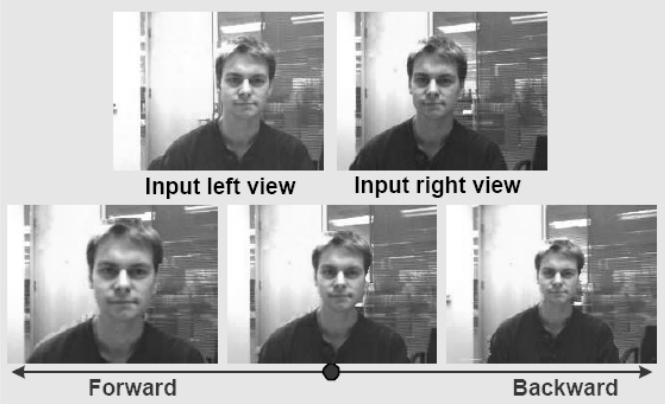


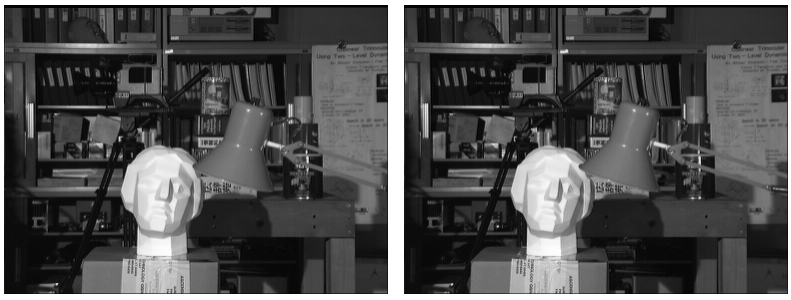
Figure 4: Tessellated spherical panorama covering the north pole (constructed from 54 images). The white triangles at the top are the parts of the texture map not covered in the 3D tesselated globe model (due to triangular elements at the poles).

Using Depth in IBR

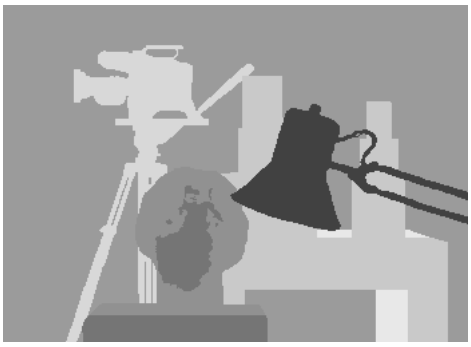
- Example, generating views for teleconfer-encing, with depth from stereo.



Stereo



Results with graph cuts



Graph Cuts

Ground truth

Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, International Conference on Computer Vision, September 1999.

(Seitz)

Approaches to Vision

Vision depends on:

- Geometry
- Physics
- The nature of objects in the world
  (This is the hardest part).

Modeling + Algorithms

- Build a simple model of the world
  (eg., flat, uniform intensity).
- Find provably good algorithms.
- Experiment on real world.
- Update model.

Problem: Too often models are simplistic or intractable.

Bayesian inference

- Bayes law: P(A|B) = P(B|A)*P(A)/P(B).
- P(world|image) =
  P(image|world)*P(world)/P(image)
- P(image|world) is computer graphics
  - Geometry of projection.
  - Physics of light and reflection.
- P(world) means modeling objects in world.
  Leads to statistical/learning approaches.

Problem: Too often probabilities can't be known and are invented.

Engineering

- Focus on definite tasks with clear requirements.
- Try ideas based on theory and get experience about what works.
- Try to build reusable modules.

Problem: Solutions that work under specific conditions may not generalize.

Marr

- Theory of Computation
- Representations and algorithms
- Implementations.

- Primal Sketch
- 2½D Sketch
- 3D Representations

Problem: Are things really so modular?

The State of Computer Vision

- Science
  - Study of intelligence seems to be hard.
  - Some interesting fundamental theory about specific problems.
  - Limited insight into how these interact.

The State of Computer Vision

- Technology
  - Interesting applications: inspection, graphics, security, internet….
  - Some successful companies. Largest ~100-200 million in revenues. Many in-house applications.
  - Commercial products beginning to emerge.
    - Cameras, photoshop, cars (soon?)
  - Future: growth in digital images exciting.

Related Fields

- Graphics. "Vision is inverse graphics".
- Visual perception.
- Neuroscience.
- AI
- Learning
- Math: eg., geometry, stochastic processes.
- Optimization.

Contact Info

Prof: David Jacobs Email: djacobs@cs.umd.edu

Office: Room 4421, A.V. Williams Building (Next to CSIC).

Office Hours: Wed., 10-11, Thurs. 3-4

TA: Matthew Mauriello Email: mattm@cs.umd.edu

Office Hours: Tues. 2-3, Wed. 4-5 (AVW 1151)

Tools Needed for Course

- Math
  - Calculus
  - Geometry
  - Probability
- Computer Science
  - Algorithms
  - Programming, we'll use Matlab.
- Signal Processing (we'll teach a little).

Syllabus

Course Organization

- Reading: Class notes, no required text.
- 8 Problem sets, six with programming.
- Midterm, Final Exam.
- Grading: Problem sets 40%, midterm 20%; final 40%.
- Web page: www.cs.umd.edu/~djacobs/CMSC426/CMSC426_12.htm

Questions?