CSE/EE 486: Computer Vision I
Fall 2007 Course Overview

Instructor: Dr. Robert Collins, *email:* rcollins@cse.psu.edu
Office: IST 354H
Office Hours: TBA

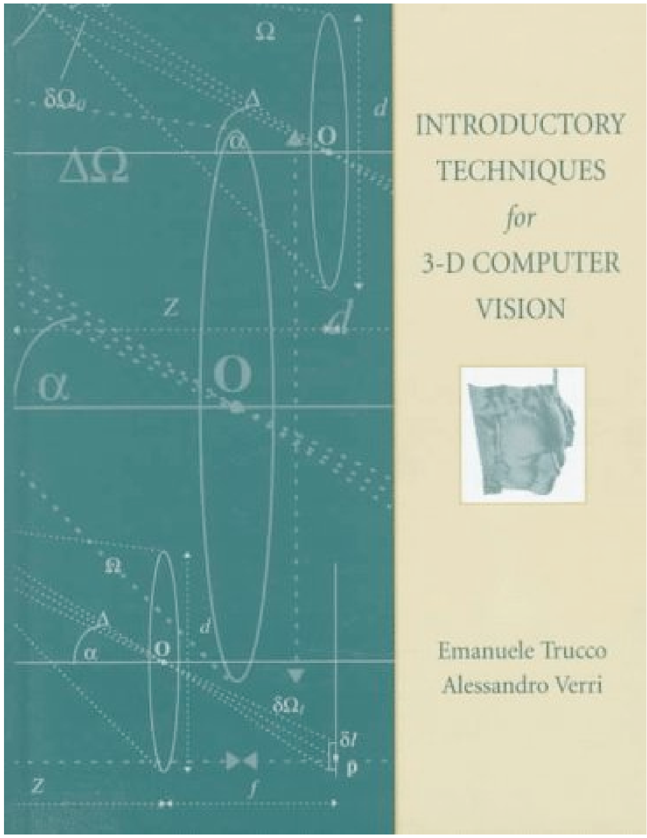
Teaching Assistant: Yaman Aksu, *email:* ya1@psu.edu
Office Hours: TBA

Class Schedule: M W F, 11:15 -- 12:05 Willard 260
Credits: 3

Robert Collins
CSE486, Penn State

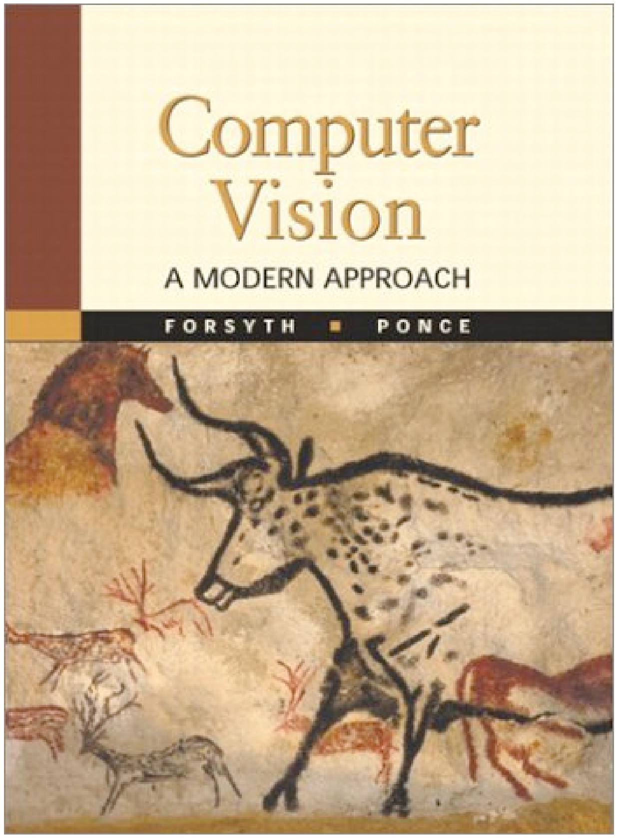
Textbook

required



Introductory Techniques for 3-D Computer Vision by E. Trucco and A. Verri, Prentice Hall, 1998.

optional



Computer Vision: A Modern Approach, by D.Forsyth and J.Ponce, Prentice Hall, 2002.

Fall 2007 Syllabus (excerpts)

Grading:

- *Homework Assignments (6):* 30%
- *Project Assignments (3):* 30%
- *In-class Midterm Exams (2):* 20%
- *Final Exam (Comprehensive):* 20%

Individual exams and assignments are not scaled or curved. However, I typically scale the overall numeric course scores (computed from everything) before assigning a letter grade at the end of the course.

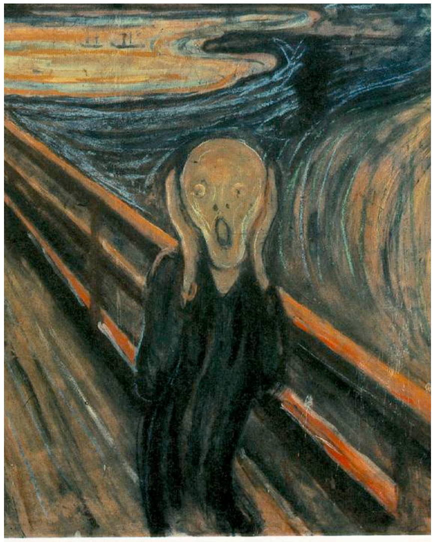
See full syllabus posted on Angel course web site

https://cms.psu.edu/

Programming Groups

excerpts from syllabus

Projects are team efforts, performed in groups of three people. **The deadline to form these groups is Sep 10.** *Students not belonging to a group will be assigned one by alphabetical order*. From past experience, things seem to work out better when you form your own group, rather than being assigned to one, even if it means you have to take the initiative to introduce yourself to someone you don't know (yikes!). For some reason, working in assigned groups for the projects seems to be a considerable source of *angst* each semester. Please heed my call, and make a serious effort to put your own teams together.



Each group submits code, a written report, and a short description of what each member of the group contributed to the project. Typically, all members of the team receive the same grade for the submission, but a member who clearly is not contributing will receive a lower grade.

Robert Collins

CSE/EE 486 Fall 2007 Course Calendar

tentative

3 days in August! intro to vision — intensity surfaces — linear operators [H1

SEPTEMBER 2007

Monday	Tuesday	Wednesday	Thursday	Friday
3	4	5 smoothing	6	7 edges H1]
10 corners [P1	11	12 template matching	13	14 intro to stereo [H2
17 stereo alg 1	18	19 stereo alg 2	20	21 pyramids H2]
24 LoG/zero-cross	25 P1]	26 review	27	28 midterm 1 M1

OCTOBER 2007

Monday	Tuesday	Wednesday	Thursday	Friday
1 extrinsics	2	3 intrinsics [H3	4	5 transformations 1
8 transformations 2	9	10 RANSAC [P2	11	12 homography 1 H3]
15 homography 2	16	17 mosaicing	18	19 suicide day
22 epipolar geom [H4	23	24 E/F matrices	25 P2]	26 8-pt alg
29 stereo reconstruct	30	31 review H4]		

NOVEMBER 2007

Monday	Tuesday	Wednesday	Thursday	Friday
			1	2 midterm 2 M2
5 camera motion	6	7 flow estimate 1 [H5	8	9 flow estimate 2
12 change detect [P3	13	14 SFM	15	16 color&light movie H5]
19	20	21	22	23
26 color&light	27	28 color histograms [H6	29	30 tracking

DECEMBER 2007

Monday	Tuesday	Wednesday	Thursday	Friday
3 tracking LK	4 P3]	5 tracking MS	6	7 ObjRec SIFT H6]
10 ObjRec PCA	11	12 texture synth	13	14 review
17	18	19	20	21
Final Exam Week *Final Exam Week *Final Exam Week				
24	25	26	27	28

Lecture 1: Introduction to Computer Vision

Readings T&V: Section 1, Section 2.1-2.2

What is Vision?

“Vision is the act of knowing what is where by looking.” --Aristotle

Special emphasis: relationship between 3D world and a 2D image. Location and identity of objects.

What is Computer Vision?

It is related to, but not equivalent to:

- Photogrammetry
- Image Processing
- Artificial Intelligence

Why study Computer Vision?

- 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
- Various deep and attractive scientific mysteries
  - how does object recognition work?
- Greater understanding of human vision

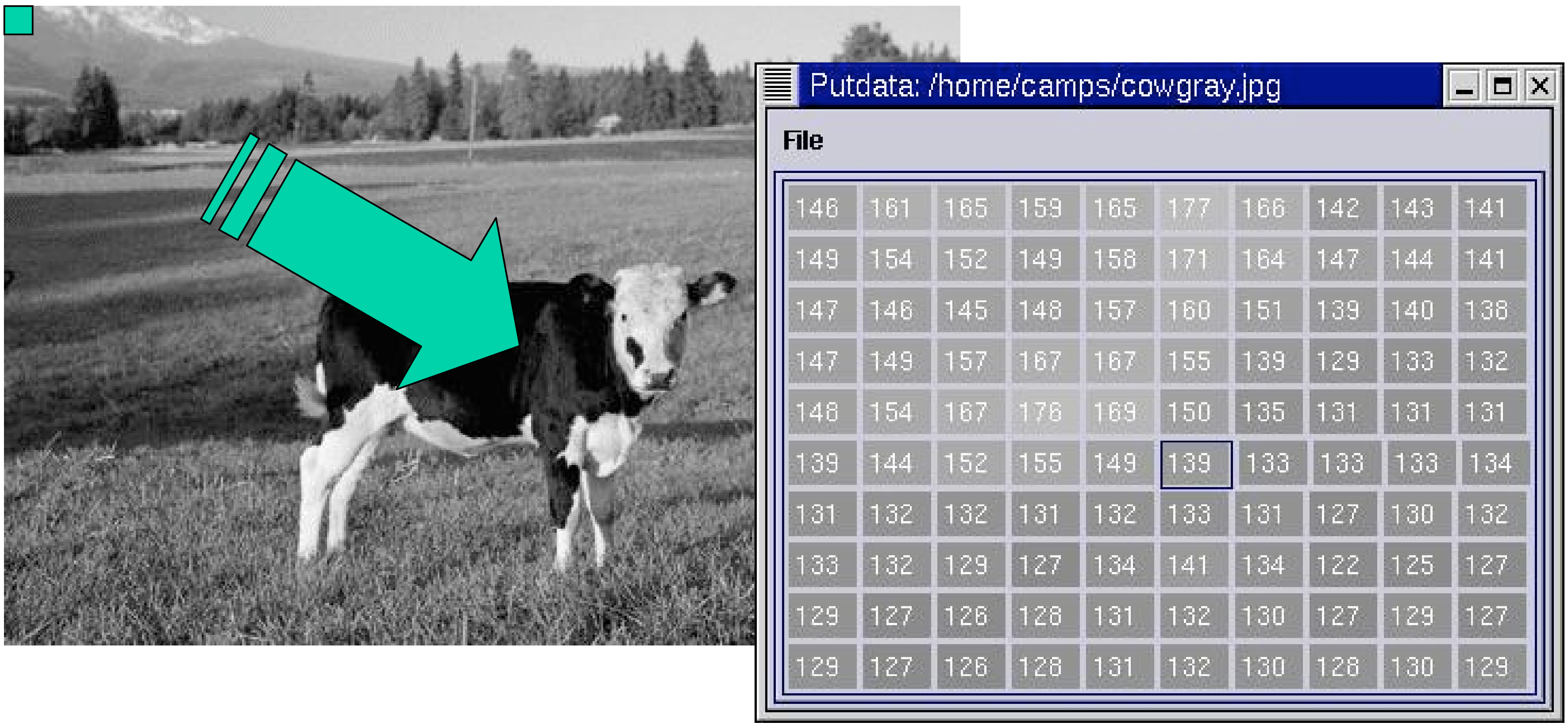
Course Goals and Objectives

- Introduce the fundamental problems of computer vision.
- Introduce the main concepts and techniques used to solve those problems.
- Enable students to implement vision algorithms
- Enable students to make sense of the vision literature

Robert Collins
CSE486, Penn State

Input: Digital Images

2D arrays (matrices) of numbers:

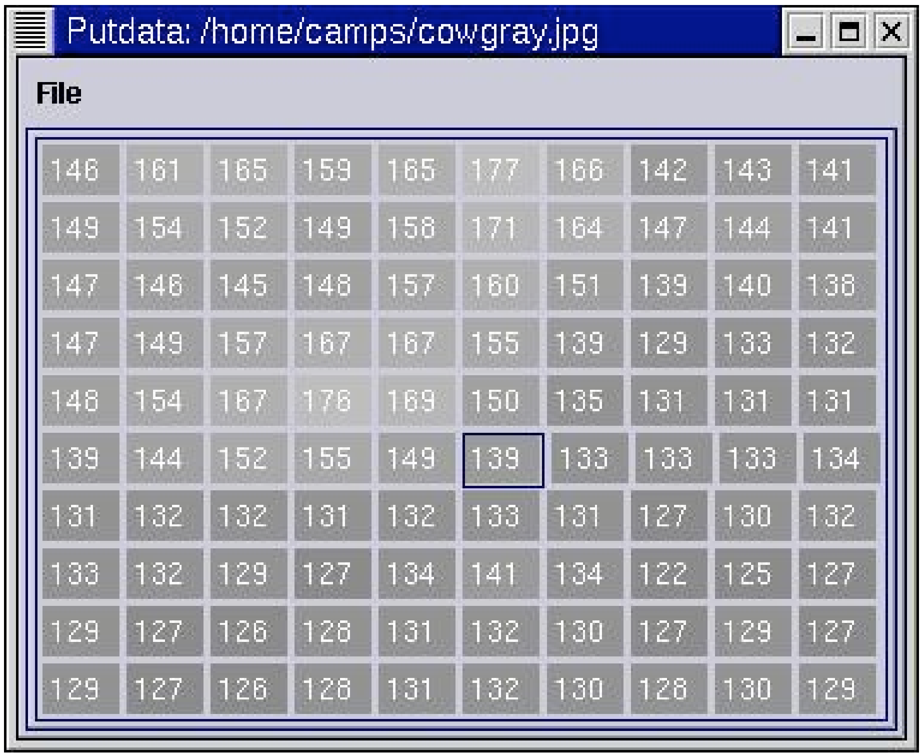


If color image, we have 3 arrays - red, green, blue

Robert Collins
CSE486, Penn State

Why is Computer Vision Hard?

We are trying to infer things about the world from an array of numbers



146	161	165	159	165	177	166	142	143	141
149	154	152	149	158	171	164	147	144	141
147	146	145	148	157	160	151	139	140	138
147	149	157	167	167	155	139	129	133	132
148	154	167	176	169	150	135	131	131	131
139	144	152	155	149	139	133	133	133	134
131	132	132	131	132	133	131	127	130	132
133	132	129	127	134	141	134	122	125	127
129	127	126	128	131	132	130	127	129	127
129	127	126	128	131	132	130	128	130	129

Shoulder of a cow…

problems: too local; lack of context; mismatch between levels of abstraction. But wait, it's even worse than that...

Why is Computer Vision Hard?

If we already know the geometry, surface material and lighting conditions, it is well-understood how to generate the value at each pixel. [this is Computer Graphics]

But this confluence of factors contributing to each pixel can not be easily decomposed. The process can not be inverted.

tell me what you see...

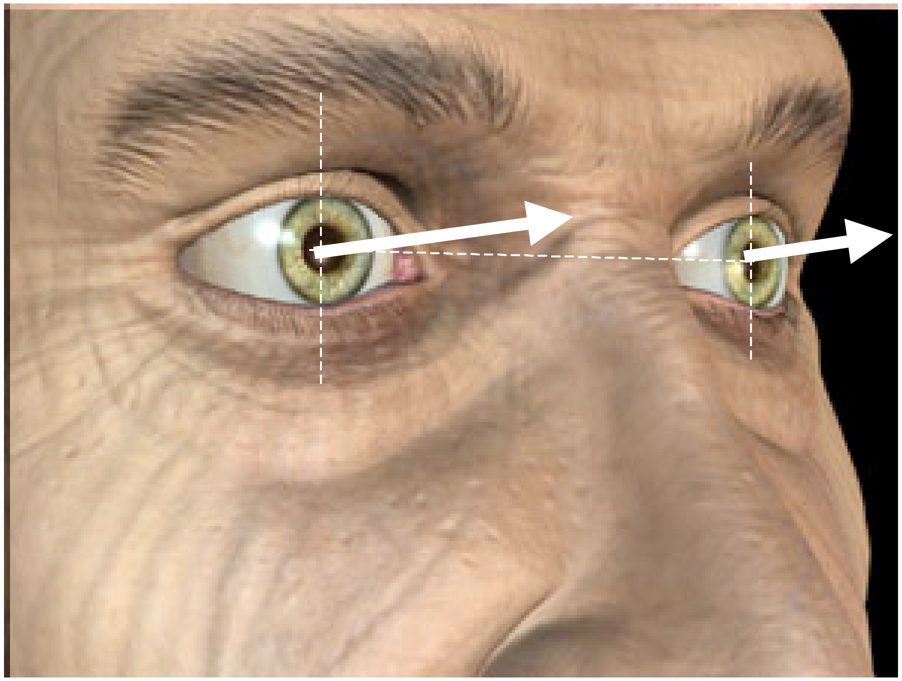
Congratulations! You just did something mathematically impossible.

How?
You used assumptions based on prior knowledge / experience about the way the world works.

Recovering 3D from a single image is a mathematically ill-posed problem.

So we can't solve the problem using only math. ☹

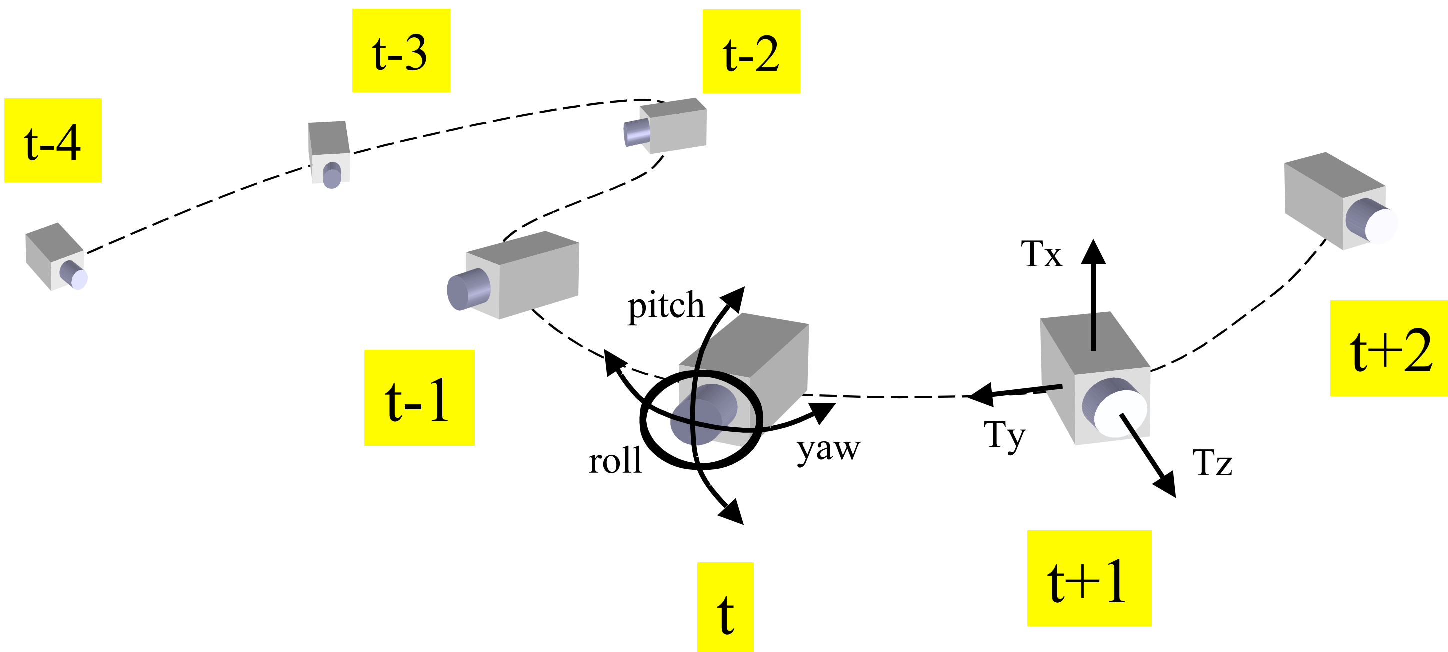
Good news: with more than one camera, we can recover 3D!



Example: Stereo Vision

Structure from Motion

We can also infer 3D from only one camera, provided we move it around "enough".



Structure from Motion

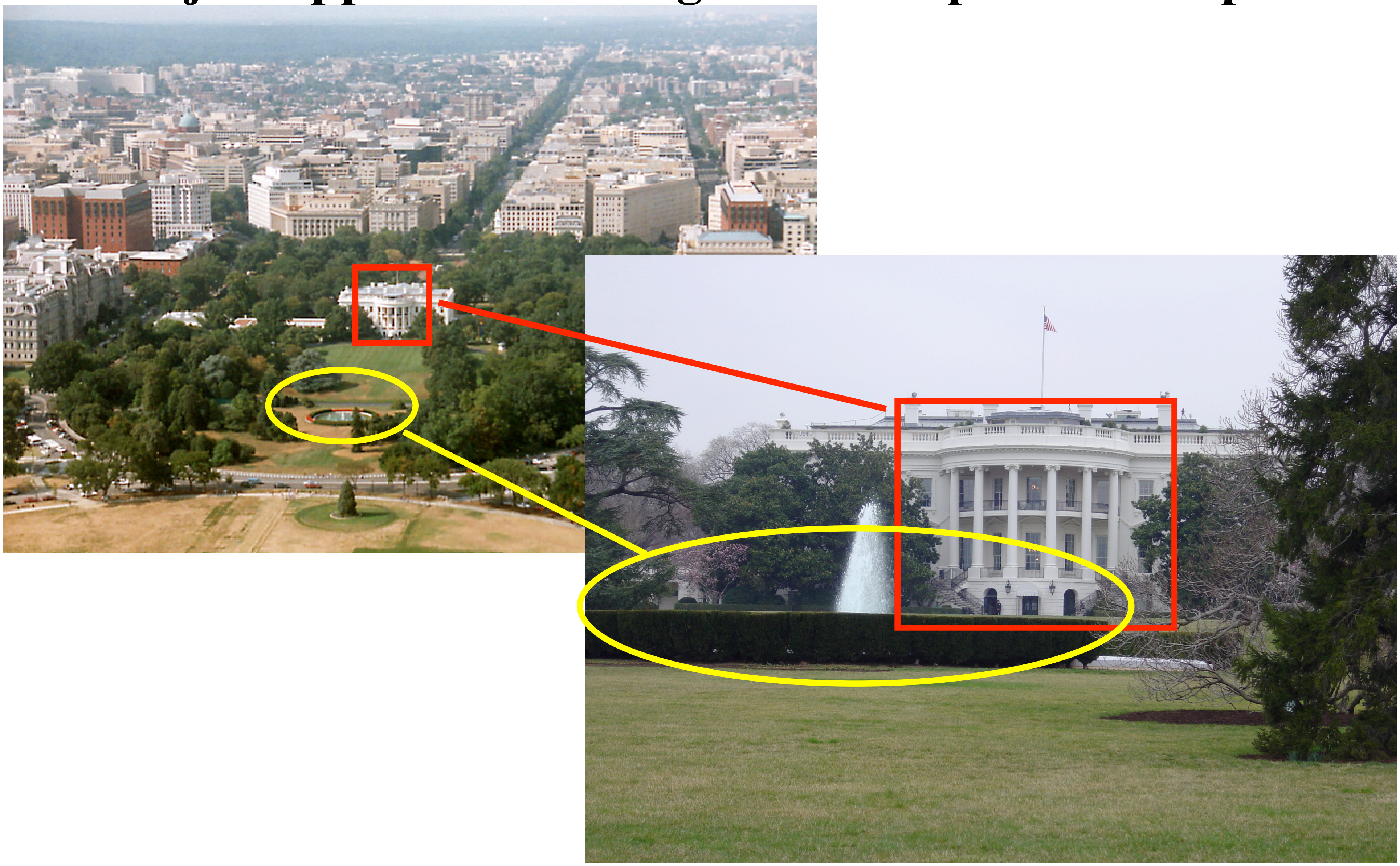
So how much is “enough” motion? It depends.

www.grand-illusions.com

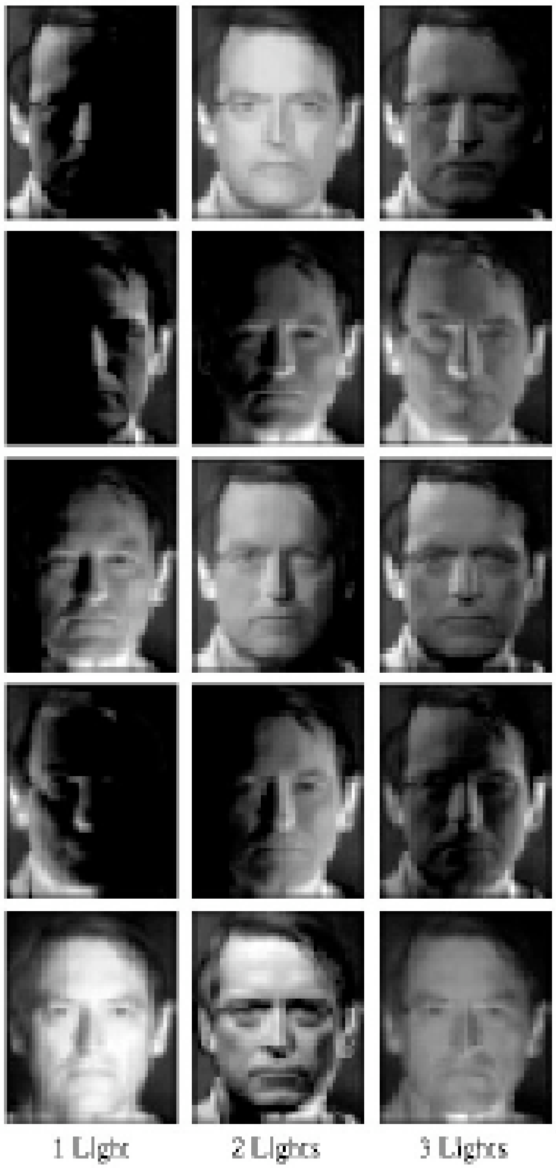
(play movie in external player)

More Difficulties

Object appearance changes with respect to viewpoint

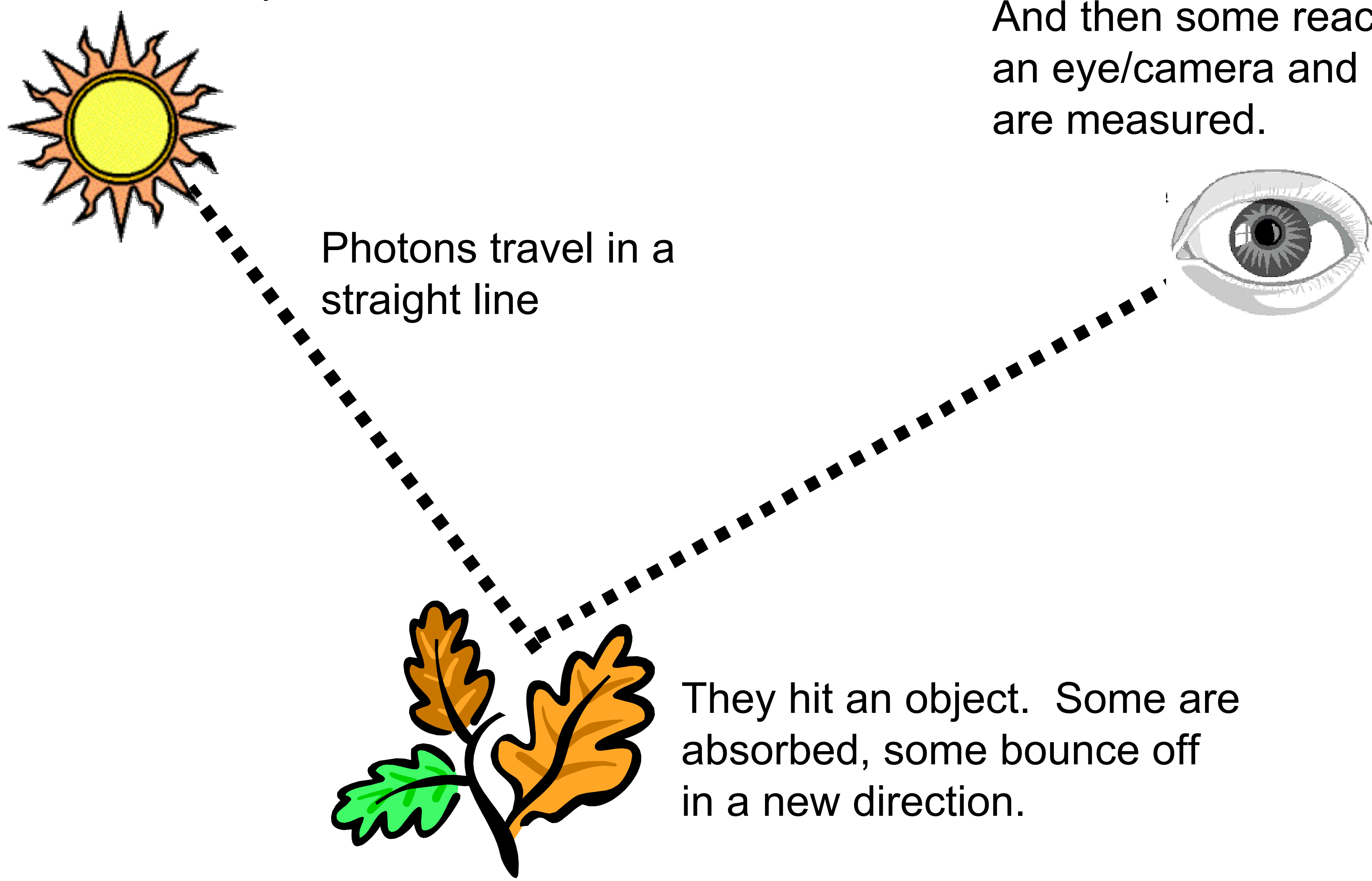


Effects of Lighting



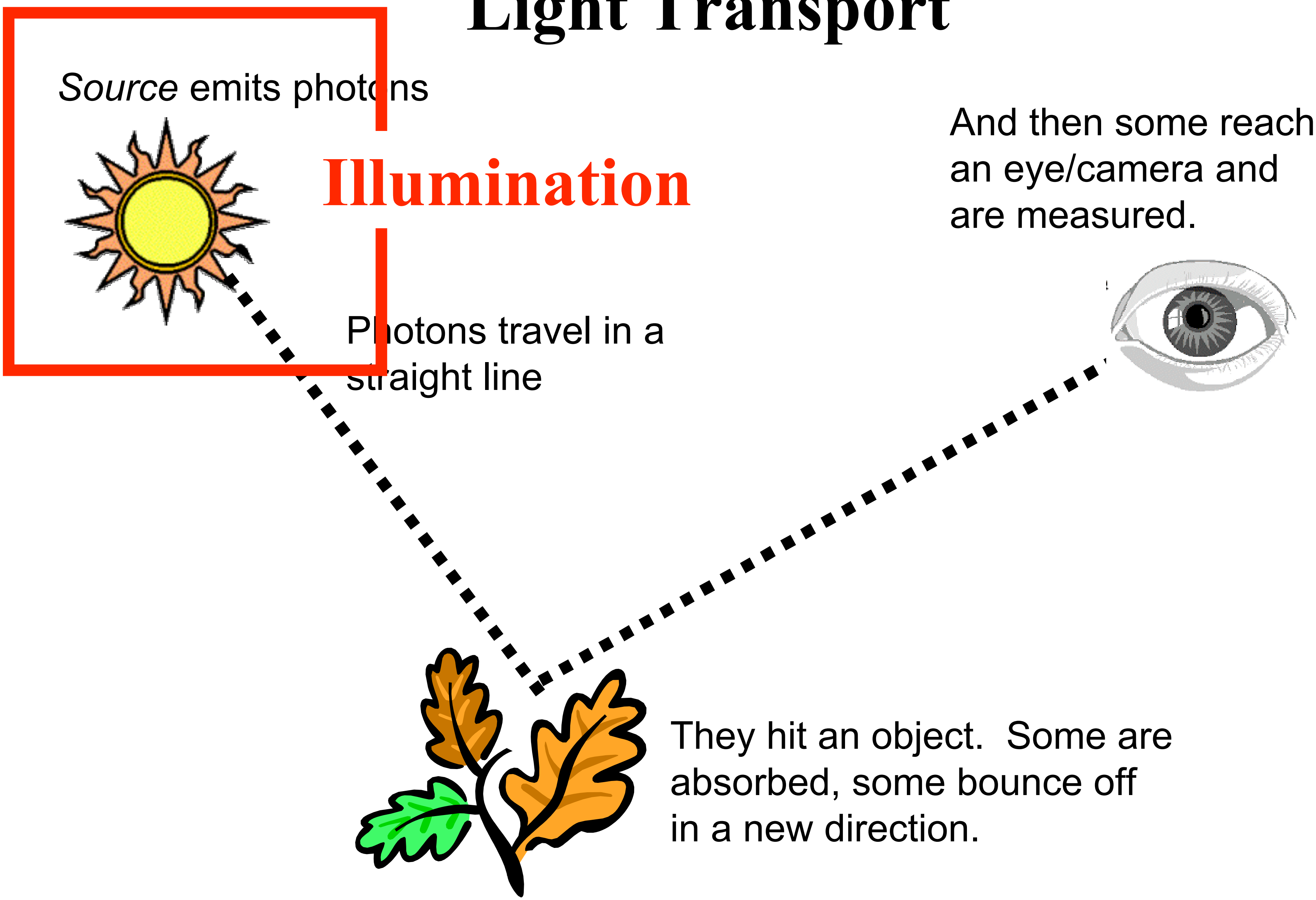
Object appearance also varies with respect to lighting magnitude and direction

Photometry Overview

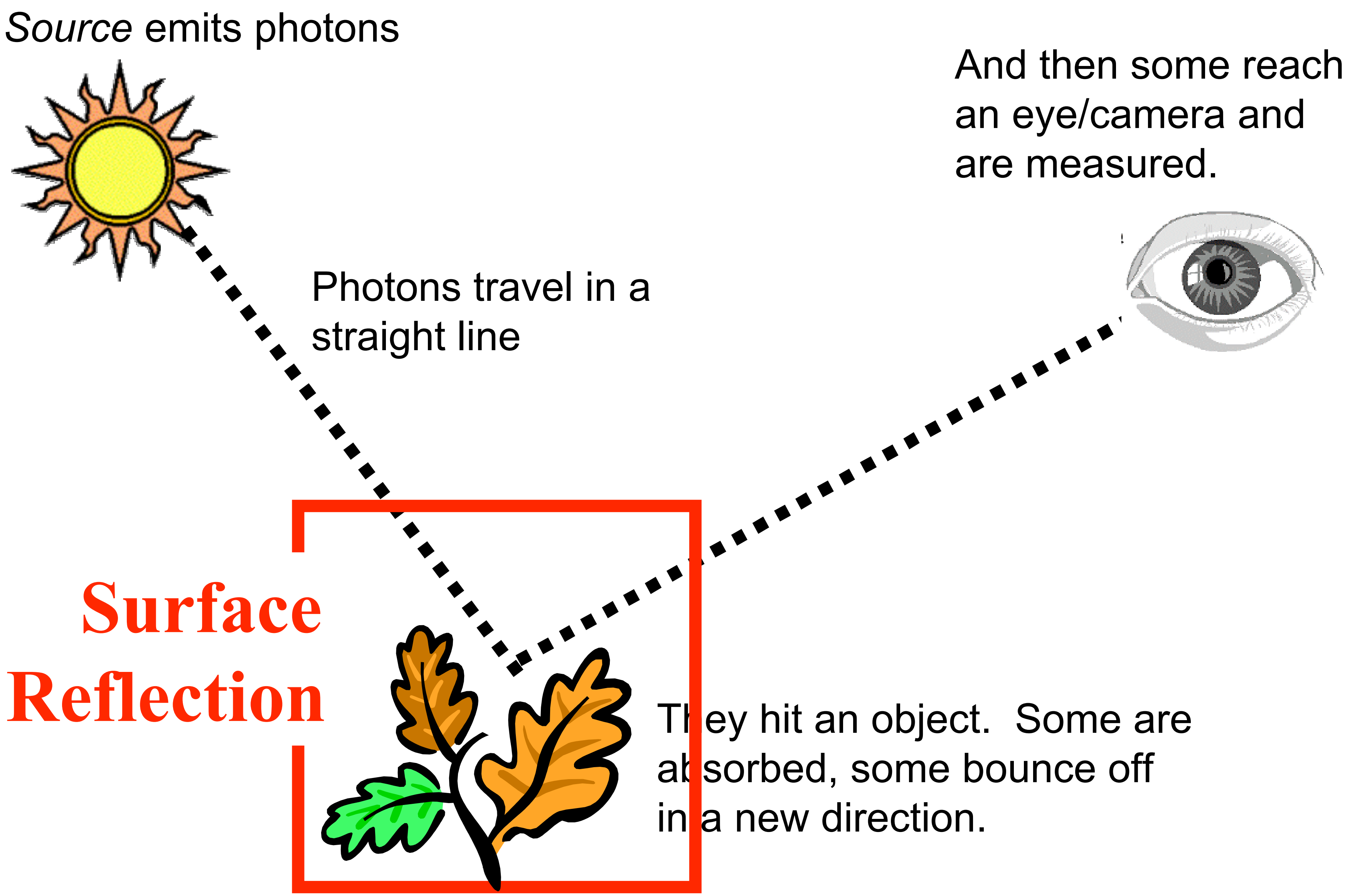


Robert Collins
CSE486, Penn State

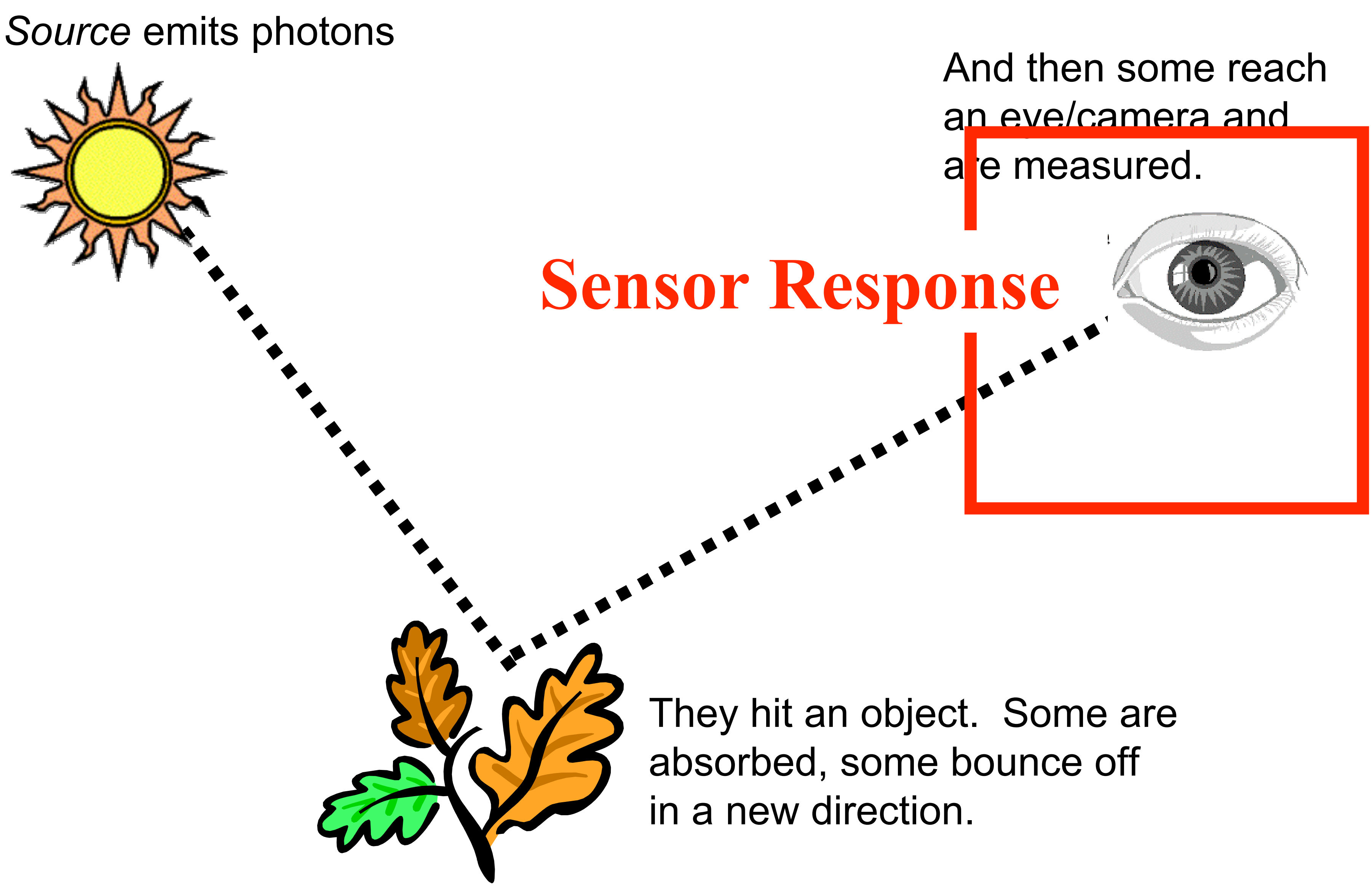
Light Transport



Light Transport



Light Transport



What do we mean by “red”

Color percepts are a composition of three factors (illumination, surface reflectance, sensor response)

We can’t easily factor the color we see in the image to infer illumination and material (even if sensor properties are fixed and known).

Again, this is counterintuitive to human experience.

Perception: Color Constancy

Humans are very good at recognizing the same material colors under different illumination. Not clear how this is achieved in the general case.

Again, math won't save us.

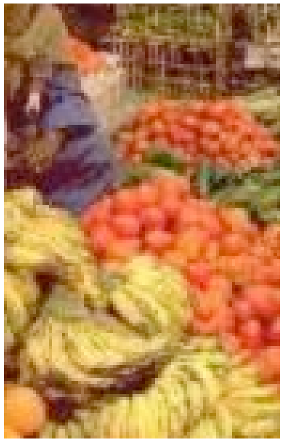
Robert Collins
CSE486, Penn State

What do we mean by "red"

Color percepts are a composition of three factors (illumination, surface reflectance, sensor response)

Some things to think about:

- "red" typically means "appears red to a human observer under white light".
- white objects appear red under a red light.
- nothing looks red if you are red/green color blind.



"normal" color perception

red/green color blind

perceived as a 3D scene

but really just a planar surface (screen) under nonuniform lighting (projector).

What we will be studying in thus course...

with a few examples of why.

Robert Collins
CSE486, Penn State

Filtering and Smoothing

Linear operators
Convolution
Smoothing

Feature Extraction

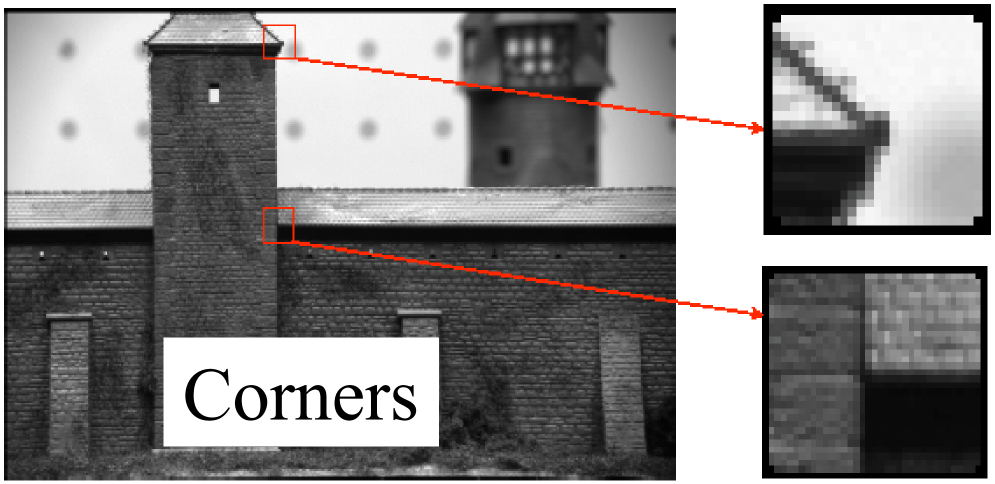
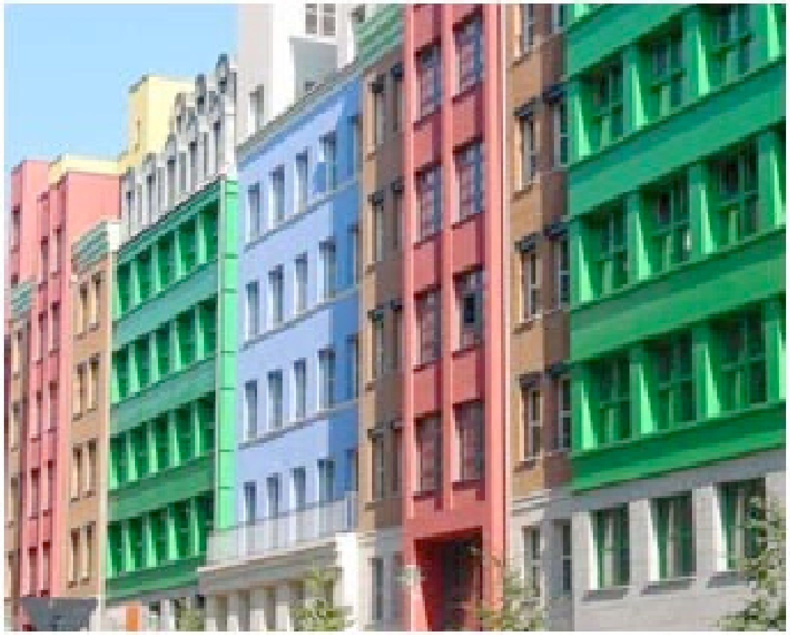


Image derivatives
Gradient operators
DoG/LoG operators
Harris corner detector

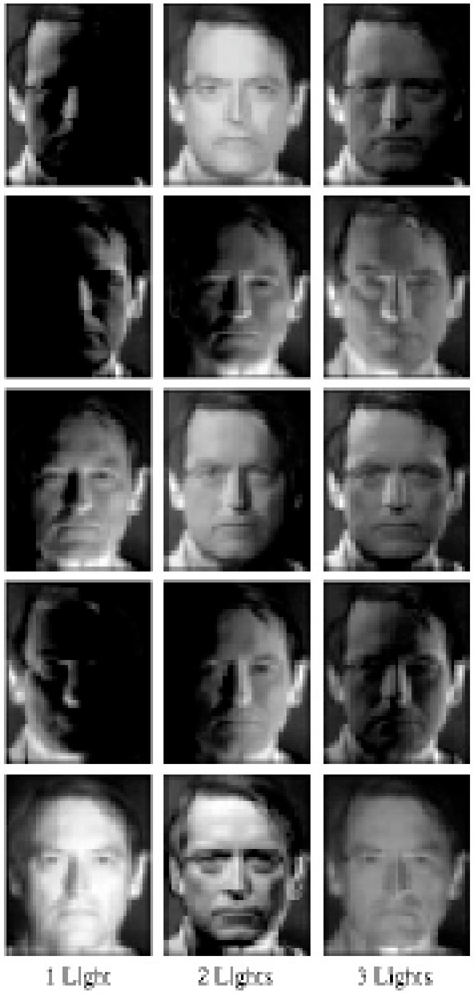
Why?
Seek more unique descriptors (than pixels) for matching

Robert Collins
CSE486, Penn State

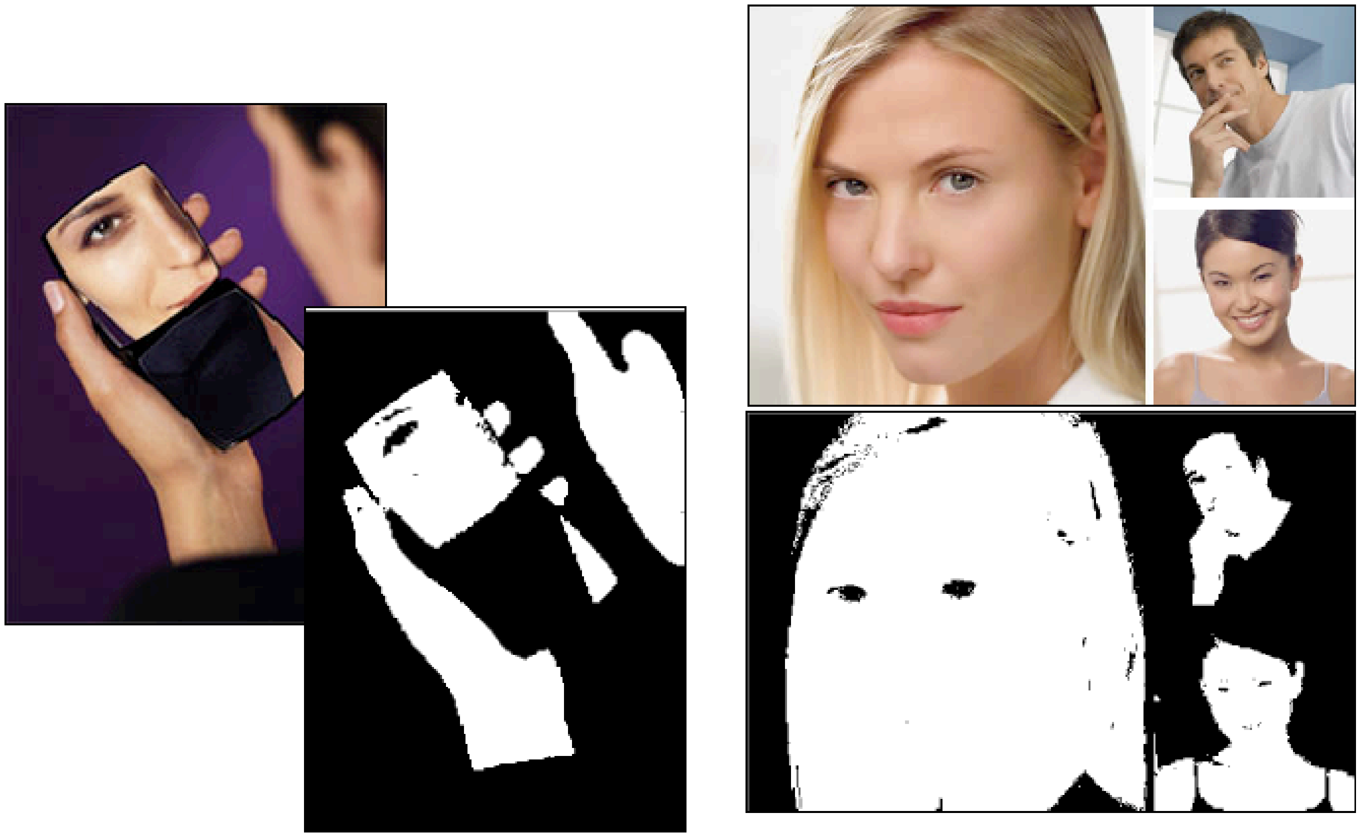
Color and Light



Radiance / Reflection
Illumination / Shading
Chromaticity
Color Constancy



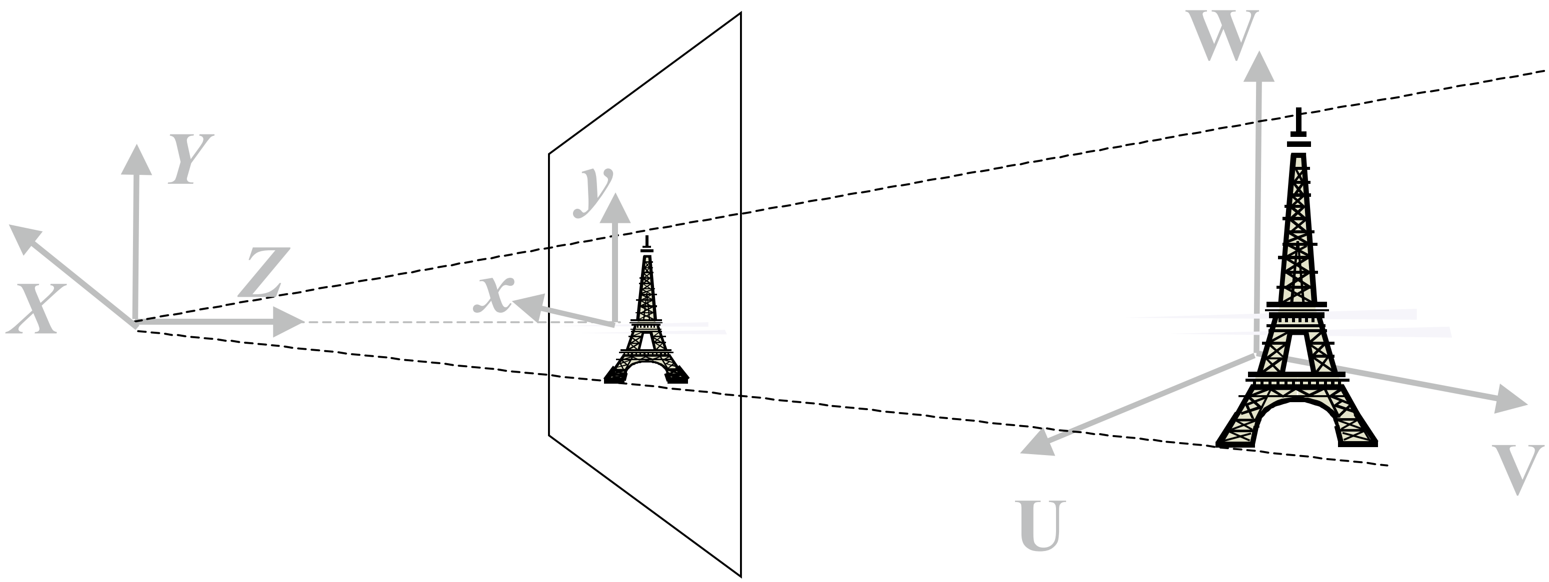
Application : Skin Detection



Skin detection has been used for web filtering based on identifying adult content

Robert Collins
CSE486, Penn State

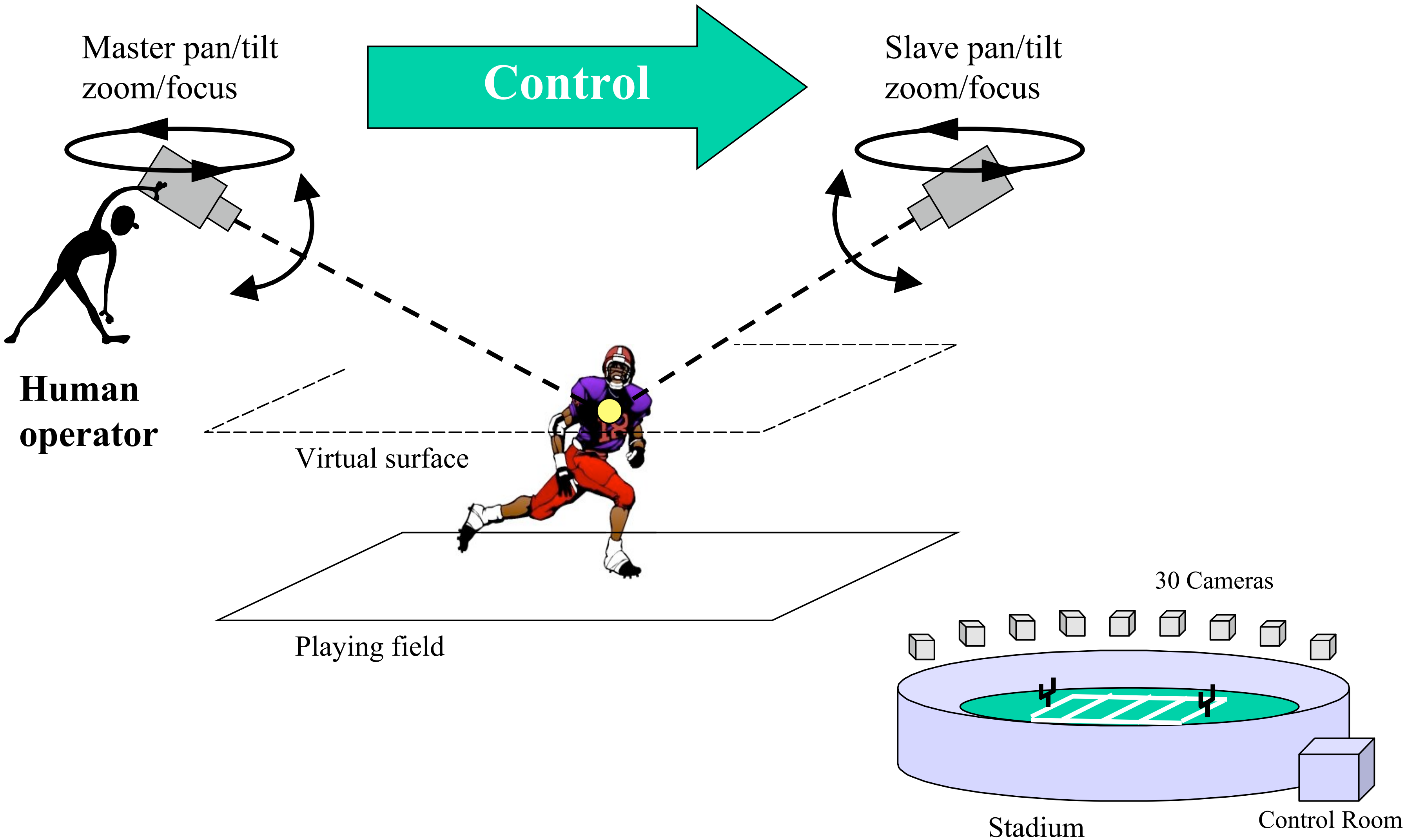
Camera Projection Models



Projection Models
Intrinsic (lens) Paramters
Extrinsic (pose) Parameters
Camera Calibration

Robert Collins
CSE486, Penn State

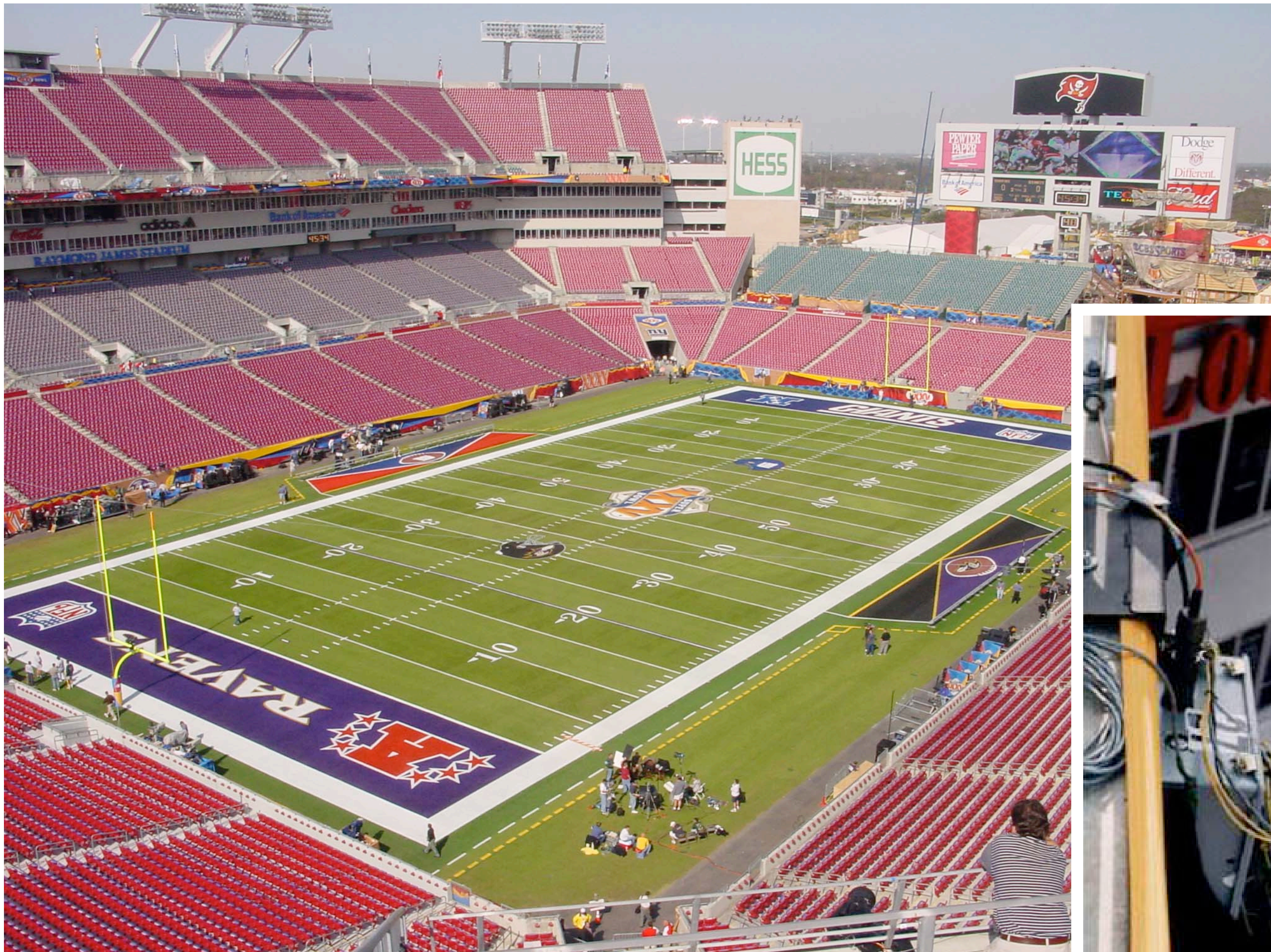
Application: Eyevision System
CAMERA CALIBRATION!

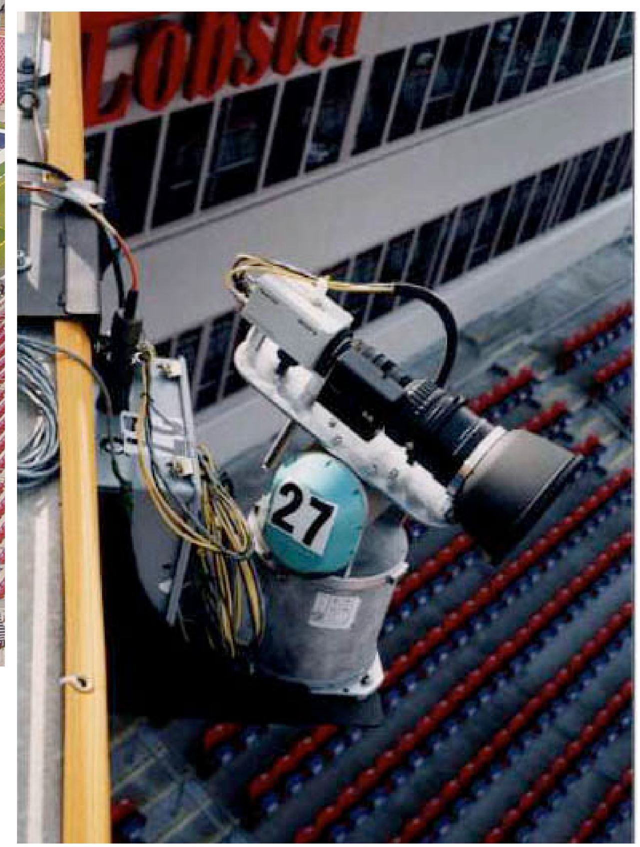


Robert Collins
CSE486, Penn State

Eyevision : SuperBowl XXXV

January 28, 2001





Robert Collins
CSE486, Penn State

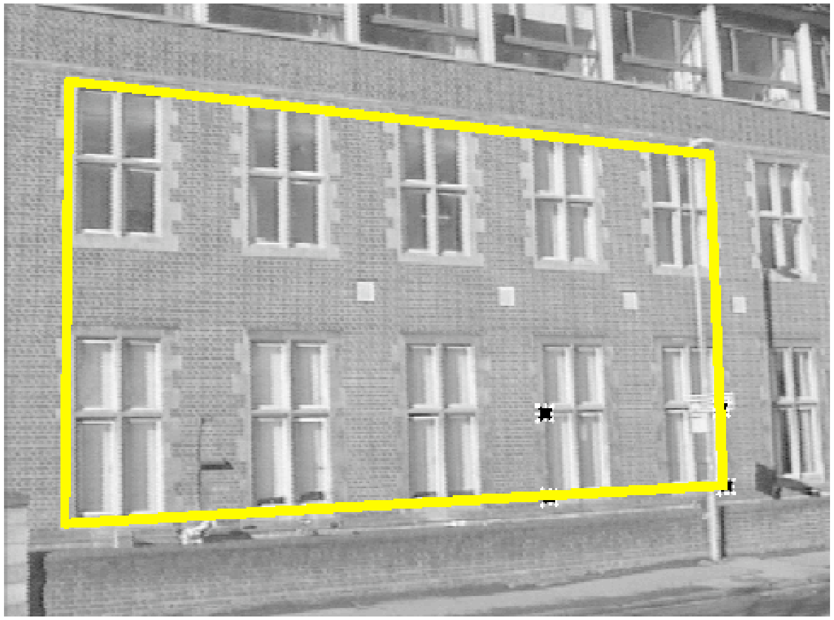
EyeVision Examples

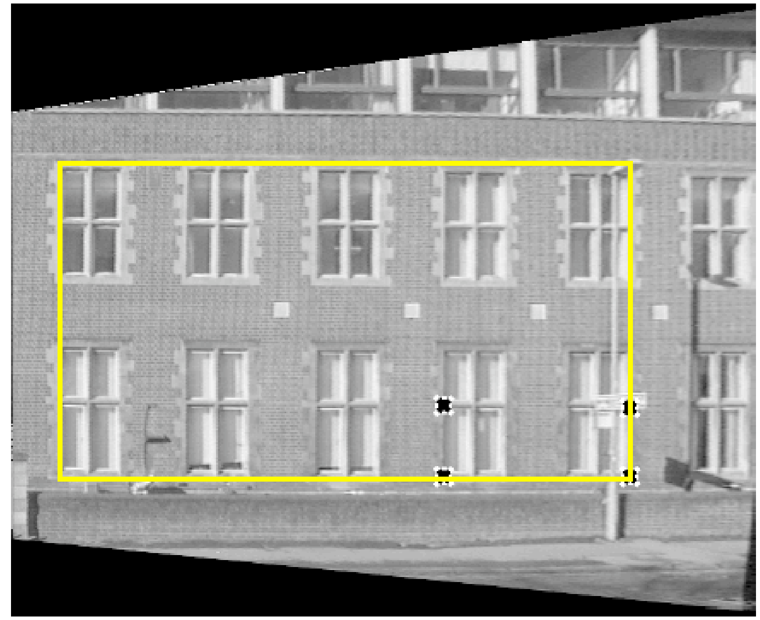
Super Bowl XXXV

January 28, 2001
Courtesy of CBS Sports

Robert Collins
CSE486, Penn State

Plane to Plane Mappings





Rigid, Similarity, Affine, & Projective Mappings
Homography Estimation
Image Warping

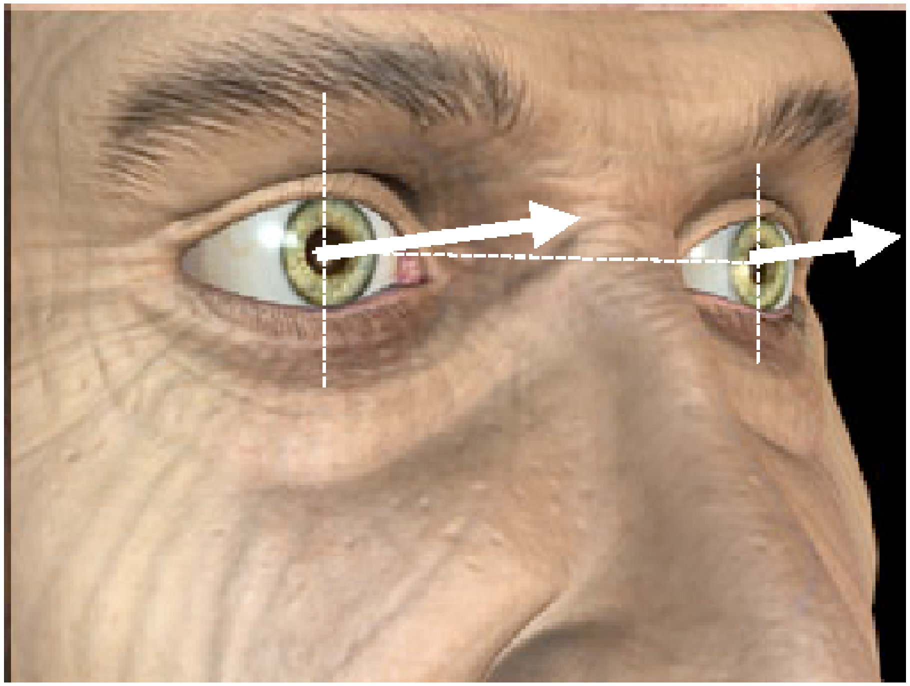
Mosaicing and Stabilization

Camera Rotation Homography
Mosaicing
Video Stabilization

Example : Quicktime VR

http://www.panoguide.com/gallery/

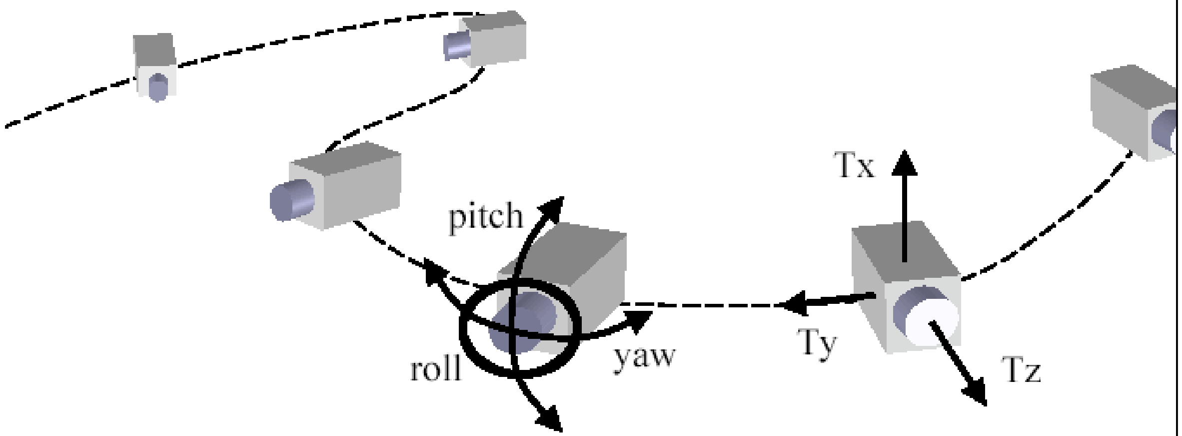
Stereo Vision

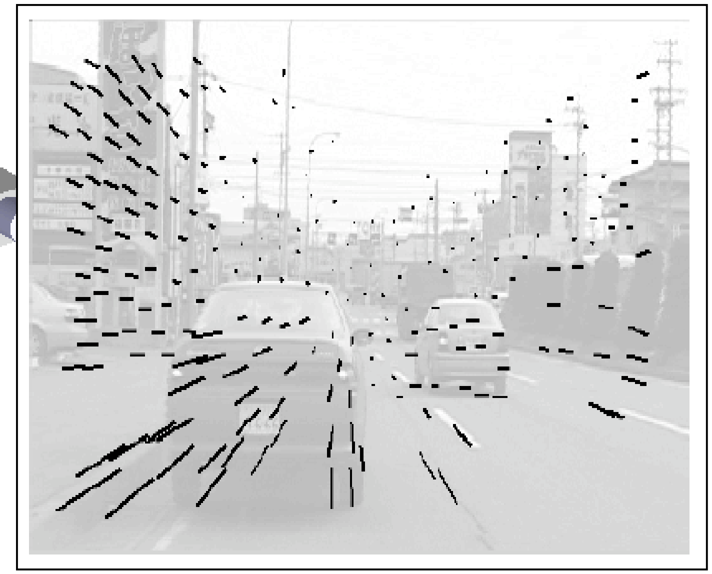


Stereo Camera Setups
Stereo Disparity / Parallax
Epipolar Geometry
Correspondence Matching
Triangulation / Depth Recovery

Robert Collins
CSE486, Penn State

Camera Motion





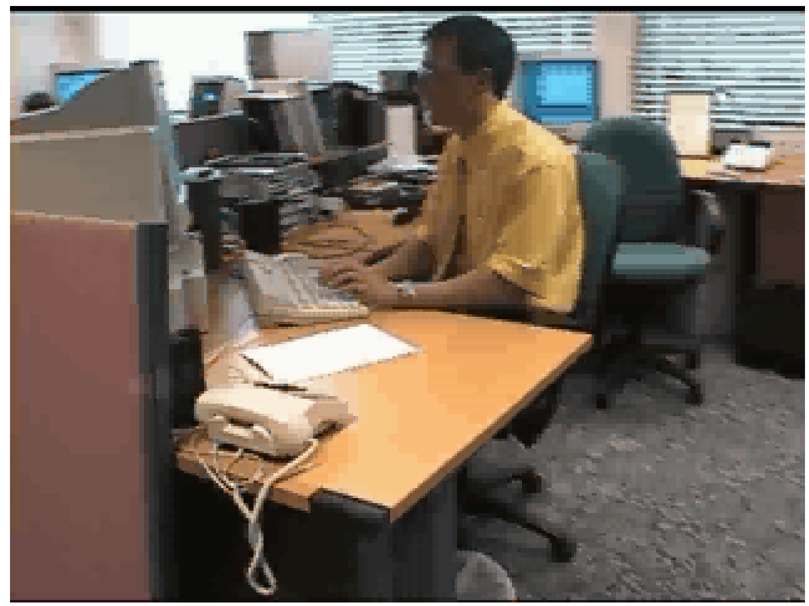
Motion Field vs Optic Flow
Flow Estimation
Egomotion Estimation
Structure from Motion

Application: Match Move

Track a set of feature points through a movie sequence

Deduce where the cameras are and the 3D locations of the points that were tracked

Render synthetic objects with respect to the deduced 3D geometry of the scene / cameras



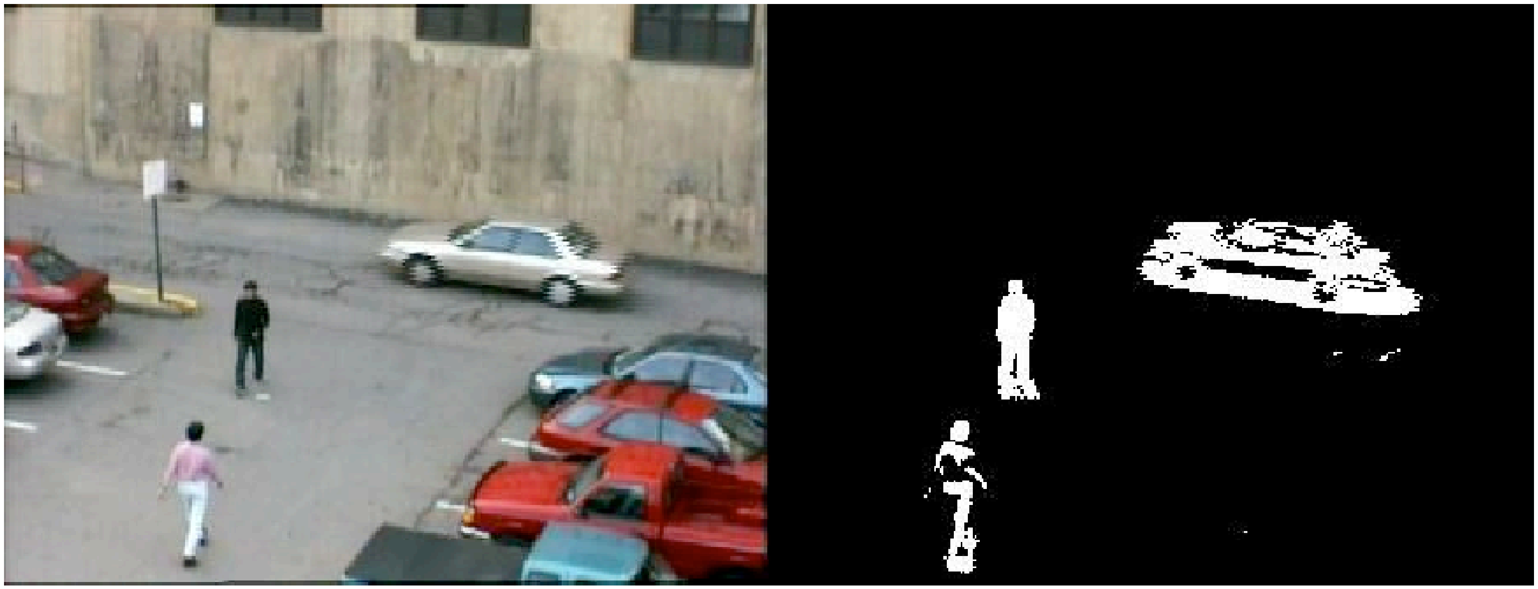
"Graham Kimpton" example from www.realviz.com MatchMover Professional gallery. Copyrighted.

Application: Autonomous Driving

Have your vehicle chauffeur you around.

Stanley, winner 2005 Darpa Grand Challenge

Video Change Detection

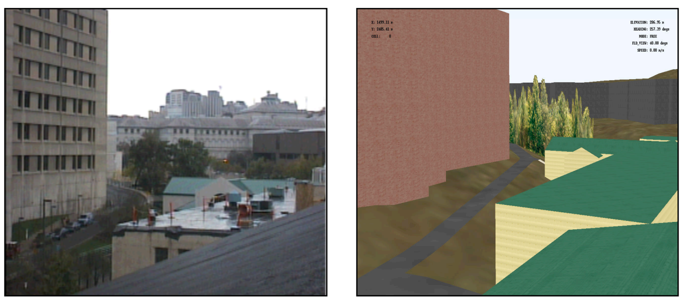


Video Sequences
Background Modeling
Change Detection

Robert Collins
CSE486, Penn State

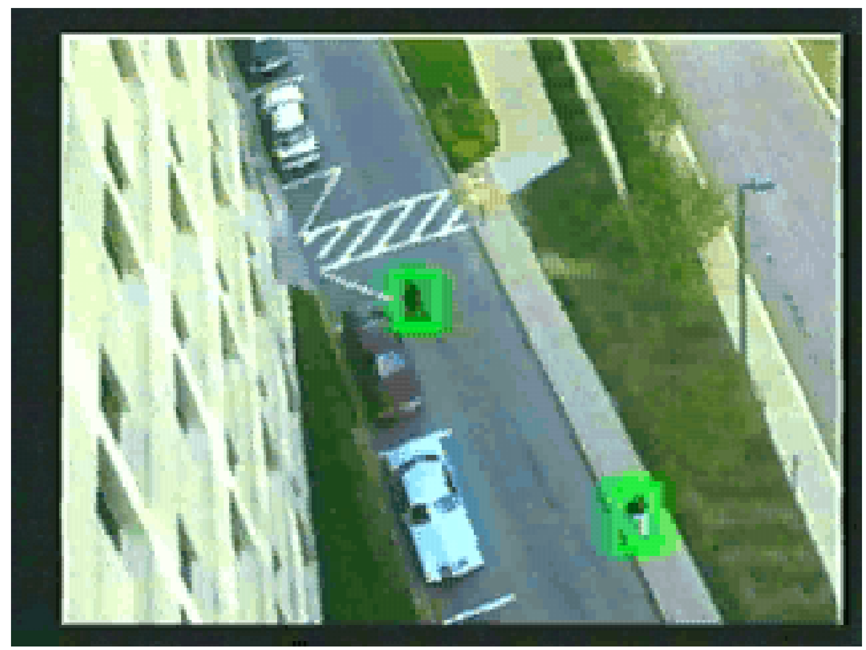
Application: Video Surveillance

Automatic detection, classification and tracking of moving people and vehicles



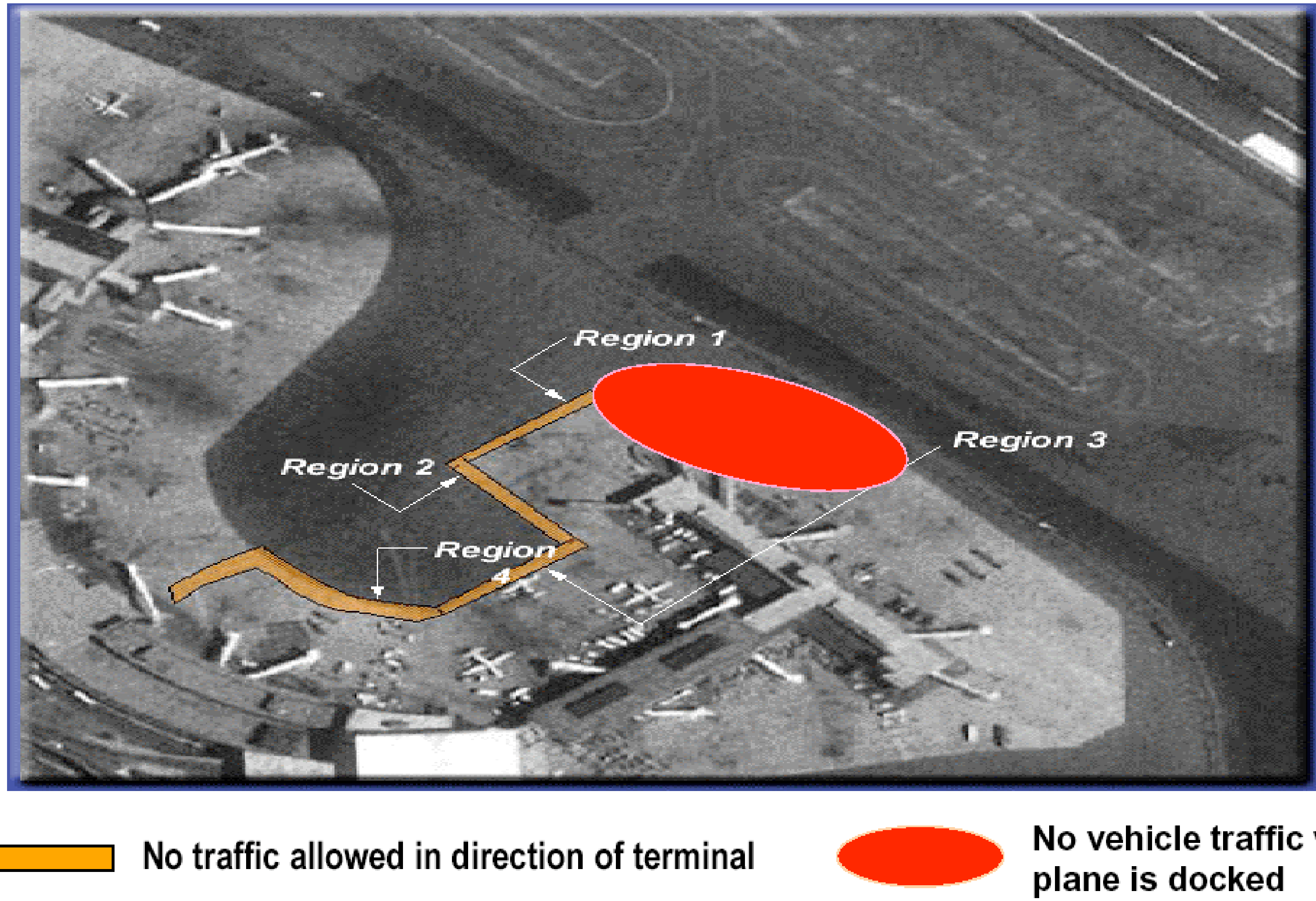
Object locations are determined with respect to a 3D campus model

A single operator can monitor results from many sensors spread over a large area



Automated Surveillance

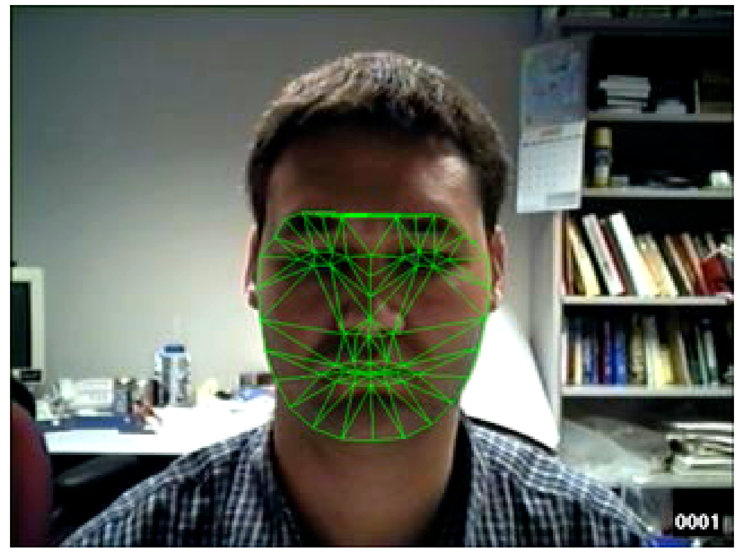
Trigger regions for detecting motion and detecting motion going in the wrong direction. These are pretty well "solved" problems.



Guardian Systems

Robert Collins
CSE486, Penn State

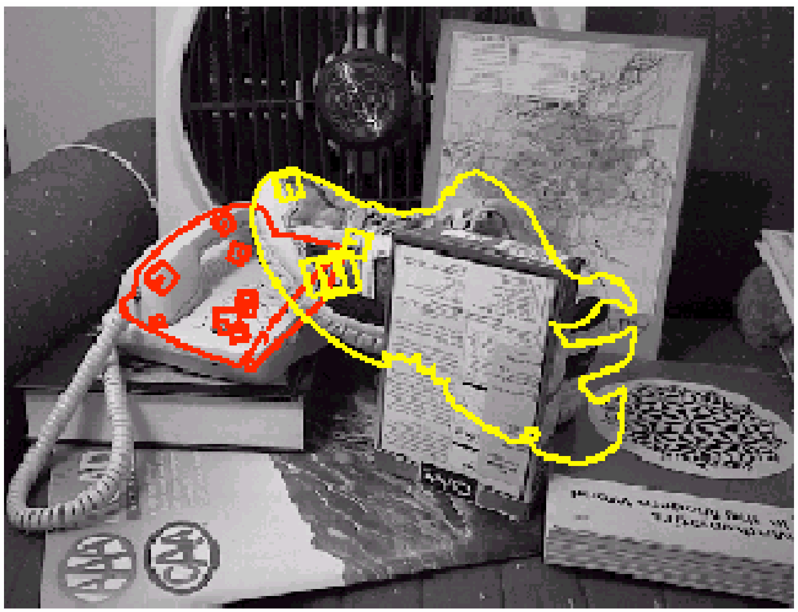
Video Tracking

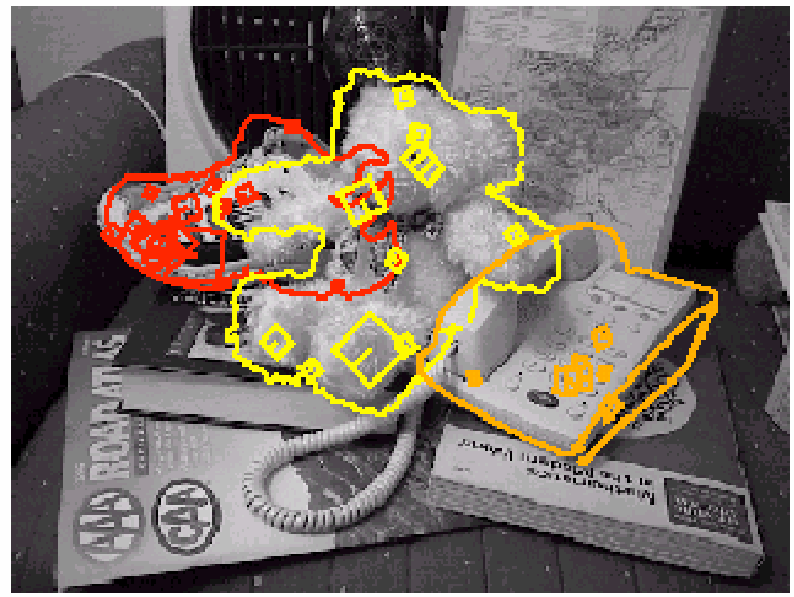


Appearance-Based Tracking
Sample Tracking Algorithms
(e.g. Mean-Shift, Lucas-Kanade)

Object Recognition

Library of models



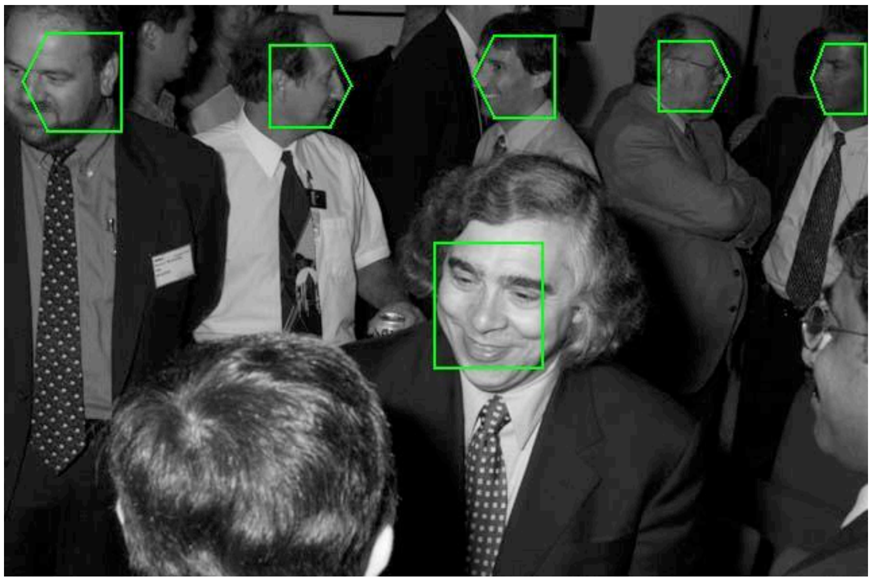


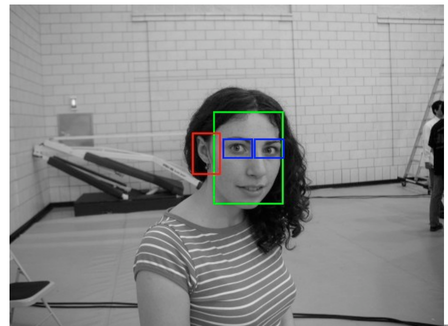
Model matches (using Lowe's SIFT keys)

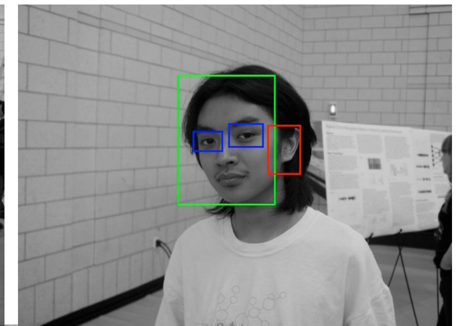
Approaches: PCA; Sift Keys; boosted cascade of detectors

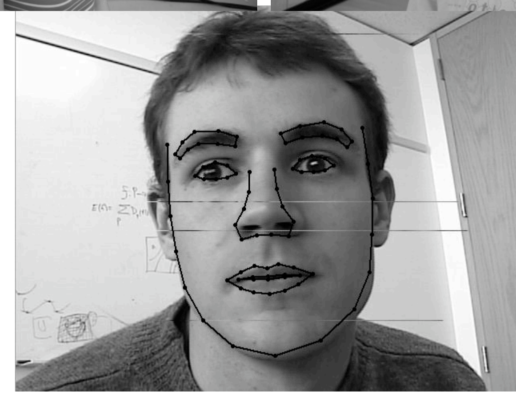
Application: Face/Eye Detection

Henry Schneiderman









Finding users to interact with.
Red-eye removal from photos.
Drowsy-driver detection.

Reminder...

“Vision is the act of knowing what is where by looking.” --Aristotle