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: yal@psu.edu
Office Hours: TBA

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

*Introductory Techniques for 3-D Computer Vision*

optional

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/](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.
<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td><strong>X</strong></td>
<td>smoothing</td>
<td></td>
<td>edges</td>
<td><strong>H1</strong></td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>corners</td>
<td>template matching</td>
<td><strong>P1</strong></td>
<td>intro to stereo</td>
<td><strong>H2</strong></td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>stereo alg 1</td>
<td>stereo alg 2</td>
<td>pyramids</td>
<td><strong>H2</strong></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>LoG/zero-cross</td>
<td>P1</td>
<td>review</td>
<td>midterm 1</td>
<td><strong>M1</strong></td>
</tr>
</tbody>
</table>

**SEPTEMBER 2007**

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>extrinsics</td>
<td><strong>H3</strong></td>
<td>intrinsics</td>
<td><strong>H3</strong></td>
<td>transformations 1</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transformations 2</td>
<td>RANSAC</td>
<td><strong>H3</strong></td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>homography 2</td>
<td>mosaicing</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>epipolar geom</td>
<td>E/F matrices</td>
<td><strong>P2</strong></td>
<td>8-pt alg</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>stereo reconstruct</td>
<td>review</td>
<td><strong>H4</strong></td>
</tr>
</tbody>
</table>

**NOVEMBER 2007**

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>camera motion</td>
<td>flow estimate 1</td>
<td><strong>[H5]</strong></td>
<td>flow estimate 2</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>change detect</td>
<td>SFM</td>
<td>color &amp; light movie</td>
<td><strong>H5</strong></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td><strong>X</strong></td>
<td><strong>X</strong></td>
<td><strong>X</strong></td>
<td><strong>X</strong></td>
<td><strong>X</strong></td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>color &amp; light</td>
<td>color histograms</td>
<td><strong>[H6]</strong></td>
<td>tracking</td>
<td></td>
</tr>
</tbody>
</table>

**OCTOBER 2007**

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>transformations 1</td>
<td><strong>H3</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RANSAC</td>
<td><strong>[P2]</strong></td>
<td><strong>H3</strong></td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mosaicing</td>
<td></td>
<td>suicide day</td>
</tr>
<tr>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>epipolar geom</td>
<td>E/F matrices</td>
<td><strong>P2</strong></td>
</tr>
<tr>
<td>29</td>
<td>30</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>stereo reconstruct</td>
<td>review</td>
<td><strong>H4</strong></td>
</tr>
</tbody>
</table>

**DECEMBER 2007**

<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>tracking LK</td>
<td>P3</td>
<td>tracking MS</td>
<td><strong>H6</strong></td>
<td>Oby Rec SIFT</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>texture synth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>final Exam Week</td>
<td>*final Exam Week</td>
<td>*final Exam Week</td>
</tr>
</tbody>
</table>

**Final Exam Week *Final Exam Week *Final Exam Week**
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
Input: Digital Images

2D arrays (matrices) of numbers:

If color image, we have 3 arrays - red, green, blue
Why is Computer Vision Hard?

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

- 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”.

\[ \text{t-4} \rightarrow \text{t-3} \rightarrow \text{t-2} \rightarrow \text{t} \rightarrow \text{t+1} \rightarrow \text{t+2} \]
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

*Source* emits photons

Photons travel in a straight line

They hit an object. Some are absorbed, some bounce off in a new direction.

And then some reach an eye/camera and are measured.
Source emits photons

Photons travel in a straight line

They hit an object. Some are absorbed, some bounce off in a new direction.

And then some reach an eye/camera and are measured.

Illumination
Light Transport

Source emits photons

Photons travel in a straight line

They hit an object. Some are absorbed, some bounce off in a new direction.

And then some reach an eye/camera and are measured.
Source emits photons

They hit an object. Some are absorbed, some bounce off in a new direction.

And then some reach an eye/camera and are measured.

Sensor Response
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. 😞 😞
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 this course...

with a few examples of why.
Filtering and Smoothing

Linear operators
Convolution
Smoothing
Feature Extraction

Corners

Edges

Image derivatives
Gradient operators
DoG/LoG operators
Harris corner detector

Why?
Seek more unique descriptors (than pixels) for matching
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.
Camera Projection Models

Projection Models
Intrinsic (lens) Parameters
Extrinsic (pose) Parameters
Camera Calibration
Application: Eyevision System
CAMERA CALIBRATION!

Master pan/tilt zoom/focus
Slave pan/tilt zoom/focus

Human operator
Virtual surface
Playing field

30 Cameras
Stadium
Control Room

Control
Eyevision: SuperBowl XXXV
January 28, 2001
EyeVision Examples

Super Bowl XXXV
January 28, 2001
Courtesy of CBS Sports
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
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

Deduze 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

Application: Autonomous Driving

Have your vehicle chauffeur you around.

Stanley, winner 2005 Darpa Grand Challenge
Video Change Detection

Video Sequences
Background Modeling
Change Detection
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.
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

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