CSCI 5561 Computer Vision

Hyun Soo Park

3D Motion Reconstruction (Trajectory Stream Association)
What This Image Can Tell Us About The Scene?
WHAT THIS IMAGE CAN TELL US ABOUT THE SCENE?

- Rainy day
- Street market
- People / role
- Interaction
- Car/fruit/shelter
- Distance
- Height of the cameraman
WHAT THIS IMAGE CAN TELL US ABOUT THE SCENE?

- Rainy day
- Street market
- People / role
- Interaction
- Car/fruit/shelter
- Distance
- Height of the cameraman

COMPUTER VISION

Def) computationally understanding the scene/image.

Cf) image processing
What This Image Can Tell Us About The Scene?

- Rainy day
- Street market
- People / role
- Interaction
- Car/fruit/shelter
- Distance
- Height of the cameraman

Computer Vision
Def) computationally understanding the scene/image.
Cf) image processing

Extremely difficult
# of pixels: 8.2M
Birth of Computer Vision

In 1966, Minsky hired a first-year undergraduate student and assigned him a problem to solve over the summer:

Connect a camera to a computer and get the machine to describe what it sees.

Marvin Minsky, MIT
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".
1960’s: interpretation of synthetic worlds

Larry Roberts
“Father of Computer Vision”

Input image

2x2 gradient operator

computed 3D model rendered from new viewpoint

Larry Roberts PhD Thesis, MIT, 1963,
Machine Perception of Three-Dimensional Solids

Slide credit: Steve Seitz
The representation and matching of pictorial structures
Fischler and Elschlager, 1973
Conference on Computer Vision and Pattern Recognition

Number of Papers

3500
3000
2500
2000
1500
1000
500
0

Takeo Kanade, CMU
Dana Ballard, U of Rochester
Conference on Computer Vision and Pattern Recognition

Number of Papers

# of submissions

# of accepted papers
Conference on Computer Vision and Pattern Recognition

Number of Papers

<table>
<thead>
<tr>
<th>Year</th>
<th>LK tracker</th>
<th>Face recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conference on Computer Vision and Pattern Recognition

Number of Papers

![Bar chart showing the number of papers from 1983 to 2010]

- LK tracker
- Face recognition
- SVM
- SIFT
- Phototourism
Conference on Computer Vision and Pattern Recognition

Number of Papers

When I started my PhD

Computer vision Known to be NOT working

LK tracker
Face recognition
SIFT
Phototourism
SVM

0
500
1000
1500
2000
2500
3000
3500

Conference on Computer Vision and Pattern Recognition

Number of Papers

- LK tracker
- Face recognition
- SVM
- SIFT
- Phototourism
- Deep learning
- When I started my PhD

Years:
- 1983
- 1984
- 1985
- 1986
- 1987
- 1988
- 1989
- 1990
- 1991
- 1992
- 1993
- 1994
- 1995
- 1996
- 1997
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015
- 2016
- 2017
- 2018
How to Make a Computer to Understand an Image?
MARR'S VISION

David Marr (1945-1980)

VISION

DAVID MARR

1982
THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)
THREE LEVELS OF VISUAL PERCEPTION

1. Computational
   e.g., what is the goal of the computation?

David Marr (1945-1980)

How to computationally find Waldo?
THREE LEVELS OF VISUAL PERCEPTION

1. Computational
e.g., what is the goal of the computation?

How to formulate the objective:

$$\text{minimize } \sum_{p} (J(w(x; p)) - T(x))^2$$

David Marr (1945-1980)
THREE LEVELS OF VISUAL PERCEPTION

1. Computational
   e.g., what is the goal of the computation?

   How to formulate the objective:

   \[
   \text{minimize} \sum_p \sum_x (J(w(x; p)) - T(x))^2 
   \]

   How to achieve the objective:

   David Marr (1945-1980)
THREE LEVELS OF VISUAL PERCEPTION

2. Algorithmic
   e.g., what representation can implement the computation?

David Marr (1945-1980)

How to represent the image?
THREE LEVELS OF VISUAL PERCEPTION

2. Algorithmic
   e.g., what representation can implement the computation?

David Marr (1945-1980)

How to represent the image?
THREE LEVELS OF VISUAL PERCEPTION

2. Algorithmic
   e.g., what representation can implement the computation?

David Marr (1945–1980)

How to represent the image?
THREE LEVELS OF VISUAL PERCEPTION

2. Algorithmic
   e.g., what representation can implement the computation?

How to represent the image?

David Marr (1945-1980)
THREE LEVELS OF VISUAL PERCEPTION

3. Implementational
   e.g., how hardware can carry out such computation?

David Marr (1945-1980)
THREE LEVELS OF VISUAL PERCEPTION

3. Implementational
   e.g., how hardware can carry out such computation?
THREE LEVELS OF VISUAL PERCEPTION

3. Implementational
e.g., how hardware can carry out such computation?

David Marr (1945-1980)
Three Levels of Visual Perception

3. Implementational
e.g., how hardware can carry out such computation?

Beyond the Scope of This Course.

David Marr (1945-1980)
WHAT WILL BE COVERED?

Basics and 4 Rs of Computer Vision
**BASICS**

Image formation
Image convolution/filtering
Feature representation
4 Rs: Registration

Optical flow
Image alignment
Tracking
4 Rs: Recognition

- Bag of feature
- Template matching
- Object proposal
- Convolutional neural network
4 Rs: Reorganization

Graph cuts
Superpixel
Semantic segmentation
4 Rs: RECONSTRUCTION

Camera geometry
Epipolar geometry
Stereo

3D Motion Reconstruction
(Trajectory Stream Association)
WHAT WILL NOT BE COVERED?
WHAT WILL NOT BE COVERED?

• Basic Machine Learning knowledge
• MATLAB / Python programming
• Linear algebra / Calculus
WHAT WILL NOT BE COVERED?

- Basic Machine Learning knowledge
- MATLAB / Python programming
- Linear algebra / Calculus

Tips:
1. Drop this course if you are not fluent on these materials---you will be embarrassed if you ask these even during office hours.
2. Drop this course if you are not confident on mathematical programming, e.g., translating math concept to code.
3. Drop this course if you are not comfortable on debugging and the usage of debugging tools.
4. Study by yourself and read relevant materials (e.g., book, wikipedia, coursera).
**EVALUATION**

- 5 programming assignments (15% each)
  - Late submission: 20% off from each extra late day
- 1 final written exam (25%): May 1st during the lecture time

No make-up assignment and exam
Office Hour

- Hyun Soo Park: MW 4-5pm @ Shepherd 261
- Jingfan Guo: TTh 4-5pm @ Shpherd 234
Course Website

https://www-users.cs.umn.edu/~hspark/csci5561/csci5561.html

Spring 2019 CSCI 5561
Computer Vision
Mon/Wed 9:45am-11:00am @ Keller Hall 3-111

Information

- Syllabus
- Schedule
- Lecture slide

Syllabus

Instructor: Hyun Soo Park (hspark at umn.edu)
Office hour: Mon/Wed 4:00pm-5:00pm (Shepherd Laboratory 251)

TA: Jingfan Guo (guo00109 at umn.edu)
Office hour: Tue/Thr 4:00pm-5:00pm (Shepherd Laboratory 234)

While only 5511 is listed as a prerequisite, I will assume that all students are fluent on the following subjects (we don't have time to cover in the lectures):
- MATLAB/Python usage for image handling
- Linear algebra
- Calculus
- Machine learning

Textbook: Not required but the following books will be frequently referred:
- "Computer Vision: A Modern Approach", David A. Forsyth and Jean Ponce

Topics:
- Visual representation
- Image formation
- Image transformation
- Image filtering/convolution
WHO AM I?
70 HD cameras
FACE DATA CAPTURE
Body Data Capture

Skeleton

Body skin
3D reconstruction of monkey movement
SENSORIMOTOR SKILL LEARNING

Time: 19.97 sec
Speed: 6.8 m/s
Air Drag: 21.13 N

Pt: -3 Nm
Th: -59 N
Lt: 35 N
No: -223 N
Yw: 6 Nm

Gravity

https://www.youtube.com/watch?v=aVJ45wlUE88

3D reconstruction