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.*
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
1970’s: some progress on interpreting selected images

The representation and matching of pictorial structures
Fischler and Elschlager, 1973
Conference on Computer Vision and Pattern Recognition

Number of Papers

<table>
<thead>
<tr>
<th>Year</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>500</td>
</tr>
<tr>
<td>1984</td>
<td>1000</td>
</tr>
<tr>
<td>1985</td>
<td>1500</td>
</tr>
<tr>
<td>1986</td>
<td>2000</td>
</tr>
<tr>
<td>1987</td>
<td>2500</td>
</tr>
<tr>
<td>1988</td>
<td>3000</td>
</tr>
<tr>
<td>1989</td>
<td>3500</td>
</tr>
</tbody>
</table>

Takeo Kanade, CMU

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

Number of Papers

LK tracker
Conference on Computer Vision and Pattern Recognition

Number of Papers

3500
3000
2500
2000
1500
1000
500
0

LK tracker
Face recognition

Conference on Computer Vision and Pattern Recognition

Number of Papers

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

Number of Papers

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

3500   3000   2500   2000   1500   1000   500    0
Conference on Computer Vision and Pattern Recognition

Number of Papers

- 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
Conference on Computer Vision and Pattern Recognition

Number of Papers

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

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

David Marr (1945-1980)

How to represent the image?
Three Levels of Visual Perception

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

David Marr (1945-1980)

How to represent the image?
THREE LEVELS OF VISUAL PERCEPTION

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

David Marr (1945-1980)

How to represent the image?
THREE LEVELS OF VISUAL PERCEPTION

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

How to represent the image?

---

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

David Marr (1945-1980)

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

How to computationally find Waldo?
THREE LEVELS OF VISUAL PERCEPTION

2. 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

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

   How to formulate the objective:
   
   \[
   \text{minimize } \sum_p \left( J(w(x; p)) - T(x) \right)^2
   \]

   How to achieve the objective:
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
- Python programming
- Linear algebra / Calculus
WHAT WILL NOT BE COVERED?

- Basic Machine Learning knowledge
- 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
• Project (up to two students) (25%)
  • Project proposal presentation 5%: Oct 22, 27
  • Written project proposal 5% (3 pages): Oct 30 midnight
  • Final project presentation 10%: Dec 10, 15 in-class
  • Written final report 5% (6 pages): Dec 18 midnight

No make-up assignment
Office Hour

- Hyun Soo Park: M 2-3pm @ Zoom
- Yasamin Jafarian: T/W 2-3pm @ Zoom
- Yang Yang: Th 2-3pm @ Zoom
Information

Syllabus
Instructor: Hyun Soo Park (hspark at umn.edu)
Office hour: Mon 2:00pm-3:00pm (Zoom)

TA1: Yasamin Jafarian (yasamin at umn.edu)
Office hour: Tue/Wed 2:00pm-3:00pm (Zoom)

TA2: Yang Yang (yang5276 at umn.edu)
Office hour: Thur 2:00pm-3:00pm (Zoom)

Teaching mode: online

Textbook: Not required but the following books will be frequently referred: