CSCI 5561 Computer Vision

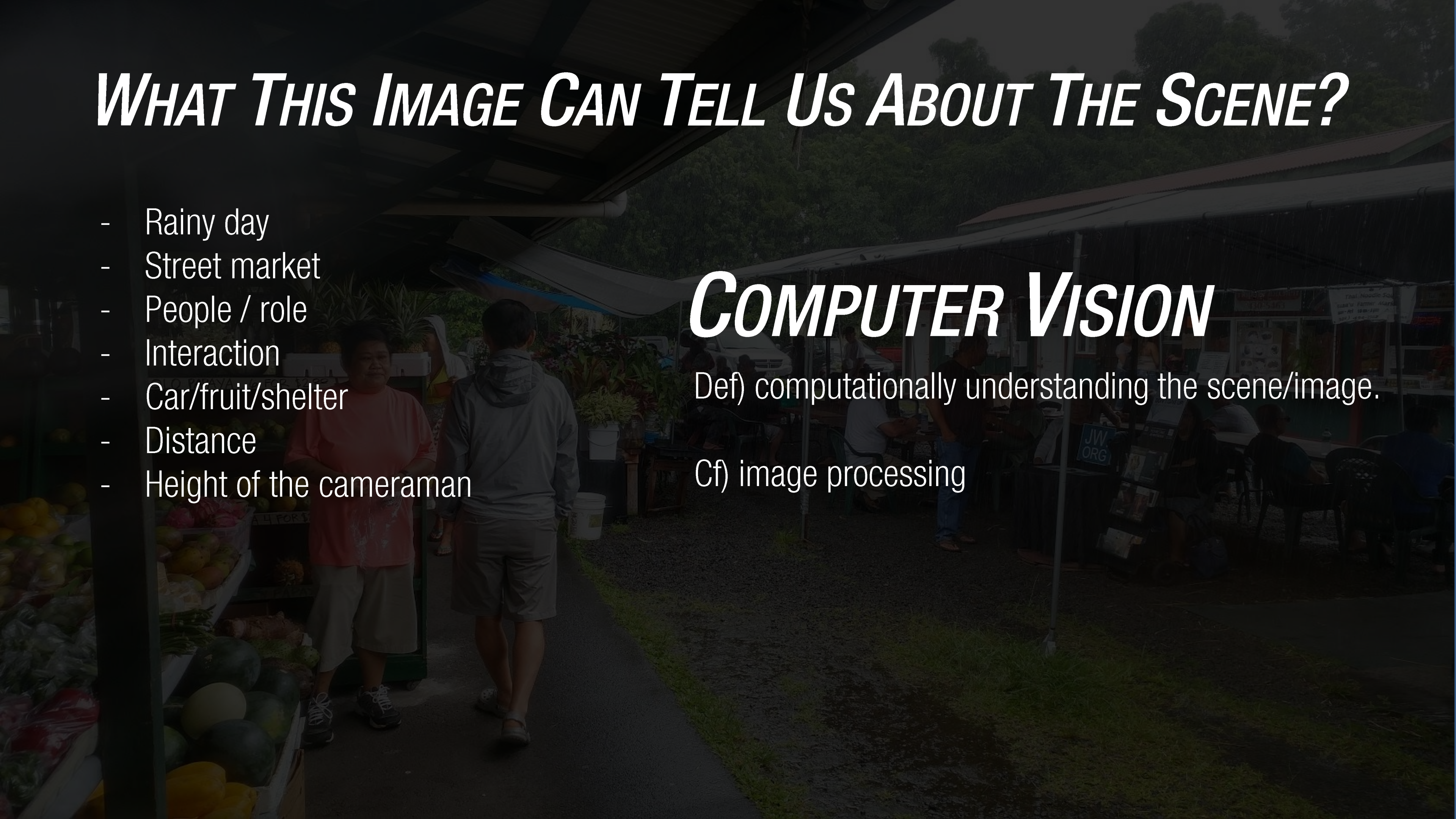
Hyun Soo Park

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

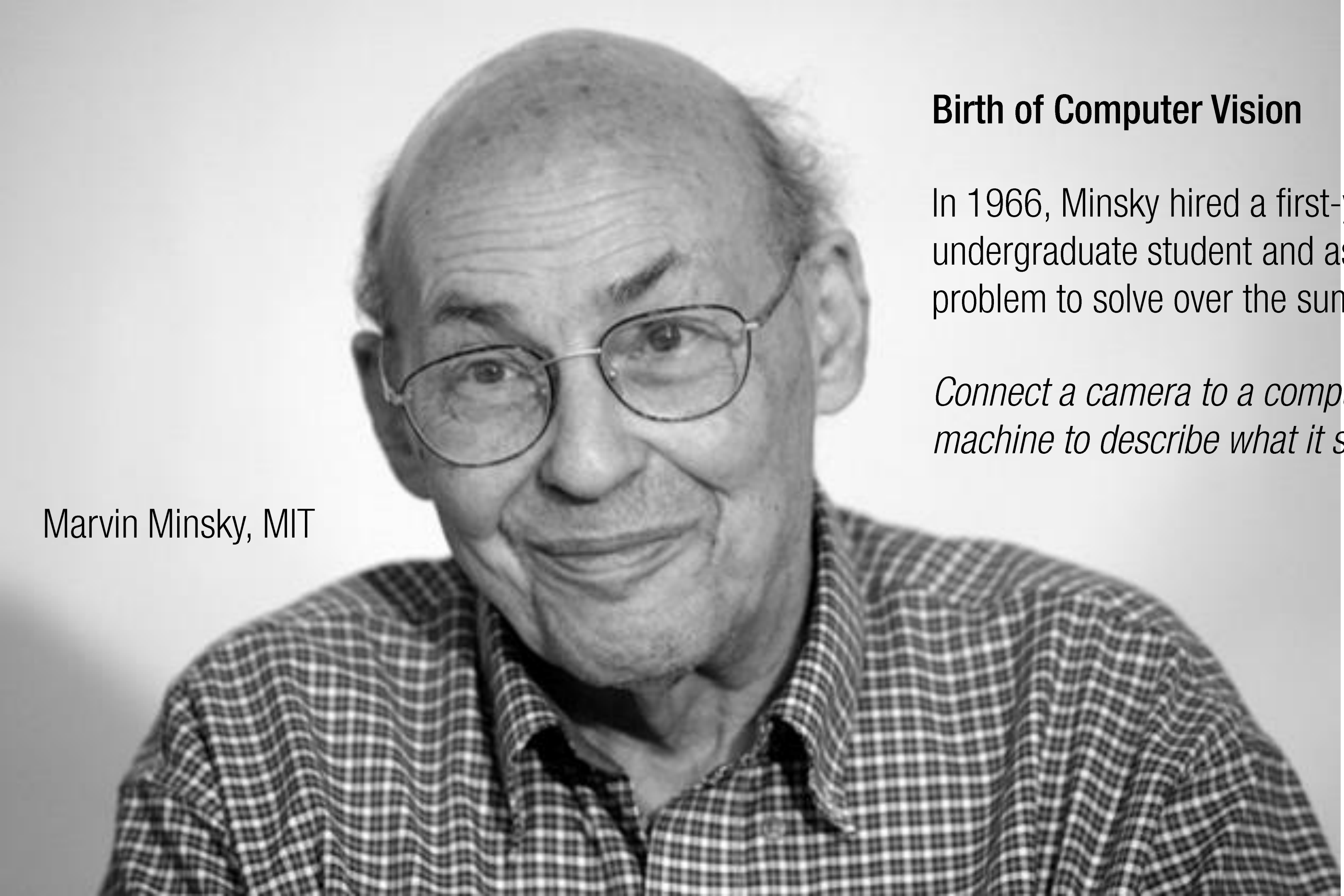
Extremely difficult

# 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

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

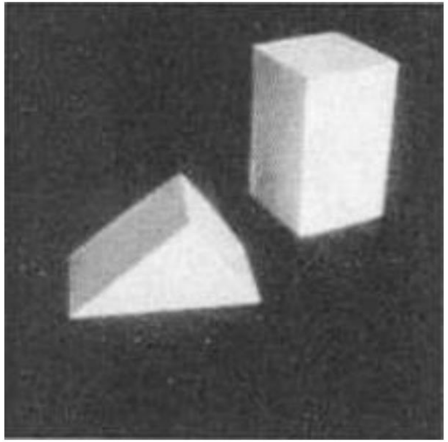
THE SUMMER VISION PROJECT

Seymour Papert

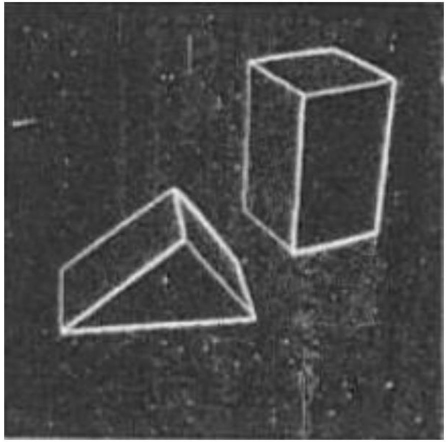
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the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

1960's: interpretation of synthetic worlds

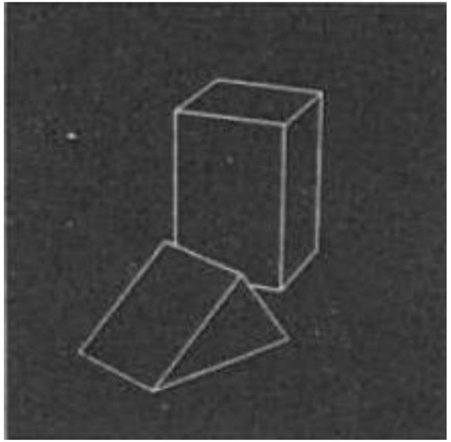
Larry Roberts
"Father of Computer Vision"



Input image



2x2 gradient operator

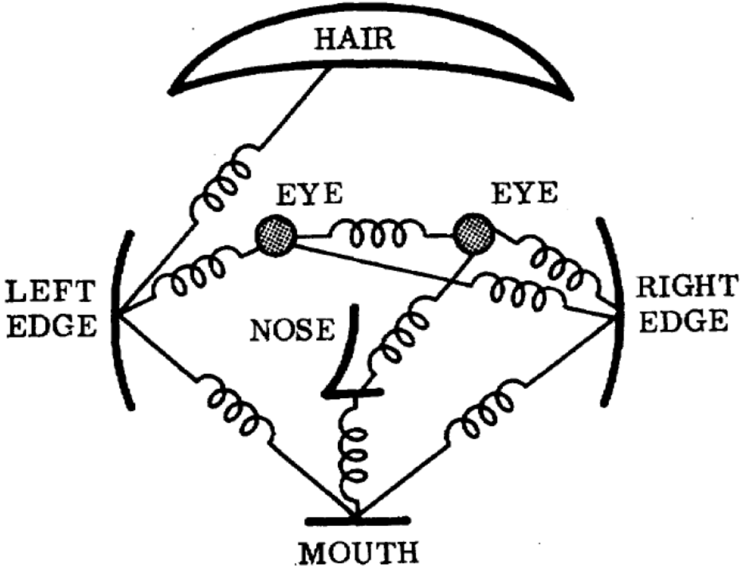


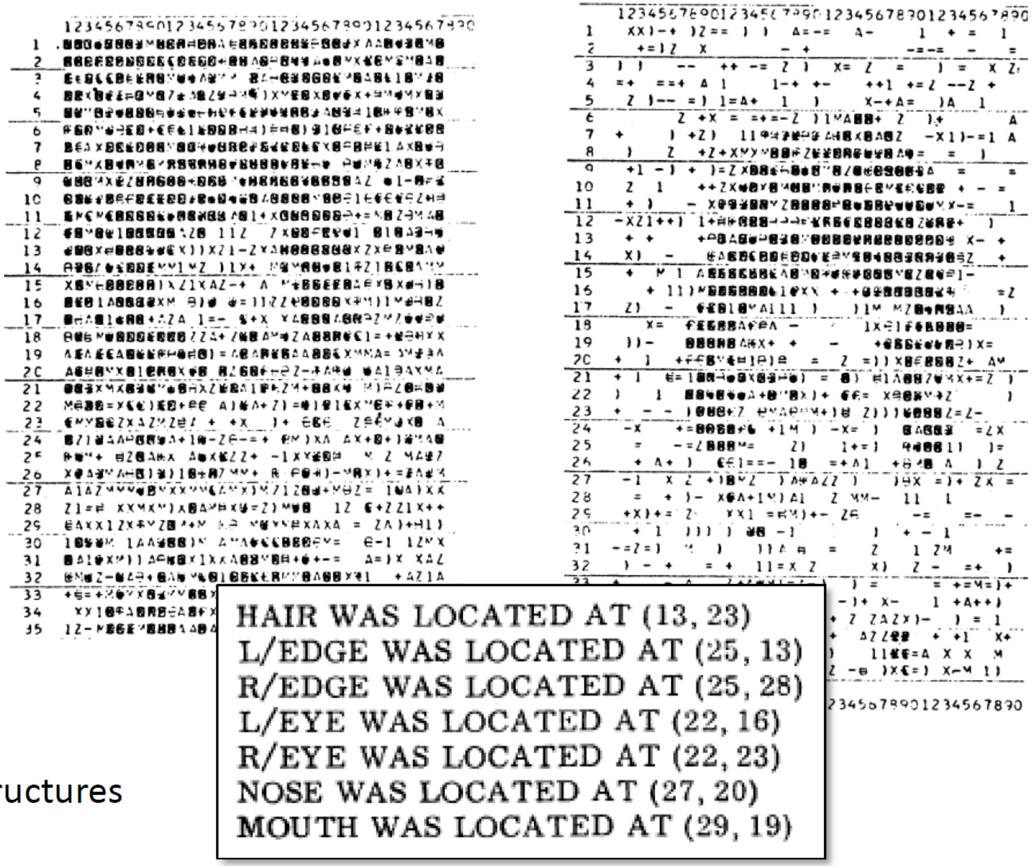
computed 3D model
rendered from new viewpoint

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

Slide credit: Steve Seitz

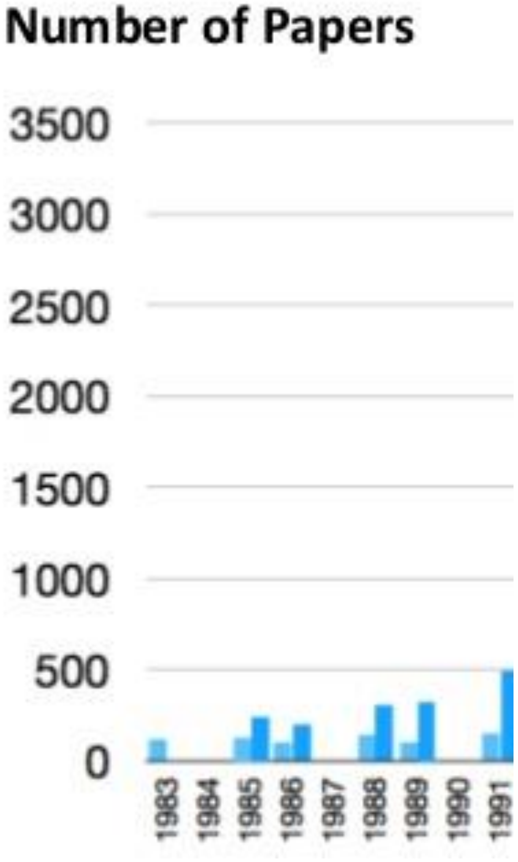
1970’s: some progress on interpreting selected images





The representation and matching of pictorial structures
Fischler and Elschlager, 1973

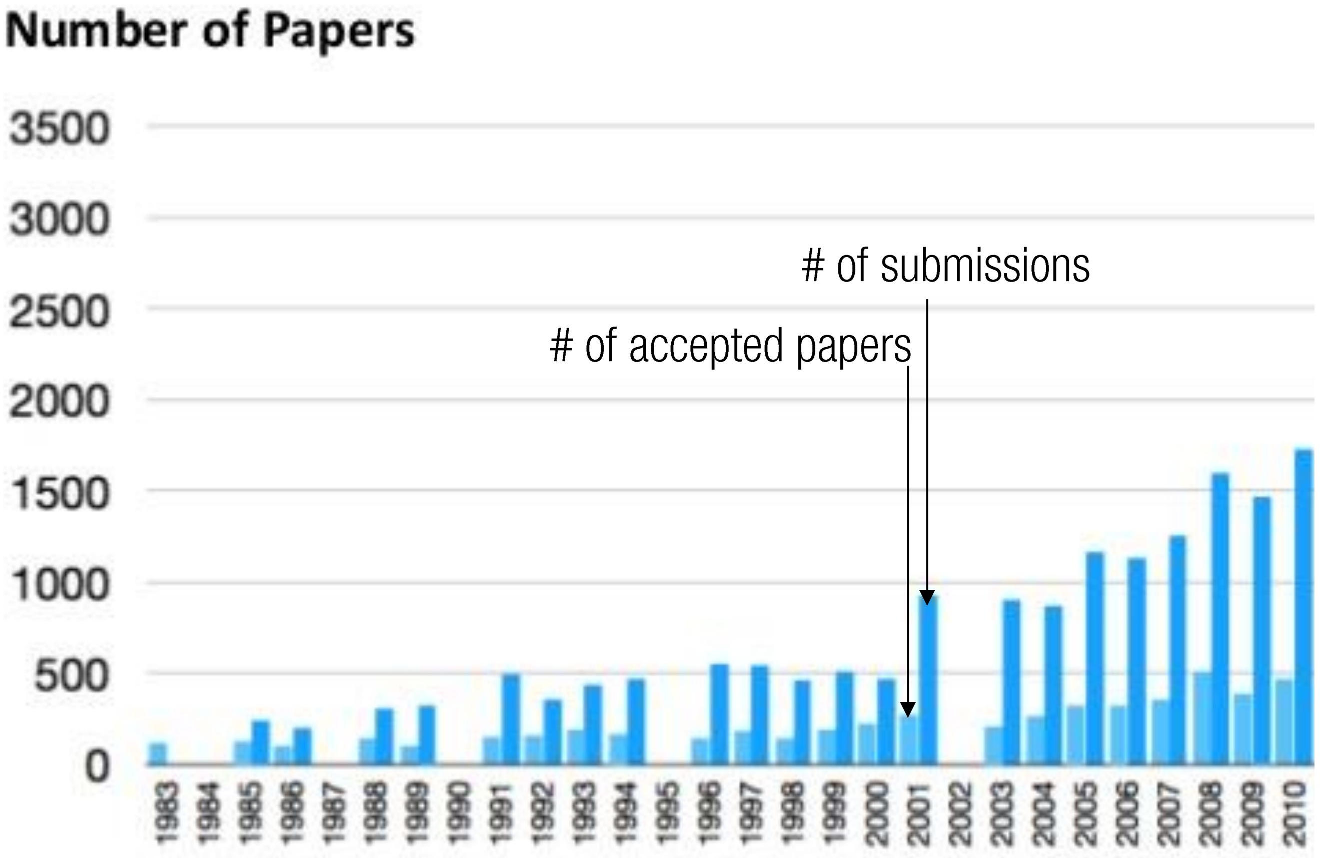
Conference on Computer Vision and Pattern Recognition



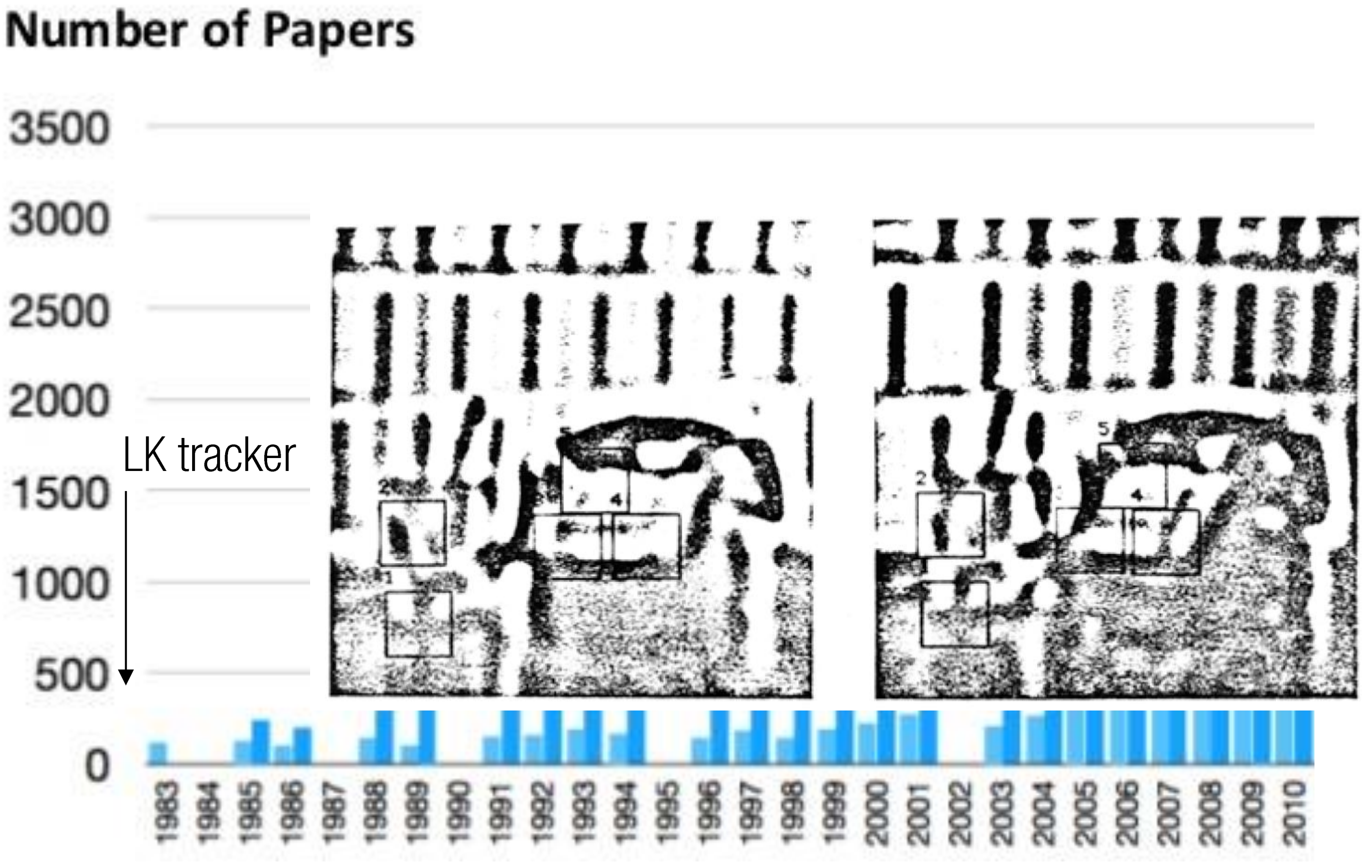
Takeo Kanade, CMU

Dana Ballard, U of Rochester

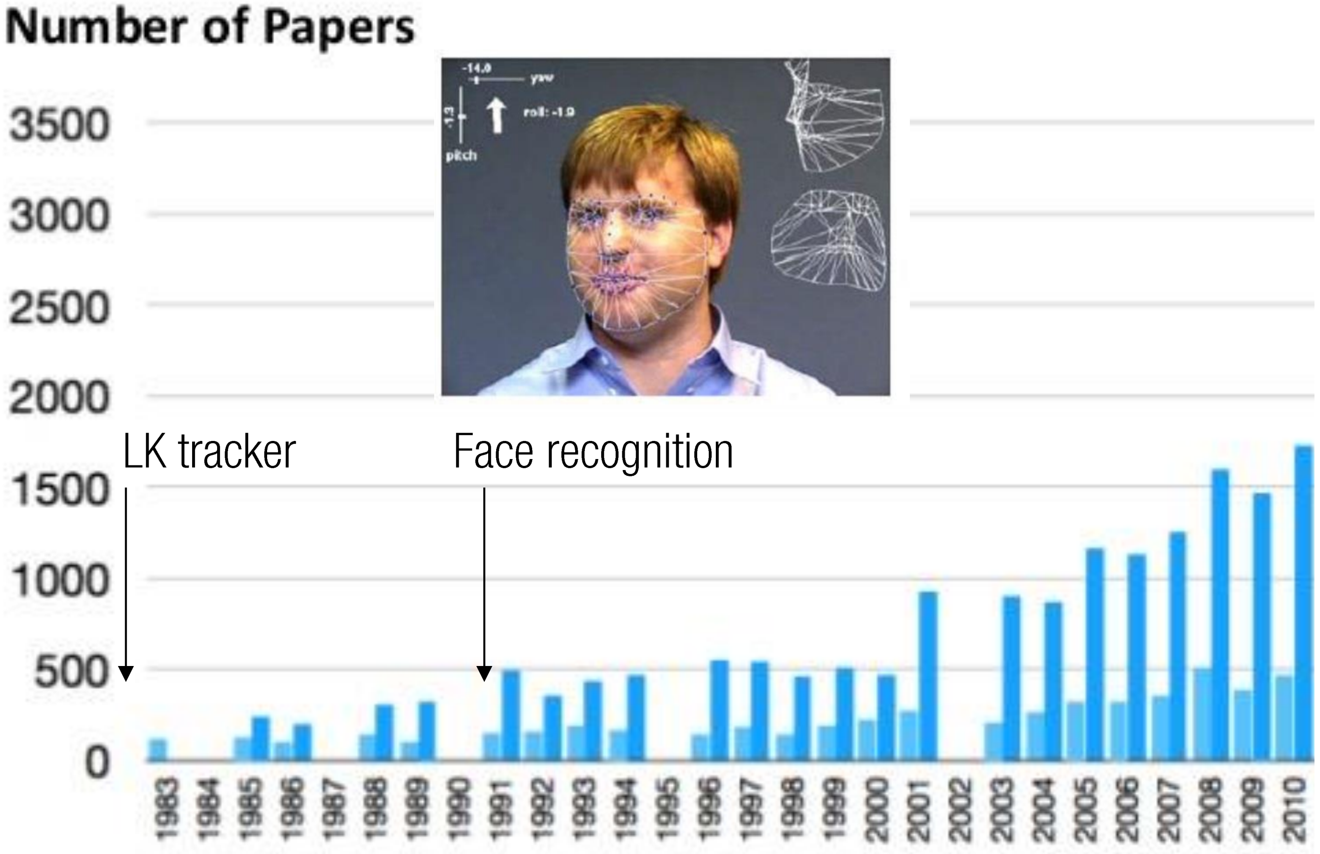
Conference on Computer Vision and Pattern Recognition



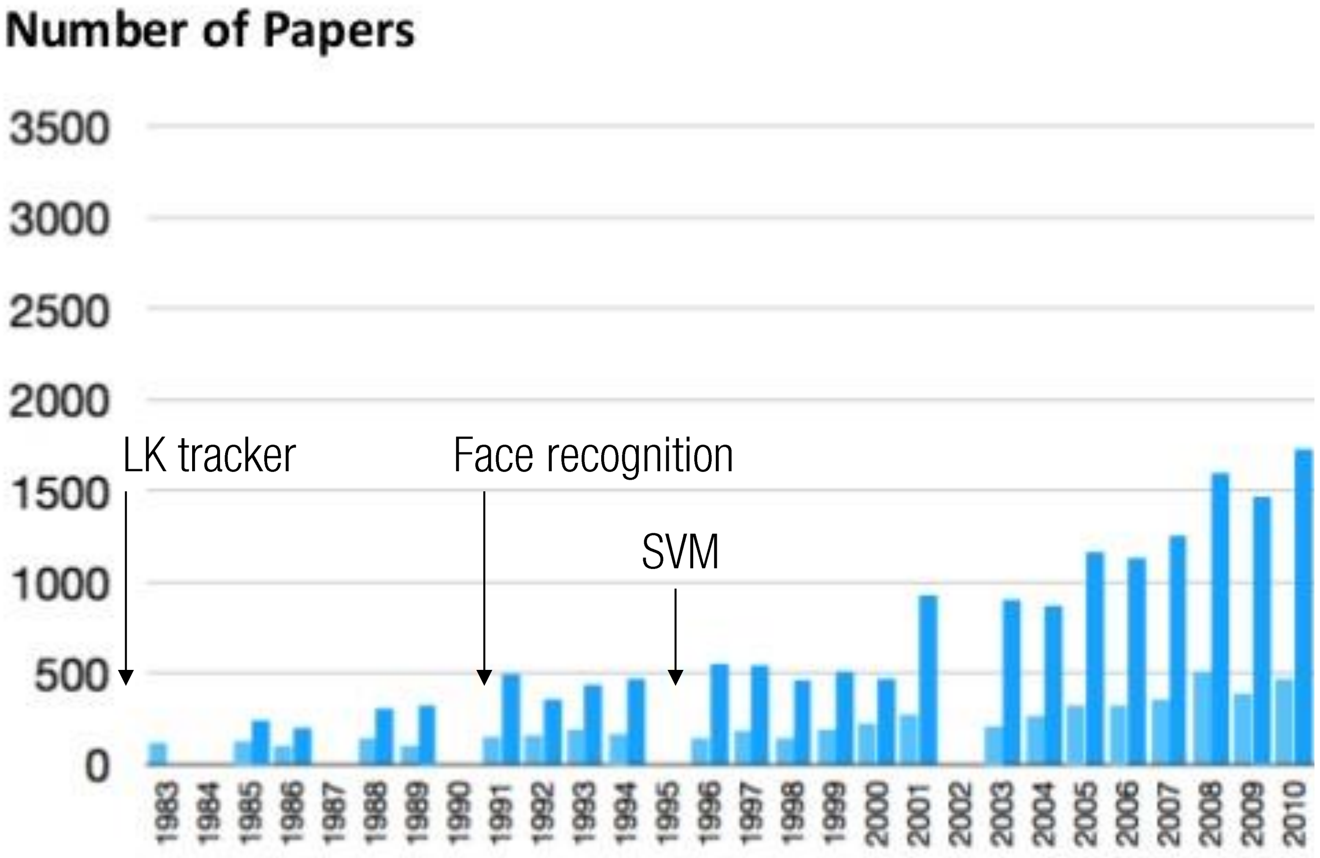
Conference on Computer Vision and Pattern Recognition



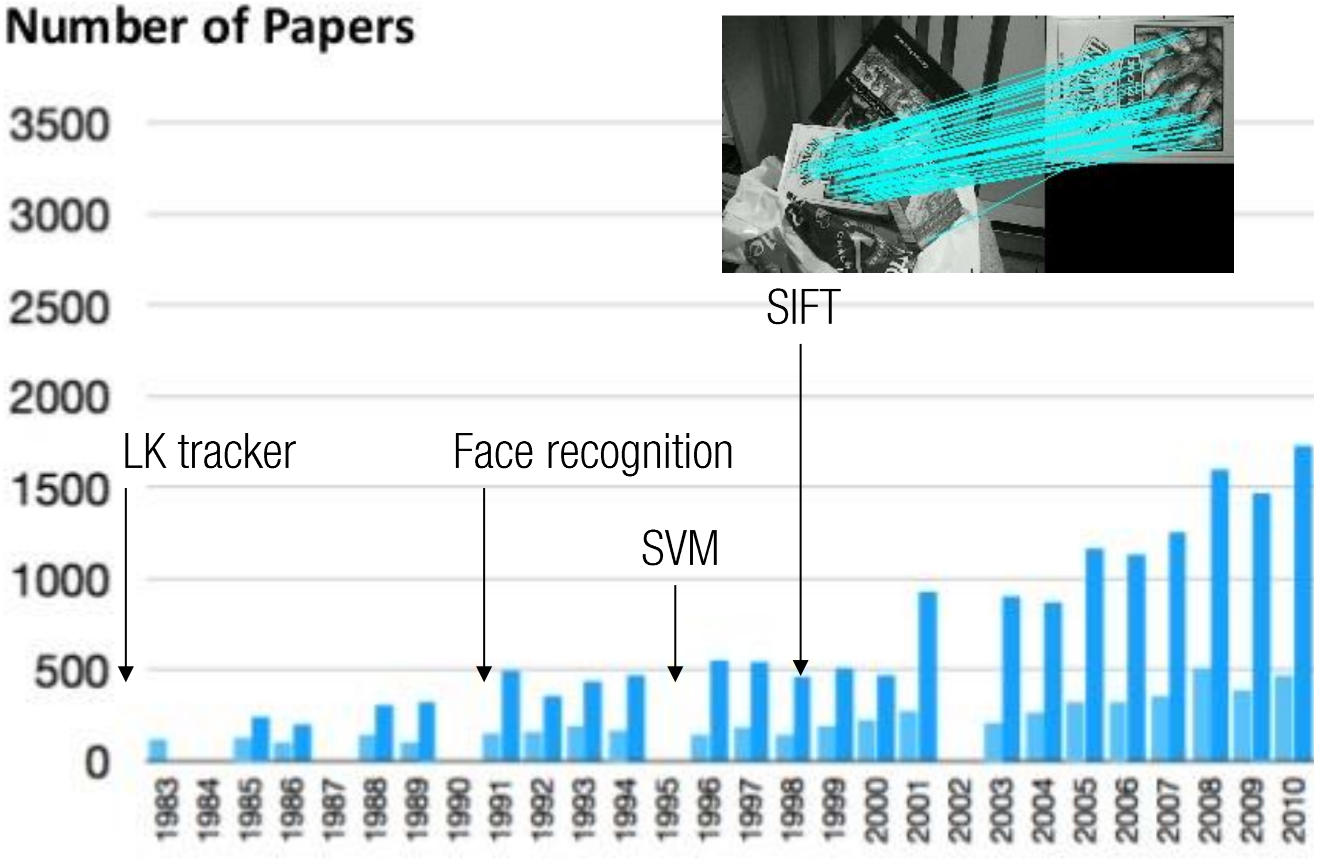
Conference on Computer Vision and Pattern Recognition



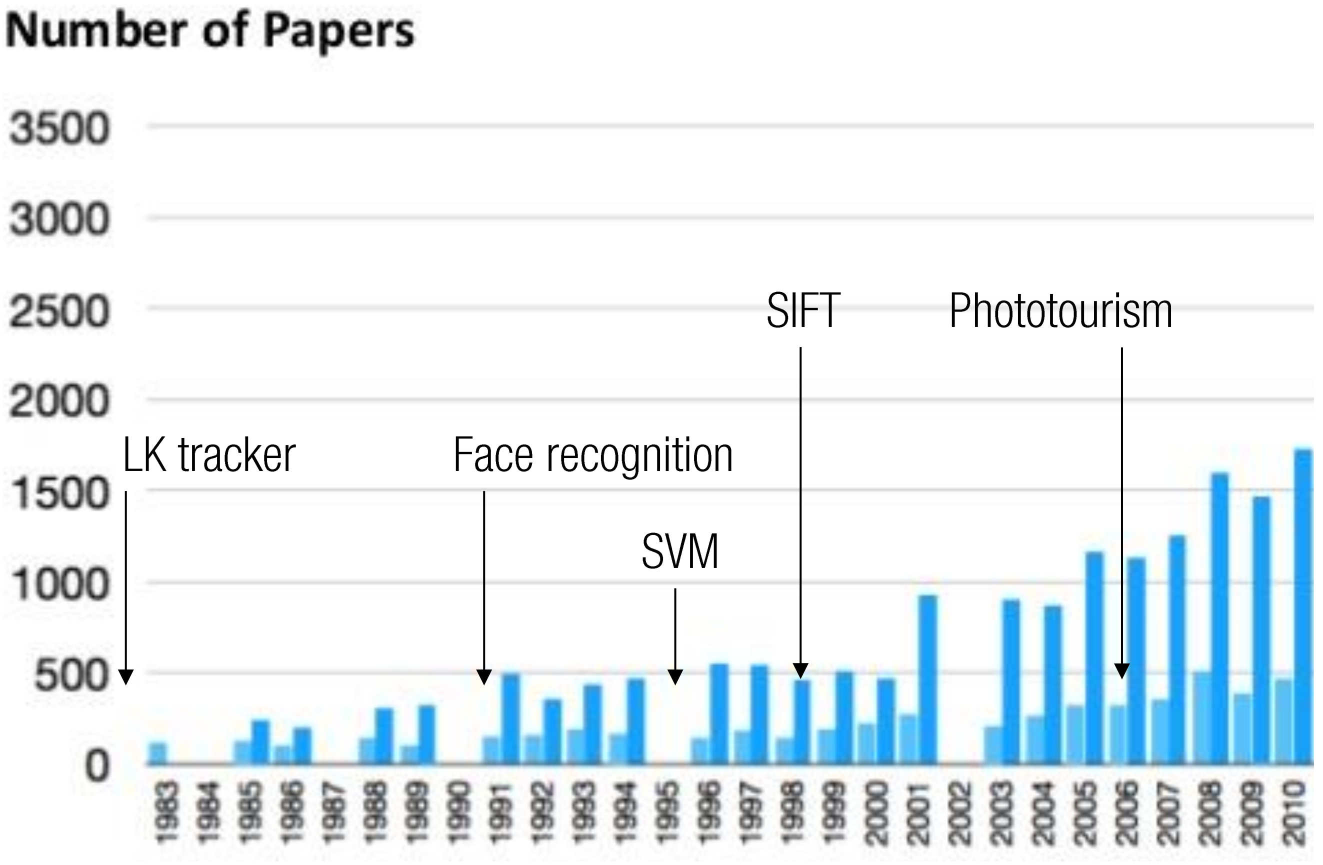
Conference on Computer Vision and Pattern Recognition



Conference on Computer Vision and Pattern Recognition



Conference on Computer Vision and Pattern Recognition



Conf

Num

350

300

250

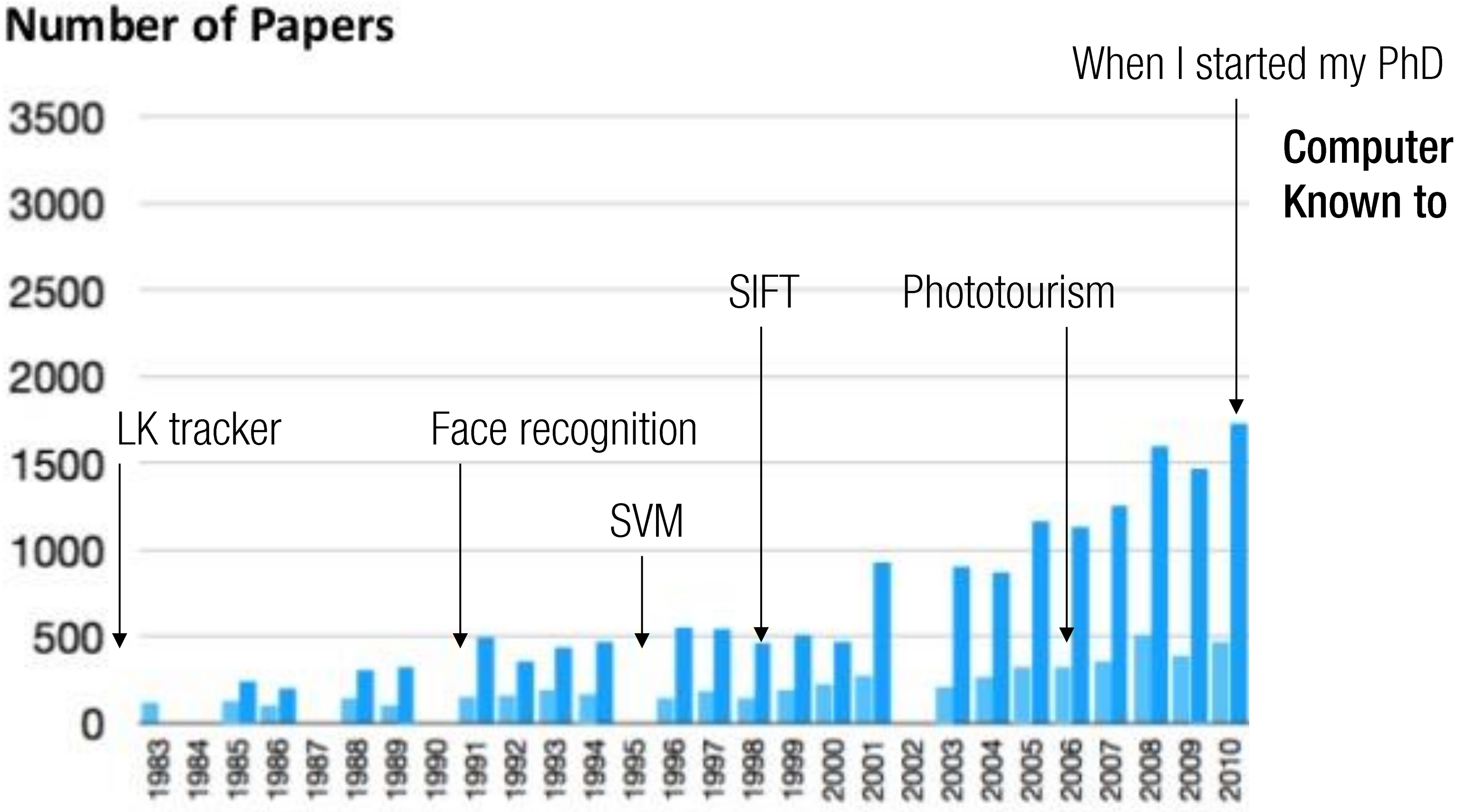
200

150

100

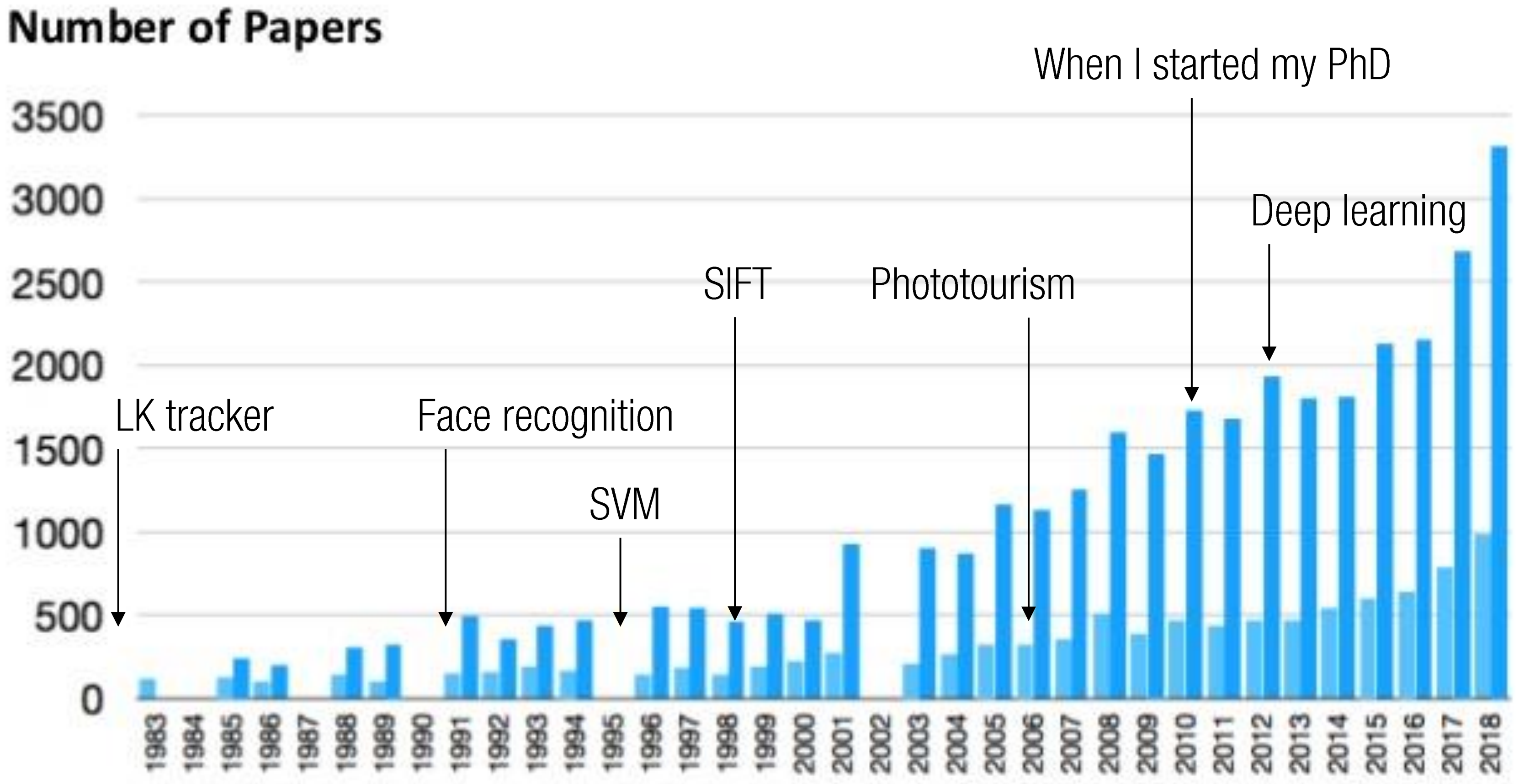
50

Conference on Computer Vision and Pattern Recognition

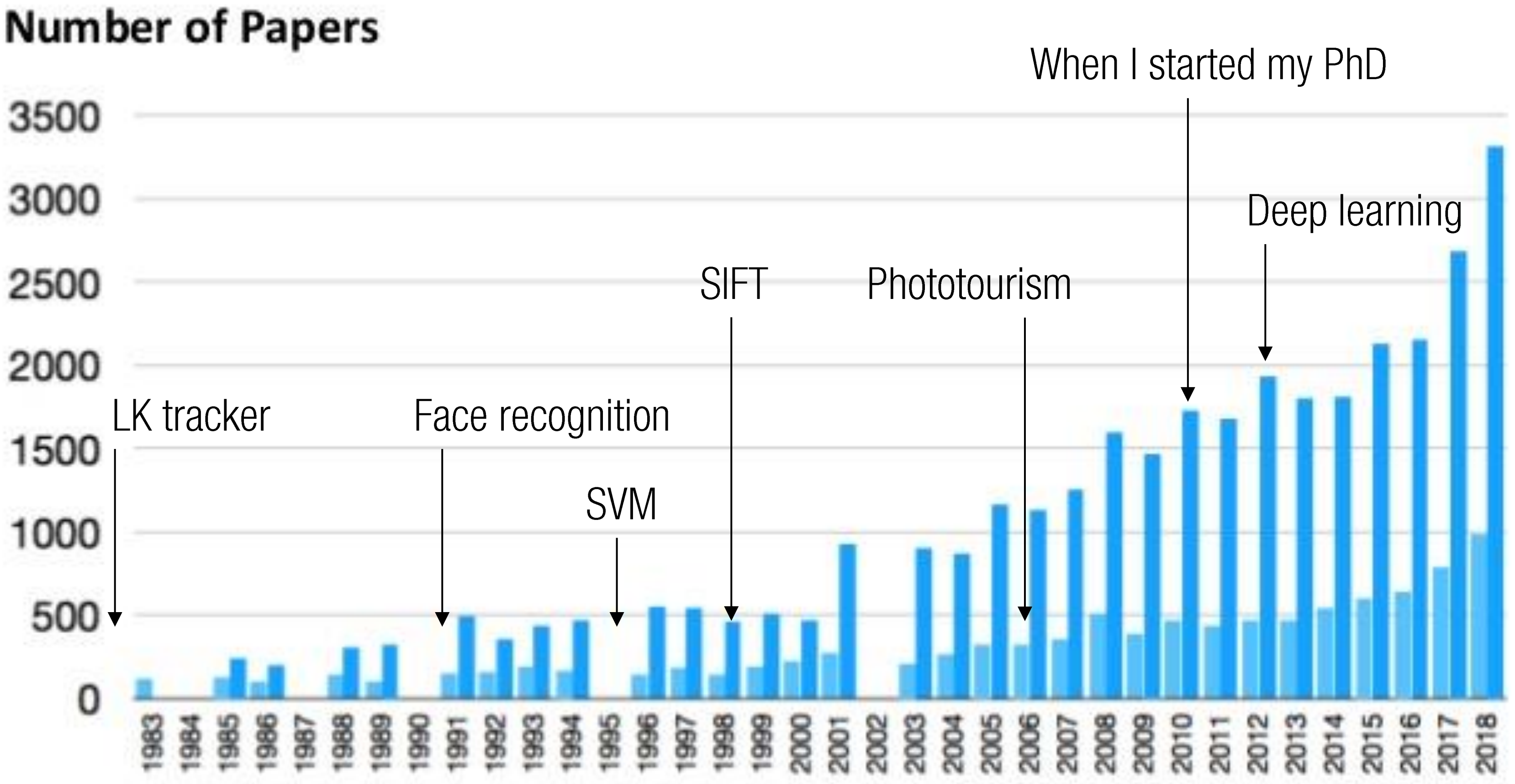


Computer vision
Known to be NOT working

Conference on Computer Vision and Pattern Recognition



Conference on Computer Vision and Pattern Recognition




12:03
3:08
LTE
SEND TO
BBC
NEWS

Platinum Sponsors

Gold Sponsors

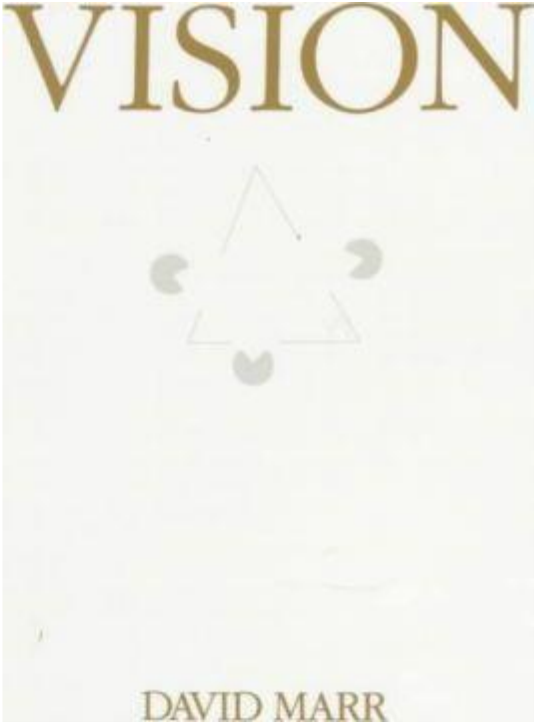
Silver Sponsors

CVPR sponsor (2018)

How to Make a Computer to Understand an Image?

MARR'S VISION

David Marr (1945-1980)



1982

Three Levels of Visual Perception

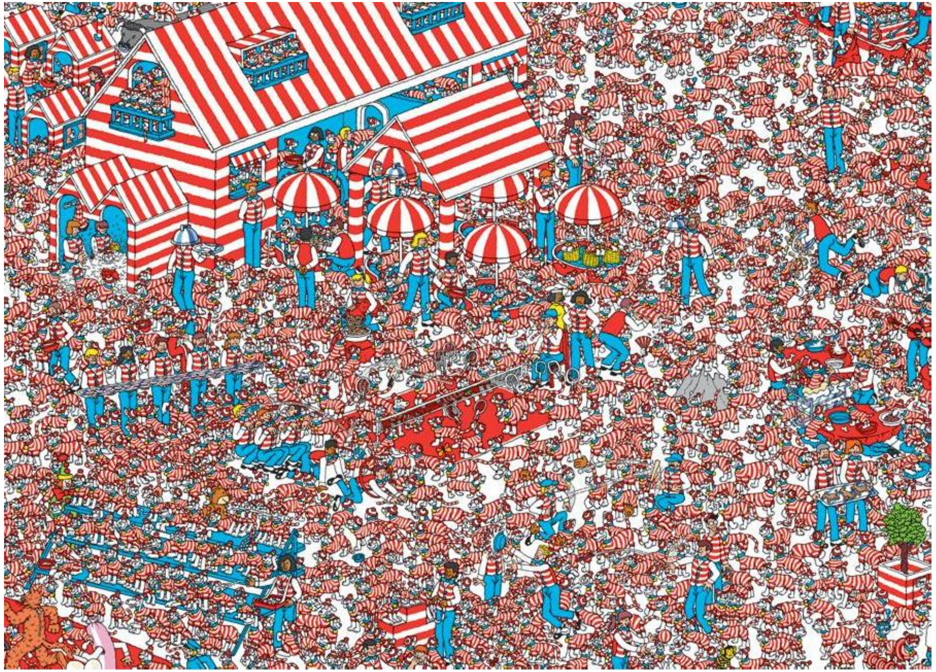
David Marr (1945-1980)

THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

1. Computational

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



How to computationally find Waldo?

THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

1. Computational

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

How to formulate the objective:

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

THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

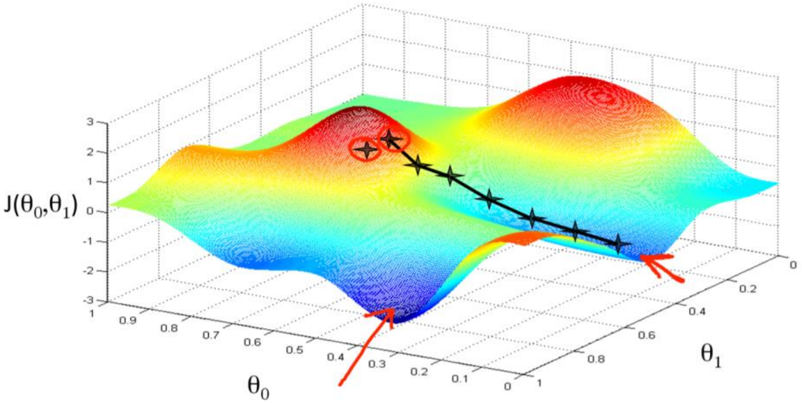
1. Computational

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

How to formulate the objective:

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

How to achieve the objective:

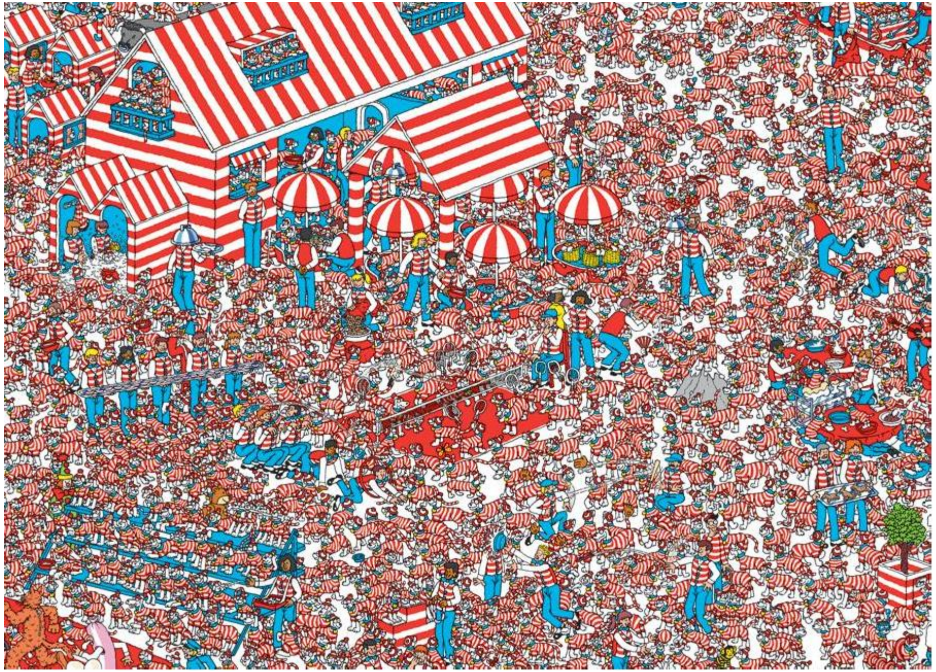


THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

2. Algorithmic

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



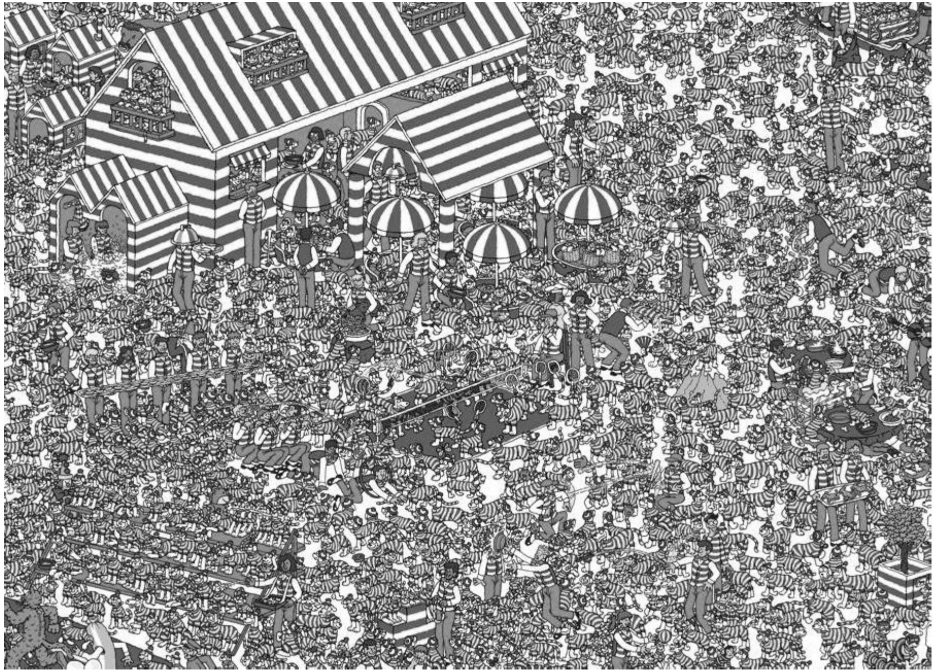
How to represent the image?

THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

2. Algorithmic

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



How to represent the image?

THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

2. Algorithmic

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

78	99	108	106	117	117	149	205	206	169	109	118	147	150	155	217	222	233
78	98	106	104	116	115	148	205	206	169	109	118	147	150	155	217	222	233
78	103	113	111	121	120	151	207	208	170	108	117	146	149	154	217	223	233
74	145	176	170	164	158	183	226	229	178	93	103	134	136	143	224	233	239
62	147	184	177	169	161	187	227	230	178	90	101	133	134	142	225	234	240
159	165	168	167	165	158	187	236	239	188	101	111	135	136	143	225	234	240
226	179	158	162	162	157	189	244	247	195	108	118	136	137	144	226	235	240
213	178	163	166	163	156	189	242	245	194	106	117	137	137	144	226	235	240
225	162	133	137	161	166	193	243	246	194	110	120	132	131	139	227	236	241
235	152	114	119	159	171	196	243	245	196	113	123	130	128	137	229	238	243
215	143	111	114	162	178	199	242	245	194	106	116	125	124	132	227	237	242
155	112	92	91	171	200	212	240	245	187	86	98	110	109	118	223	234	239
161	114	91	91	170	198	211	241	247	187	86	98	108	107	117	223	234	239
141	110	96	93	174	205	212	234	239	184	90	101	119	119	127	223	234	240
85	100	106	99	187	224	216	213	220	176	102	111	147	150	155	226	234	240
87	100	105	98	186	223	216	215	221	176	102	112	146	148	154	226	234	240
111	140	152	146	206	236	210	175	184	144	77	88	139	143	148	218	225	234
117	151	166	160	213	240	209	164	173	135	69	81	135	140	146	215	222	232
115	148	162	156	211	239	210	167	176	138	70	81	136	141	146	216	223	233
197	215	223	223	201	198	159	89	94	101	108	113	169	175	178	225	229	238
222	238	246	248	202	190	147	66	71	88	115	118	174	180	183	228	232	240
206	202	200	203	172	165	132	72	73	100	139	141	203	211	210	225	226	236
192	160	146	149	134	129	110	73	70	110	171	173	241	250	246	222	219	232
201	171	157	160	143	139	117	76	73	110	165	166	237	247	243	226	223	235
151	136	130	132	113	108	90	57	54	105	190	186	232	238	233	202	196	218

255	255	245	215
223	222	213	197
111	109	123	145
114	112	125	145
222	222	226	233
253	252	254	255
248	248	250	254
246	246	248	251
246	246	247	249
245	245	248	251
245	244	248	253
245	243	248	254
249	248	248	246
254	255	248	235
253	253	248	238

How to represent the image?

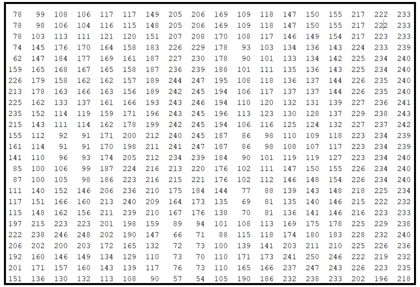
THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

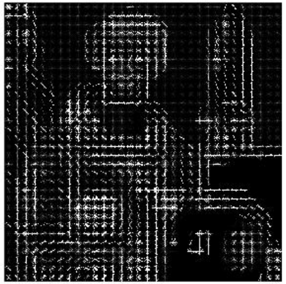
2. Algorithmic

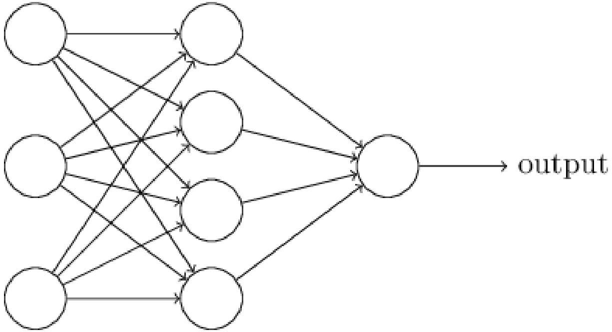
e.g., what representation can implement the computation?

How to represent the image?



78	99	108	106	117	117	149	205	206	169	109	118	147	150	155	217	222	233
78	98	106	104	116	115	148	205	206	169	109	118	147	150	155	217	222	233
78	103	113	111	121	120	151	207	208	170	108	117	146	149	154	217	223	233
74	145	176	170	164	158	183	226	229	178	93	103	134	136	143	224	233	239
62	147	184	177	169	161	187	227	230	178	90	101	133	134	142	225	234	240
159	165	168	167	165	158	187	236	239	188	101	111	135	136	143	225	234	240
226	179	158	162	162	157	189	244	247	195	108	118	136	137	144	226	235	240
213	178	163	166	163	156	189	242	245	194	106	117	137	137	144	226	235	240
225	162	133	137	161	166	193	243	246	194	110	120	132	131	139	227	236	241
235	152	114	119	159	171	196	243	245	196	113	123	130	128	137	229	238	243
215	143	111	114	162	178	199	242	245	194	106	116	125	124	132	227	237	242
155	112	92	91	171	200	212	240	245	187	86	98	110	109	118	223	234	239
161	114	91	91	170	198	211	241	247	187	86	98	108	107	117	223	234	239
141	110	96	93	174	205	212	234	239	184	90	101	119	119	127	223	234	240
85	100	106	99	187	224	216	213	220	176	102	111	147	150	155	226	234	240
87	100	105	98	186	223	216	215	221	176	102	112	146	148	154	226	234	240
111	140	152	146	206	236	210	175	184	144	77	88	139	143	148	218	225	234
117	151	166	160	213	240	209	164	173	135	69	81	135	140	146	215	222	232
115	148	162	156	211	239	210	167	176	138	70	81	136	141	146	216	223	233
197	215	223	223	201	198	159	89	94	101	108	113	169	175	178	225	229	238
222	238	246	248	202	190	147	66	71	88	115	118	174	180	183	228	232	240
206	202	200	203	172	165	132	72	73	100	139	141	203	211	210	225	226	236
192	160	146	149	134	129	110	73	70	110	171	173	241	250	246	222	219	232
201	171	157	160	143	139	117	76	73	110	165	166	237	247	243	226	223	235
151	136	130	132	113	108	90	57	54	105	190	186	232	238	233	202	196	218



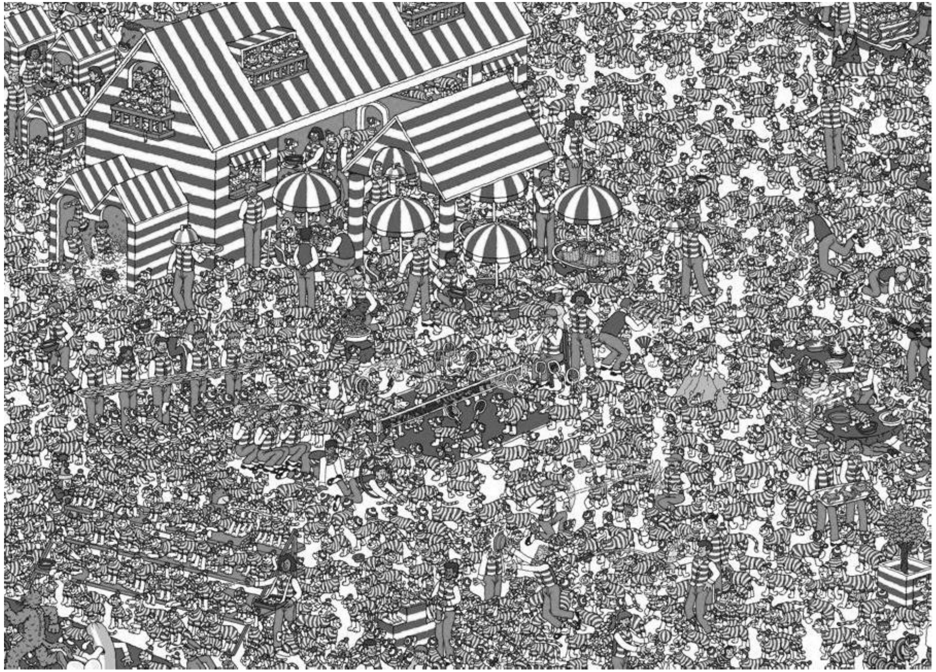


THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

3. Implementational

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

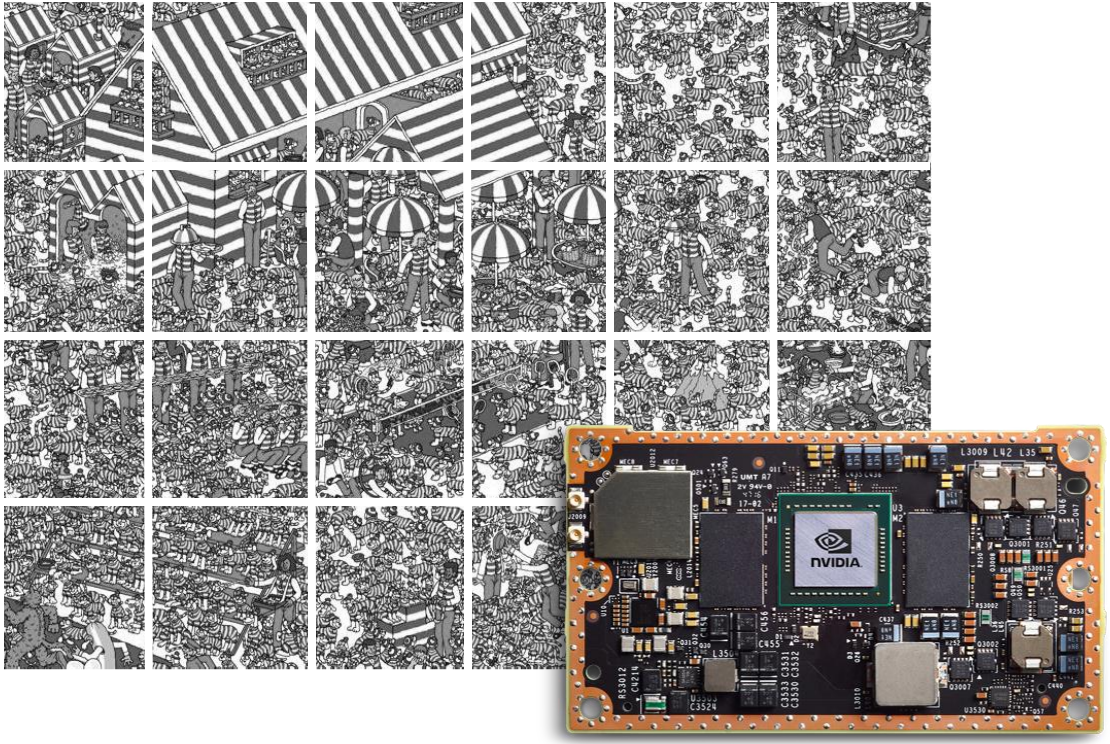


THREE LEVELS OF VISUAL PERCEPTION

David Marr (1945-1980)

3. Implementational

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

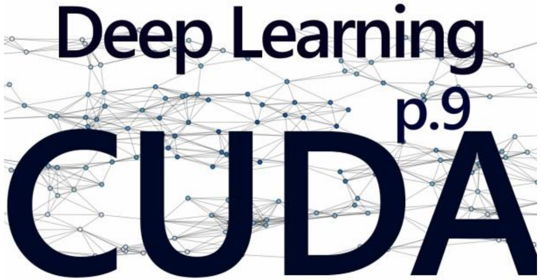


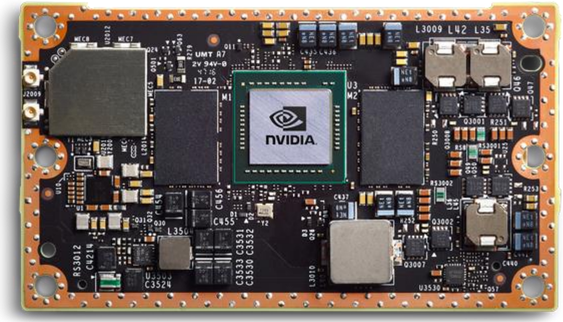
Three Levels of Visual Perception

David Marr (1945-1980)

3. Implementational

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



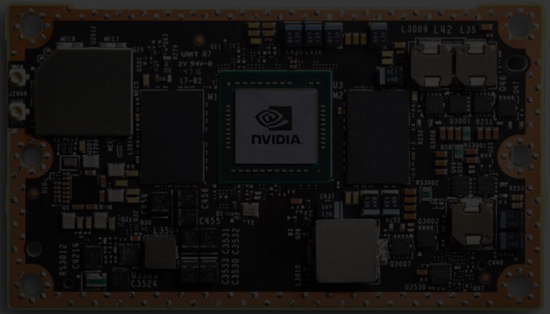


Three Levels of Visual Perception

David Marr (1945-1980)

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

Beyond the Scope of This Course.



What Will Be Covered?

Basics and 4 Rs of Computer Vision

BASICS

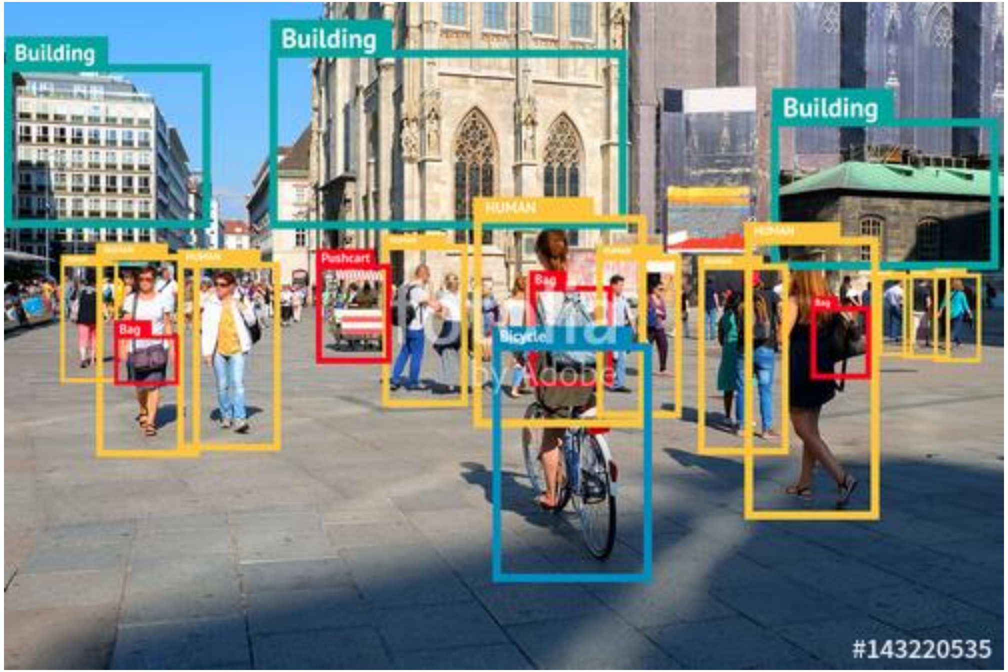
Image formation
Image convolution/filtering
Feature representation

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

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 embrassed 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 confortable 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 (10%): Dec 10th in-class
- Project (two students) (15%)
  - Written project proposal 5% (3 pages): Nov 1 midnight
  - Poster presentation 5%: Dec 12th in-class
  - Written final report 5% (6 pages): Dec 18th midnight

No make-up assignment and exam

Office Hour

- Hyun Soo Park: MW 2-3pm @ Shepherd 261
- Zhixuan Yu: TTh 1-2pm @ Shpherd 234

COURSE WEBSITE

https://www-users.cs.umn.edu/~hspark/csci5561_F2019/csci5561.html

Fall 2019 CSCI 5561
Computer Vision

Tue/Thr 4:00pm-05:15pm @ Keller Hall 3-125

Information

Syllabus

Instructor: Hyun Soo Park (hspark at umn.edu)
Office hour: Mon/Wed 2:00pm-3:00pm (Shepherd Laboratory 261)

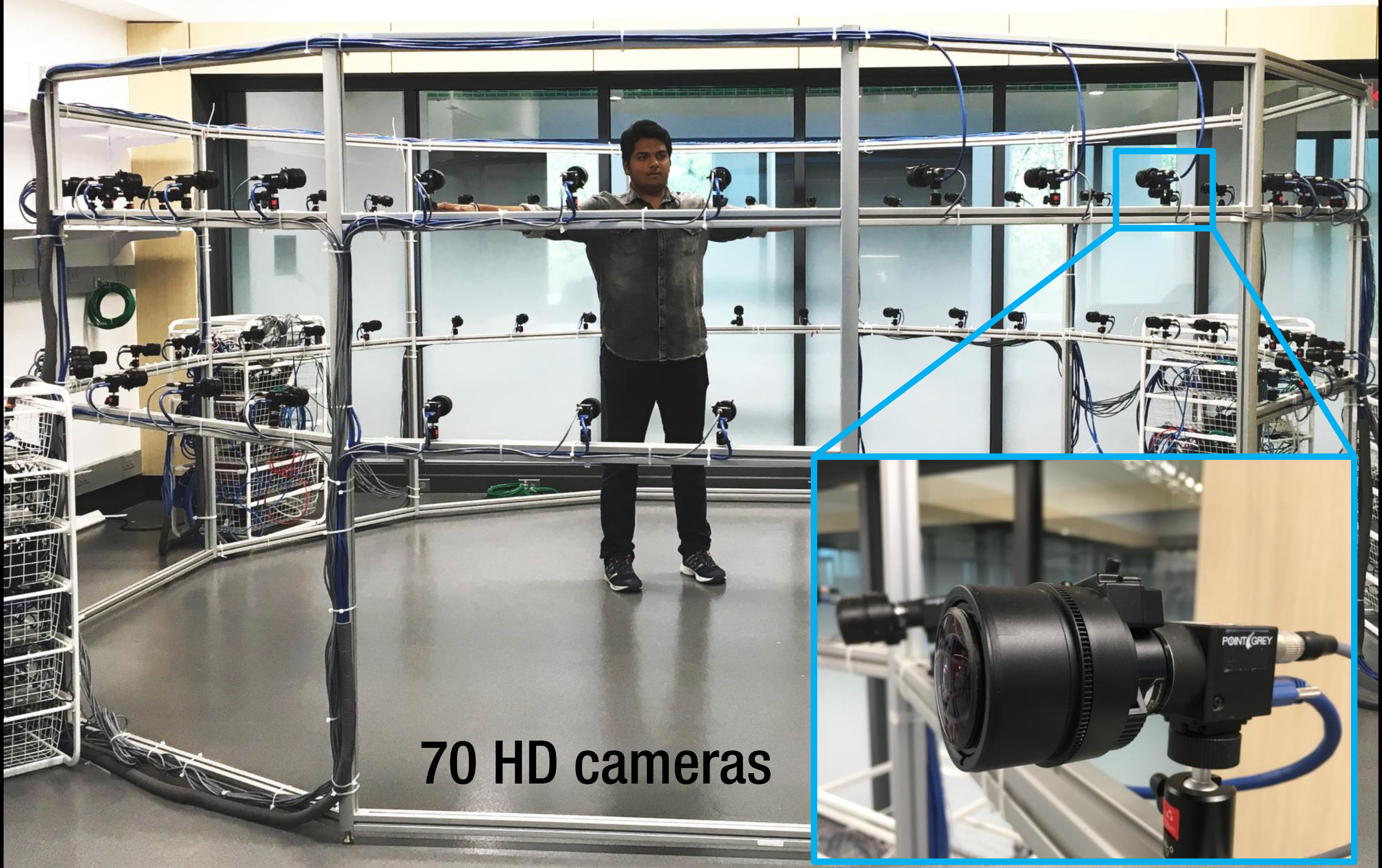
TA: Zhixuan Yu (yu000064 at umn.edu)
Office hour: Tue/Thr 1:00pm-2:00pm (Shepherd Laboratory 234)

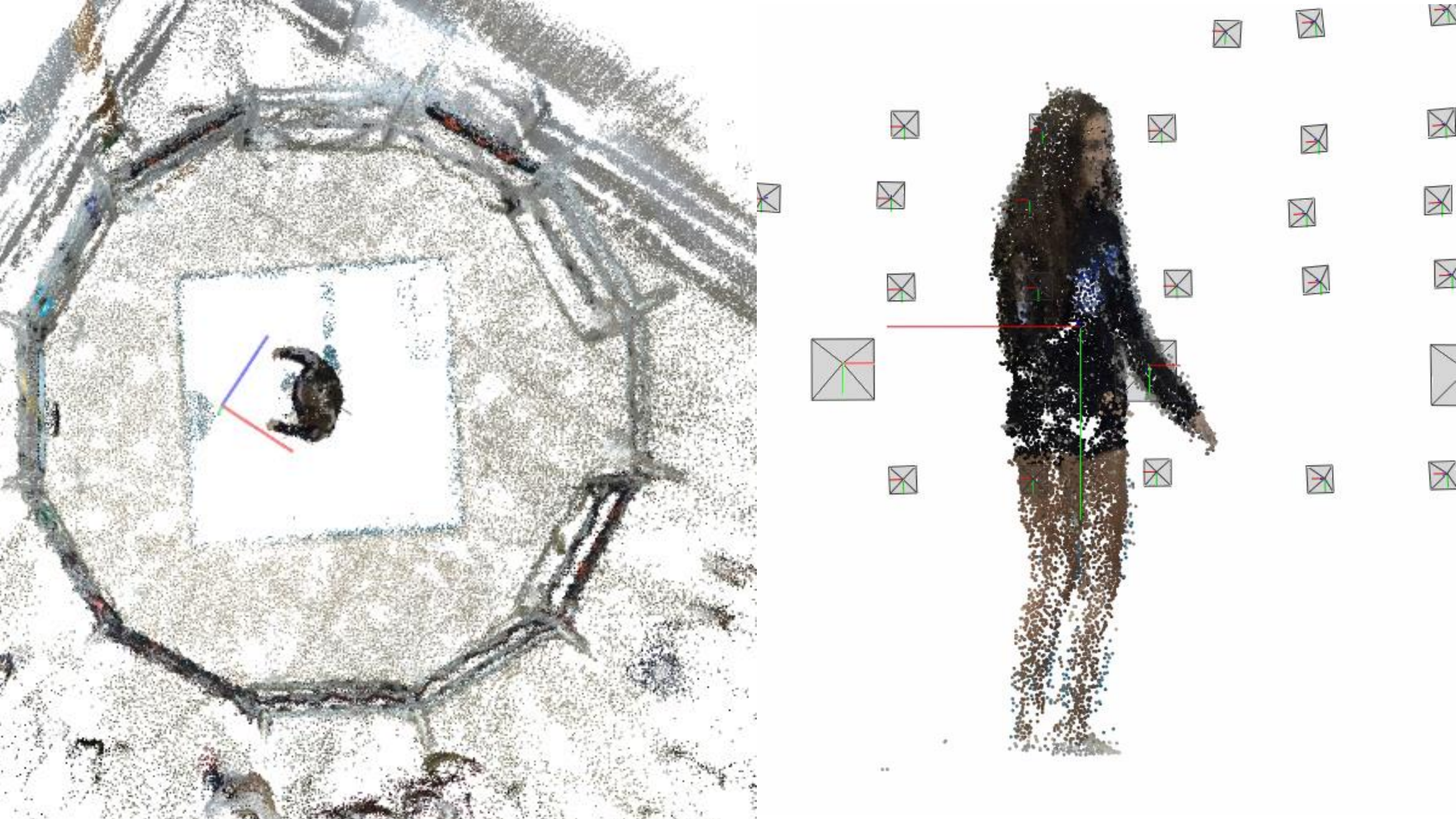
Textbook: Not required but the following books will be frequently referred:
+ "Computer Vision: Algorithms and Applications", Richard Szeliski
+ "Computer Vision: A Modern Approach", David A. Forsyth and Jean Ponce

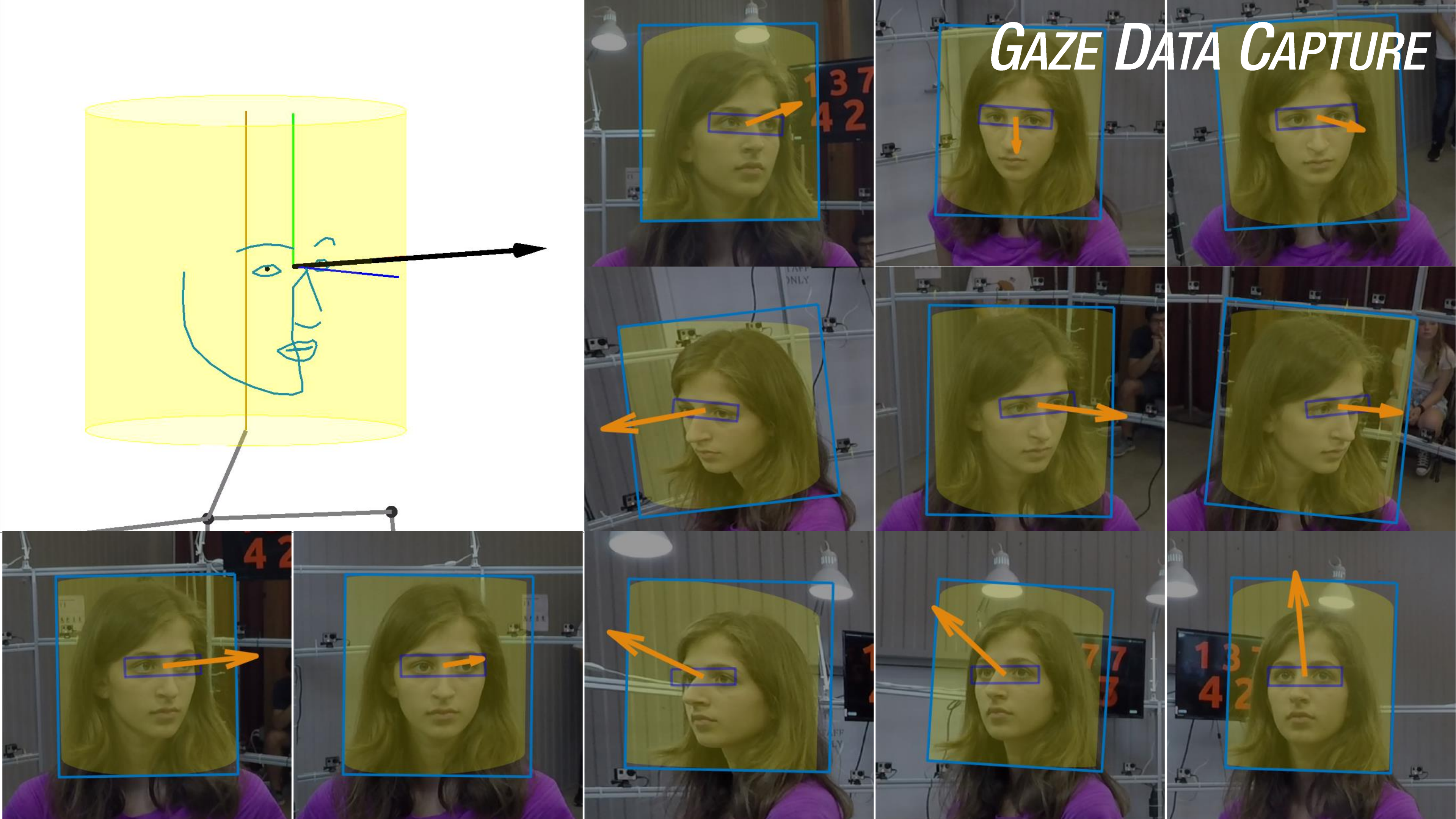
The course materials are inspired by the slides of Jianbo Shi (UPenn), Kris Kitani (CMU), James Hays (GATECH), Kristen Grauman (UT Austin), Steve Seitz (UW), and Robert Collins (PSU).

- Syllabus
- Schedule
- Lecture slide

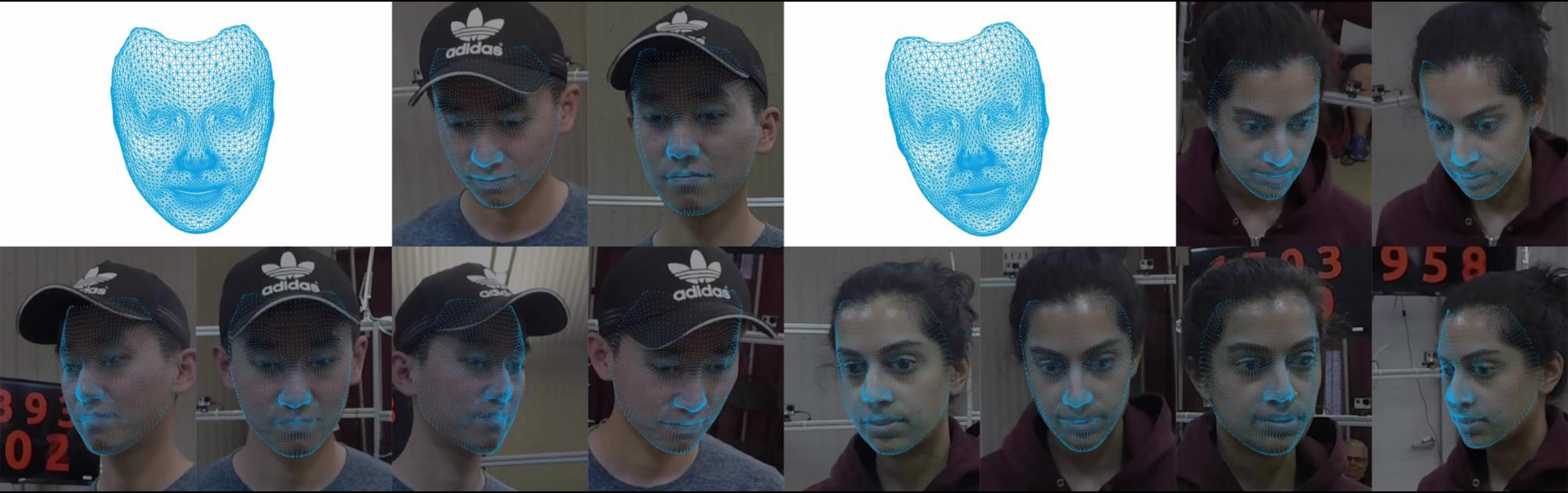
Who am I?


70 HD cameras
POINT GREY



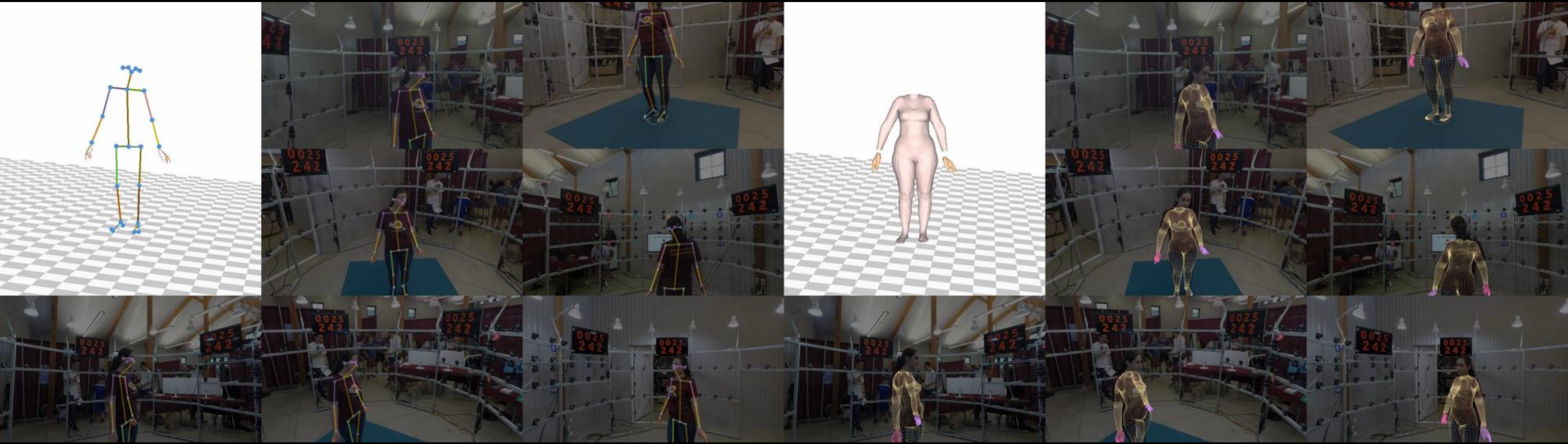

GAZE DATA CAPTURE

Face Data Capture



Body Data Capture

Body Data Capture

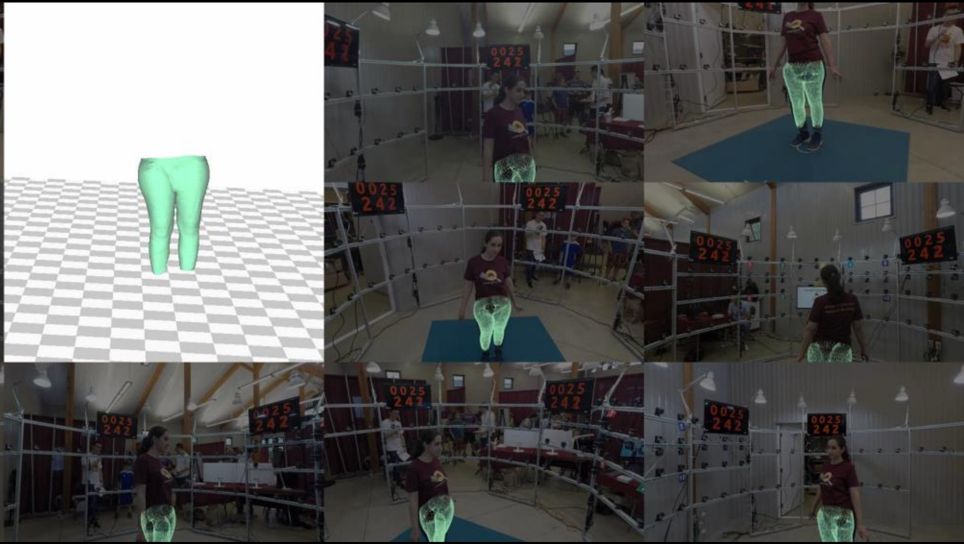


Skeleton

Body skin

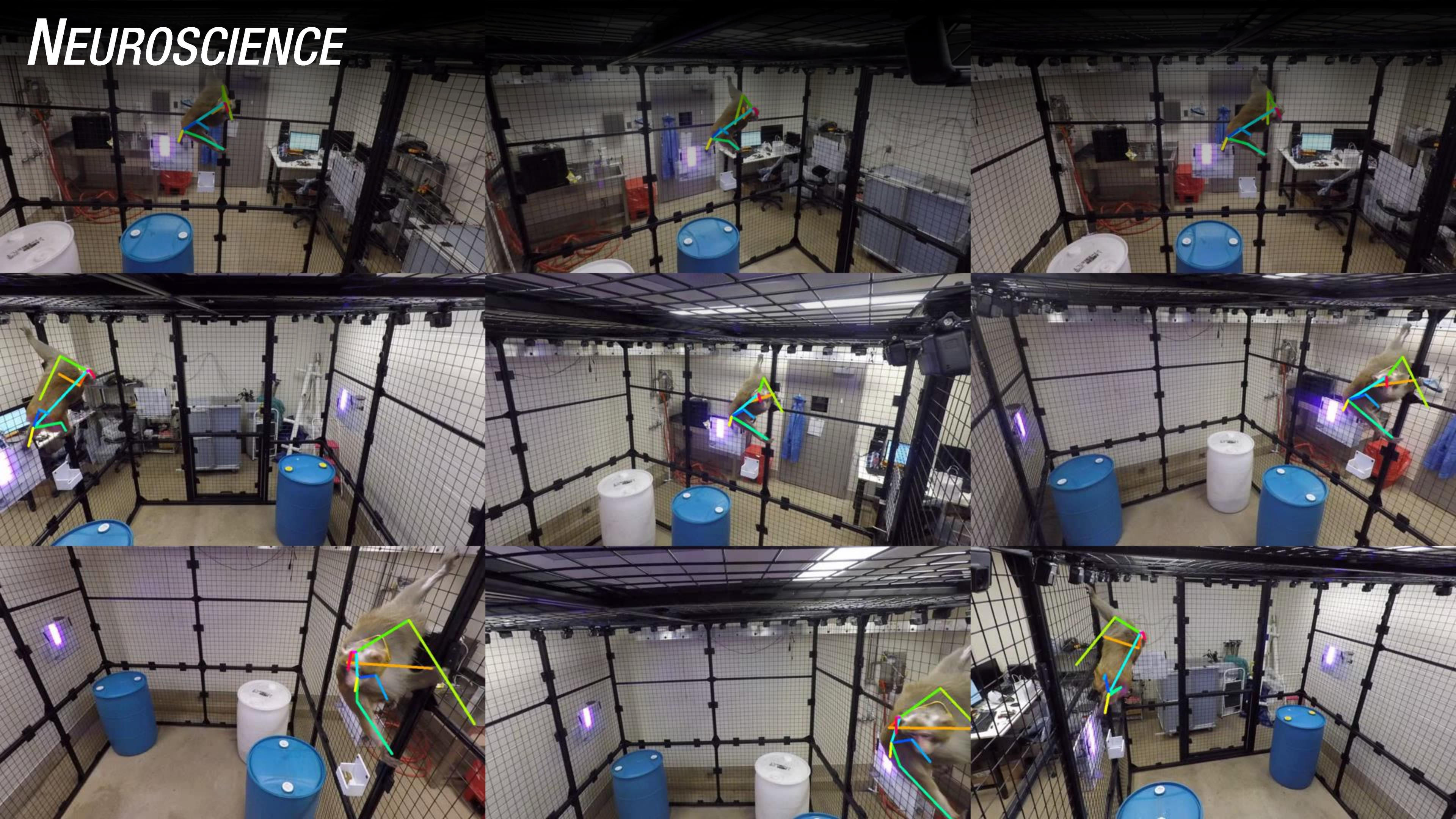
Cloth Data Capture

Top

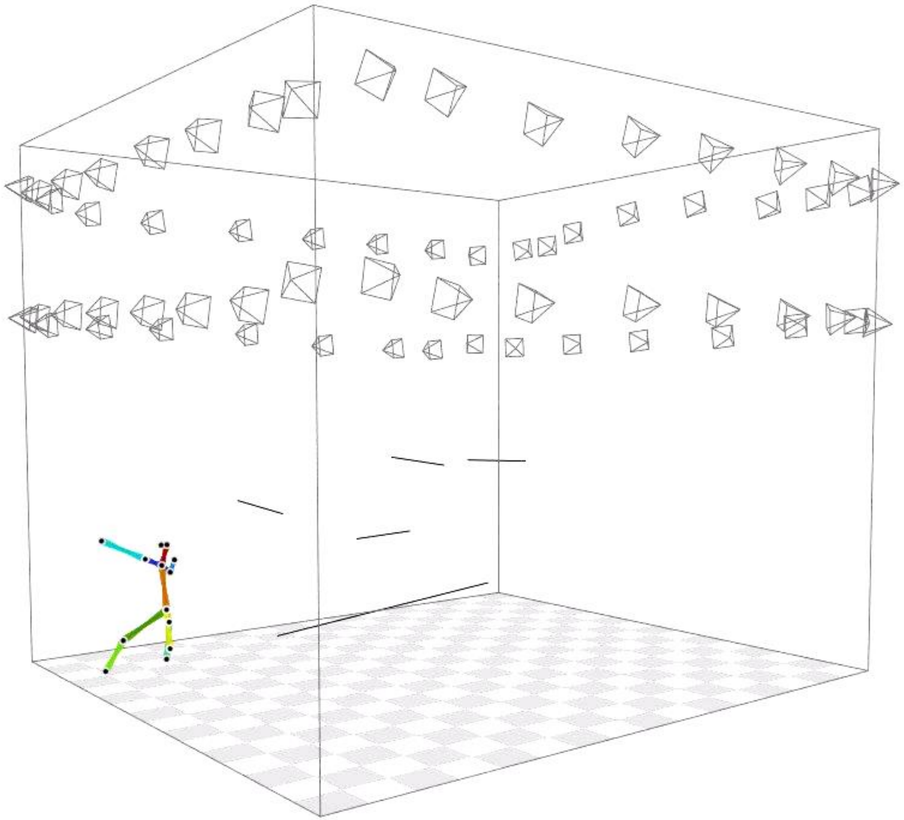


Pants

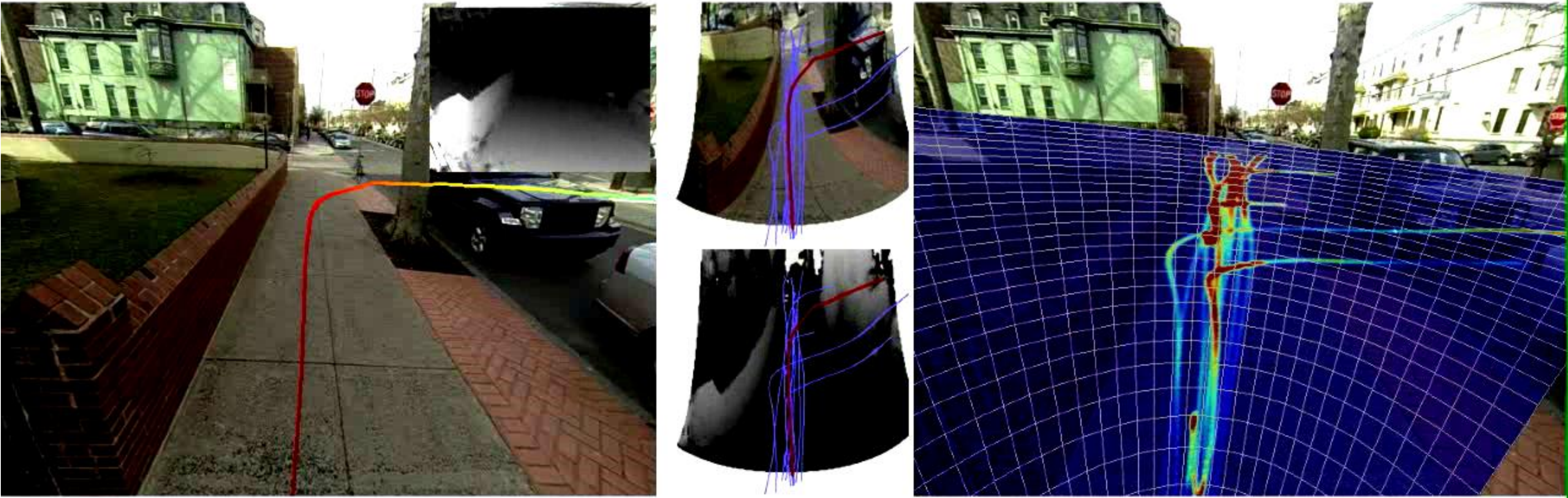
NEUROSCIENCE



NEUROSCIENCE



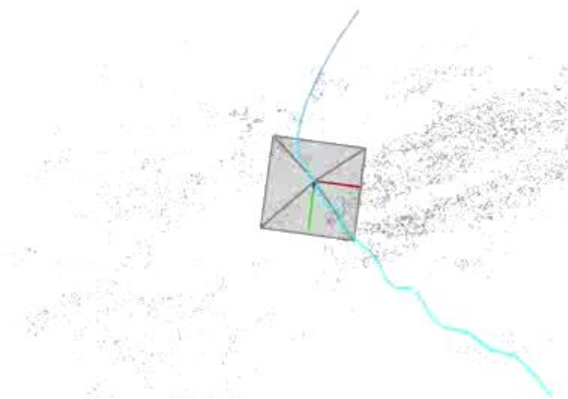
3D reconstruction of monkey movement



Ground truth EgoRetinal map Predicted trajectories

SENSORIMOTOR SKILL LEARNING

https://www.youtube.com/watch?v=aVJ45wIUE88



3D reconstruction