

# SE 461 Computer Vision

Nazar Khan

PUCIT

Lecture 1: Course Introduction and  
Basics

# Disclaimer!

- Many of the lecture slides during this course will heavily borrow material from people I learned from, especially
  - Dr. Sohaib Khan
  - Dr. Joachim Weickert

# Lecture Etiquette

## On Questions

- There is no such thing as a 'stupid question'.
- Your questions will help your class-mates.
- Your questions will make sure I do not go too fast.
- Your questions provide feedback for me.

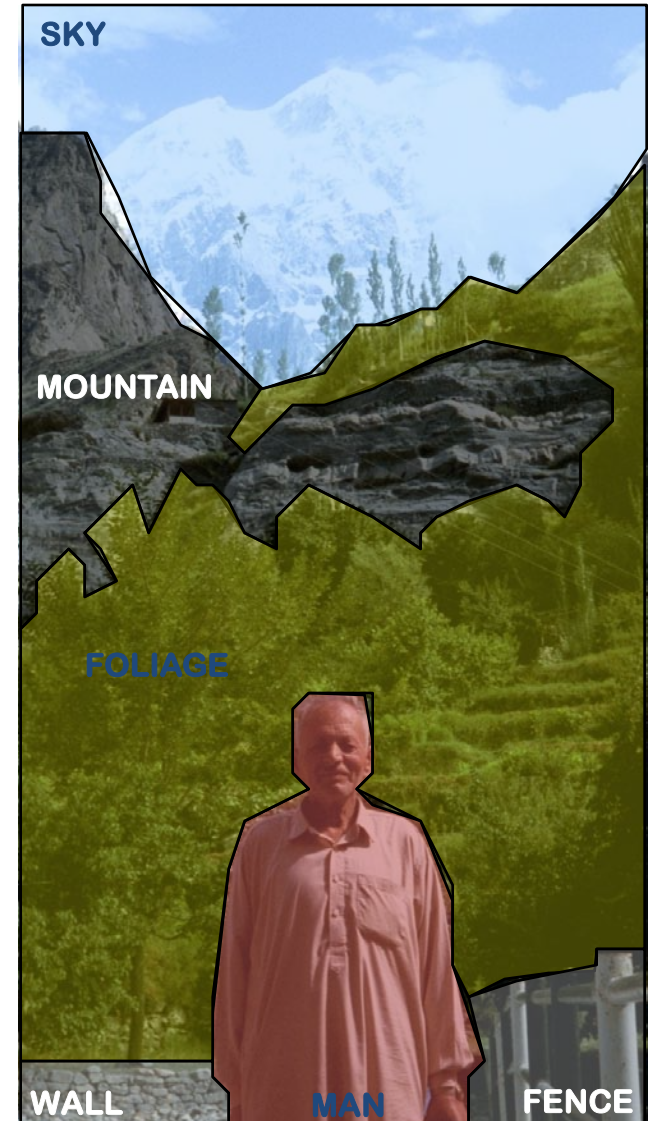
# Introduction

- Sight is our primary sensation
  - 80% of our first 12 years of learning is through vision
  - 40% of the brain is dedicated to visual processing
- Human (and Animal) Experience



# What does it mean to see?

- “To know what is where by looking”
  - Aristotle
- **Computer Vision Scientist:** “How do we build a machine to do that?”
- Scientifically unsatisfactory
  - What is “what”?
  - What is “where”?
- Ongoing investigation
  - Identification
  - Localization



What do you see?



What colors do you see?





What letters do you see?



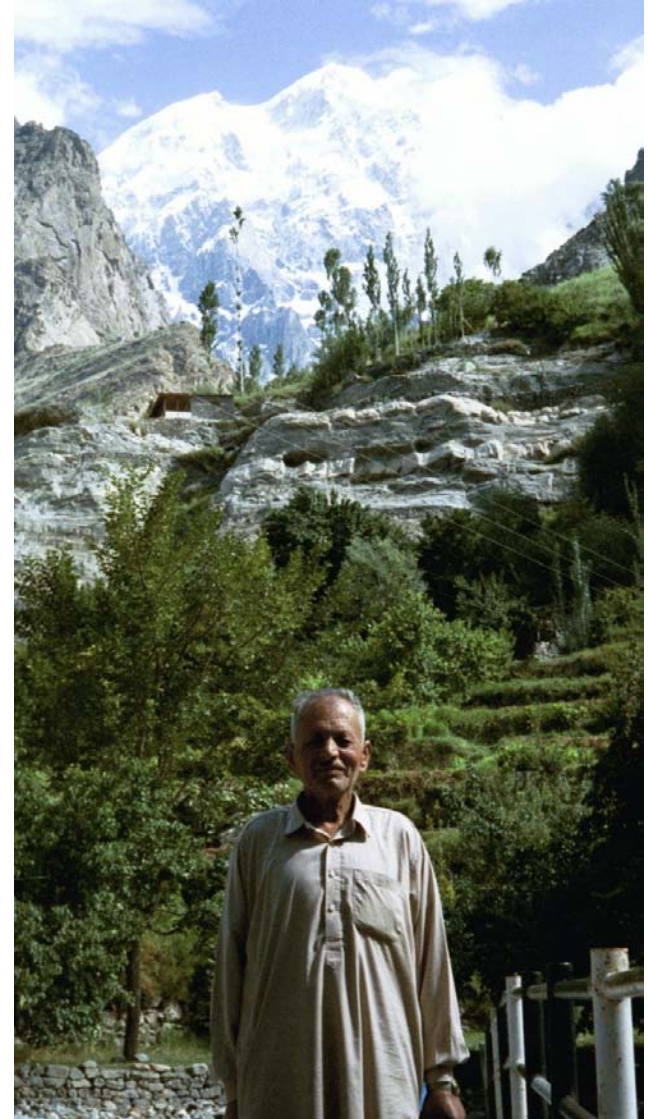


Do you see a grid of tiles?

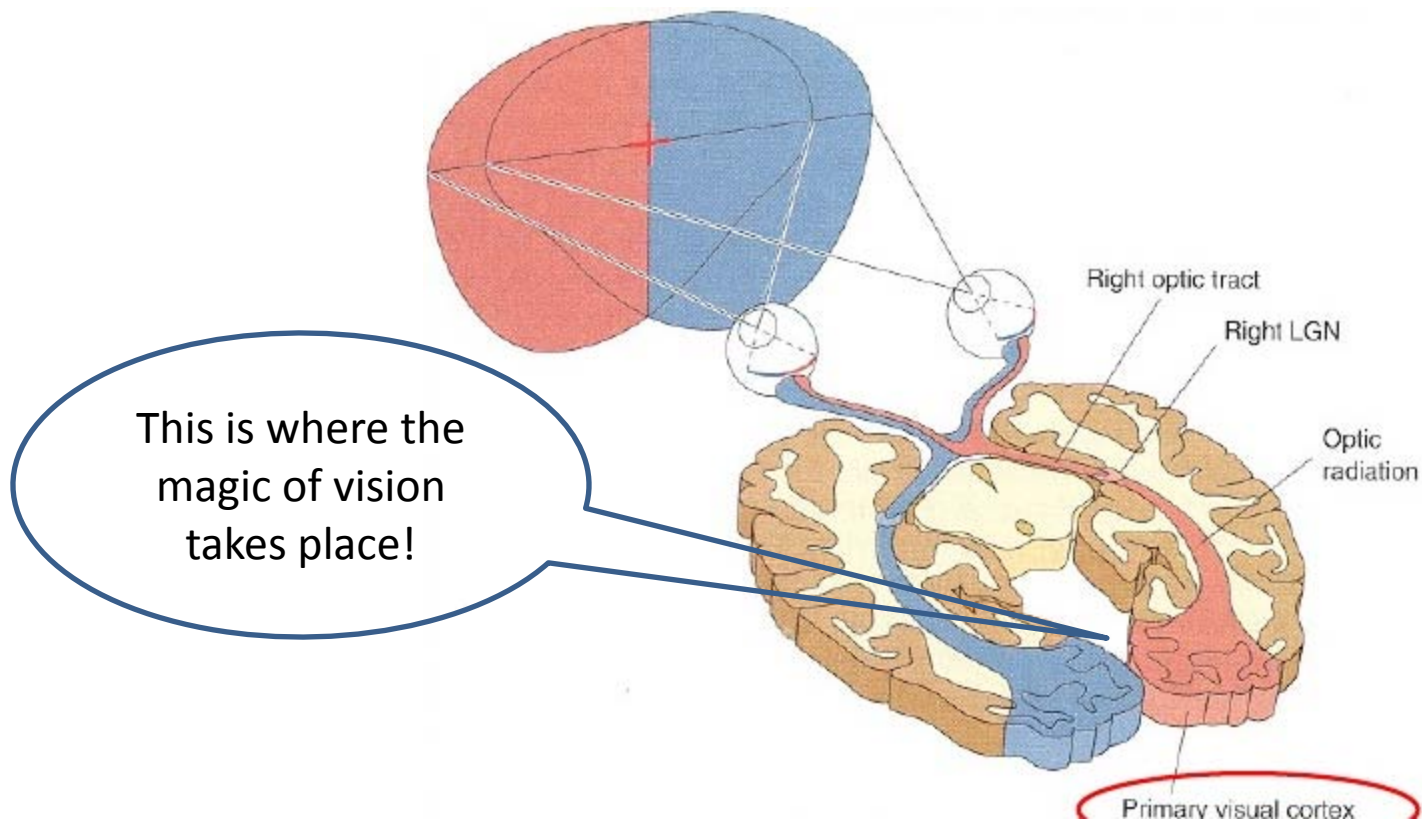


# Biological Vision

- Biological Vision is a very sparse process.
- Everything entering your eye  $\neq$  What you see.
- Our brains choose only the most crucial information
  - and *that* is what we “see”



# Where in the brain do we see?

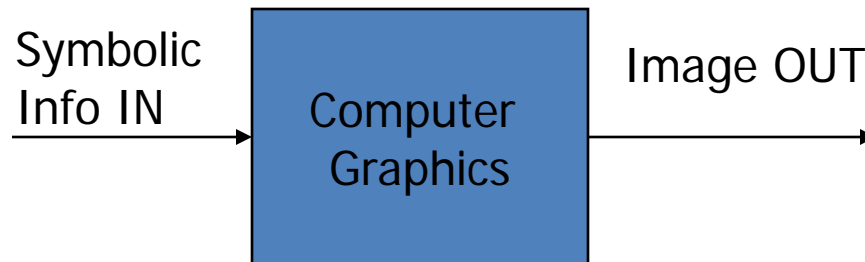
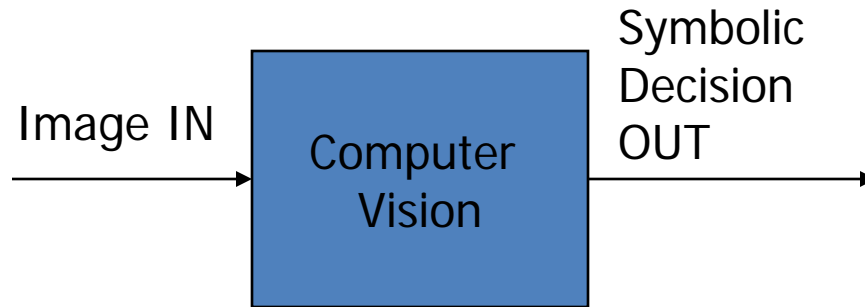


The visual pathway (<http://www.slideshare.net/Codeye/visual-pathway>)

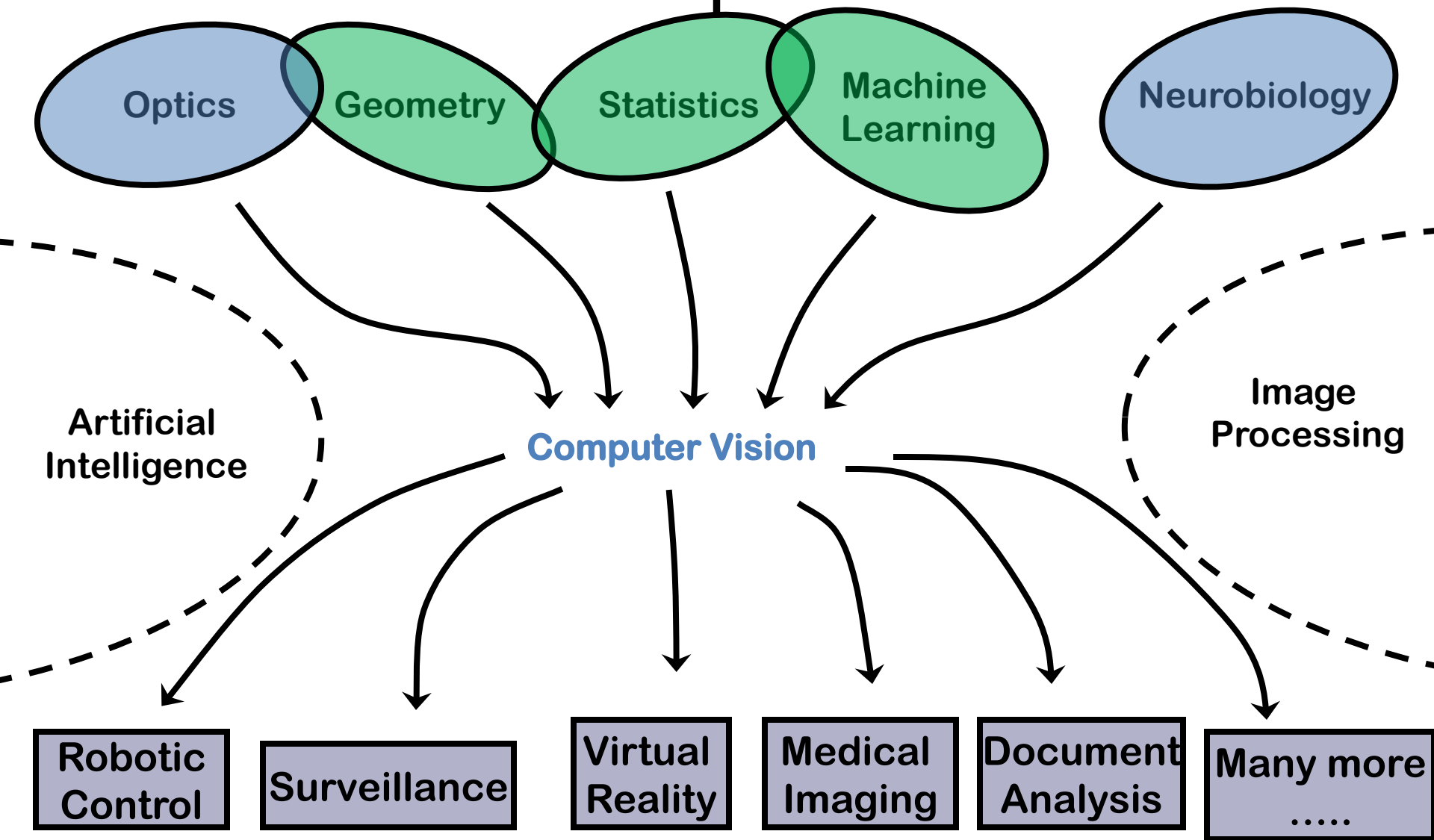
# What is Computer Vision?

- “The goal of Computer Vision is to make useful decisions about real physical objects and scenes from their images”
- “What’s in the image?”
- “What’s happening in the video?”

# CV vs IP vs CG



# What is computer vision?





# Why is computer vision hard?

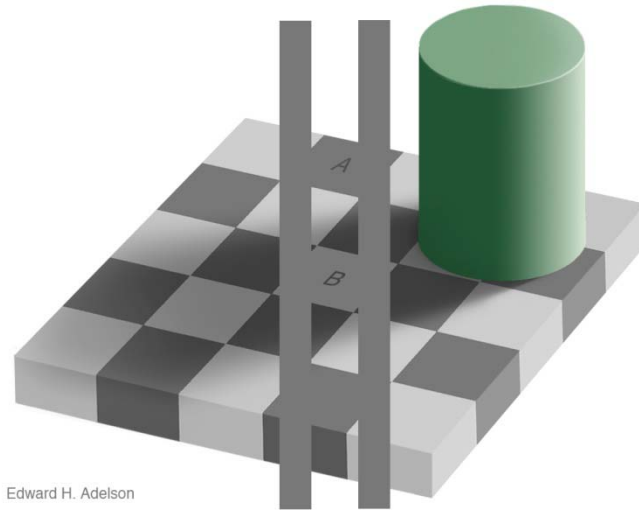
- Computers are good at numerical processing
- Humans are good at perceptual processing
- We want to use a computer to mimic human perception... which is complex to understand

# The Complexity of Perception

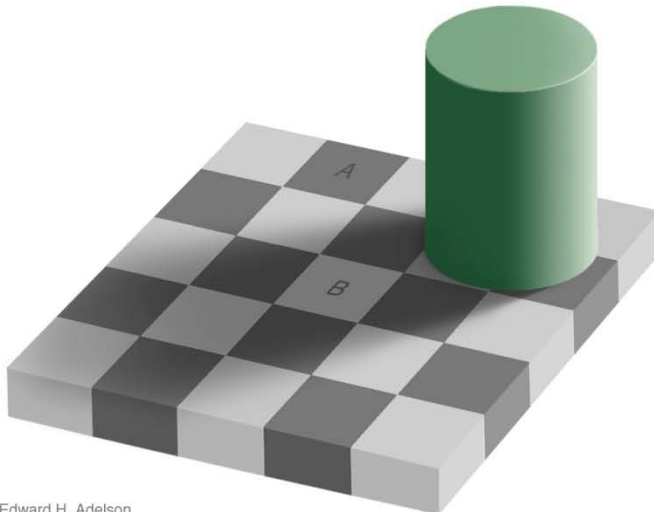
# The Complexity of Perception



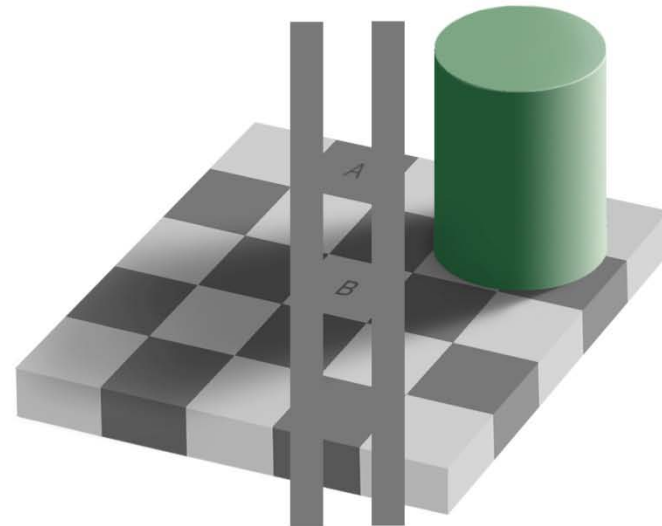
# The Complexity of Perception



Edward H. Adelson

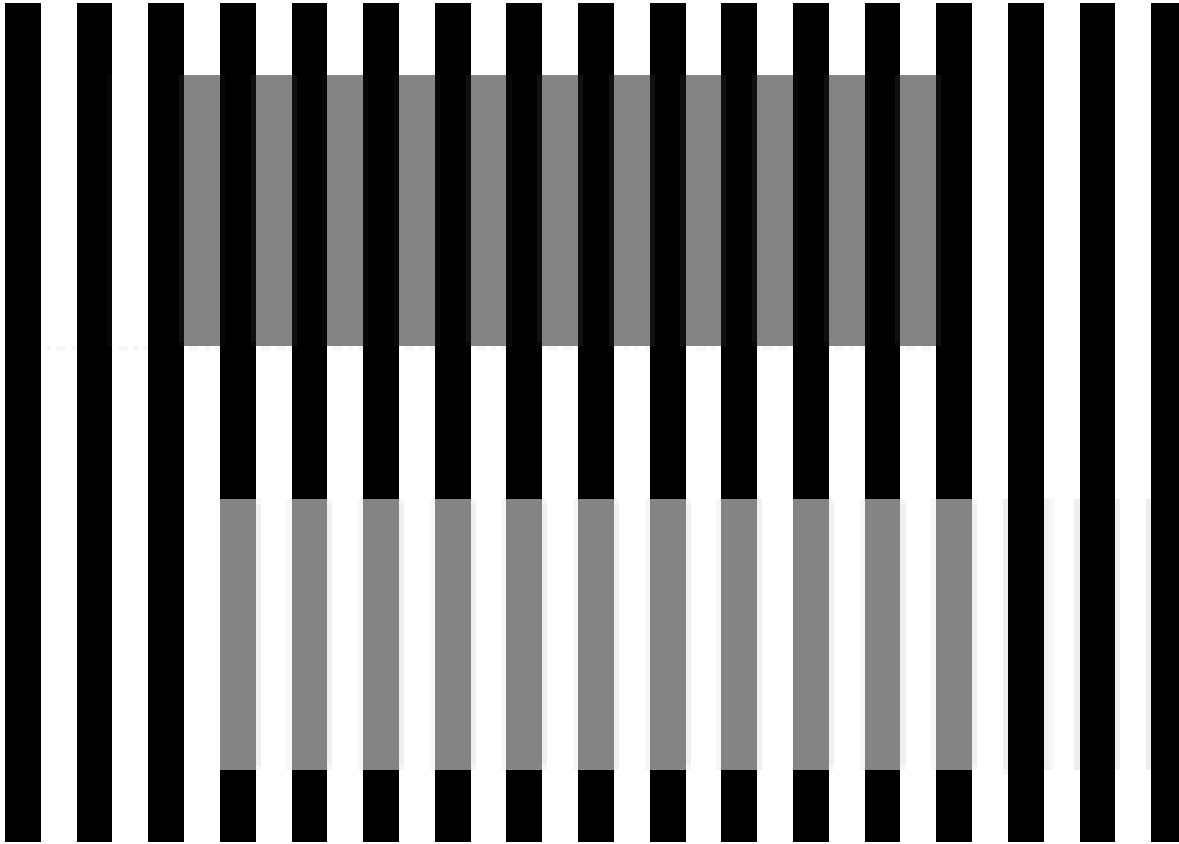


Edward H. Adelson

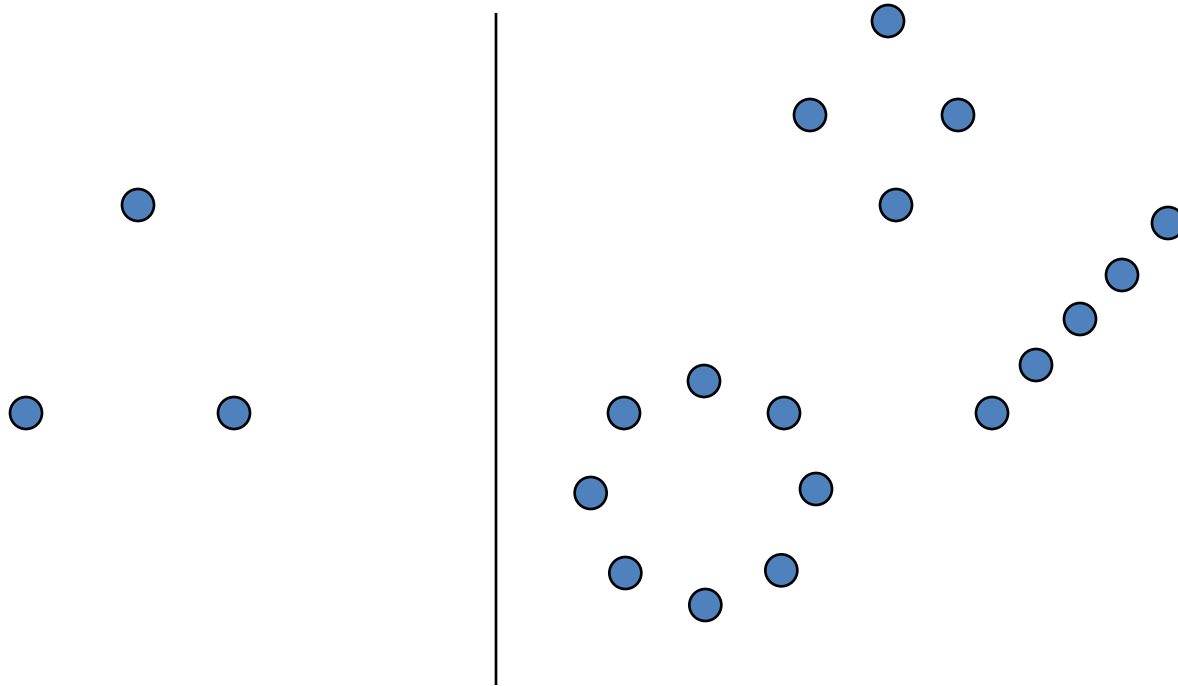


By Edward Adelson, 1995

# Perception



# Perception



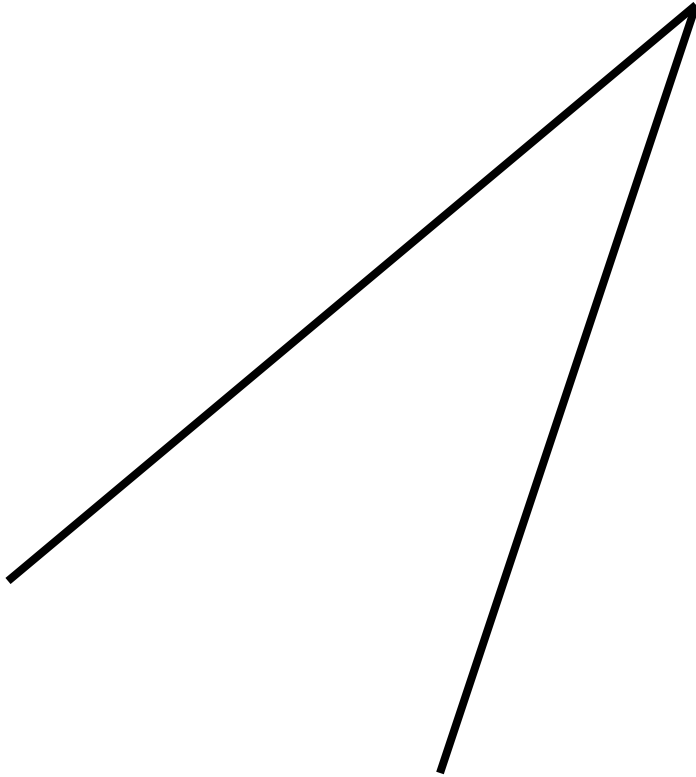


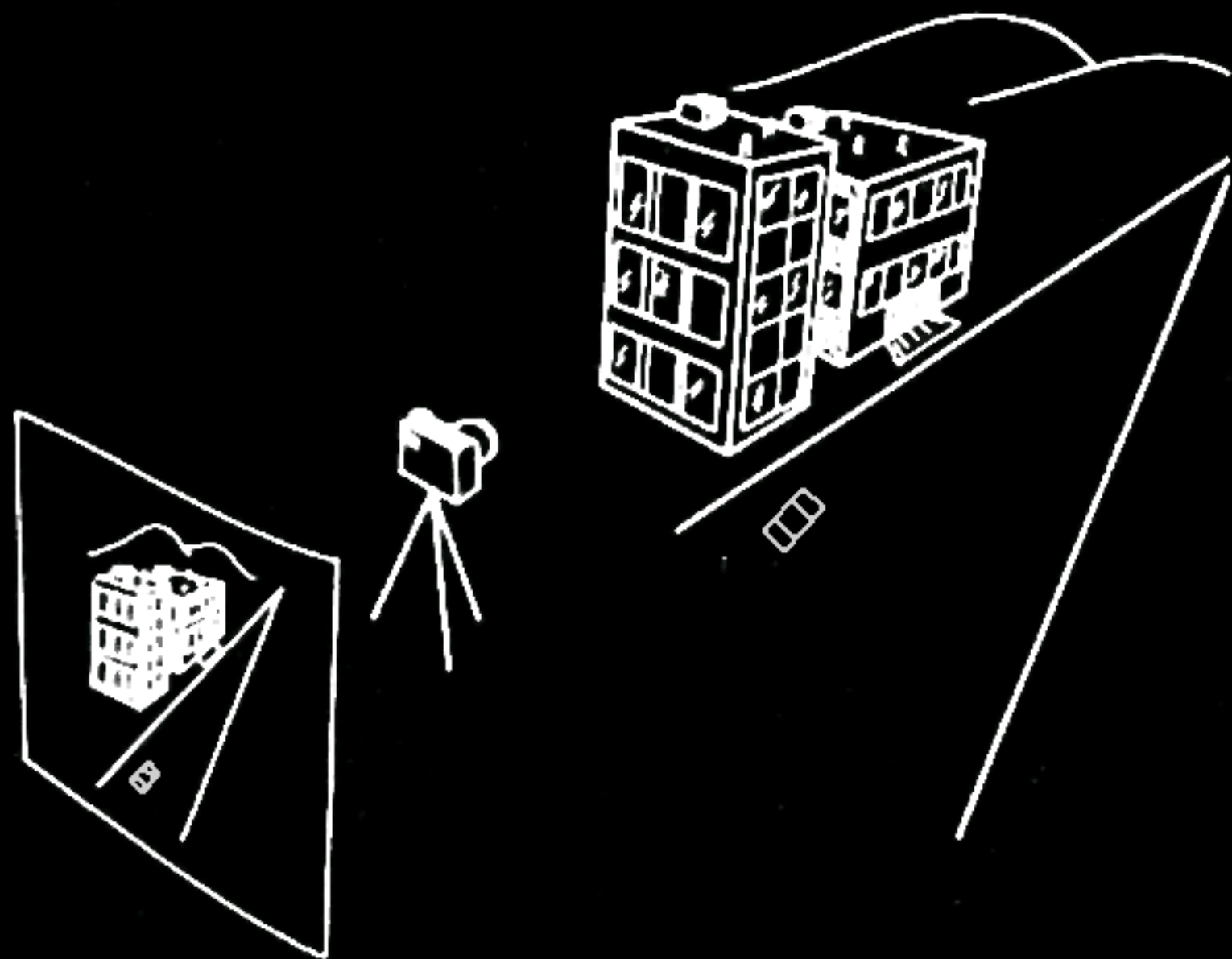
# Perception



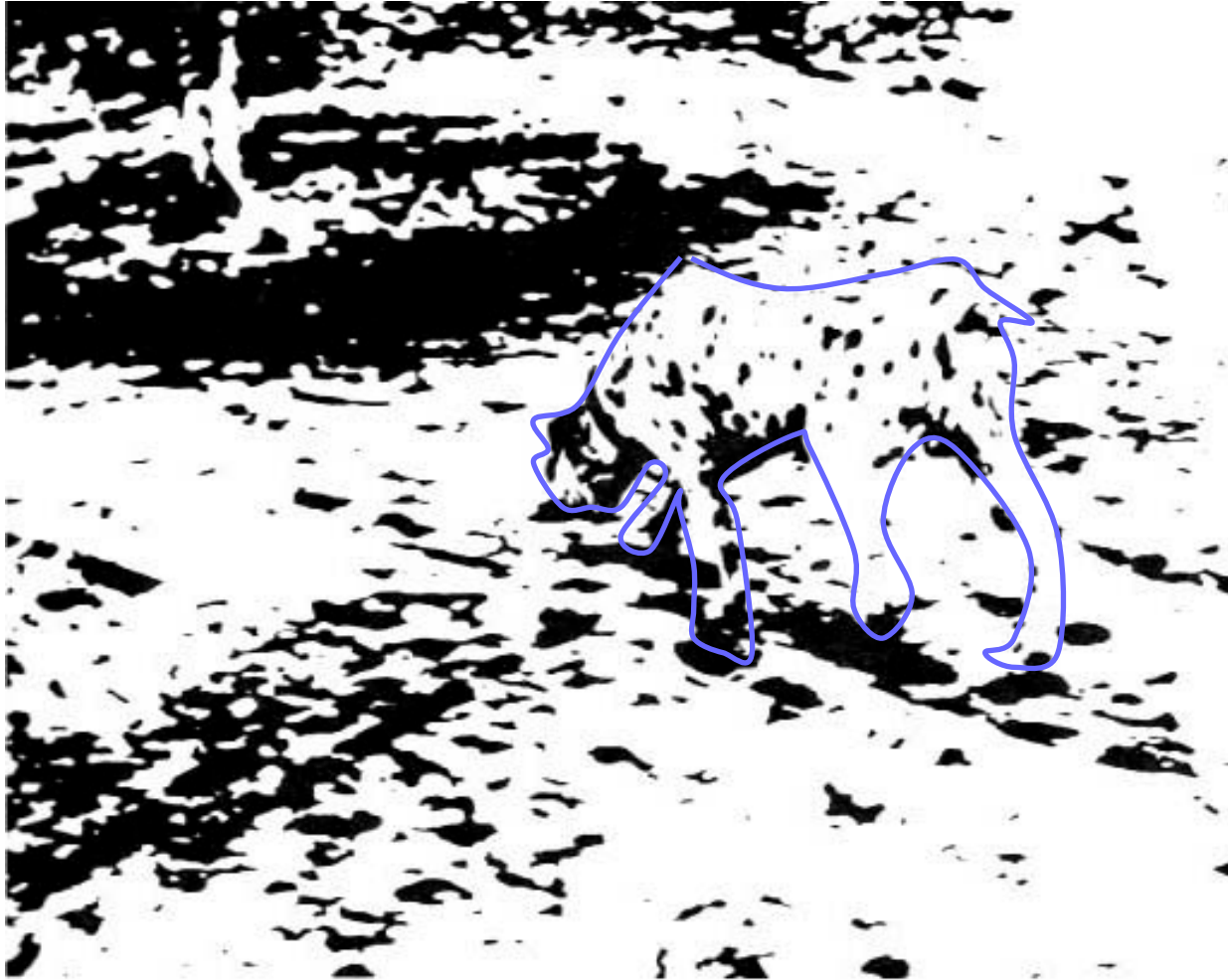
Ref: Light and Vision: LIFE Science Library

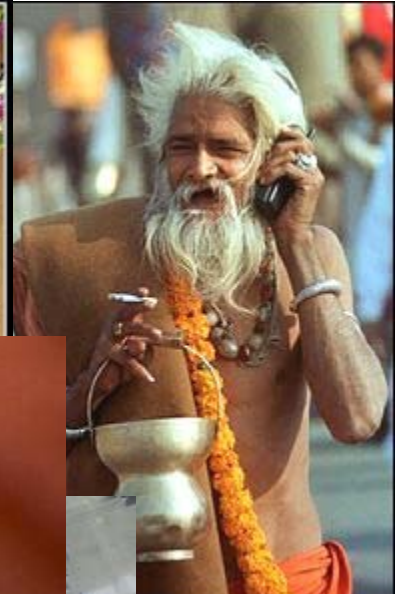
What is this?





# The Complexity of Perception





It seems like a hopeless task to be able to write a program to interpret these images



# Writing Programs that “See”

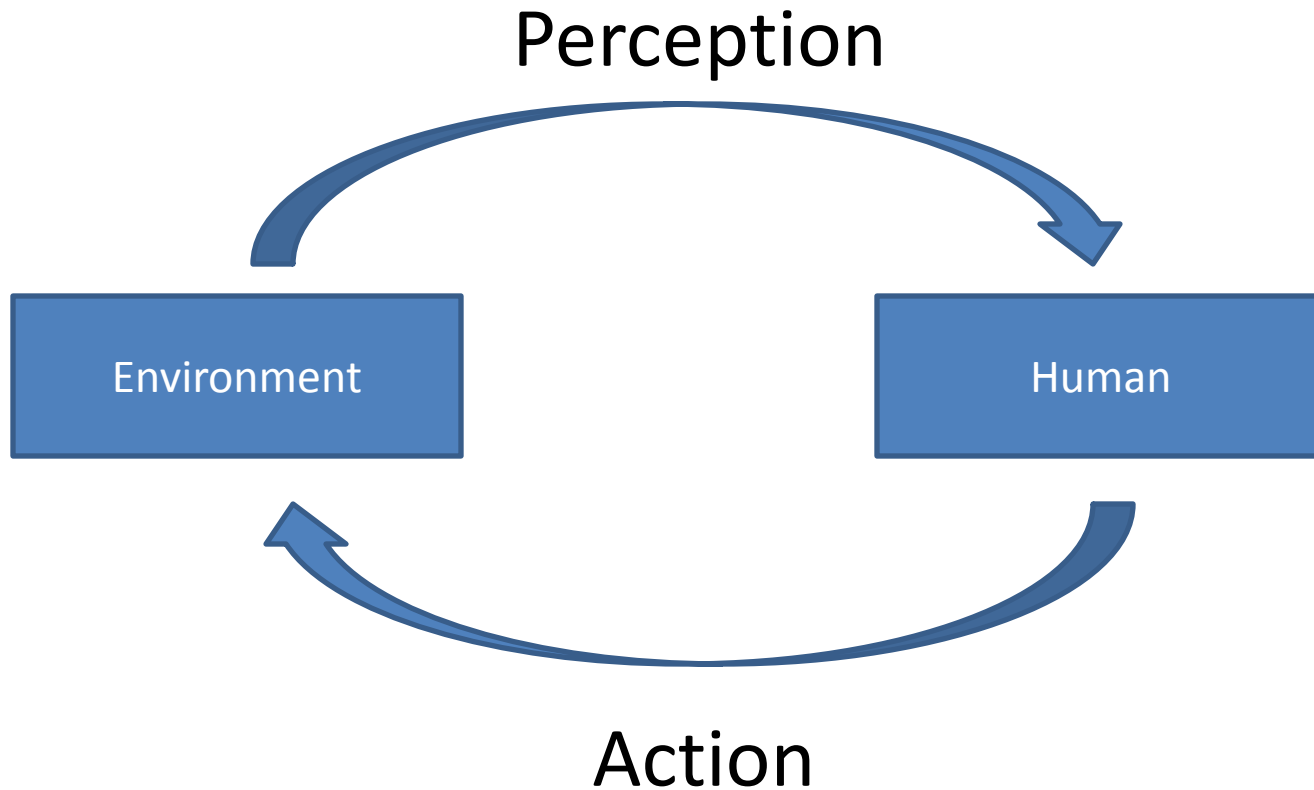
An Example

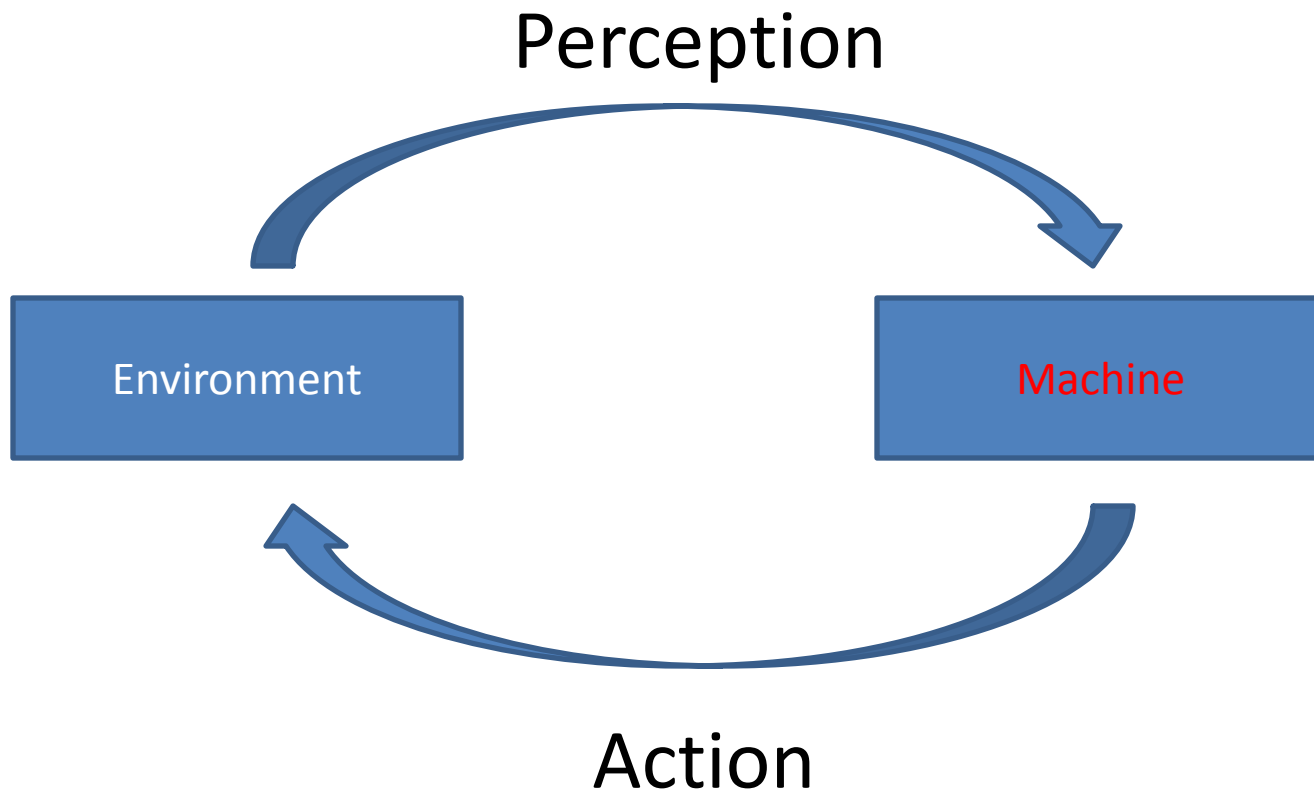


# Motivation

- Humans have highly sophisticated capabilities of sensing the environment, interpreting it and taking actions accordingly.
- Human Perception: *“the process of attaining awareness or understanding of sensory information”*
- Can machines have similar capabilities?
  - Can we write algorithms for perception?

# The Perception-Action Cycle





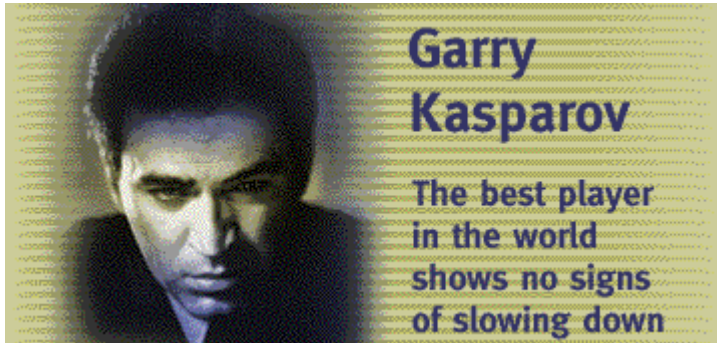
# Motivation

Machines *Compute*

Humans *Perceive*

Can *perception* arise from *computation*?

# Gary Kasparov vs. Deep Blue (1997)



- Can evaluate about 3 positions per second
- Large chess 'knowledge', low computation ability
- Uses a lot of intuition and feeling
- Learns and adapts very quickly from his mistakes
- Can get bored, fatigued, loss of concentration
- Highly intelligent
- Selective searching of positions



- Can evaluate about 200,000,000 positions per second
- Small chess knowledge, high computation ability
- Uses only computations
- Not a learning system (machine learning algorithms were not used)
- Not affected by feelings such as Kasparov's stare
- Dumber than a 2 year old
- Brute-force evaluation of all moves

<http://www.research.ibm.com/deepblue/>

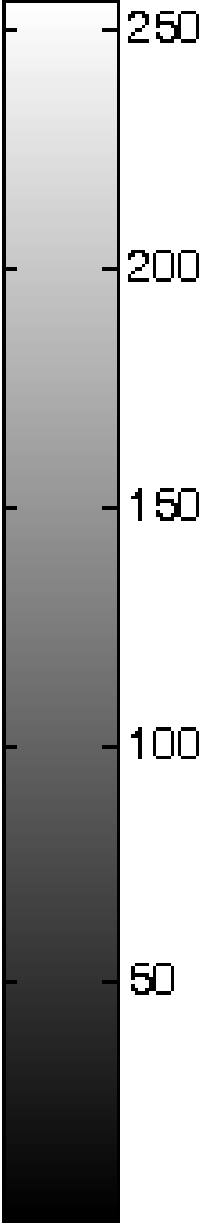
# Why is Perception Difficult for Computational Machines?

- Example Problem:
  - Given an image, search through all possible sub-windows and identify those which contain a human face



# Face Detection Problem





62	87	119	142	156	164	167	168	176	182	188	188	187	189	194	199	195	194	192	195	195	193	193	193	197	193	194	198	202	199	195	192	182	191	186	143	75	24	21	42	44
92	114	138	152	159	166	173	176	178	180	184	187	189	190	191	193	194	192	194	197	197	194	194	195	192	201	203	197	190	192	200	208	202	184	187	192	144	60	17	25	37
113	129	147	155	159	167	175	180	181	181	183	189	195	195	195	193	197	196	199	202	203	202	201	203	200	207	210	208	205	207	205	198	201	192	193	199	180	118	52	11	31
123	133	145	153	157	163	173	181	183	181	185	191	194	191	191	194	194	193	198	203	203	202	203	207	211	208	206	200	186	172	172	180	205	202	198	192	188	171	110	37	23
125	133	141	147	152	159	170	178	178	182	190	192	191	191	195	187	197	204	204	202	206	208	207	210	189	160	132	106	90	91	99	137	157	180	191	195	190	152	100	29	
125	131	137	142	147	156	167	175	175	176	181	188	191	190	190	191	189	200	202	196	200	209	207	193	148	118	83	60	54	62	83	102	98	107	125	147	175	199	188	156	61
123	128	135	142	149	158	168	174	185	189	191	189	188	187	193	195	193	191	194	187	160	131	83	67	57	67	89	113	136	153	139	130	127	137	159	186	187	168	126		
119	125	134	146	156	165	172	175	187	193	195	192	192	190	190	188	187	189	133	132	61	68	68	81	130	133	132	138	146	153	158	173	184	181	167						
115	121	134	147	161	170	174	175	189	191	193	195	196	195	195	197	199	121	93	84	71	86	86	122	131	131	111	110	118	137	153	174	190	192	178						
115	120	130	145	160	171	174	174	185	187	191	191	190	191	191	166	198	111	97	99	101	96	98	110	113	131	128	121	117	98	100	121	146	173	191	193	196				
114	118	127	142	158	169	172	172	181	177	181	183	180	174	175	164	121	100	79	89	99	124	87	90	117	138	164	193	207	200											
110	115	125	144	163	174	175	171	162	164	158	153	159	163	166	175	177	174	171	164	149	126	102	89	82	81	93	76	52	65	100	145	100	63	82	138	167	182	198	201	212
116	124	136	150	156	149	134	123	106	104	96	98	113	129	149	172	184	187	188	177	154	125	98	82	78	69	50	38	53	64	91	161	143	114	130	173	190	201	211	210	209
115	121	125	119	100	80	64	57	53	54	56	69	88	103	127	158	184	196	202	190	157	122	95	81	68	70	65	42	32	35	69	133	152	138	157	189	198	206	214	212	209
121	116	100	73	45	31	32	39	39	43	52	67	79	85	108	144	176	195	207	196	161	126	107	100	60	72	113	97	48	65	106	118	125	130	157	183	193	204	212	210	214
102	83	57	37	32	37	49	56	58	59	59	62	64	62	86	123	165	190	201	197	167	139	127	125	91	86	115	108	81	104	120	104	119	137	167	188	200	210	215	215	215
65	47	32	38	60	80	85	81	85	85	80	75	68	56	81	141	172	181	161	143	141	141	120	93	102	102	92	113	145	166	188	200	209	213	215	214	210				
52	51	57	73	93	106	108	104	101	99	91	86	84	72	62	61	117	153	181	194	179	161	156	152	140	140	135	130	171	176	190	200	206	211	212	210	212	210			
45	63	82	90	88	87	96	105	101	93	75	70	81	85	75	104	145	181	201	181	171	171	171	188	182	173	181	183	188	193	203	205	208	214	214	214	217	214			
58	83	82	66	67	65	56	54	76	67	96	88	54	75	102	138	181	207	181	186	186	186	186	181	189	192	198	198	206	208	211	215	215	212	210	210	213				
74	71	63	59	60	47	31	27	43	66	121	126	94	95	99	88	99	135	178	201	204	201	197	195	194	189	190	197	204	206	203	202	207	208	209	210	210	210	209	207	
79	61	48	45	50	57	49	31	25	59	114	127	118	123	120	117	105	141	182	202	206	207	205	200	191	185	183	189	191	189	194	201	205	208	208	205	204	206	207	205	202
71	61	48	33	38	75	87	59	67	84	107	110	119	131	119	116	111	145	181	197	204	204	195	182	184	195	202	194	171	155	161	175	187	200	204	200	195	198	199	196	194
68	63	55	39	38	71	90	72	90	101	113	115	134	145	121	112	110	131	153	166	175	179	168	152	156	177	190	182	166	154	144	133	153	171	187	188	185	185	186	185	179
71	65	66	66	63	70	80	85	90	106	124	129	145	155	131	126	112	111	113	120	134	144	141	132	91	105	134	160	150	122	126	140	158	169	172	170	171	172	166		
81	81	82	92	98	94	96	110	123	131	149	148	151	157	131	121	109	98	93	101	111	114	111	60	50	45	118	160	165	148	125	125	135	151	160	158	156	158	157		
91	103	99	99	115	122	119	125	136	139	161	167	173	179	151	126	91	63	81	89	89	89	91	118	147	169	175	172	138	126	123	141	154	151	146	146	146				
102	105	107	108	119	133	145	151	160	165	168	172	179	173	141	105	32	21	51	73	83	96	108	123	149	171	182	192	189	168	130	103	122	149	147	141	150	146			
103	113	121	125	129	140	153	161	170	173	175	177	176	160	121	91	61	61	61	61	61	61	61	98	163	181	178	182	187	187	160	123	107	122	145	147	136	154			
100	116	133	138	139	146	158	168	176	175	176	177	169	144	107	80	58	47	41	48	52	54	65	81	98	127	155	159	149	147	153	160	161	158	139	109	102	130	152	146	160
93	113	131	137	138	144	155	164	171	169	170	171	160	133	103	83	67	56	54	58	62	77	99	110	128	132	133	110	90	87	87	95	106	123	140	133	110	111	137	153	158
90	105	120	128	133	141	150	155	163	161	161	162	153	131	108	94	87	82	80	81	86	102	113	109	109	87	86	87	92	93	78	84	64	68	84	97	100	108	129	149	160
86	98	111	118	126	135	143	147	154	154	154	152	146	132	113	100	102	105	104	99	95	90	82	77	70	98	145	153	171	200	188	183	132	98	49	36	73	115	137	150	164
77	88	102	111	118	125	133	136	141	145	145	141	138	131	114	97	108	112	105	92	79	62	70	103	124	171	234	234	235	246	238	238	223	179	93	51	85	119	130	149	161
65	80	96	104	109	116	123	127	127	135	137	133	132	130	112	92	106	106	89	74	62	52	98	180	235	213	242	248	243	213	176	201	185	170	116	92	117	118	116	152	154
48	72	89	89	92	106	113	112	120	120	123	129	126	113	100	92	94	83	70	60	67	112	179	223	246	237	238	233	168	84	65	99	87	83	103	141	153	137	130	141	147
54	64	76	83	87	91	98	104	109	109	113	118	121	113	95	79	85	78	63	66	114	177	211	211	233	171	106	81	78	80	102	130	138	145	156	164	162	153	142	135	144
66	59	64	77	81	78	83	95	96	99	101	106	113	113	93	70	65	50	33	69	151	189	150	101	80	67	61	80	119	158	180	186	182	171	158	157	169	176	162	141	145
72	65	64	73	78	76	80	89	87	93	95	97	105	111	95	73	52	30	25	81	147	131	69	42	75	115	160	183	191	193	188	180	164	144	134	151	174	179	166	154	149
66	68	69	69	73	80	85	86	81	90	94	93	99	106	100	87	46	37	38	74	99	75	68	107	147	171	197	200	194	182	170	162	122	132	15						



# Why is pattern recognition hard?



- To write an algorithm, you may need to precisely define a human face
- Your description should be invariant to pose, 3D rotation, occlusion, race, gender
- The description should hold for all faces and nothing which is not a face should match the description

# IMPOSSIBLE?

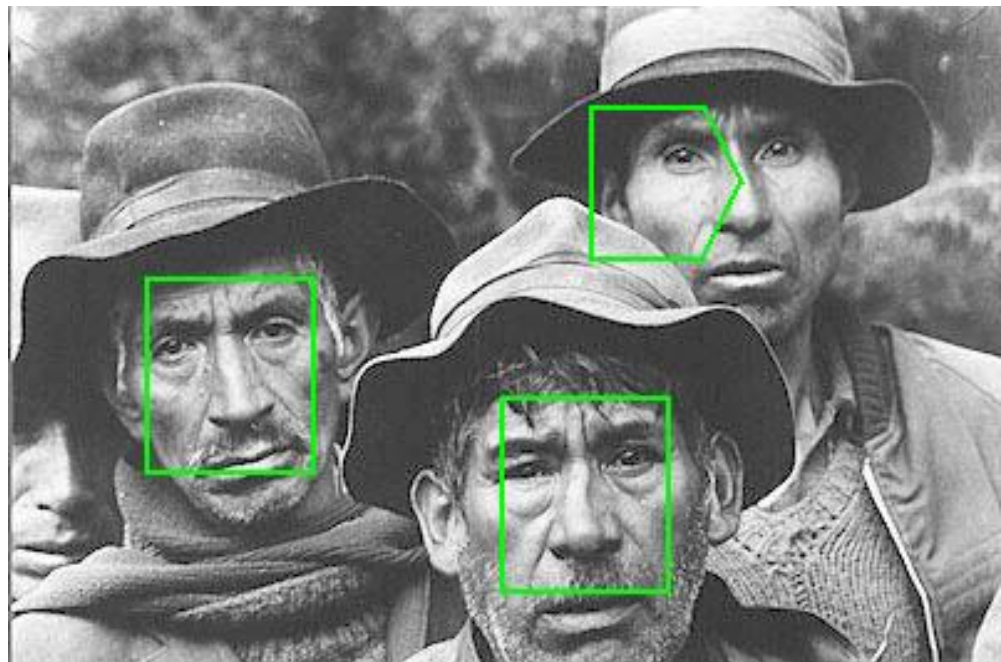
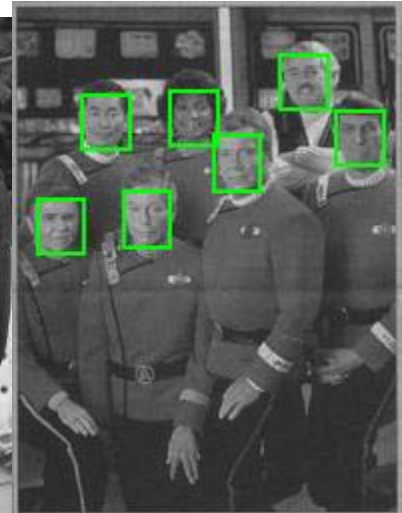
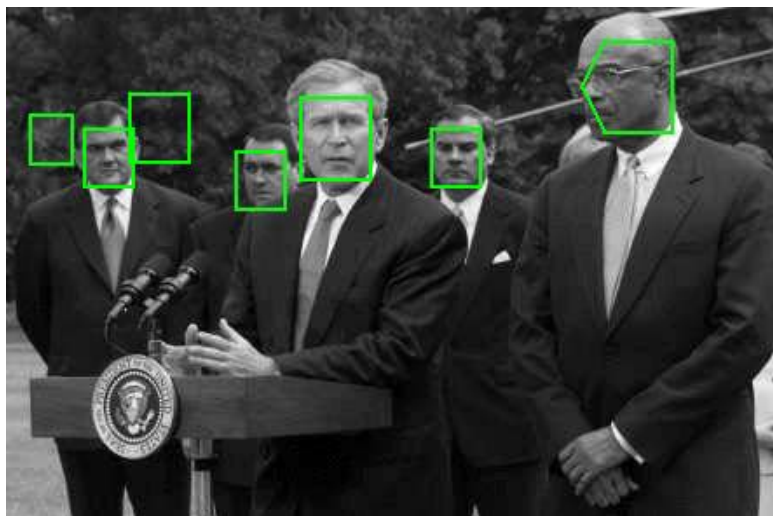
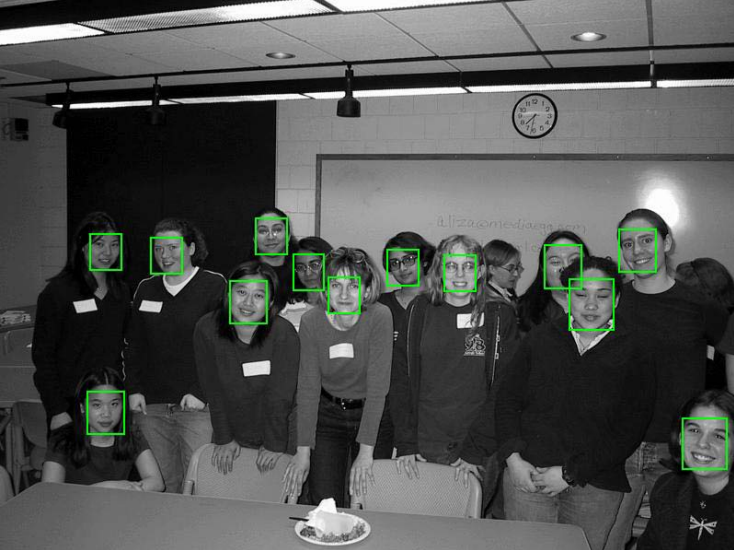
- “If it were not for the human existence proof, we would have given up a long time ago”
  - We do it, so there MUST be some method behind it.
- Computer Vision usually does not follow Biological Vision
  - It’s just too complex!!!
  - Our brains are massively parallel processors

# State of the Art in Face Detection



Viola/Jones Face Detector (2001): Using implementation in OpenCV





Results of Schneiderman/Kanade Face Detector



# Face Detection Demo



# Solving Sub-problems

- The study of Computer Vision started with the effort of trying to make complete visual understanding systems
  - Initially, it was considered an easy problem
  - Marvin Minsky gave the visual perception problem as a summer project to a sophomore
- Complete Scene Understanding Systems
  - [VISIONS](#), Hansen and Riseman, 1978
  - [ACRONYM](#), Brooks and Binford, 1979
- Did not succeed
  - Too complex a problem
  - Lack of computational power
  - Lack of data
- We have been trying to solve sub-problems
  - The researchers will return to the complete problem at a later date

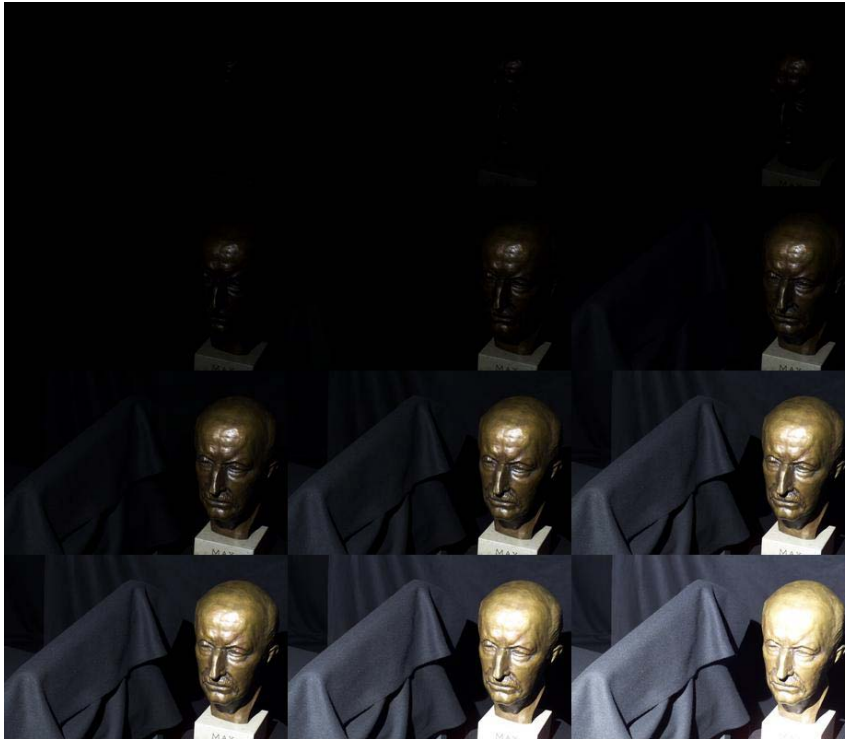
# Sample Problems from Image Processing and Computer Vision

# Computational Photography



[http://en.wikipedia.org/wiki/High-dynamic-range\\_imaging](http://en.wikipedia.org/wiki/High-dynamic-range_imaging)

# Computational Photography

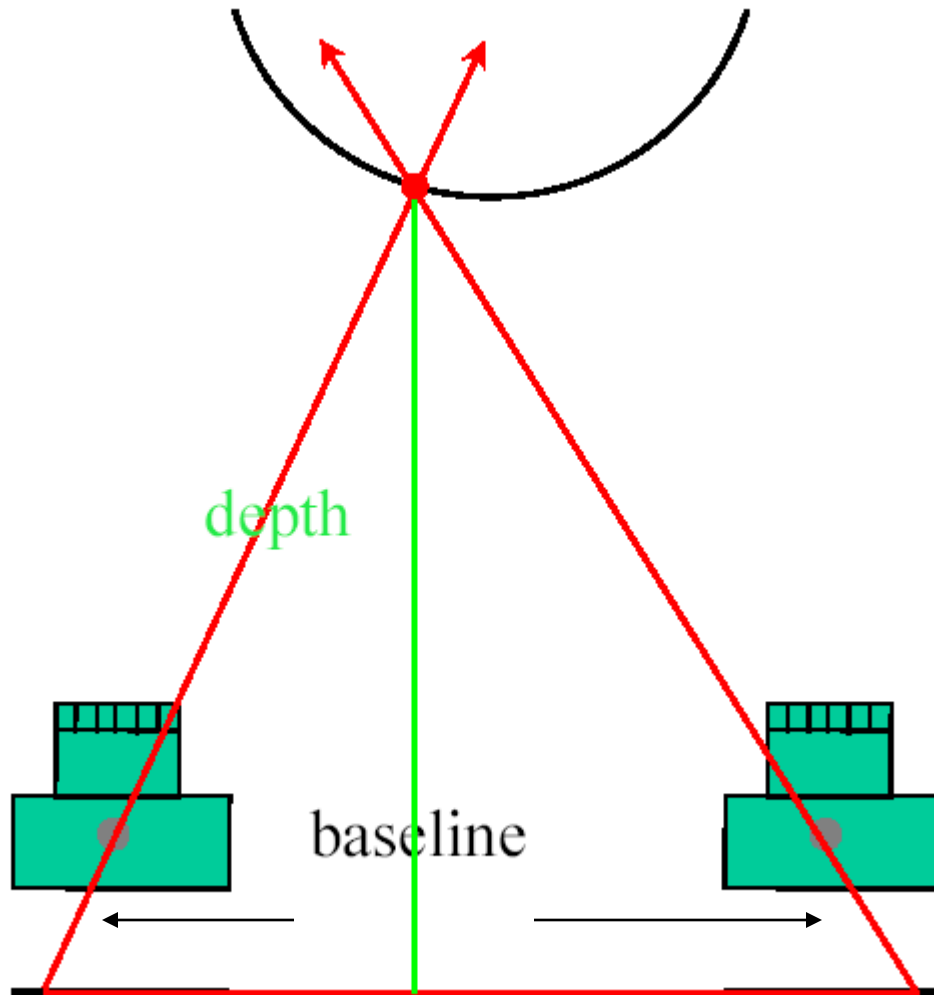


Author: Nazar Khan (2004)

# Reconstruction of 3D Structure

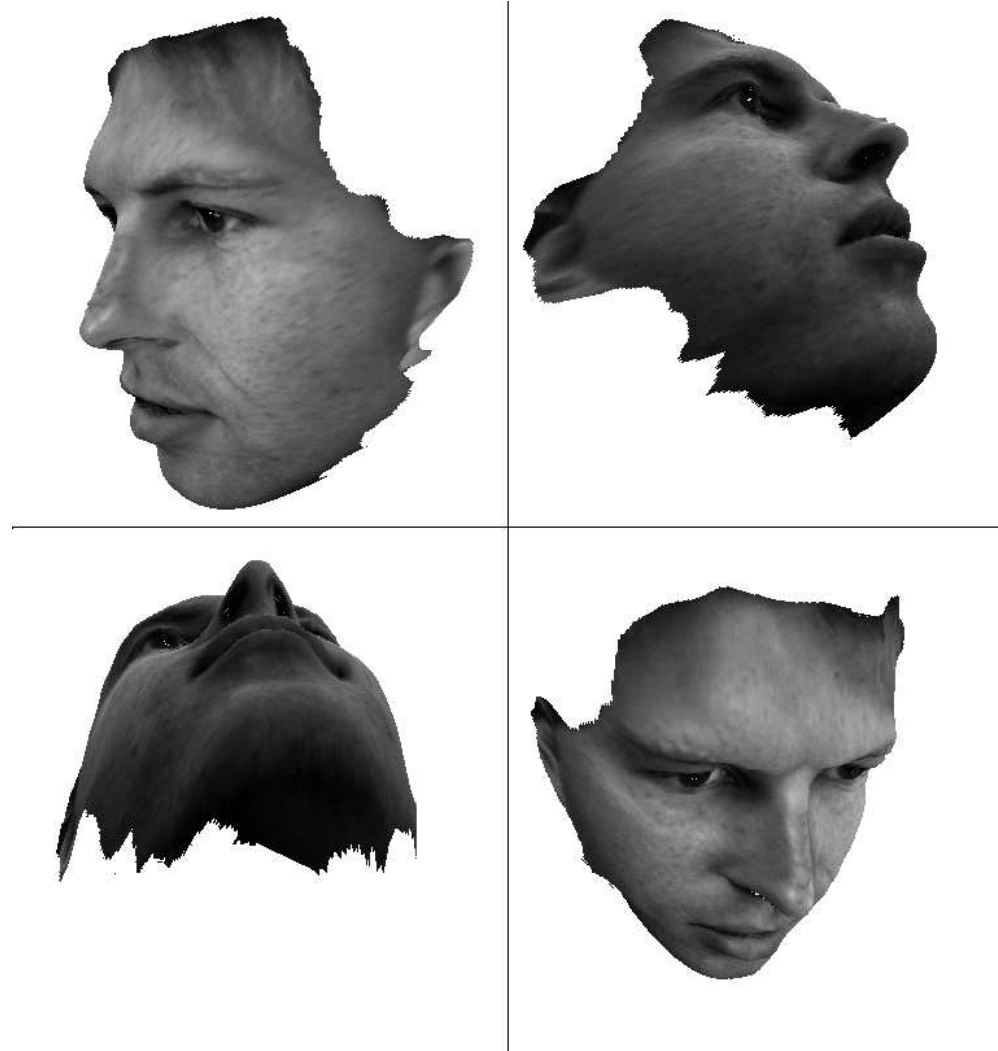
- An Image is a 2-Dimensional projection of 3D World
- 3D can be reconstructed from
  - Two images
    - Stereo Problem
  - Video with moving camera
    - Structure from Motion Problem
  - Some understanding about what is being viewed
    - Geometrical inference
    - Shape from shading or texture

# Stereo





Stereo image pair.



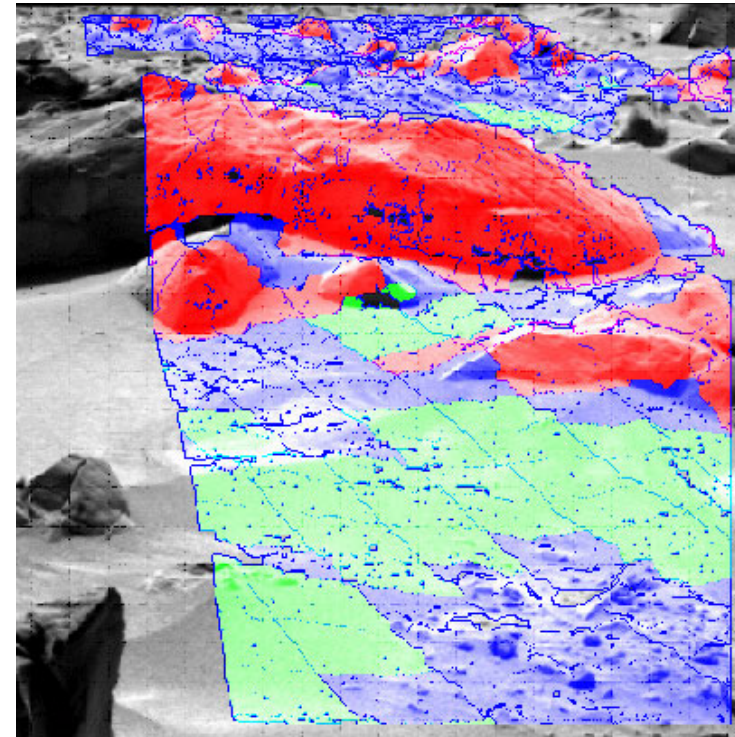
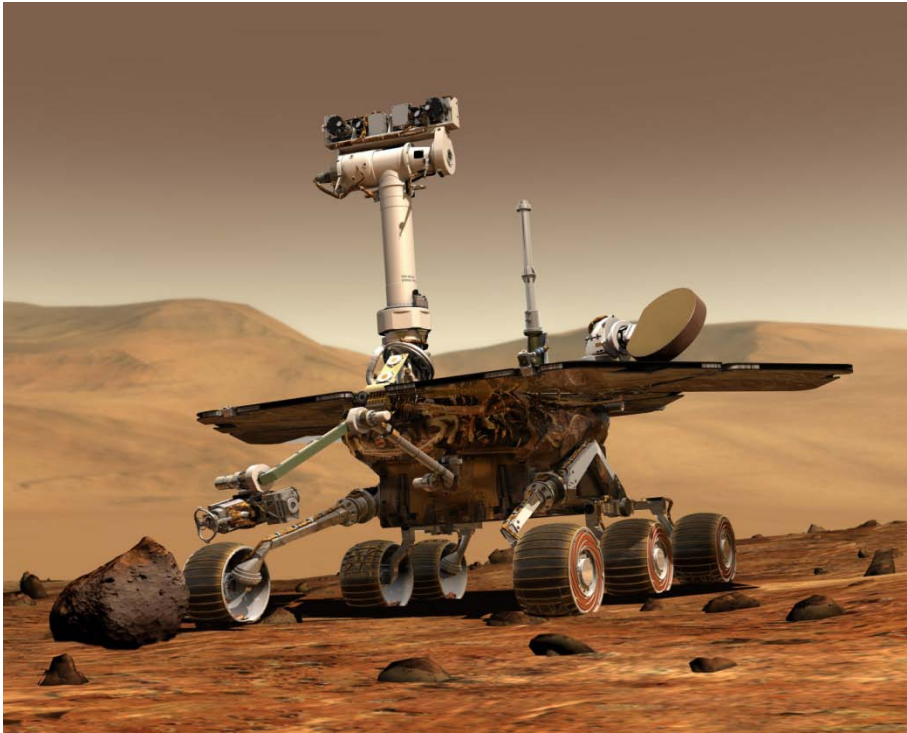
3-D reconstructions

L. Alvarez, R. Deriche, J. S´anchez, J. Weickert (2002).



# Stereo in Space Exploration

- Mar Exploration Rover



Source: <http://www-robotics.jpl.nasa.gov>

# Stereo in Book Scanning

# Pose Estimation

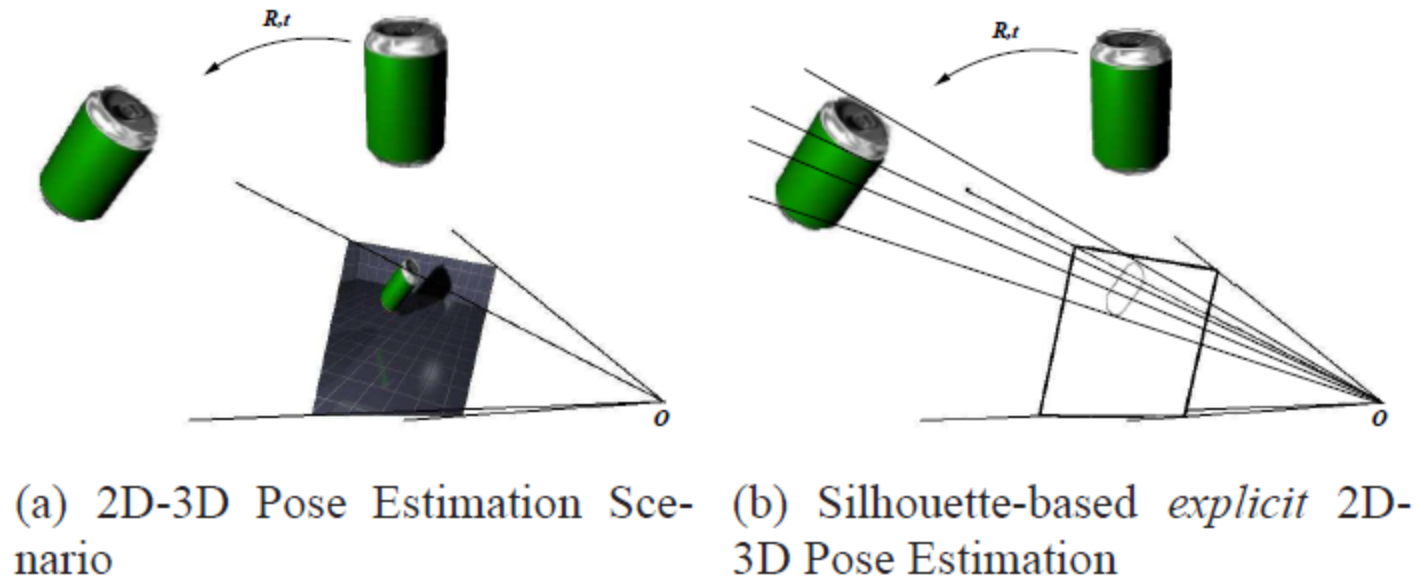
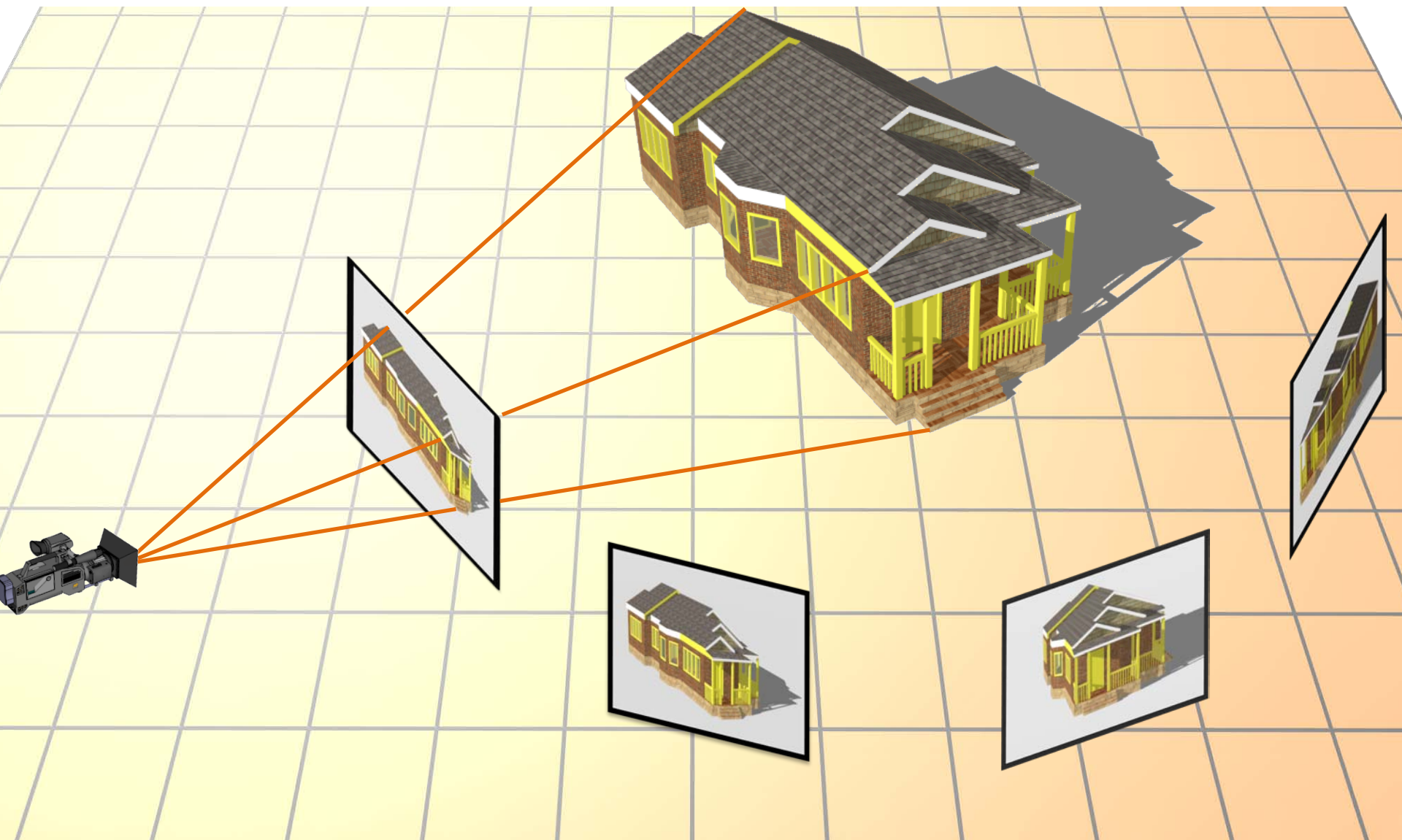


Figure 1. 2D-3D Pose Estimation

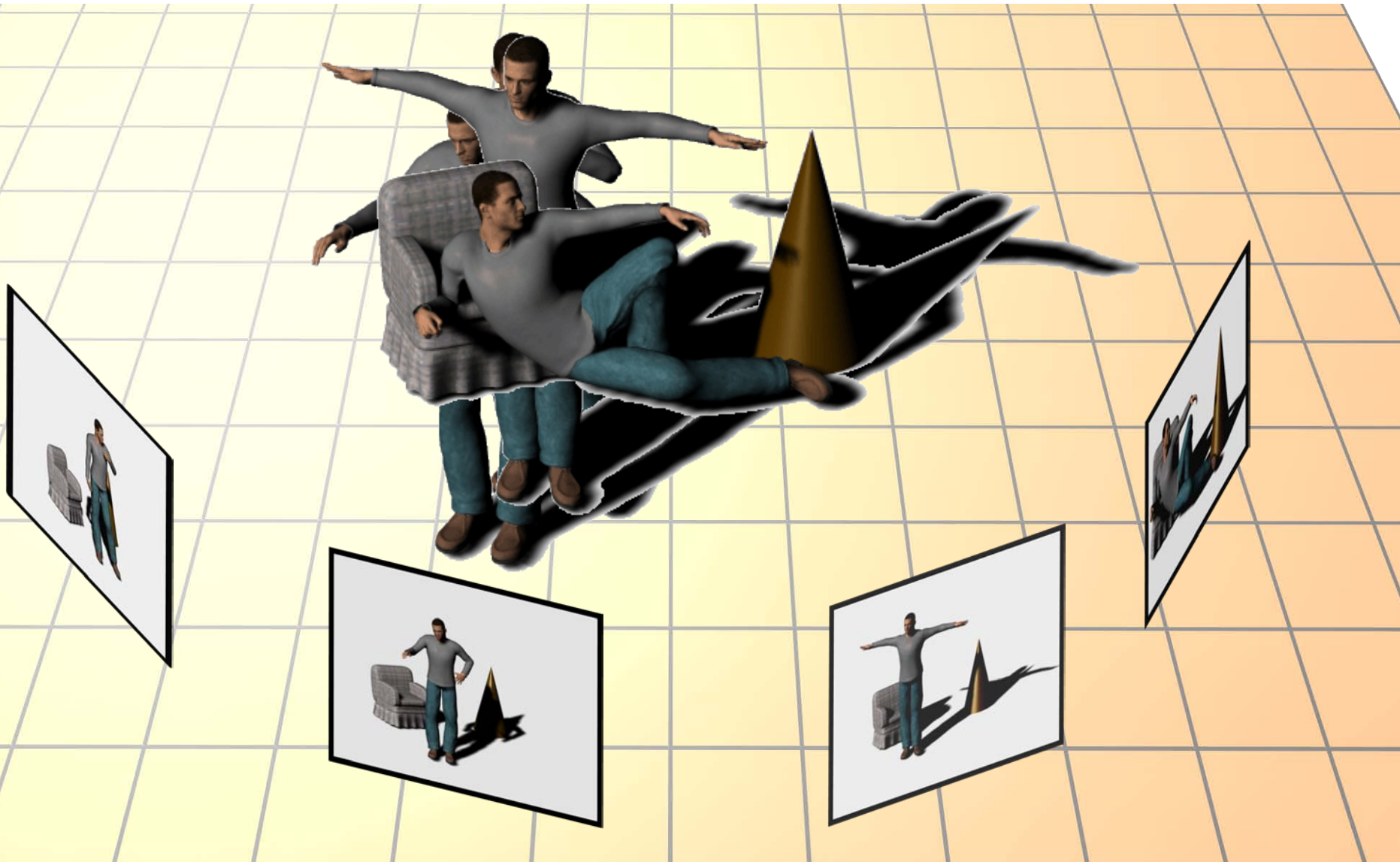
Author: Nazar Khan (2007)

Find the rotation  $\mathbf{R}$  and translation  $\mathbf{t}$  that aligns a 3D object with its 2D image.

# Rigid Structure from Motion

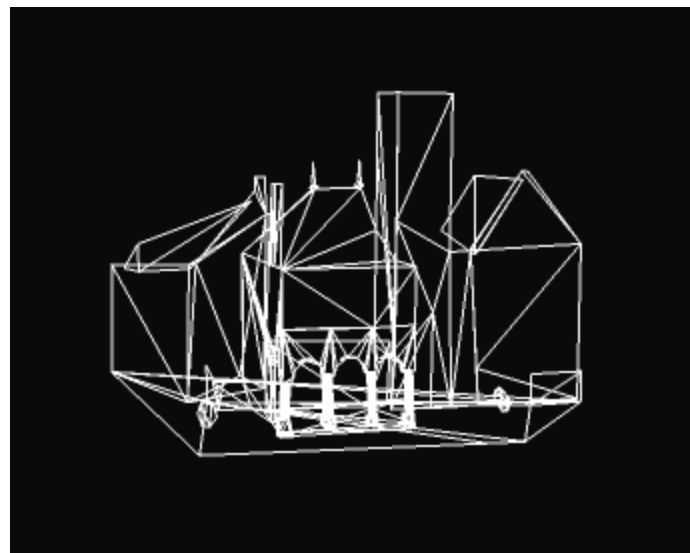


# Nonrigid Structure from Motion





# 3D Reconstruction from a Single View

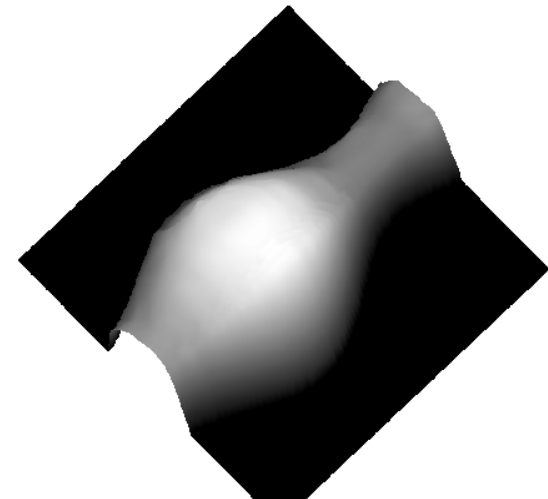
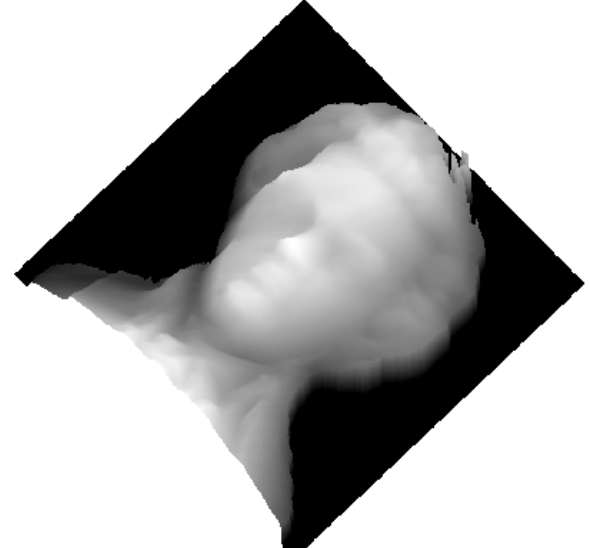
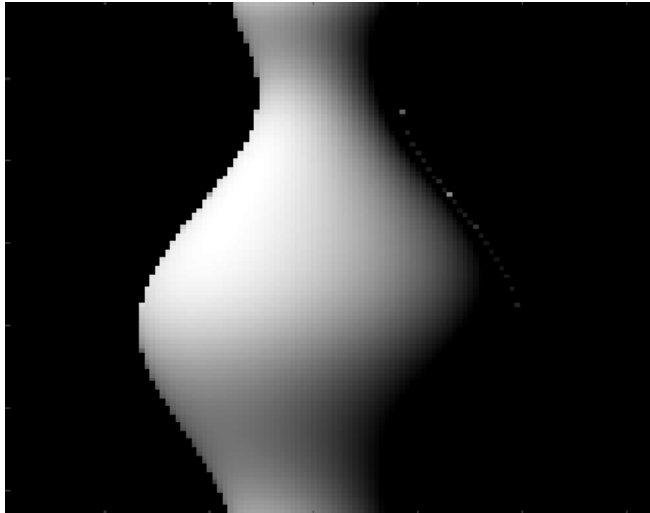


# 3D Reconstruction from a Single View



*Saint Jerome in his study* (1630) Joseph R. Ritman Collection  
by Henry V. Steinwick (1560-1649)

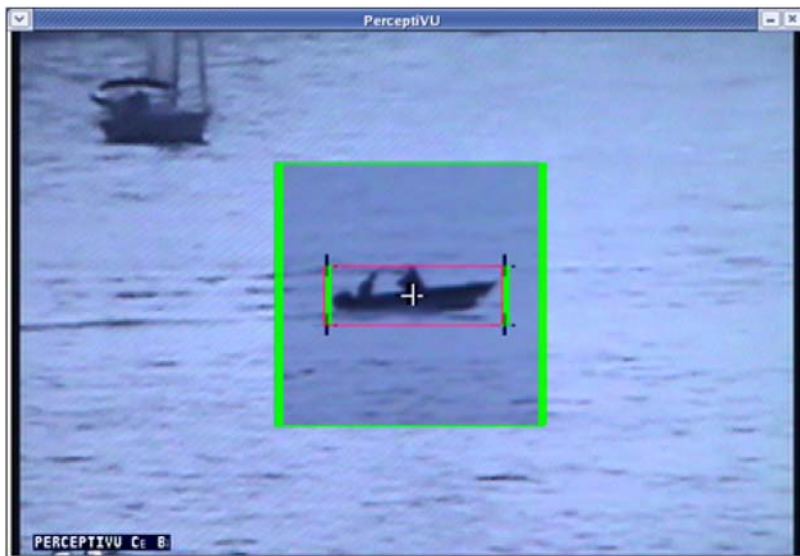
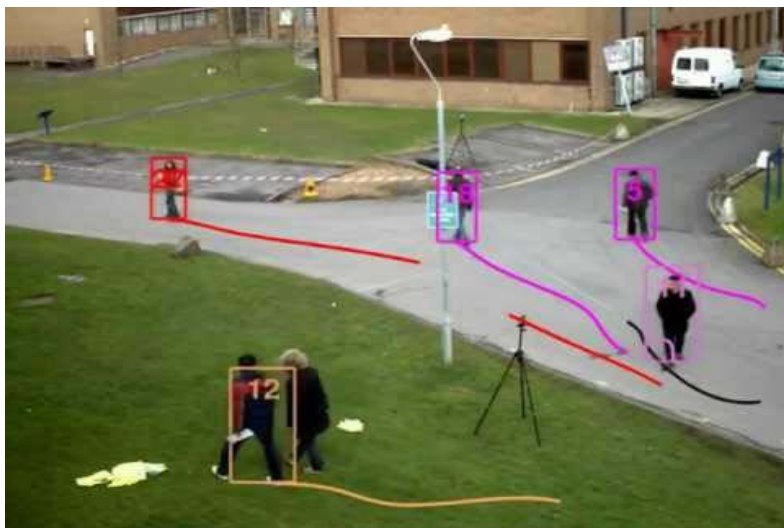
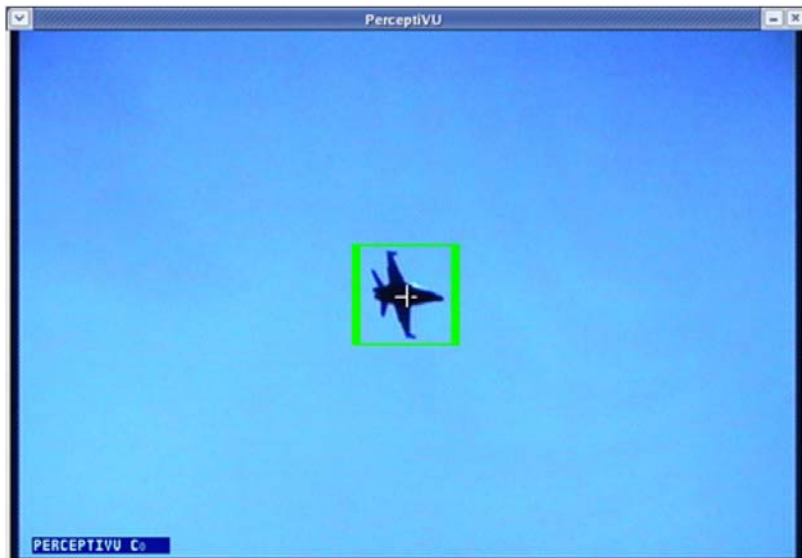
# 3D Reconstruction from a Single View



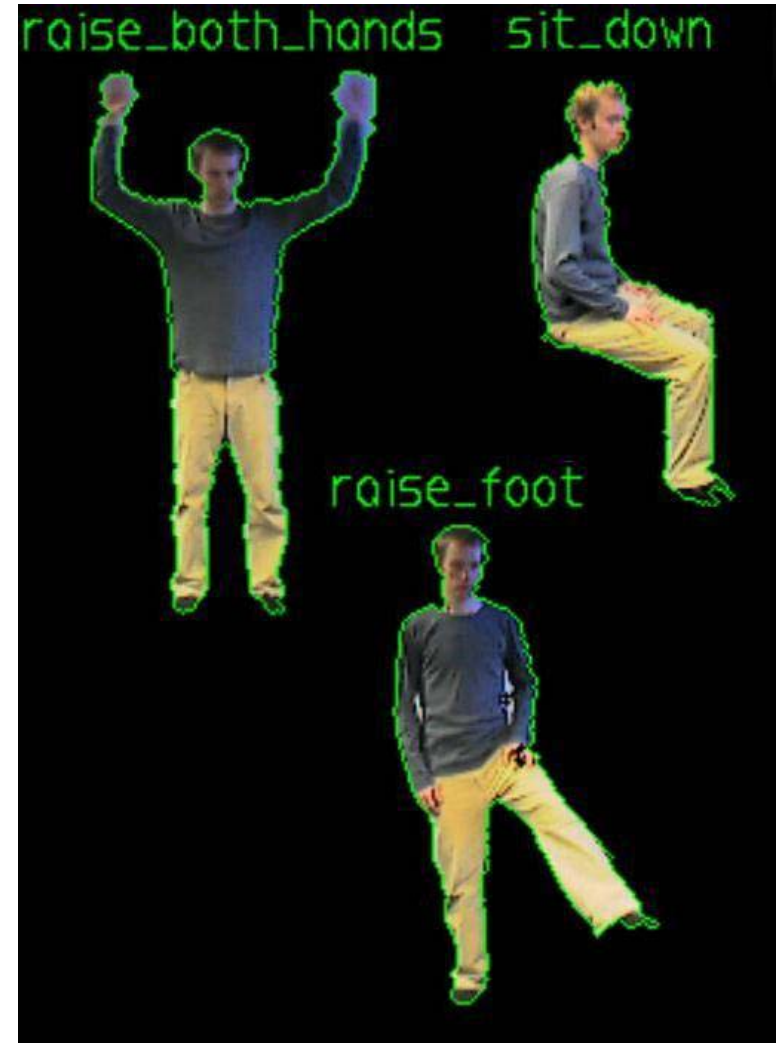
Author: Nazar Khan (2009)



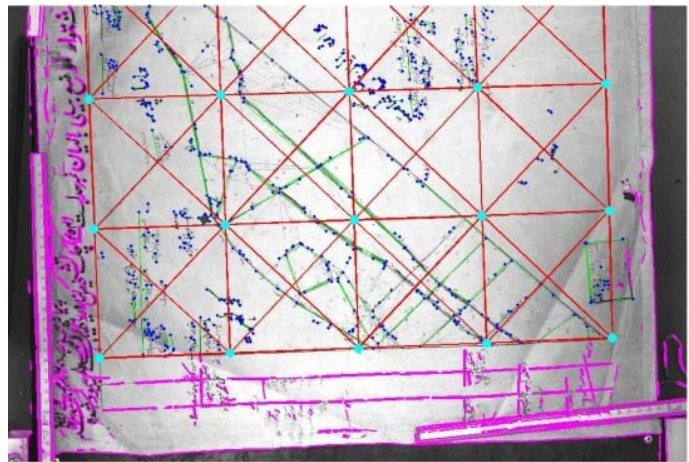
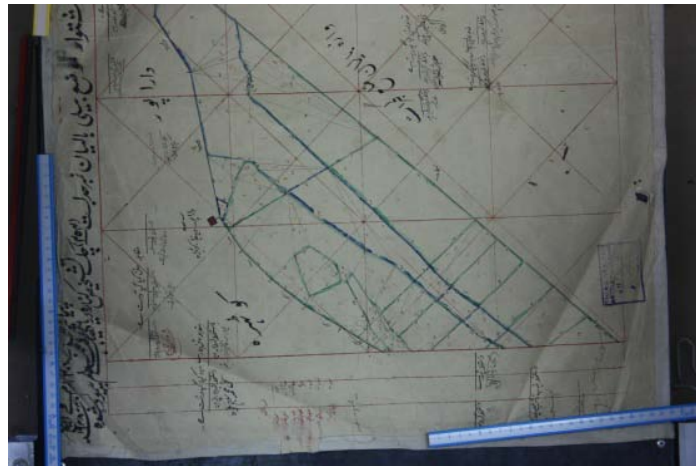
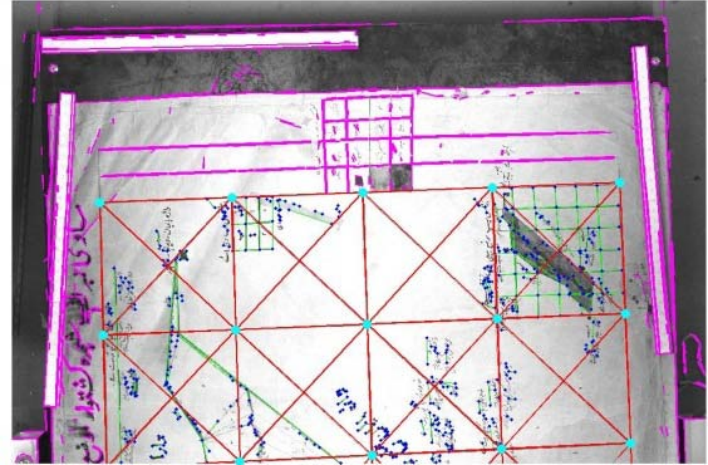
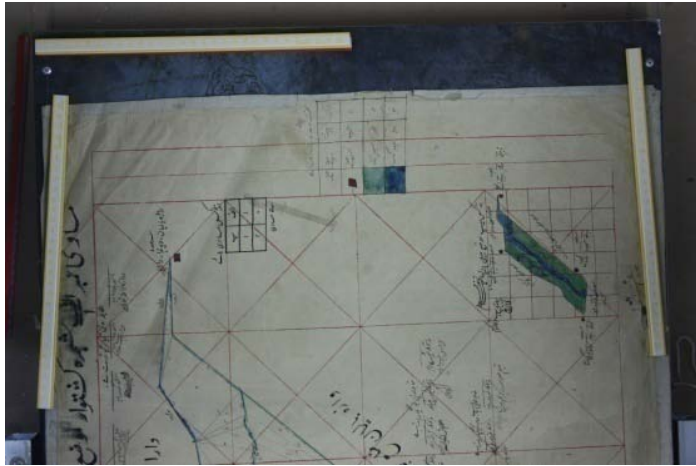
# Tracking



# Action Recognition

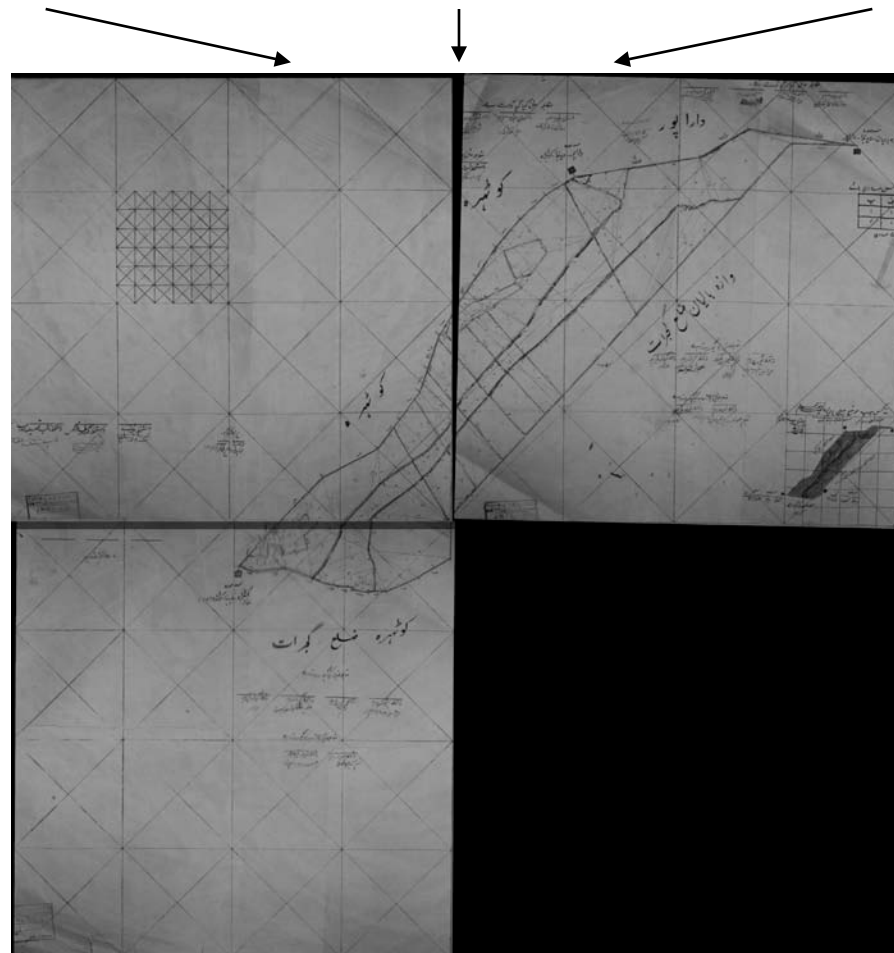
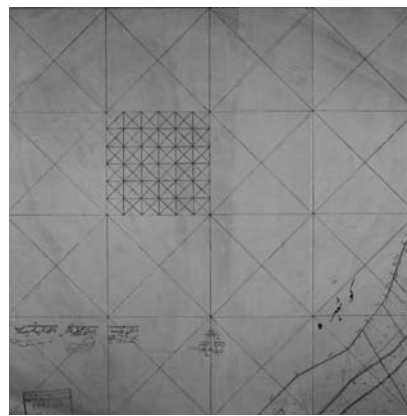
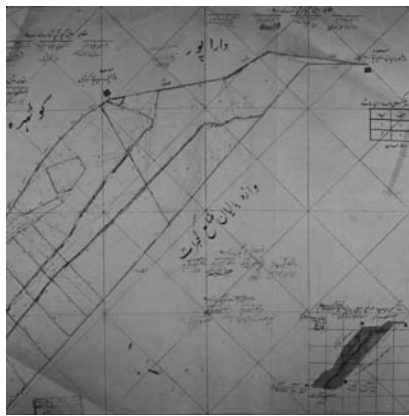


# Document Processing

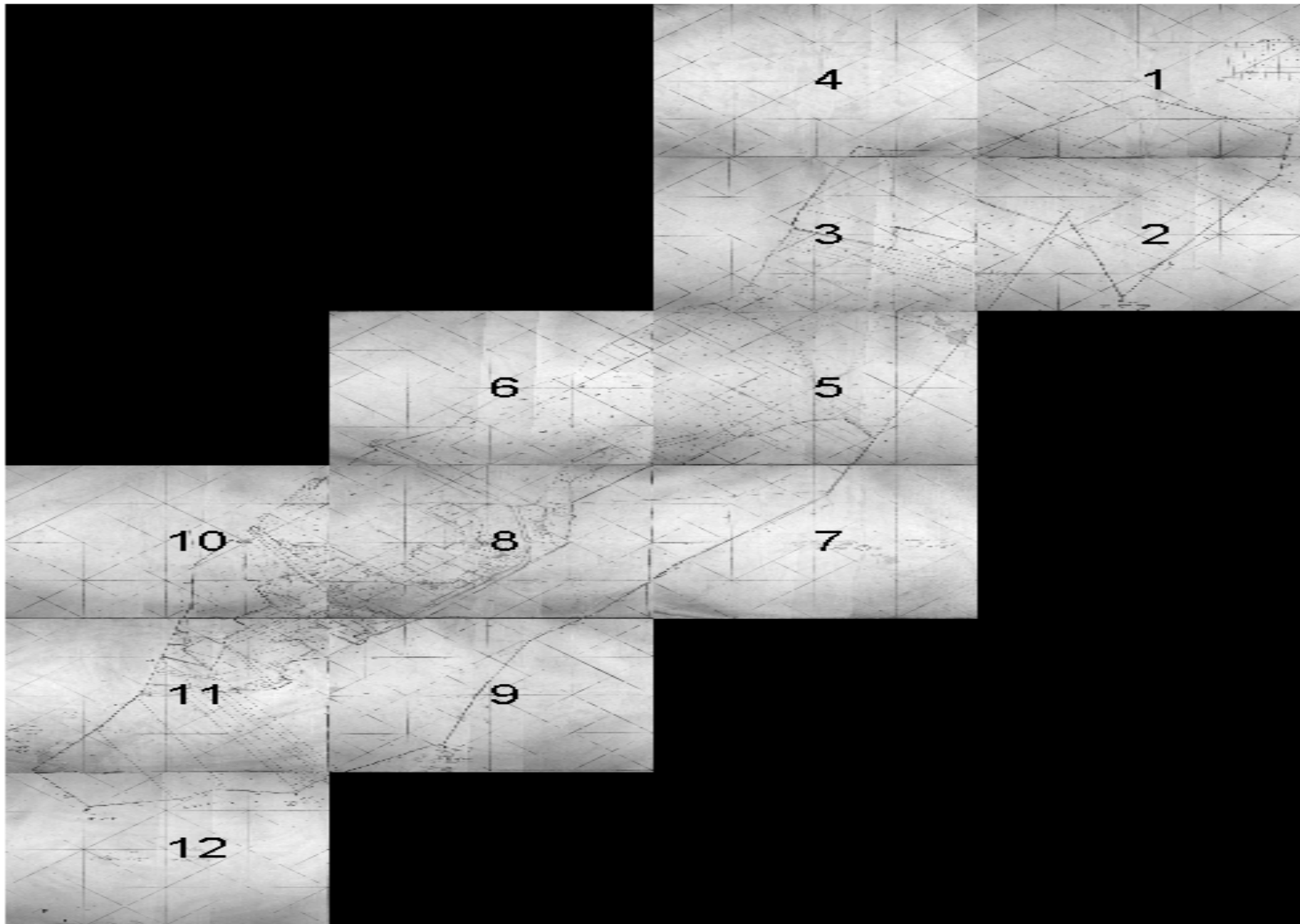




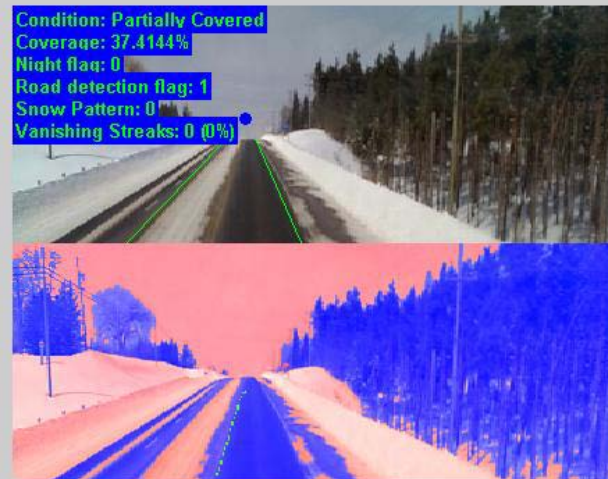
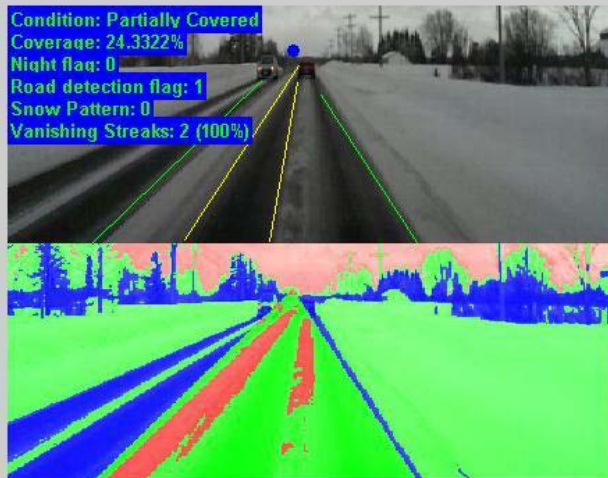
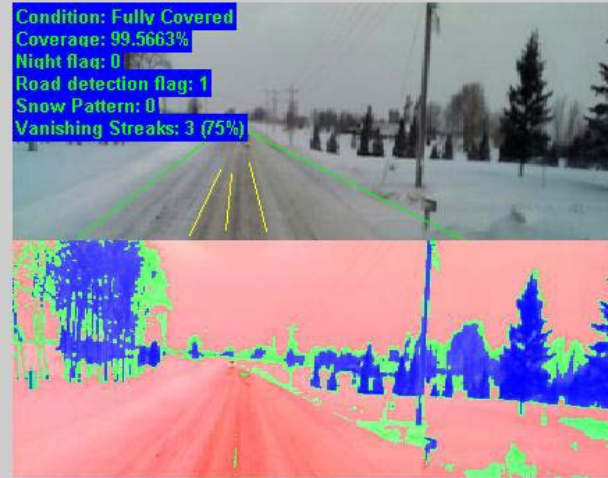
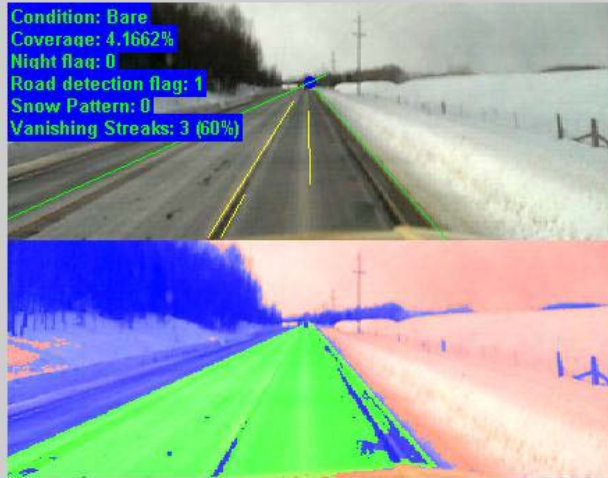




Author:  
Nazar Khan  
(2014)



# Road Condition Classification



Author:  
Nazar Khan  
(2014)