CS 565 – 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

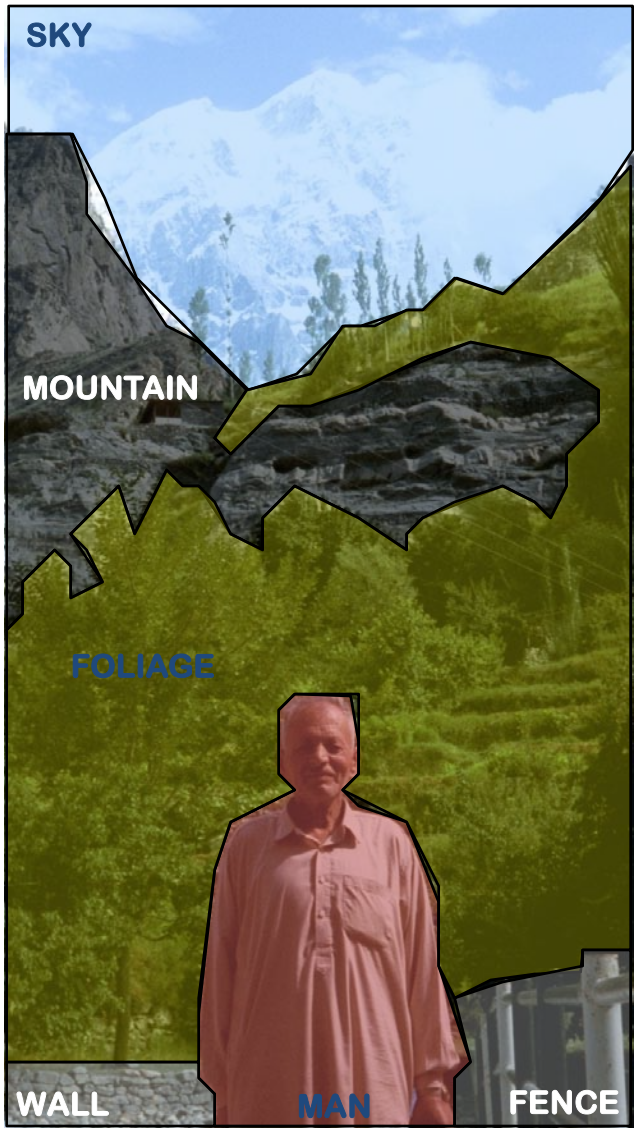
- There is no such thing as a ‘stupid question’.
- Your questions will help your class-mates.
- Your questions will make sure I do not go too fast.
- Your questions provide feedback for me.

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?

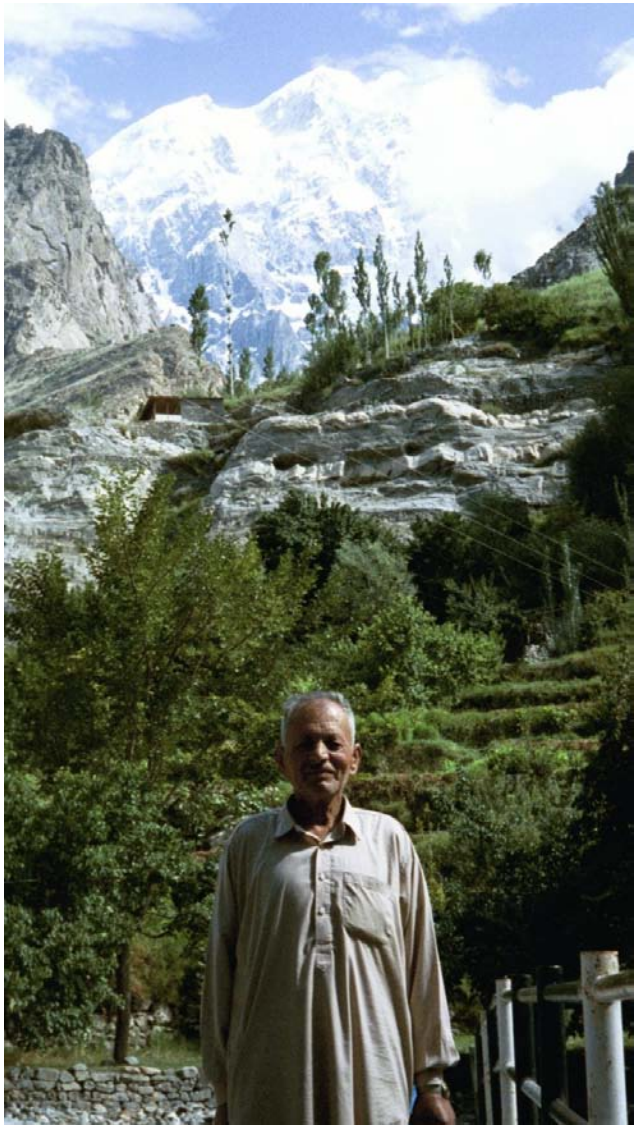
What *colors* do you see?

What *letters* do you see?

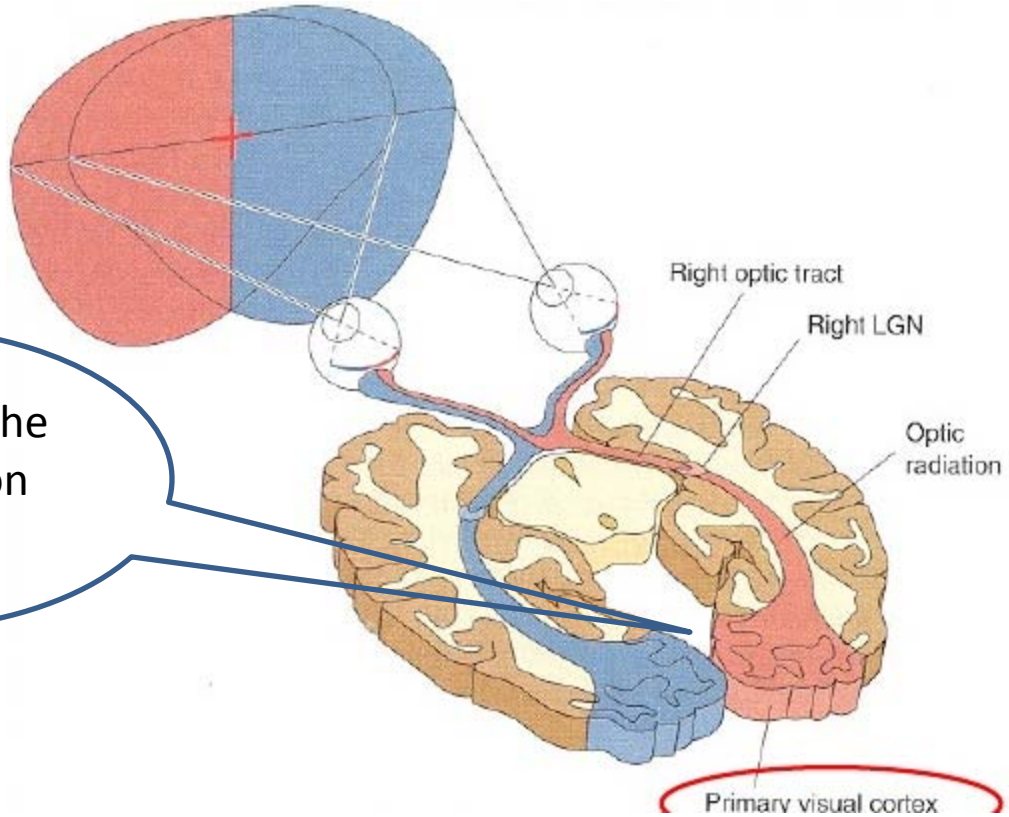
Do you see a grid of tiles?

Biological Vision

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



Where in the brain do we see?



This is where the magic of vision takes place!

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

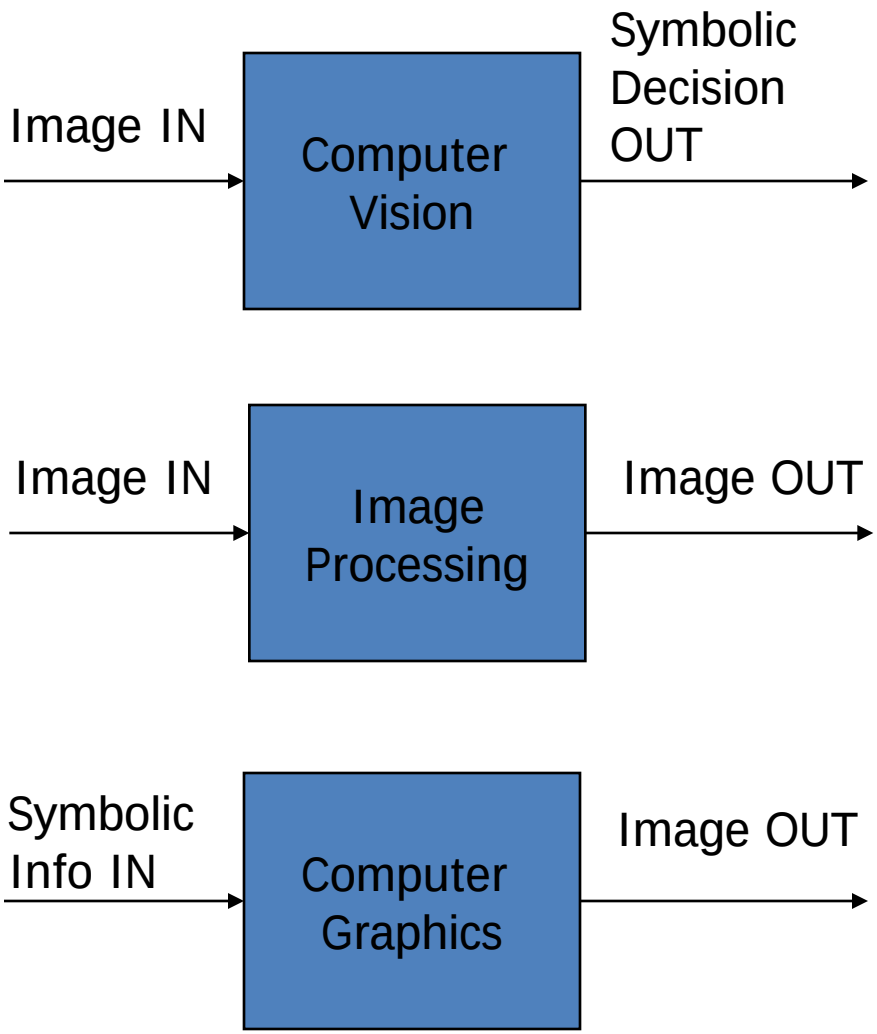
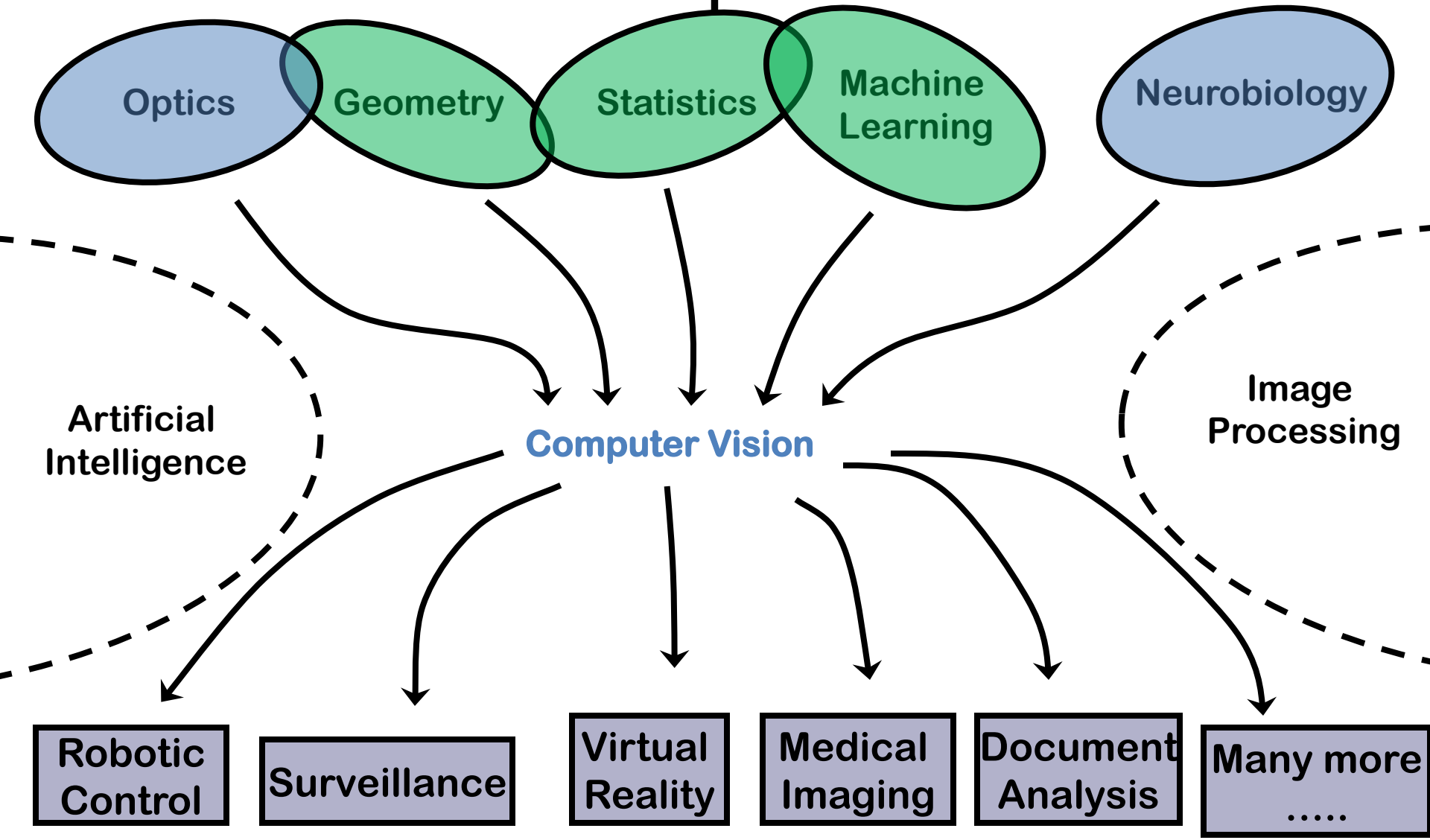


Image IN
Computer Vision
Symbolic Decision OUT
Image IN
Image Processing
Image OUT
Symbolic Info IN
Computer Graphics
Image OUT

What is computer vision?



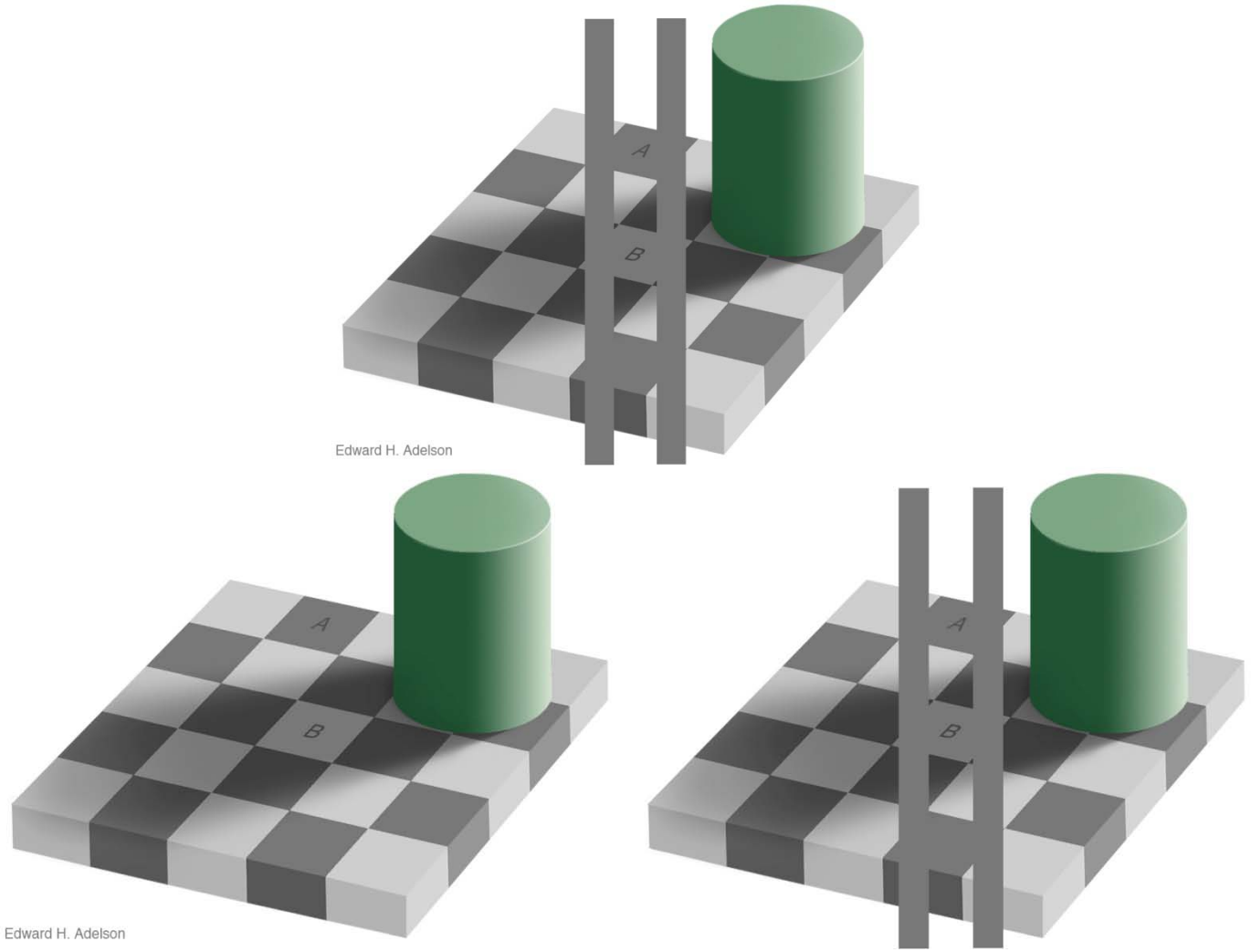
Why is computer vision hard?

- 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

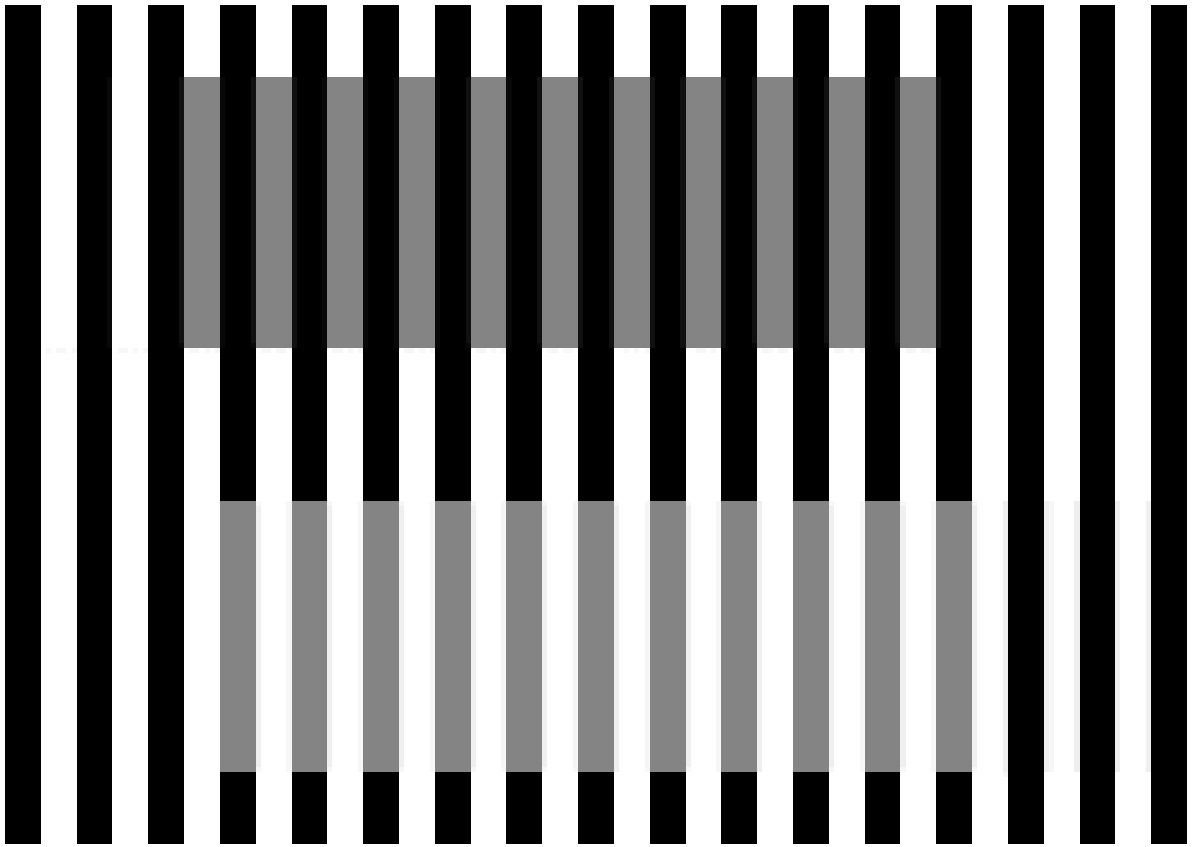
The Complexity of Perception

The Complexity of Perception

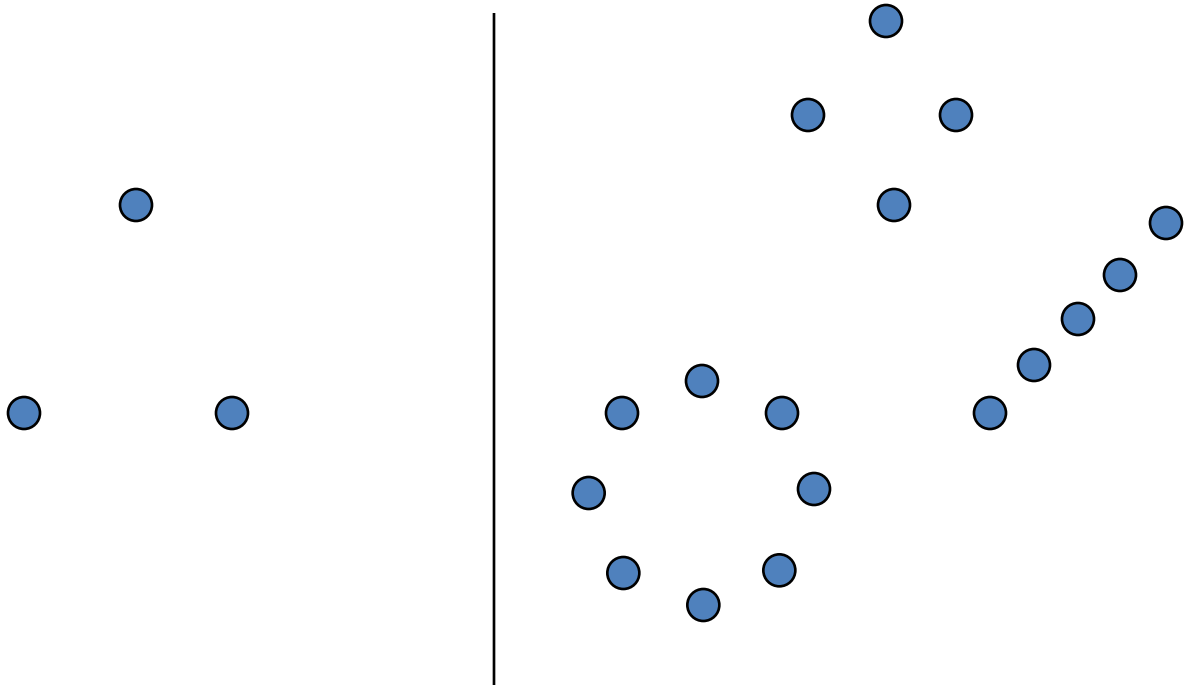


By Edward Adelson, 1995

Perception



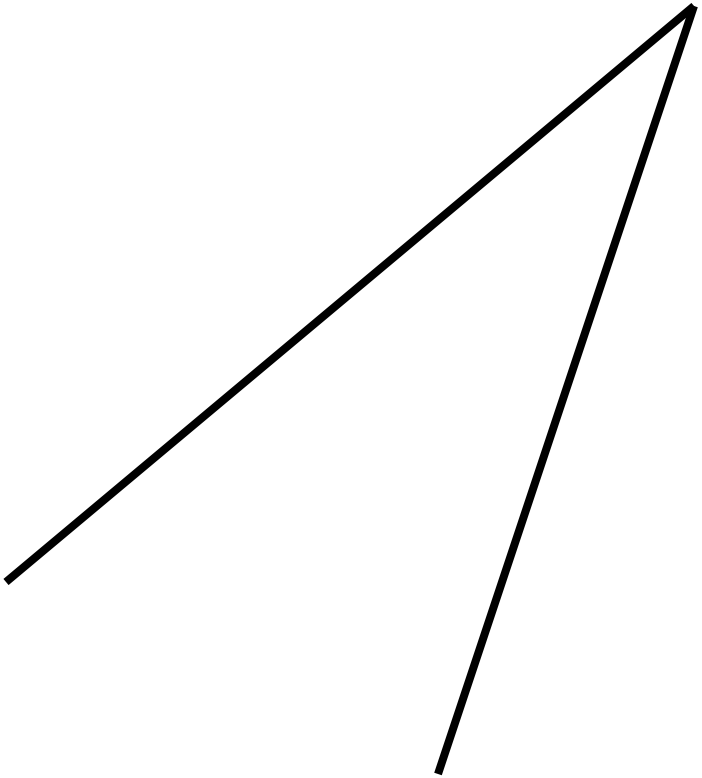
Perception

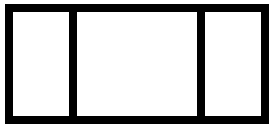


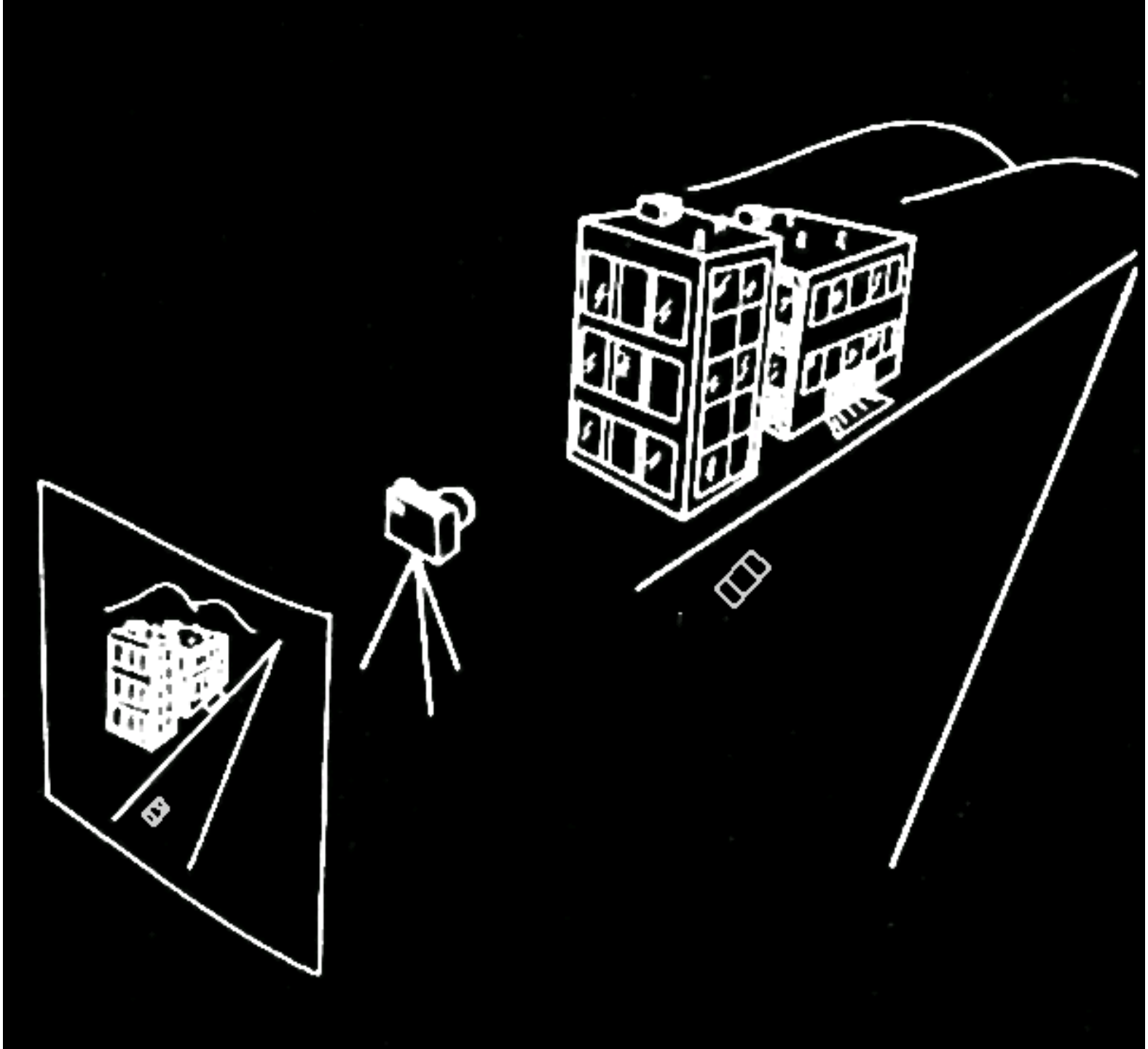
Perception

Ref: Light and Vision: LIFE Science Library

What is this?

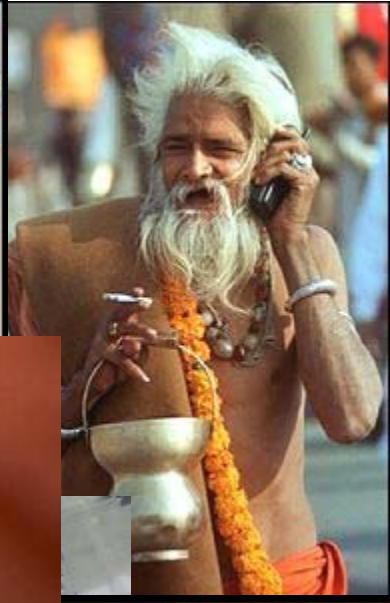




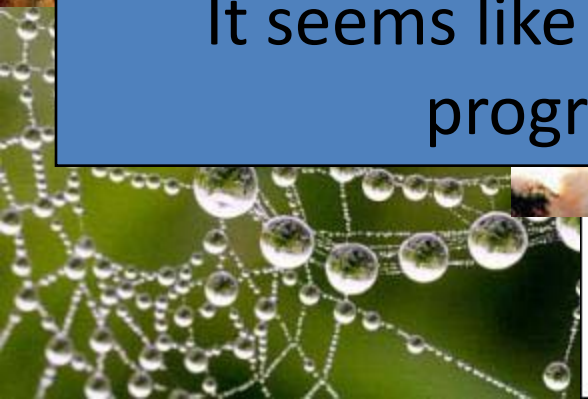


The Complexity of Perception




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



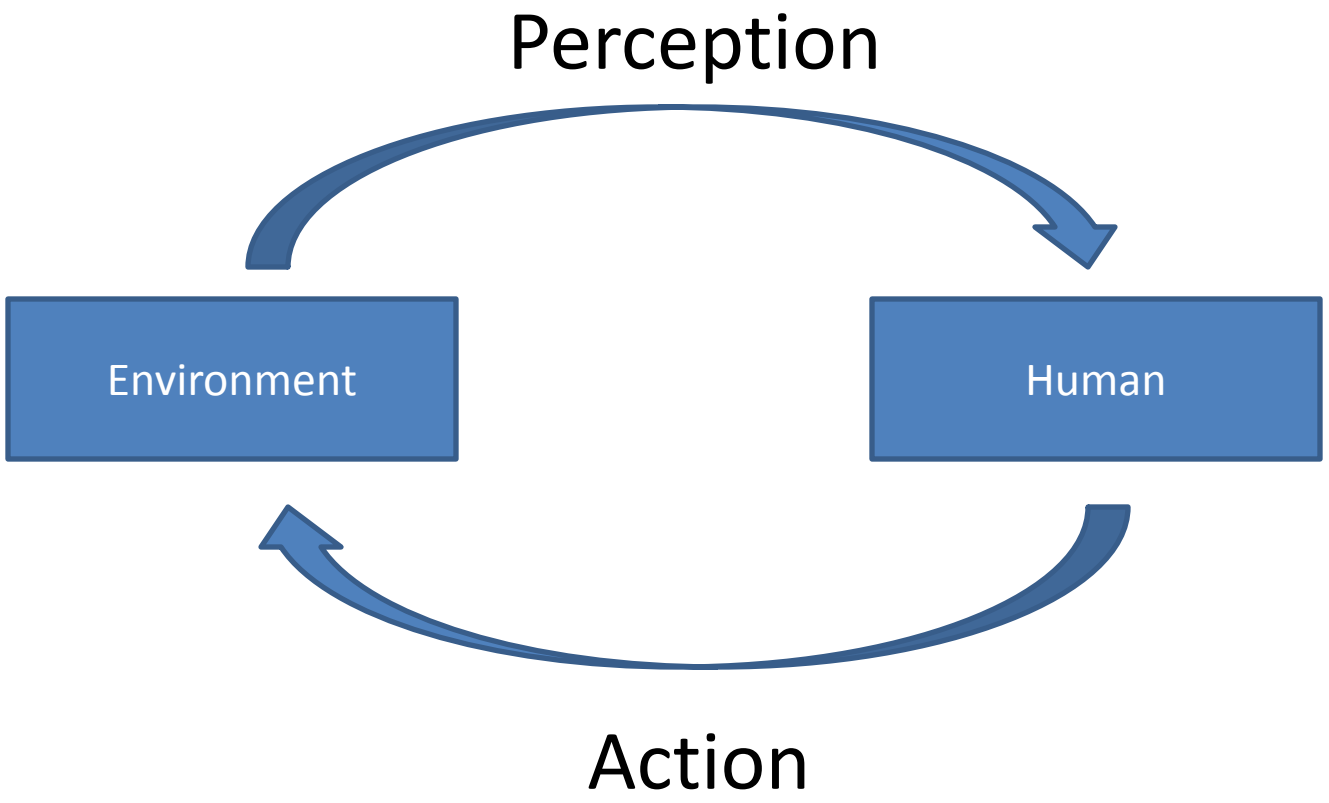
Writing Programs that “See”

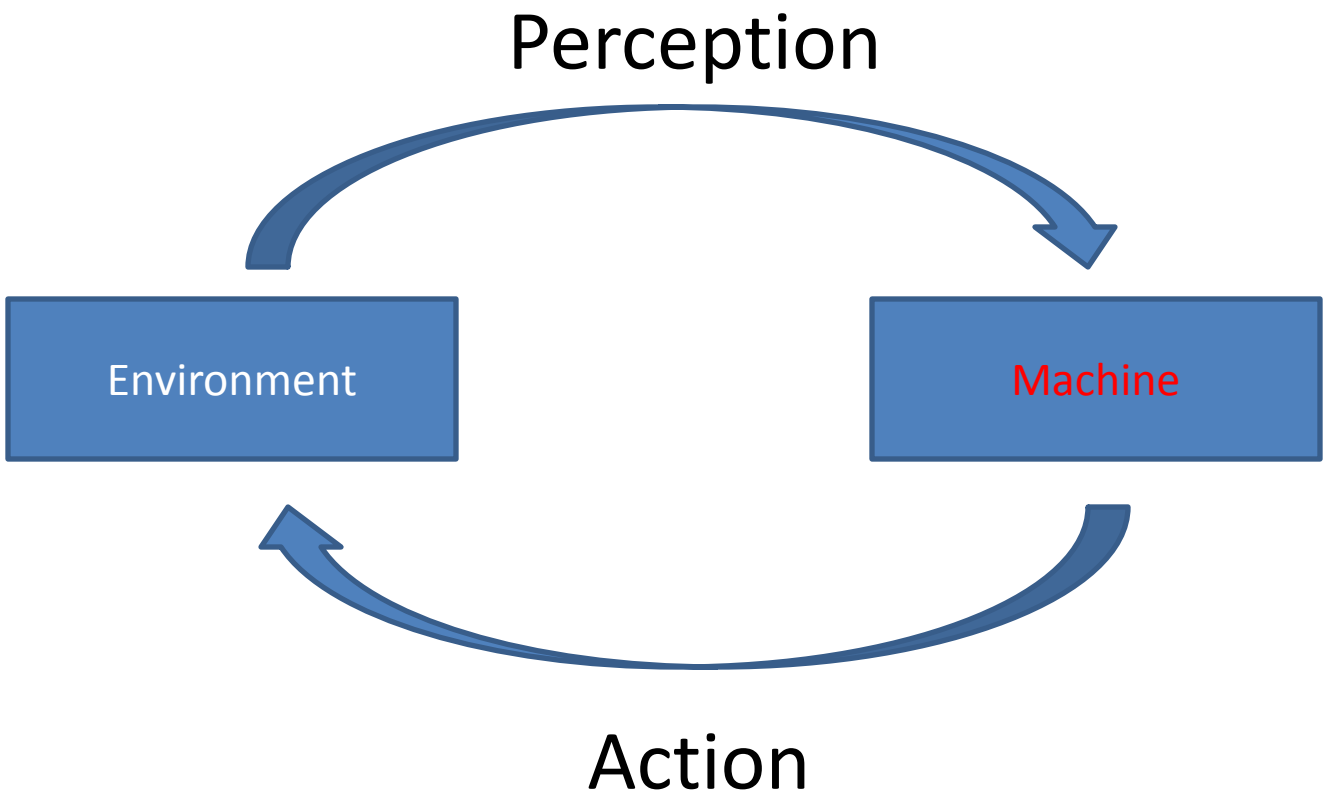
An Example

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 inform ation*"
- Can machines have similar capabilities?
  - Can we write algorithms for perception?

The Perception-Action Cycle




Perception
Environment
Machine
Action

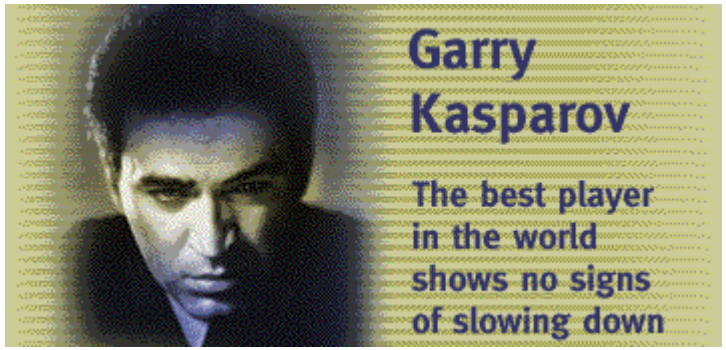
Motivation

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

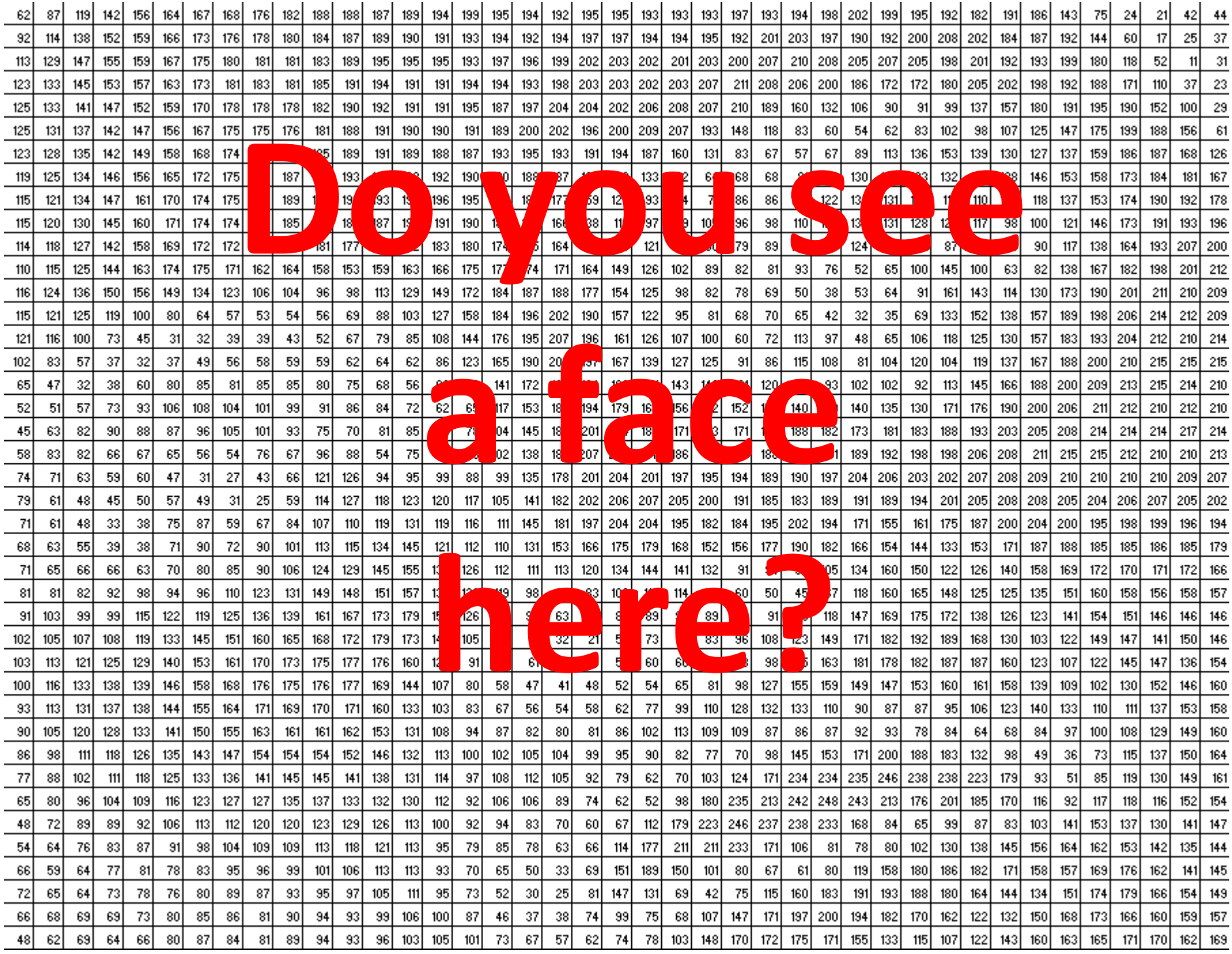
http://www.research.ibm.com/deepblue/

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


250
200
150
100
50


Do you see
a face
here?

Why is pattern recognition hard?

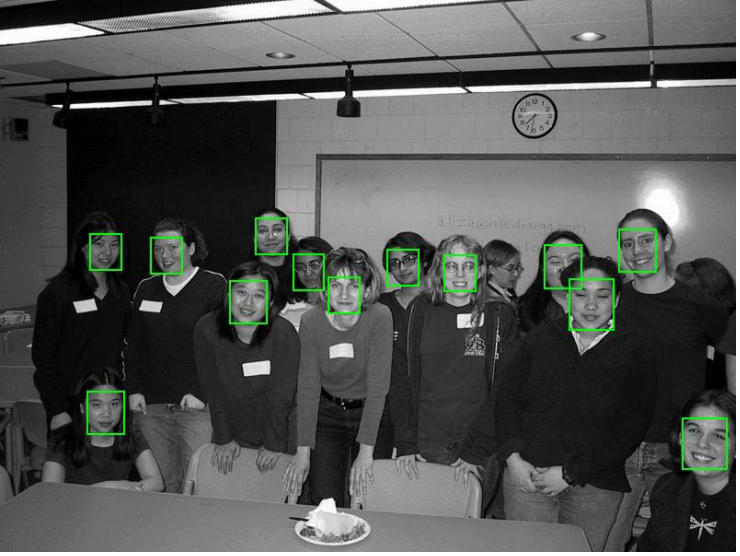
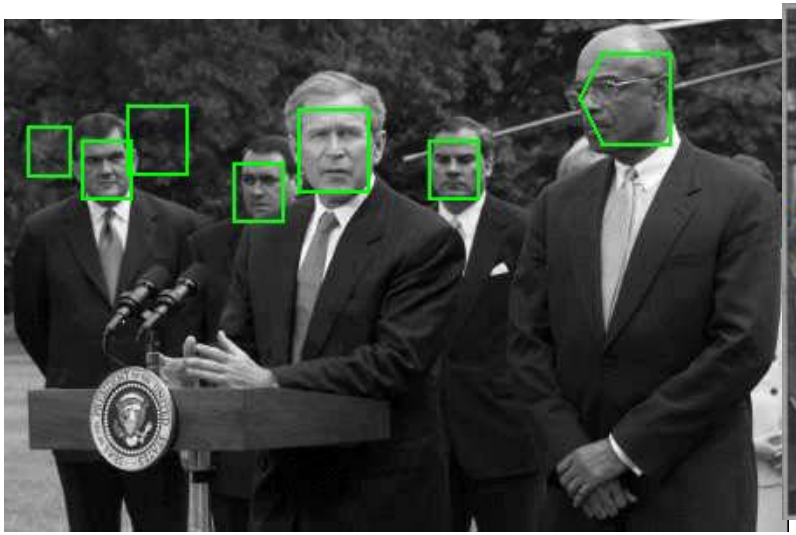
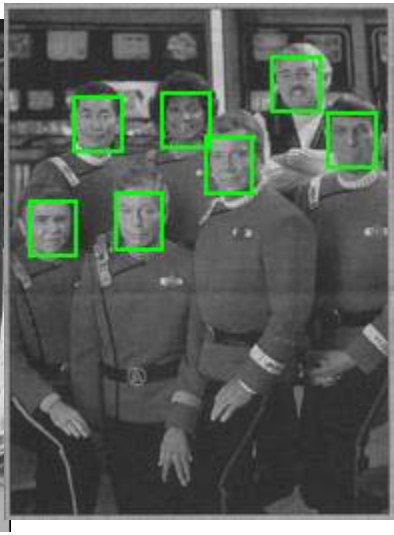
- To write an algorithm, you may need to precisely define a human face
- Your description should be invariant to pose, 3D rotation, occlusion, race, gender
- The description should hold for all faces and nothing which is not a face should match the description

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

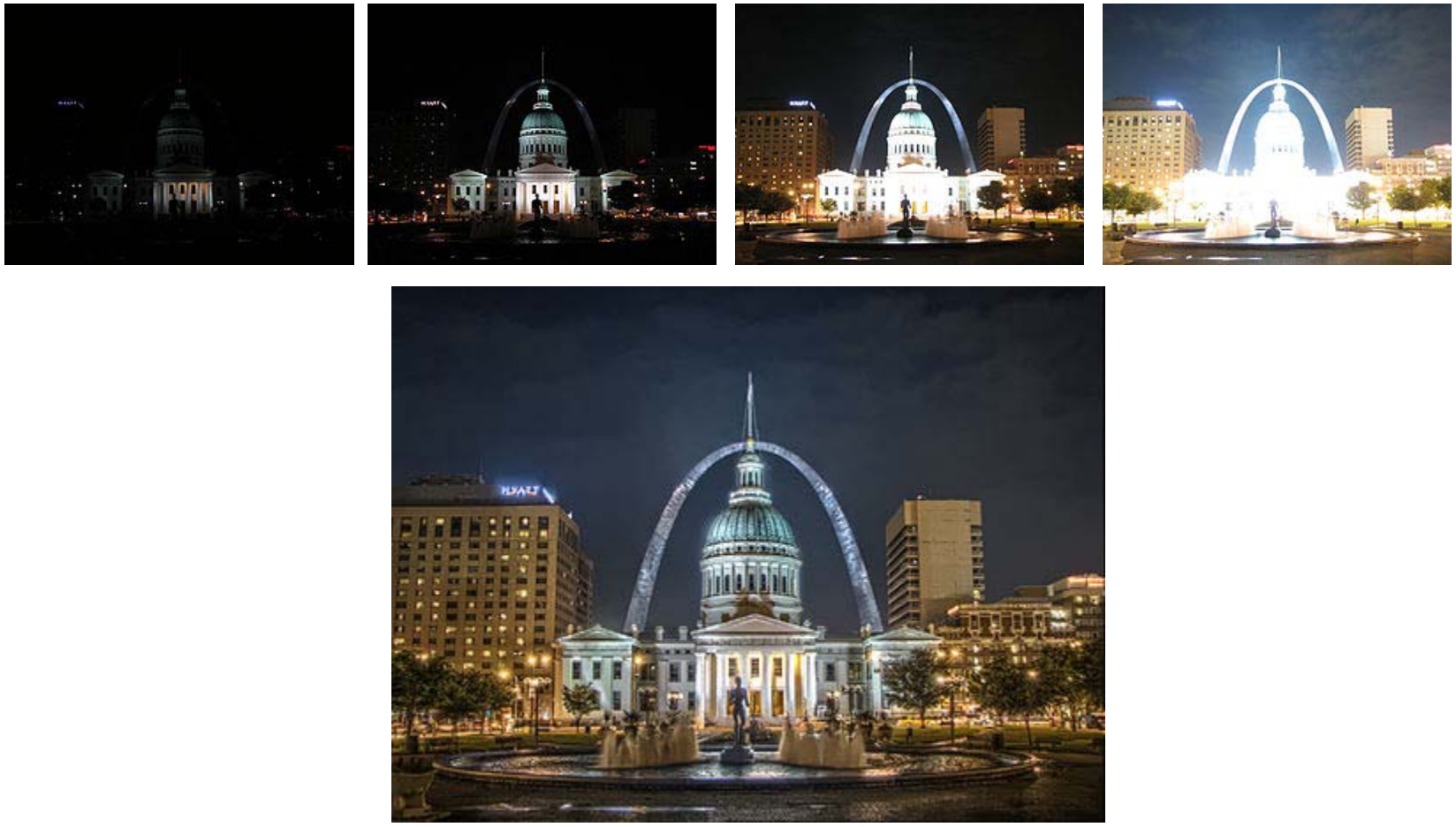
Results of Schneiderman/Kanade Face Detector

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

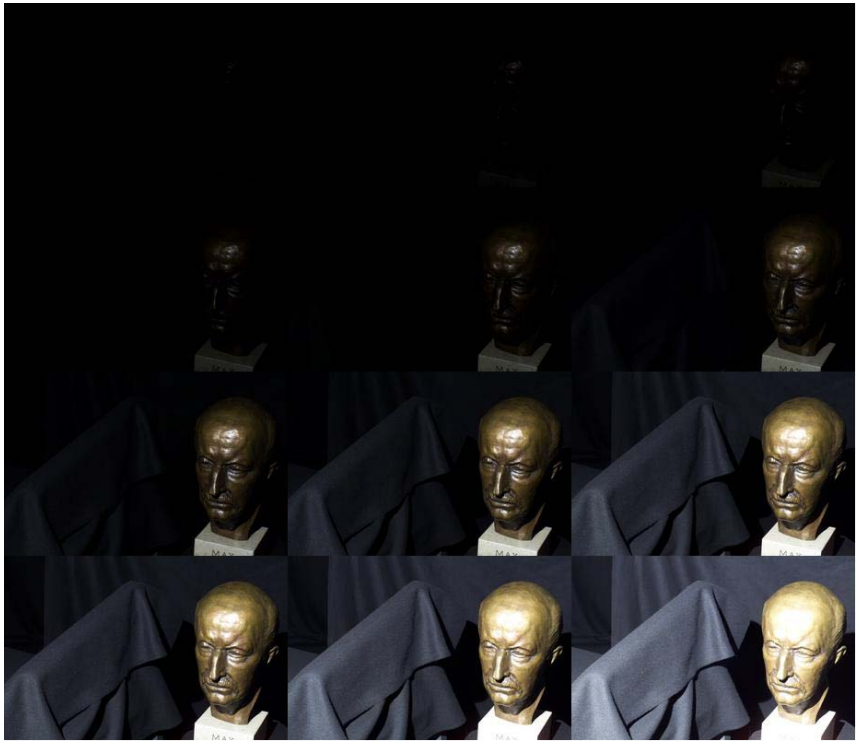
SAMPLE PROBLEMS FROM IMAGE PROCESSING AND COMPUTER VISION

Computational Photography



http://en.wikipedia.org/wiki/High-dynamic-range_imaging

Computational Photography



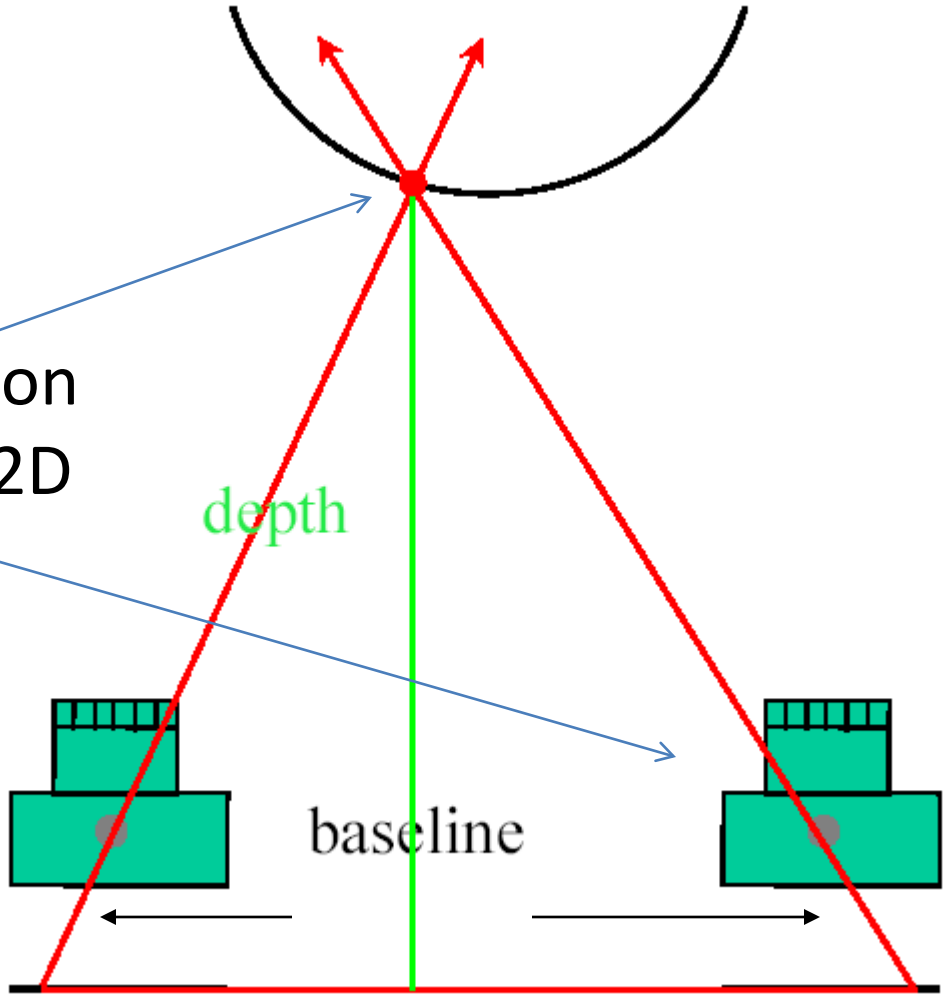
Author: Nazar Khan (2004)

Reconstruction of 3D Structure

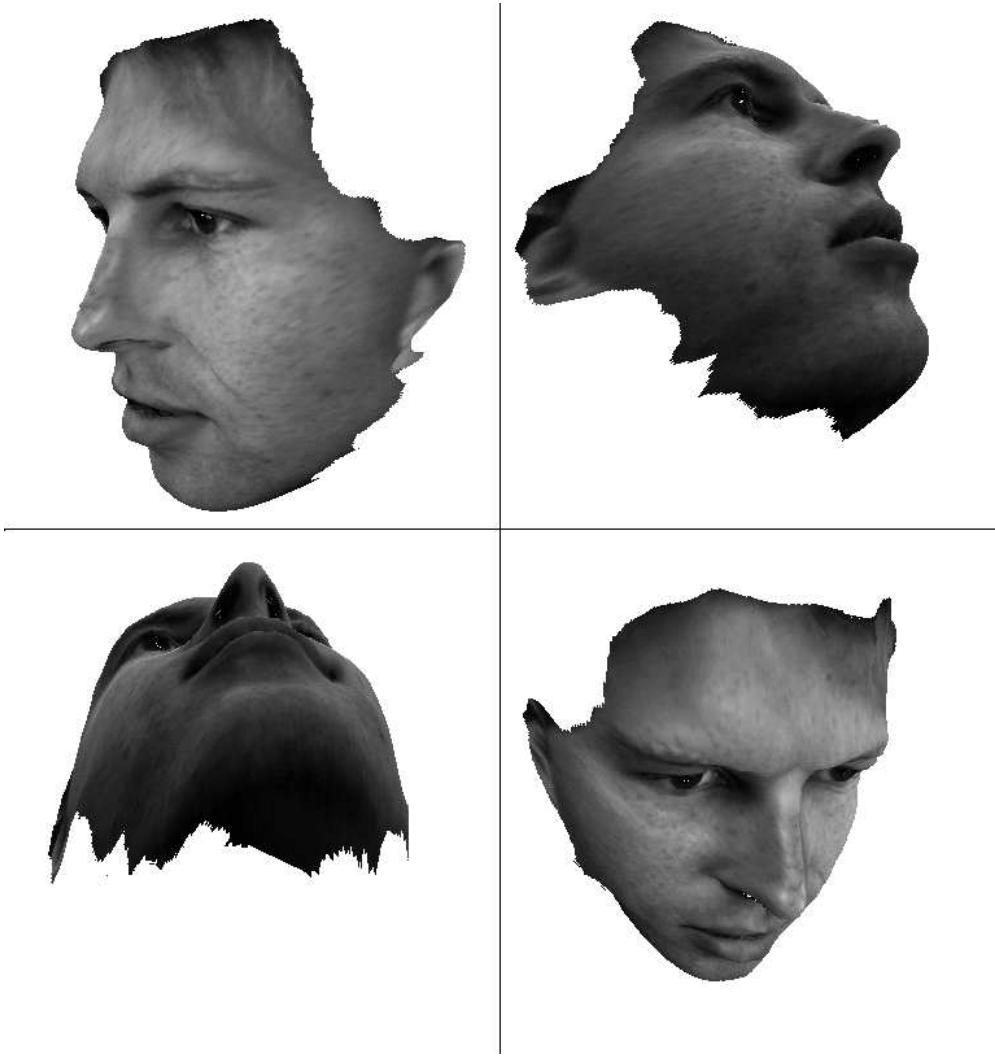
- An image is a 2D projection of the 3D world
- 3D can be reconstructed from
  - Two images
    - Stereo Problem
  - Video with moving camera
    - Structure from Motion Problem
  - Some understanding about what is being viewed
    - Geometrical inference
    - Shape from shading or texture

Stereo

3D point location from a pair of 2D images



Stereo image pair.

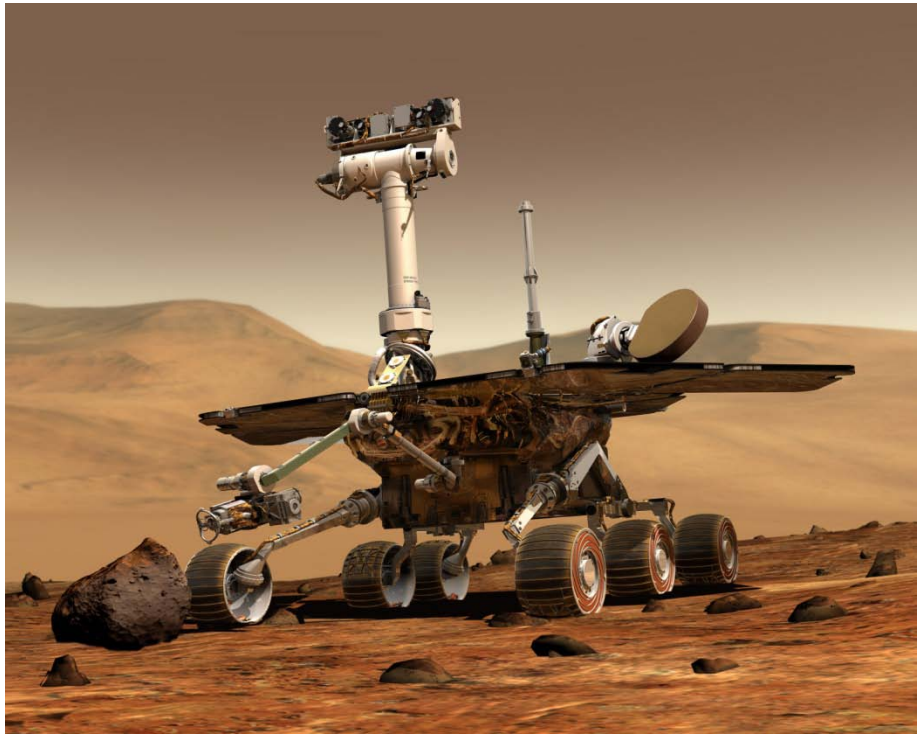


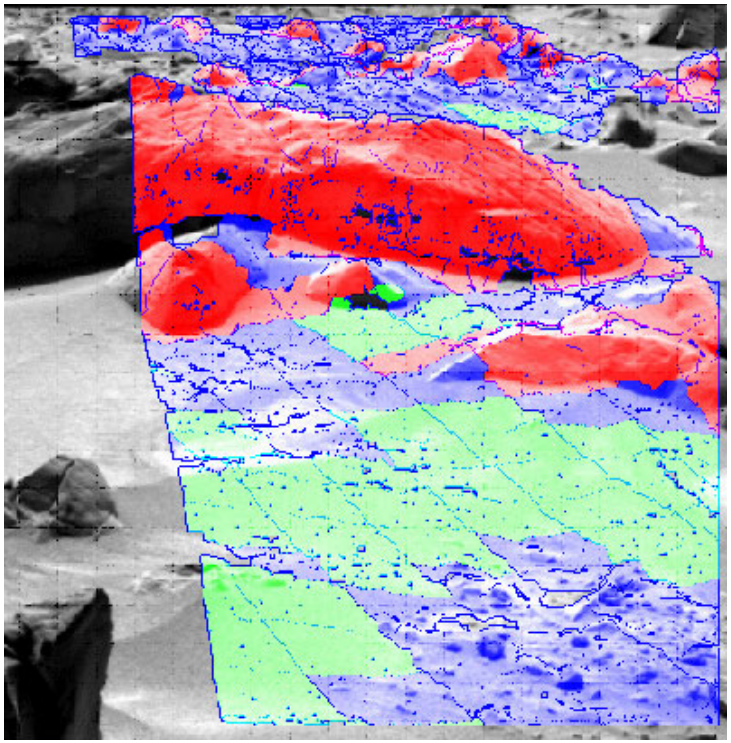
3-D reconstructions

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

Stereo in Space Exploration

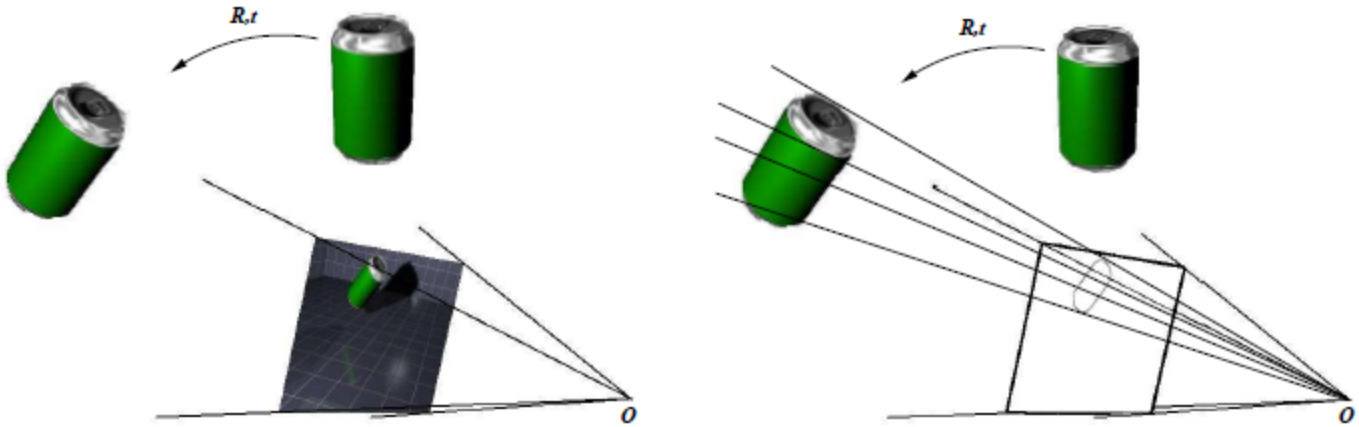
- Mar Exploration Rover





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

Pose Estimation



(a) 2D-3D Pose Estimation Scenario

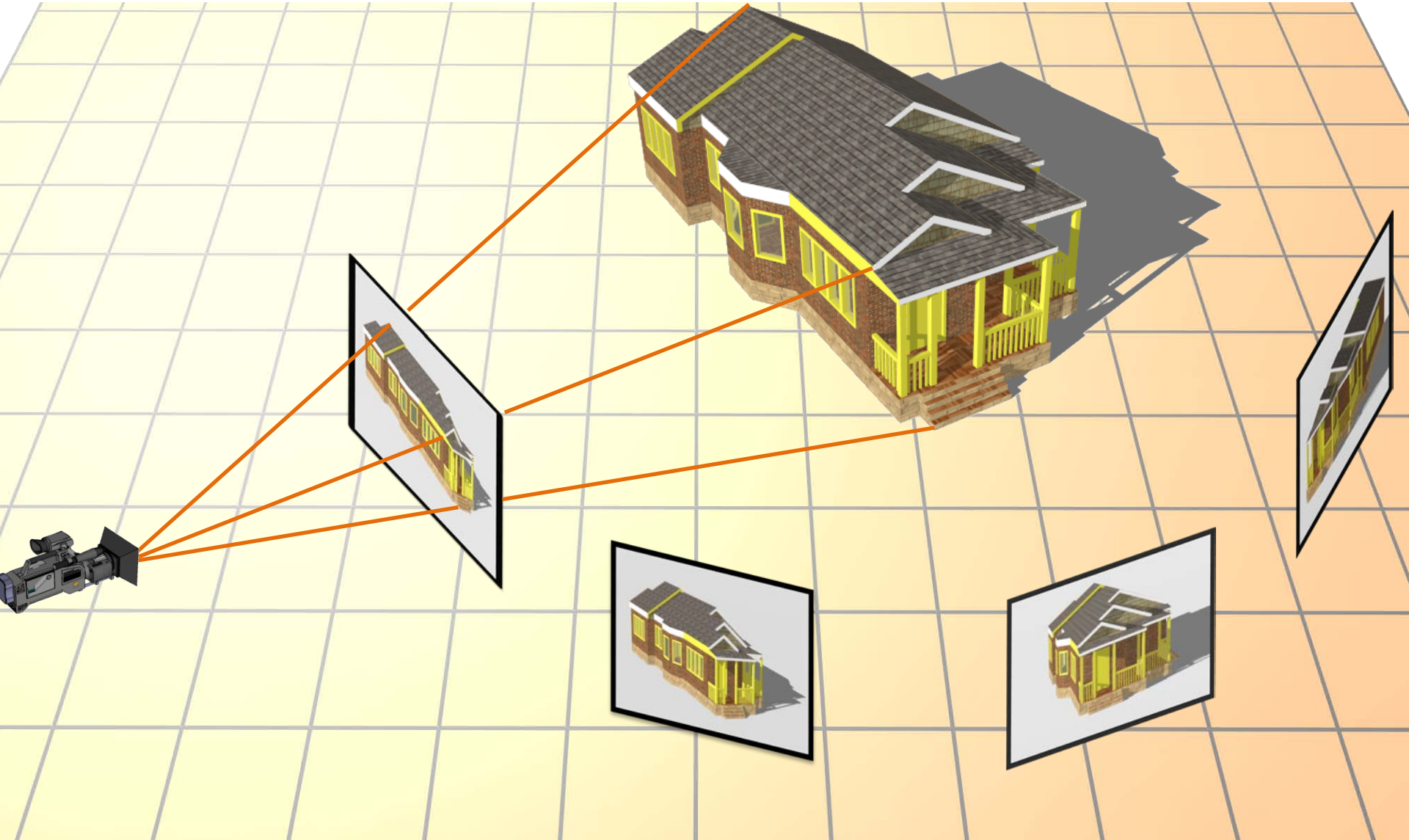
(b) Silhouette-based *explicit* 2D-3D Pose Estimation

Figure 1. 2D-3D Pose Estimation

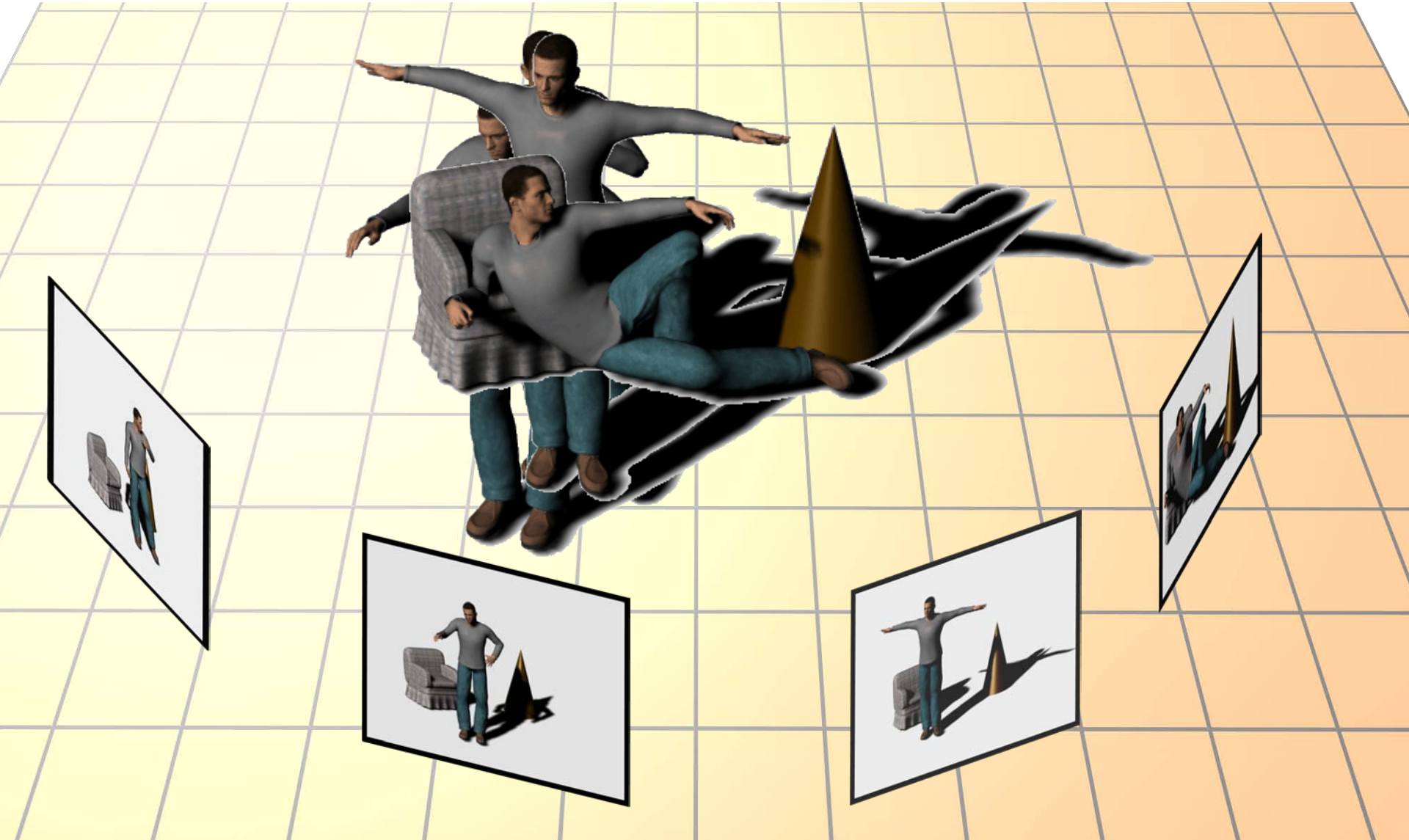
Author: Nazar Khan (2007)

Find the rotation ***R*** and translation ***t*** that aligns a 3D object with its 2D image.

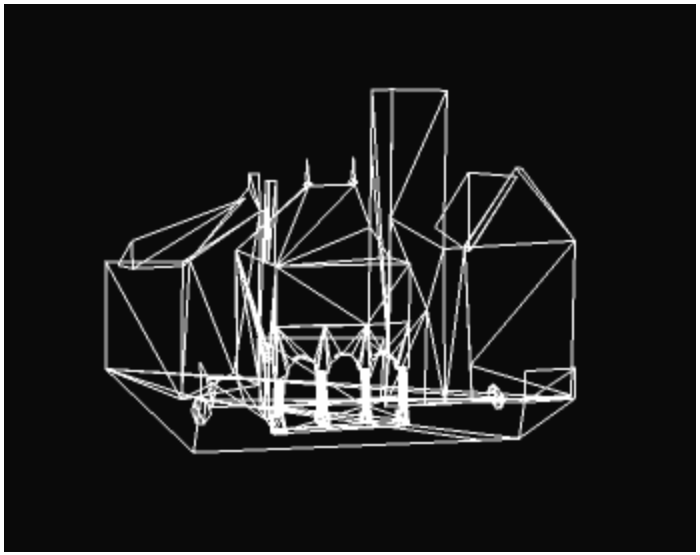
Rigid Structure from Motion



Nonrigid Structure from Motion



3D Reconstruction from a Single View

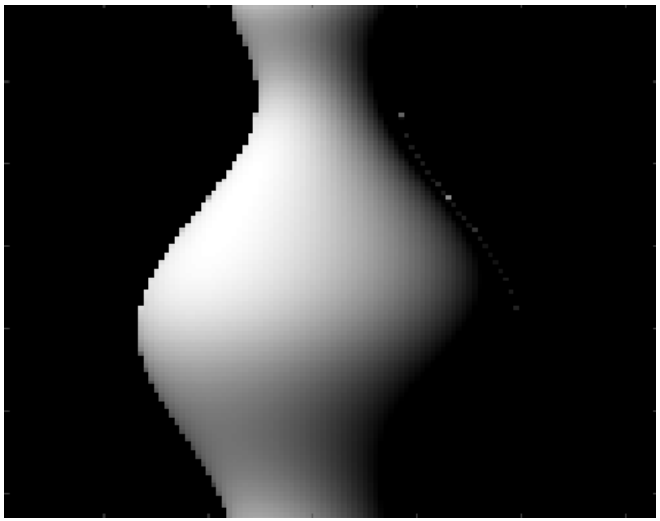


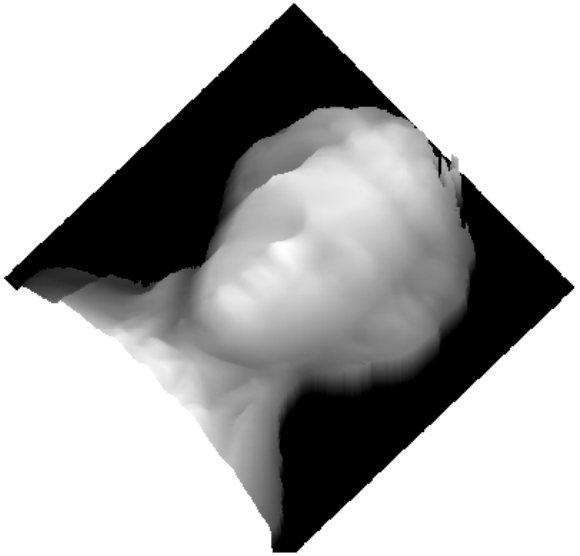
3D Reconstruction from a Single View

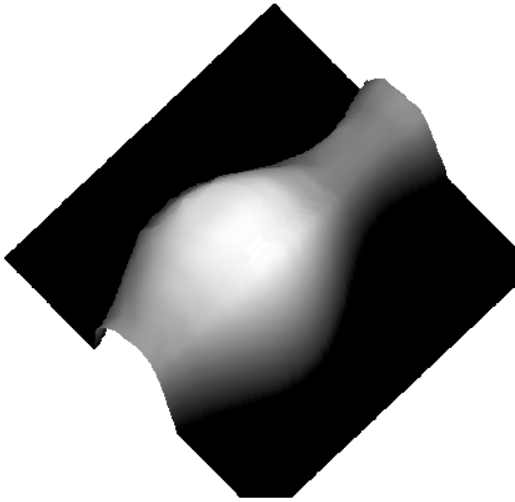
Saint Jerome in his study (1630) Joseph R.Ritman Collection
by Henry V Steinwick (1560-1649)

3D Reconstruction from a Single View

3D Reconstruction from a Single View

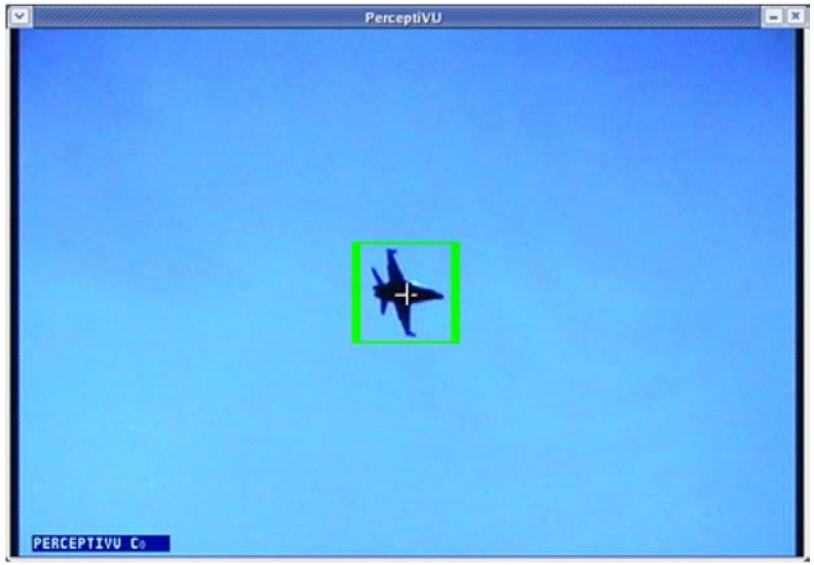




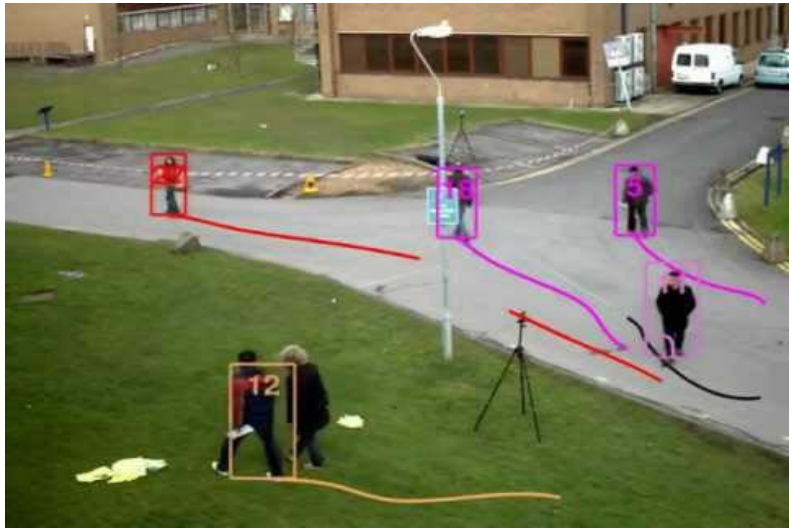


Author: Nazar Khan (2009)

Tracking

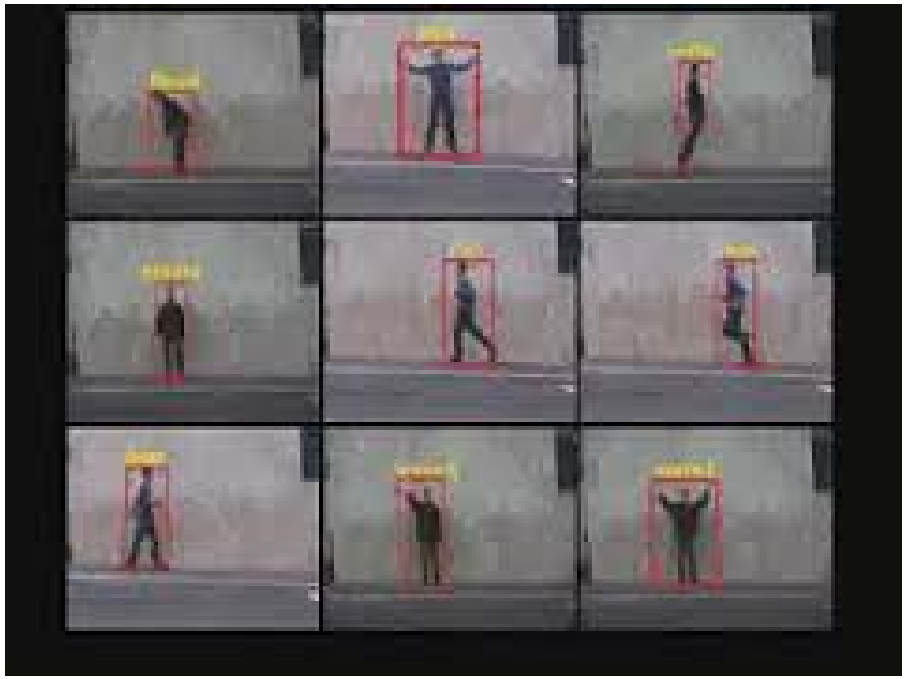

PerceptiVU
PERCEPTIVU

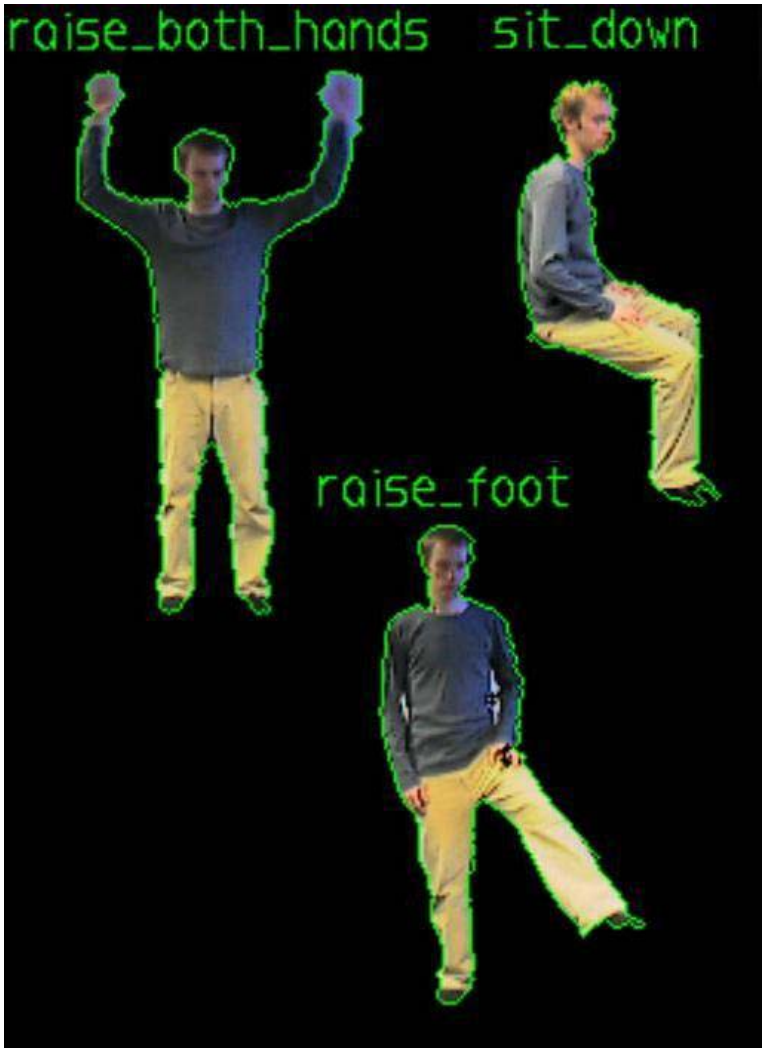

PerceptiVU
PERCEPTIVU


12

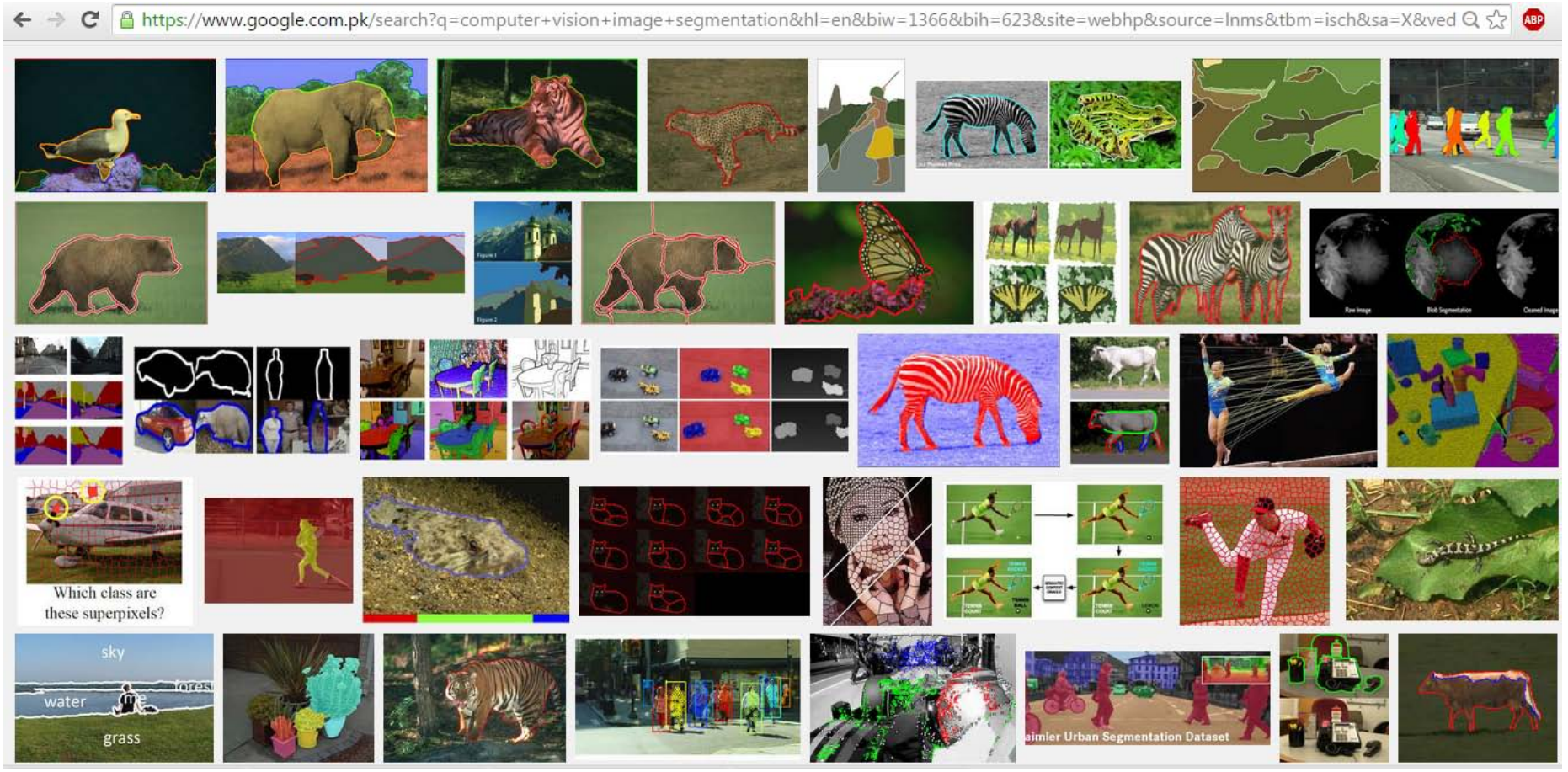

PerceptiVU
PERCEPTIVU

Action Recognition



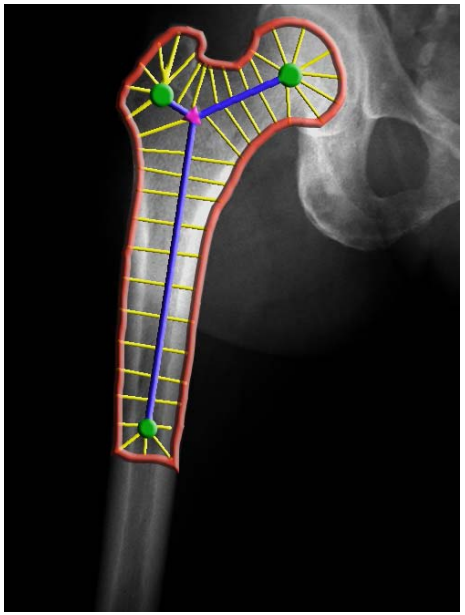


Segmentation



Source: www.google.com

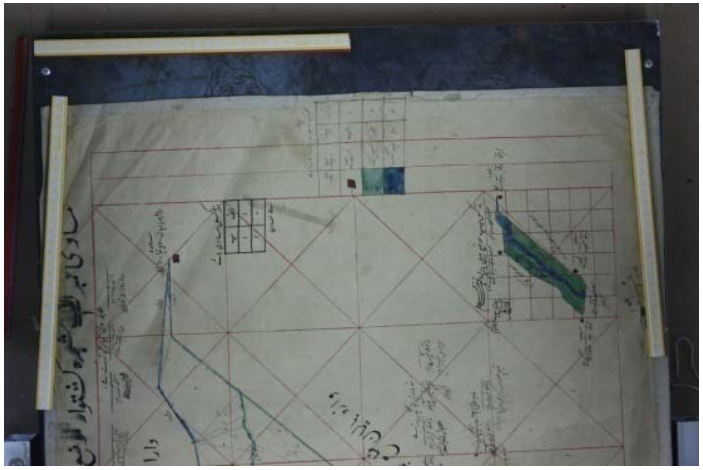
Segmentation

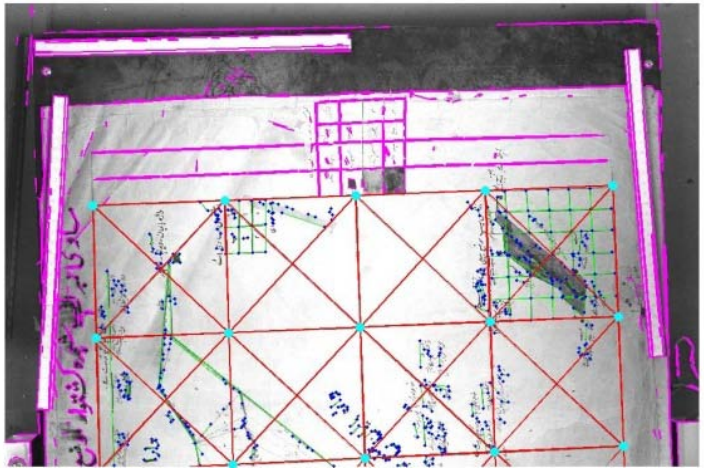


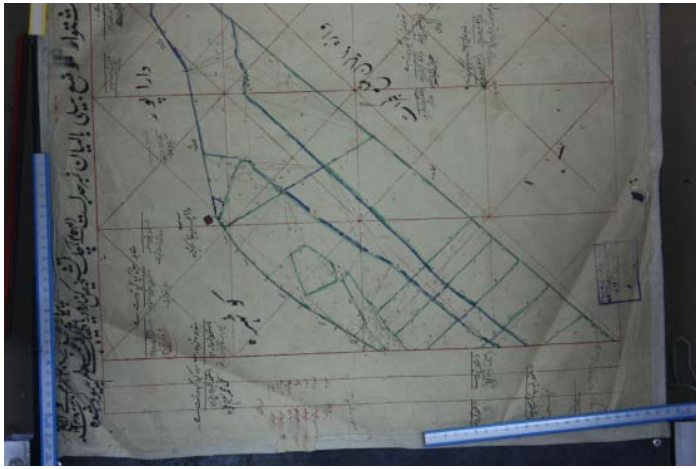
http://lsro.epfl.ch/page-68376-en.html

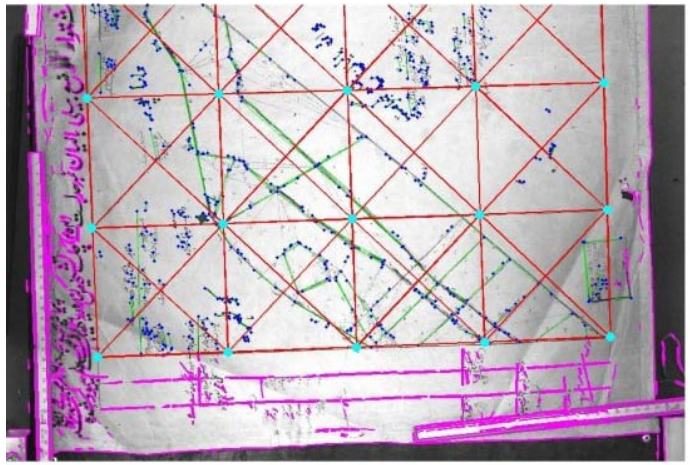
COMPUTER VISION LAB @ PUCIT

Map Processing

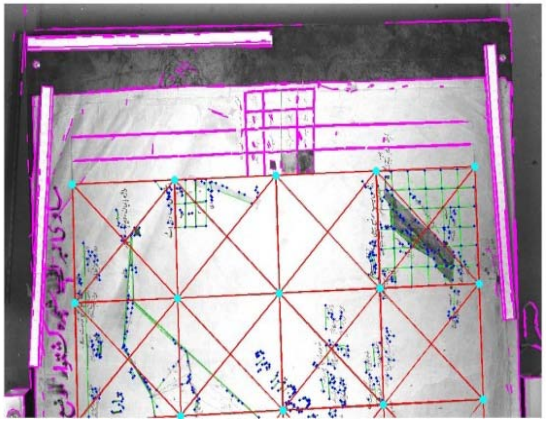


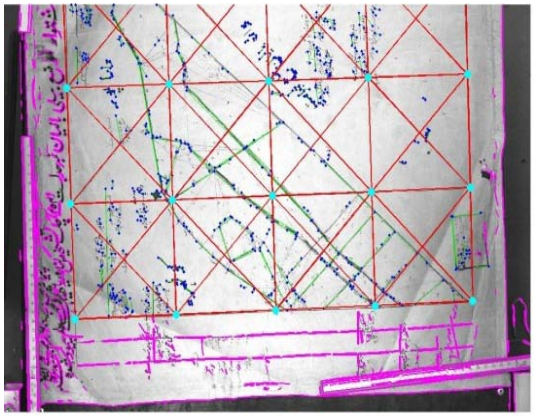


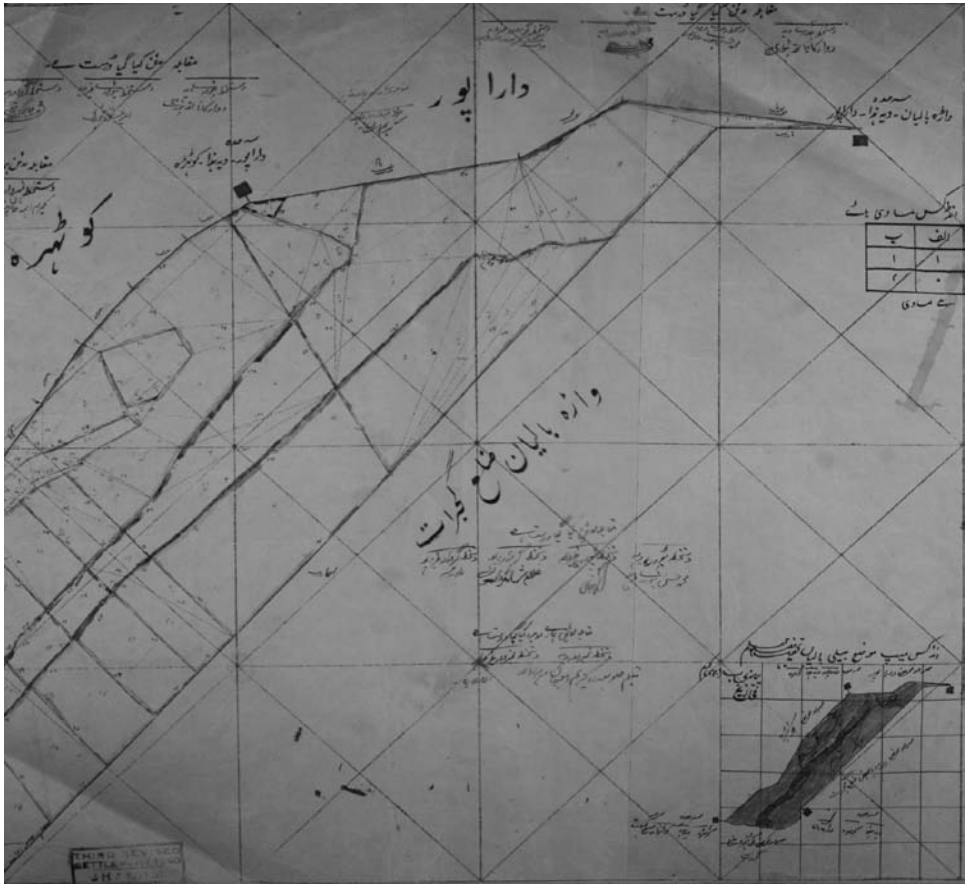




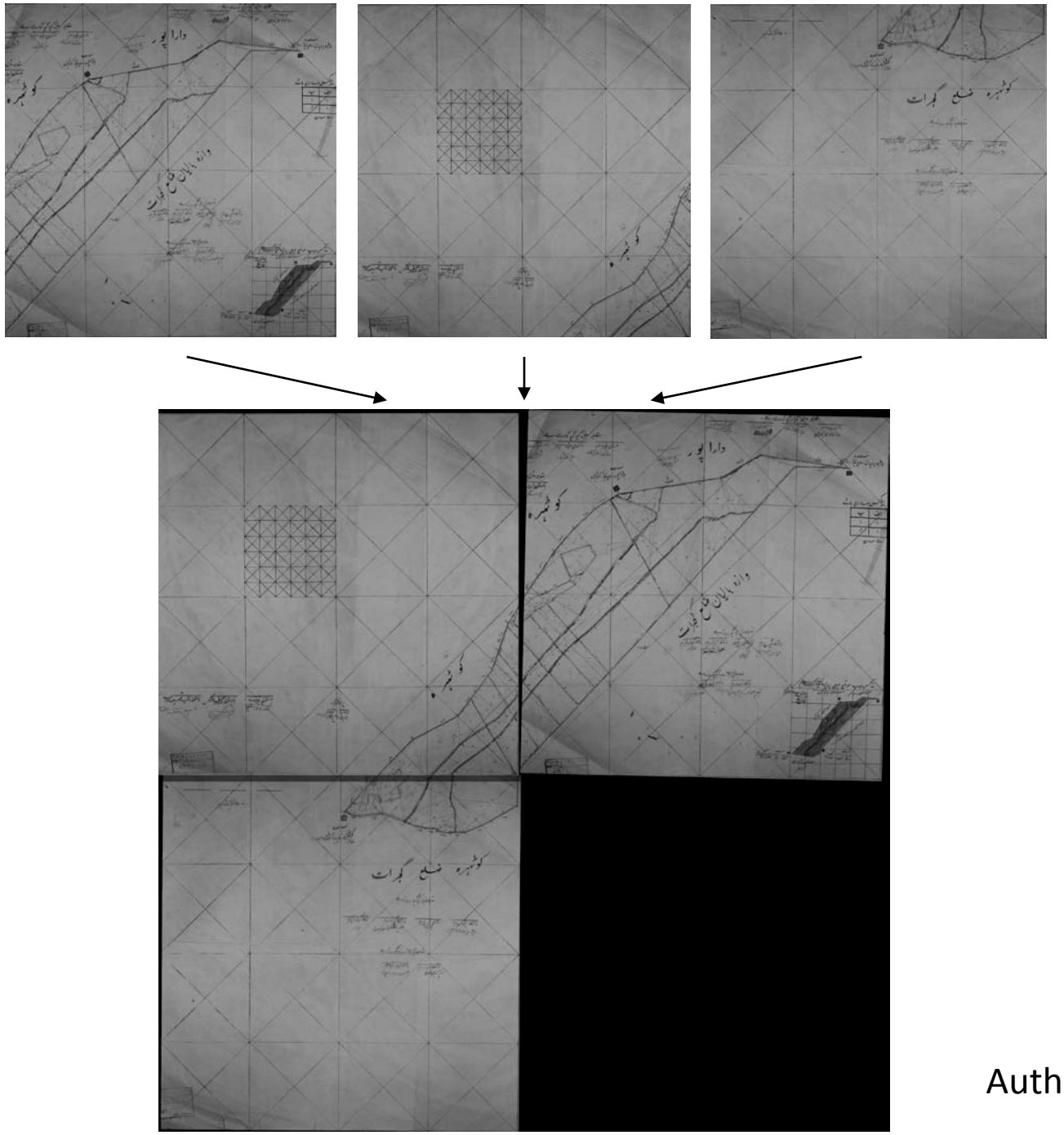
Author: Nazar Khan (2014)



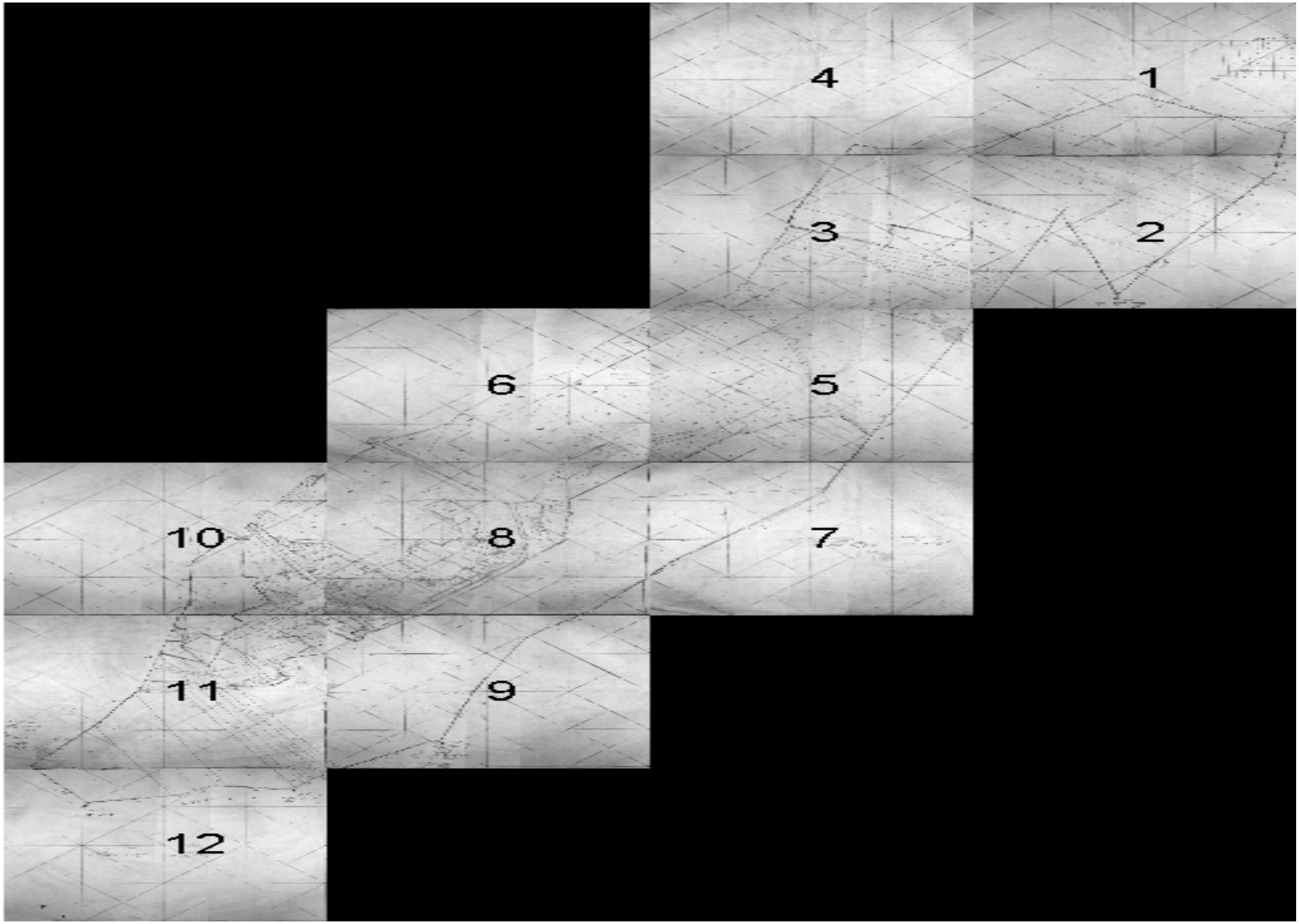




Author: Nazar Khan (2014)

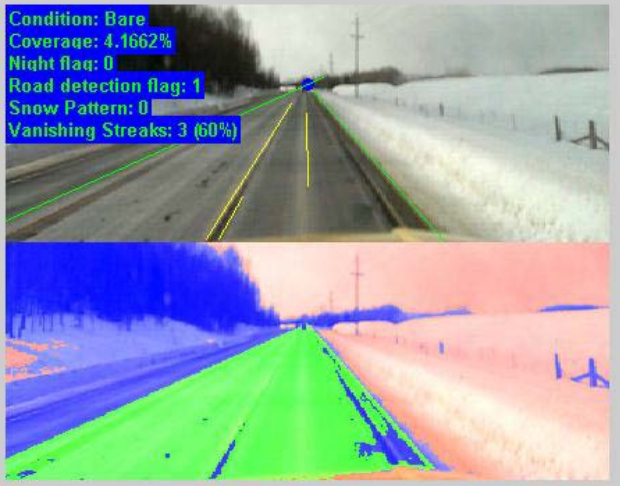


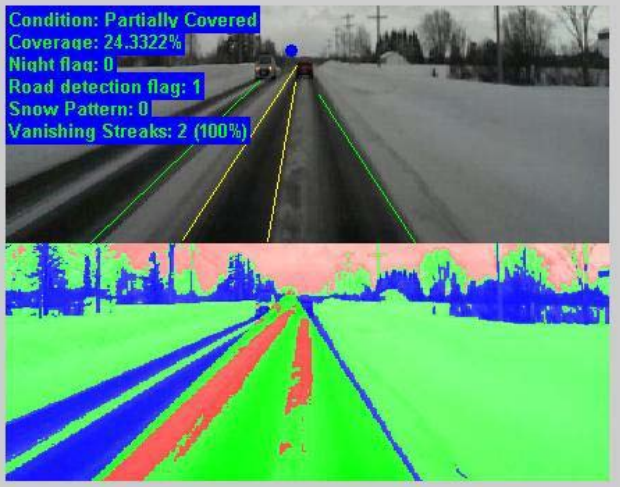
Author: Nazar Khan (2014)

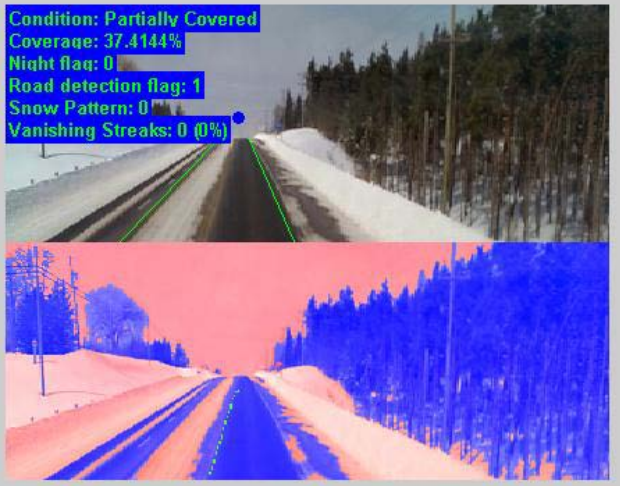


Author: Nazar Khan (2014)

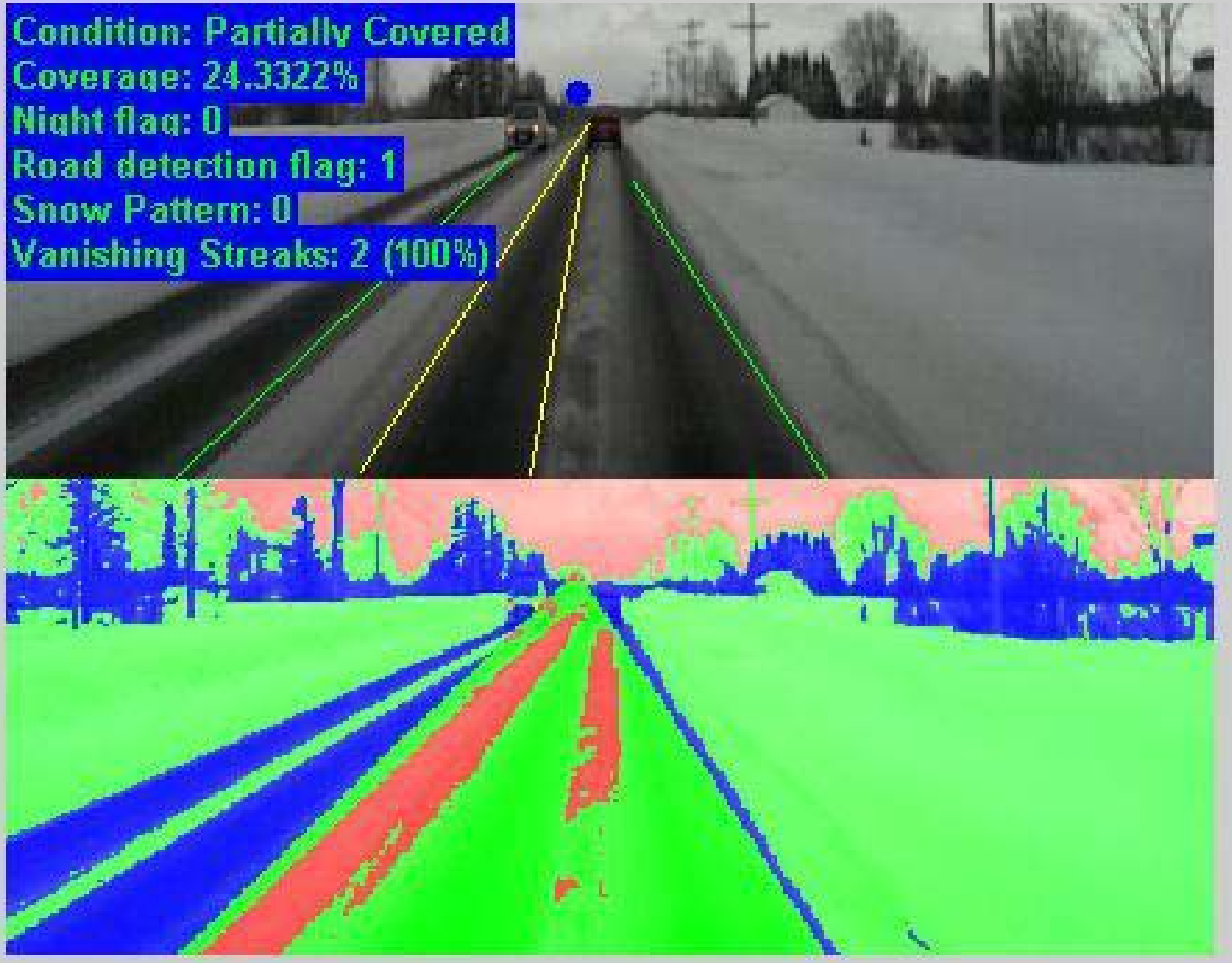
Road Condition Classification

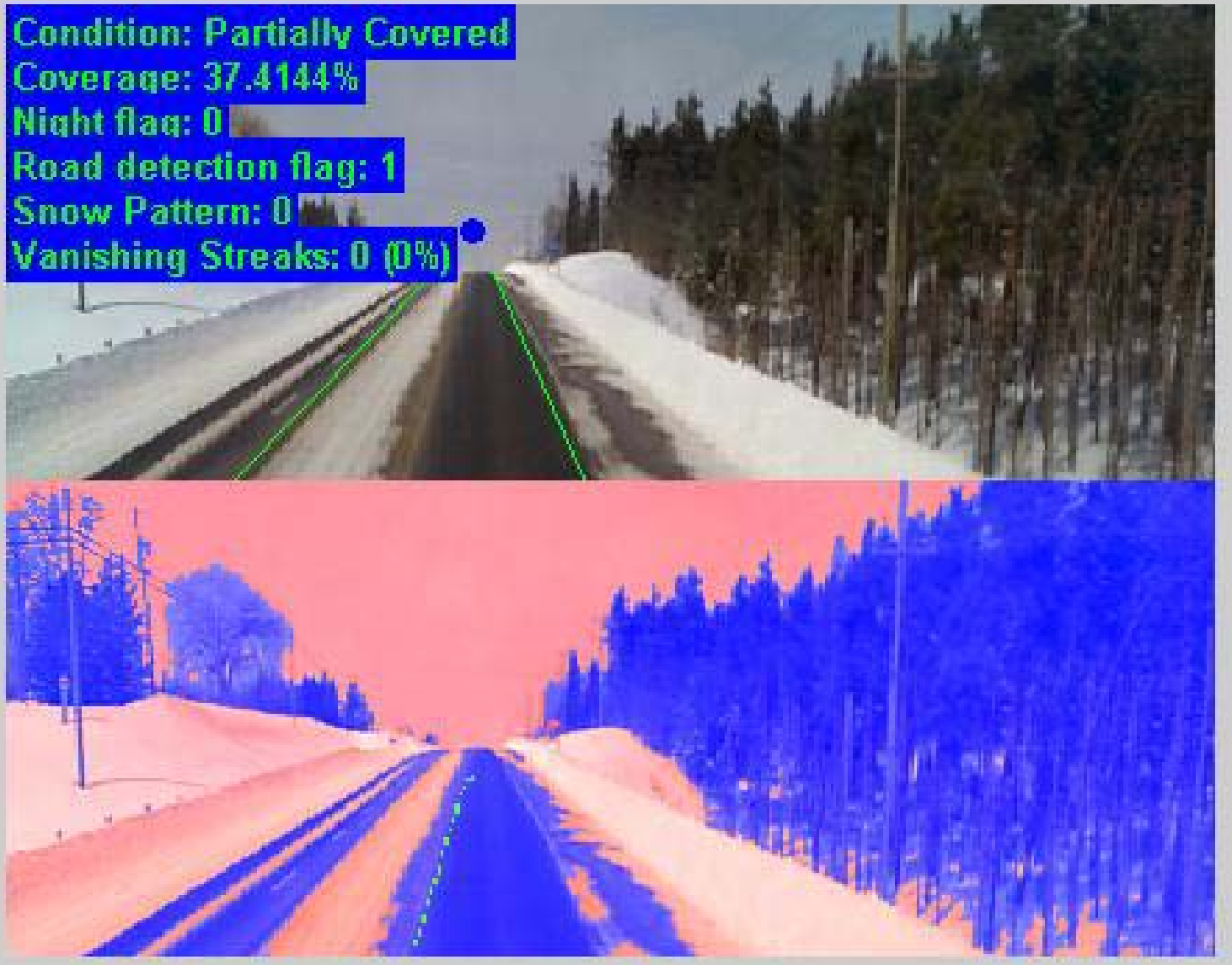




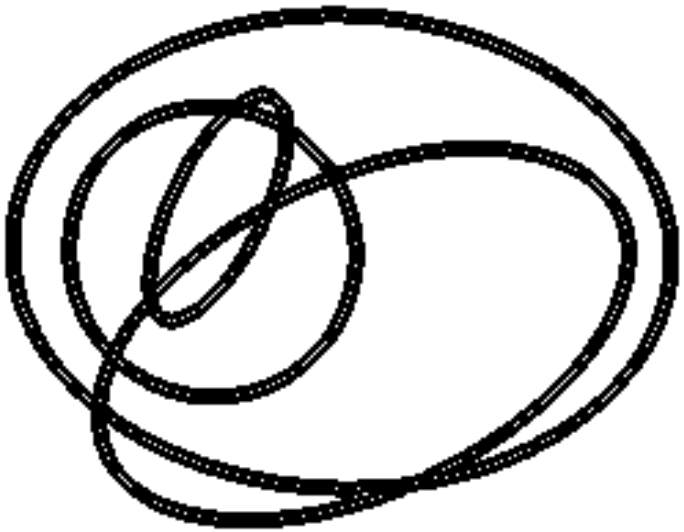


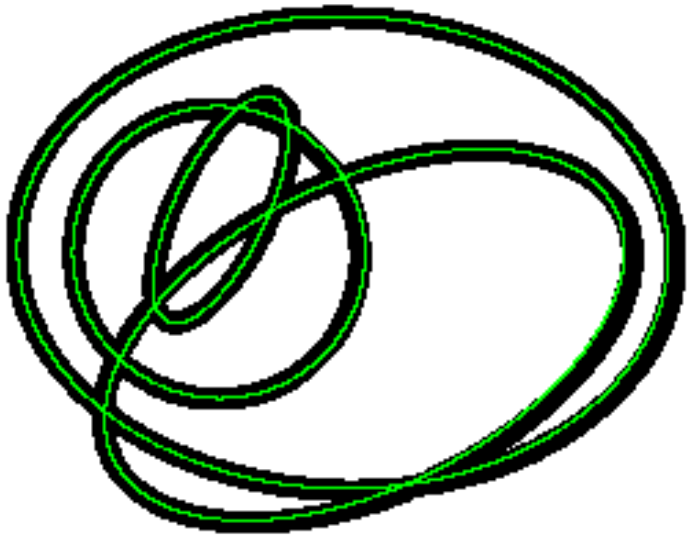
Author:
Naila Hamid,
Nazar Khan
(2014)


Condition: Partially Covered
Coverage: 24.3322%
Night flag: 0
Road detection flag: 1
Snow Pattern: 0
Vanishing Streaks: 2 (100%)


Condition: Partially Covered
Coverage: 37.4144%
Night flag: 0
Road detection flag: 1
Snow Pattern: 0
Vanishing Streaks: 0 (0%)

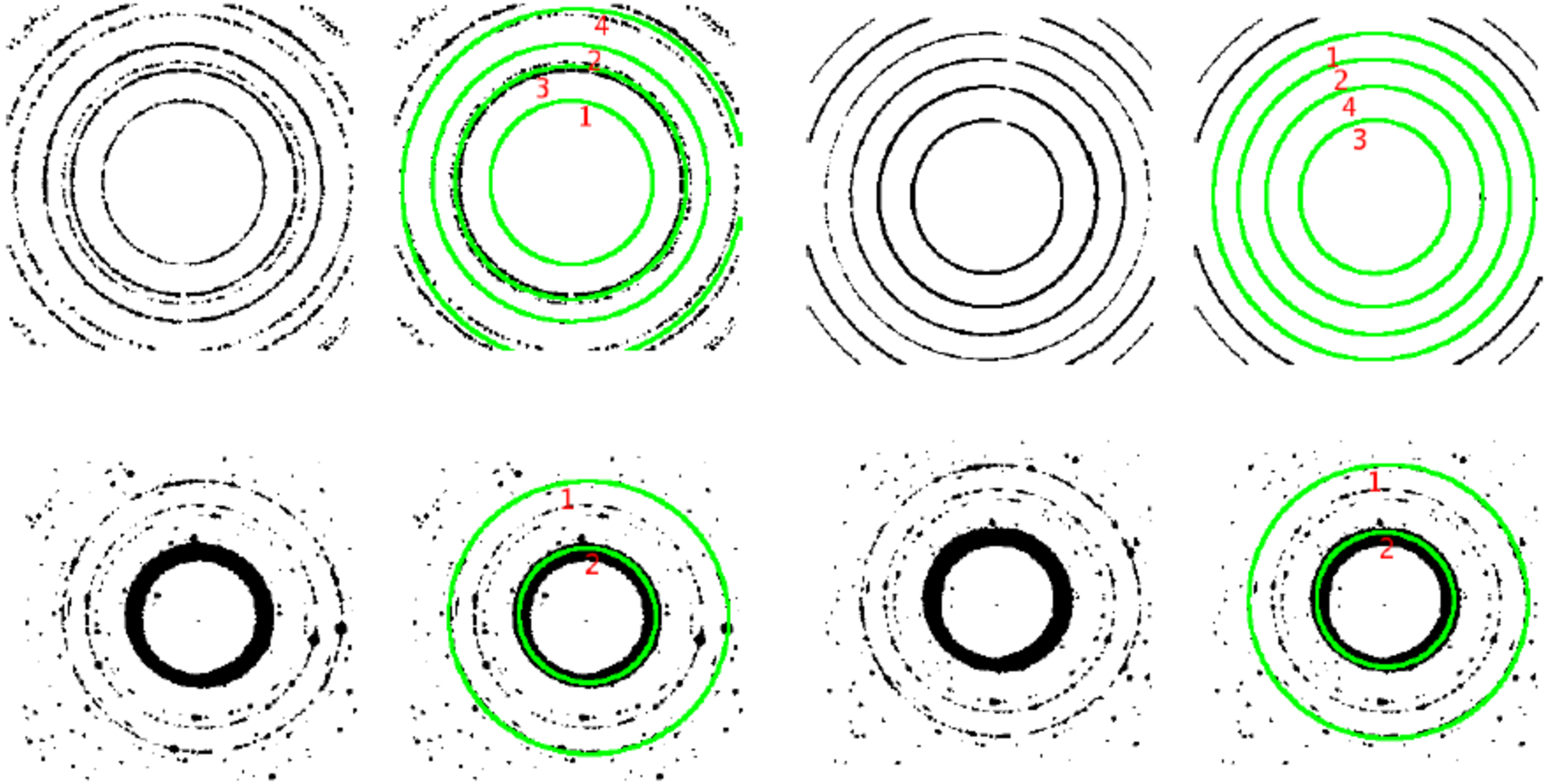
Ellipse Detection





Author: Nazar Khan (2015)

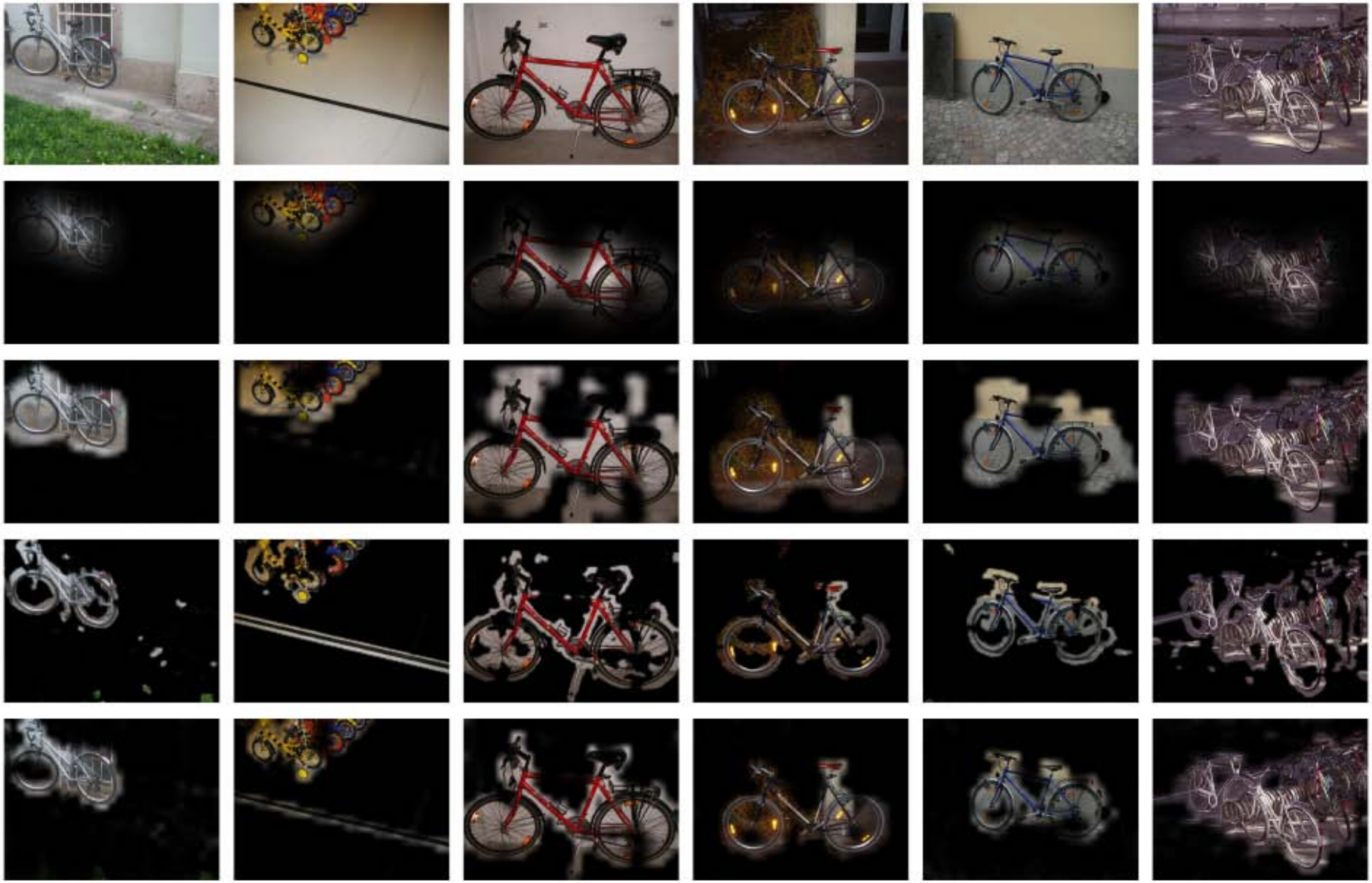
Ellipse Detection



Author: Saadia Shahzad (2015)

Segmentation via Sparse Coding

Author: Nazar Khan (2012)



Author: Nazar Khan (2013)

Expression Mapping via Neural Networks

Author: Nazar Khan (2015)

Undergraduate Projects

- PUCIT SoftExpo ’15
  - Expression Modeling -- Winners (Research)
  - Video Surveillance -- Runners-up (Development)
- FAST Softec ’15
  - Automatic Attention Analyser -- Runners-up
- Automated Exam Checking
  - Your entrance exam was checked by a Computer Vision system developed by Kashif Murtaza and his undergraduate students.
- Book digitisation
- Automated presentation builder
- …