Introduction to Computer Vision

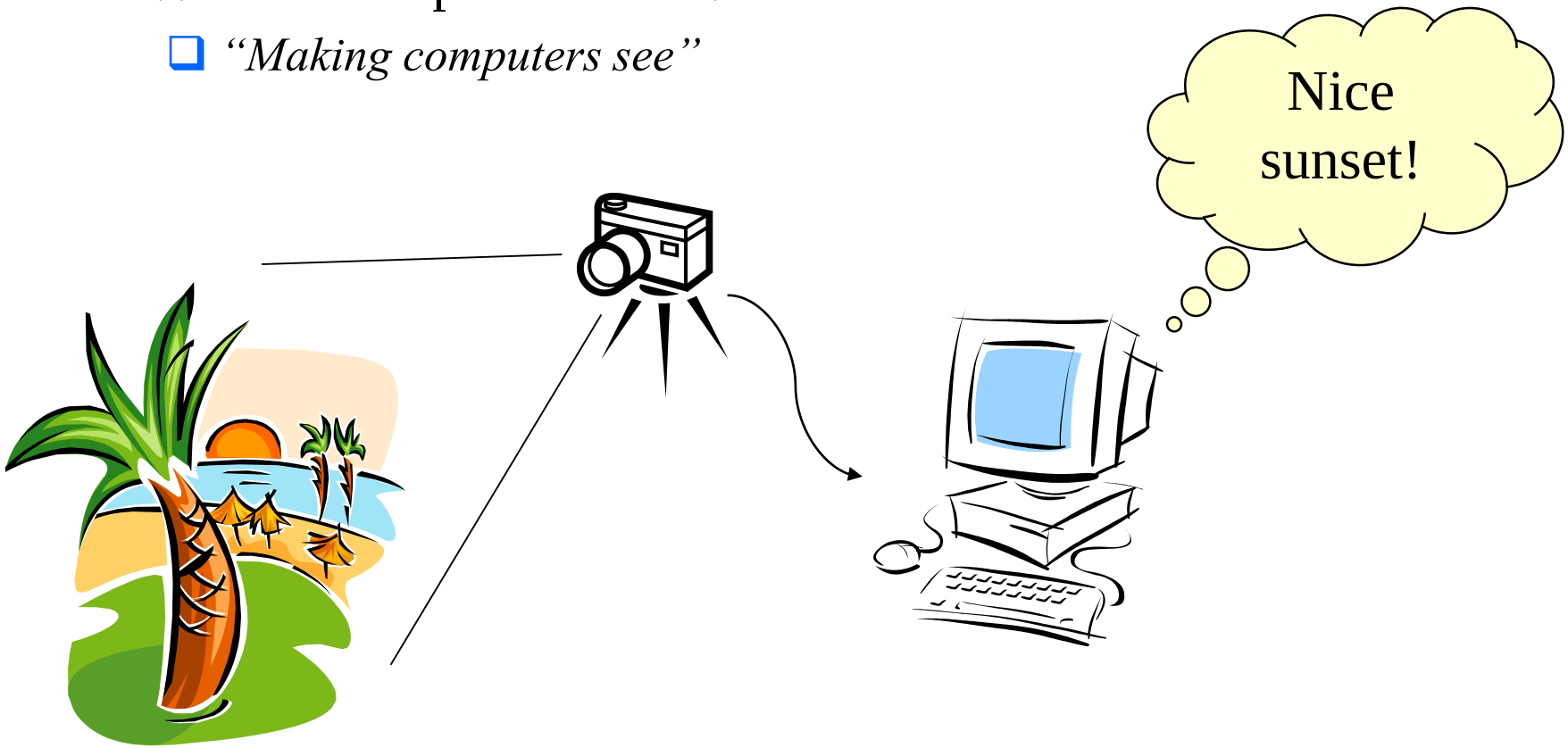
CS / ECE 181B

Prof. Yuan-Fang Wang

Computer Science Dept.

Computer Vision

- What is computer vision?
  - *"Making computers see"*



"Extracting descriptions of the world from pictures or sequences of pictures"

Computer Vision

- Also known as ***image understanding***, ***machine vision***, ***computational vision***

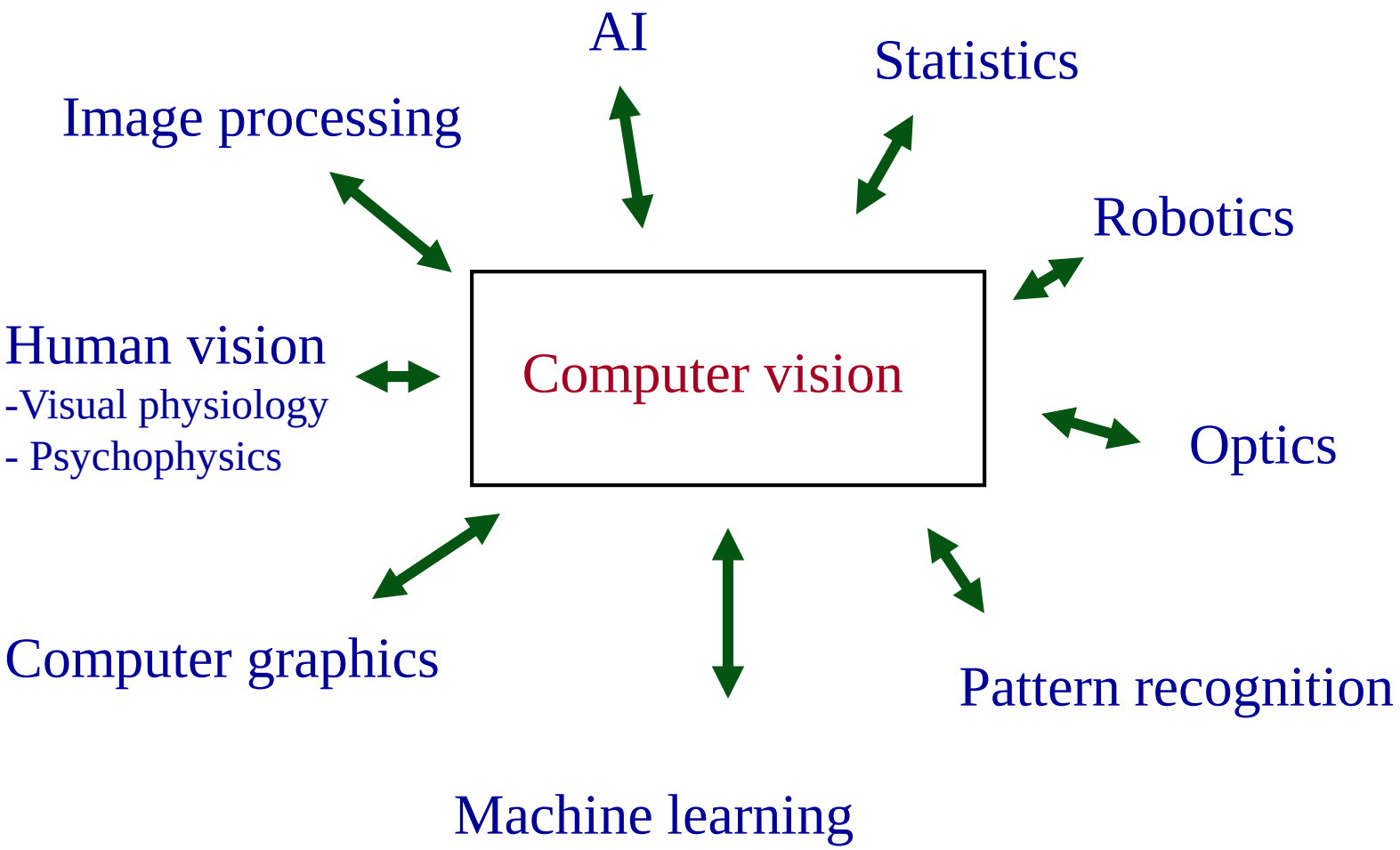
- CV is about interpreting the *content* of images and videos
  - Field is 35-40 years old

- Vision is easy, right? Just open your eyes!
  - No, it's a hard problem!
  - Much of your very complex brain is devoted to doing vision
  - It involves cognition, navigation, manipulation and learning
    - Not just simple "match a feature vector to a database" tasks

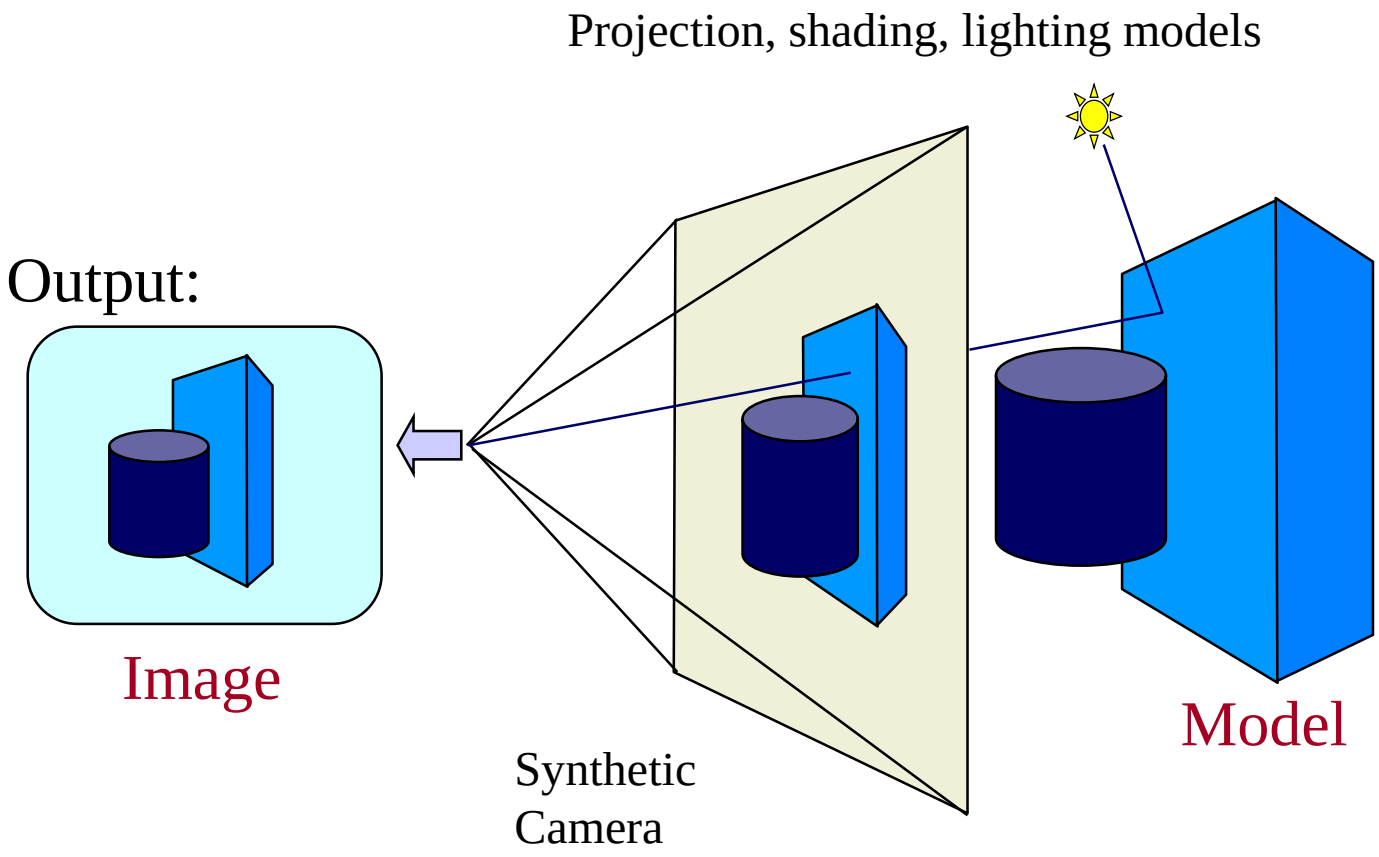
Why Difficult?

- Visual perception comes naturally for human beings
- We know we can do it
- But do you know *how we go about doing it*?
- *Observation:* things that are hard (e.g. chess playing) for humans may be easy for machines, things that are easy (e.g. vision) for humans may be hard for machines

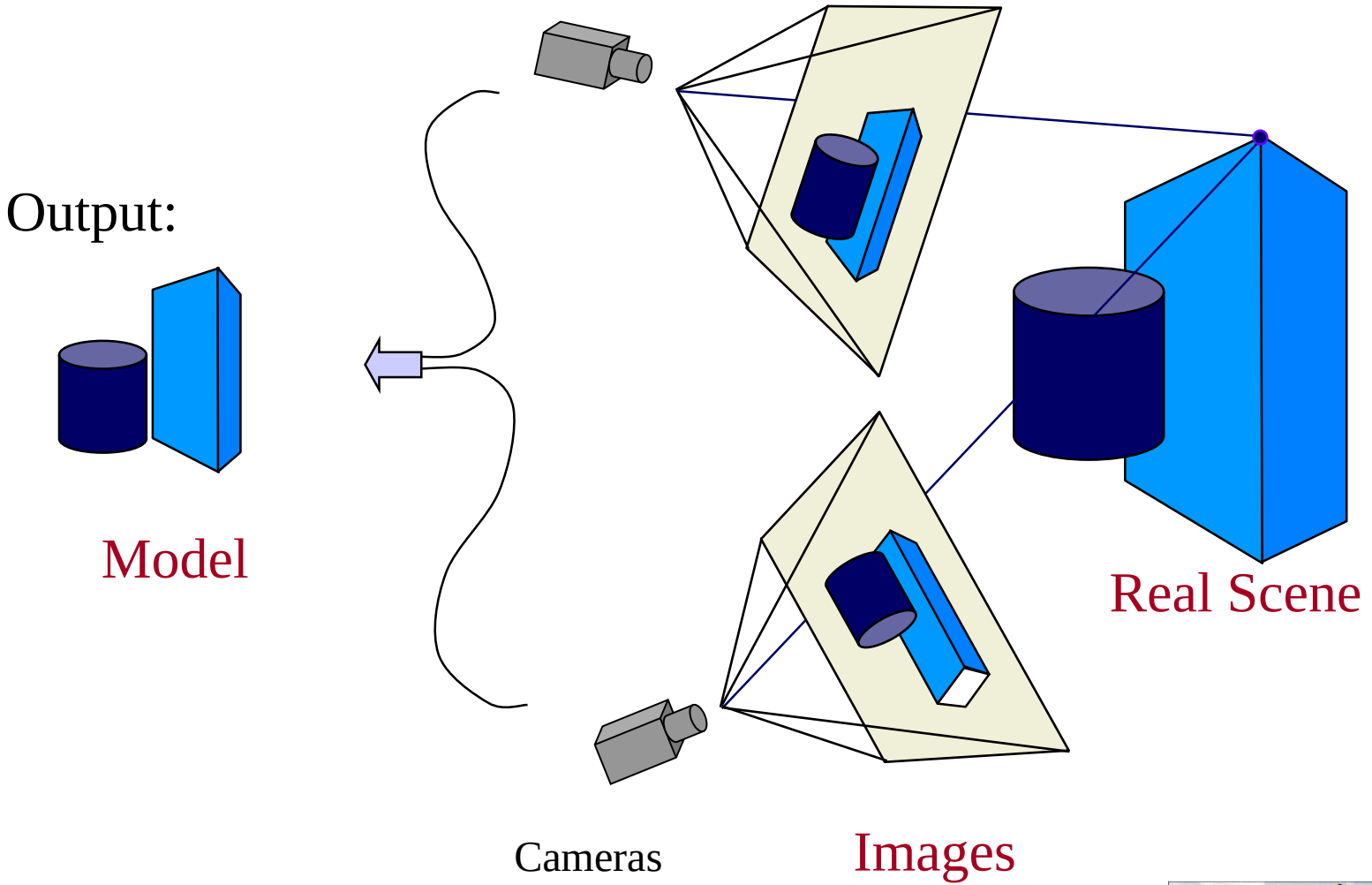
Relation to other fields



Computer Graphics - Synthesis



Computer Vision -Analysis



Computer Vision

- Imaging process:

$$image = f\,(world)$$

including lighting, camera location and parameters, ...

- Computer vision

$$world = f^{-1}\,(image)$$

So a goal of computer vision is to recover the inverse function that creates *images* from the *world*

Is this possible? Why or why not?

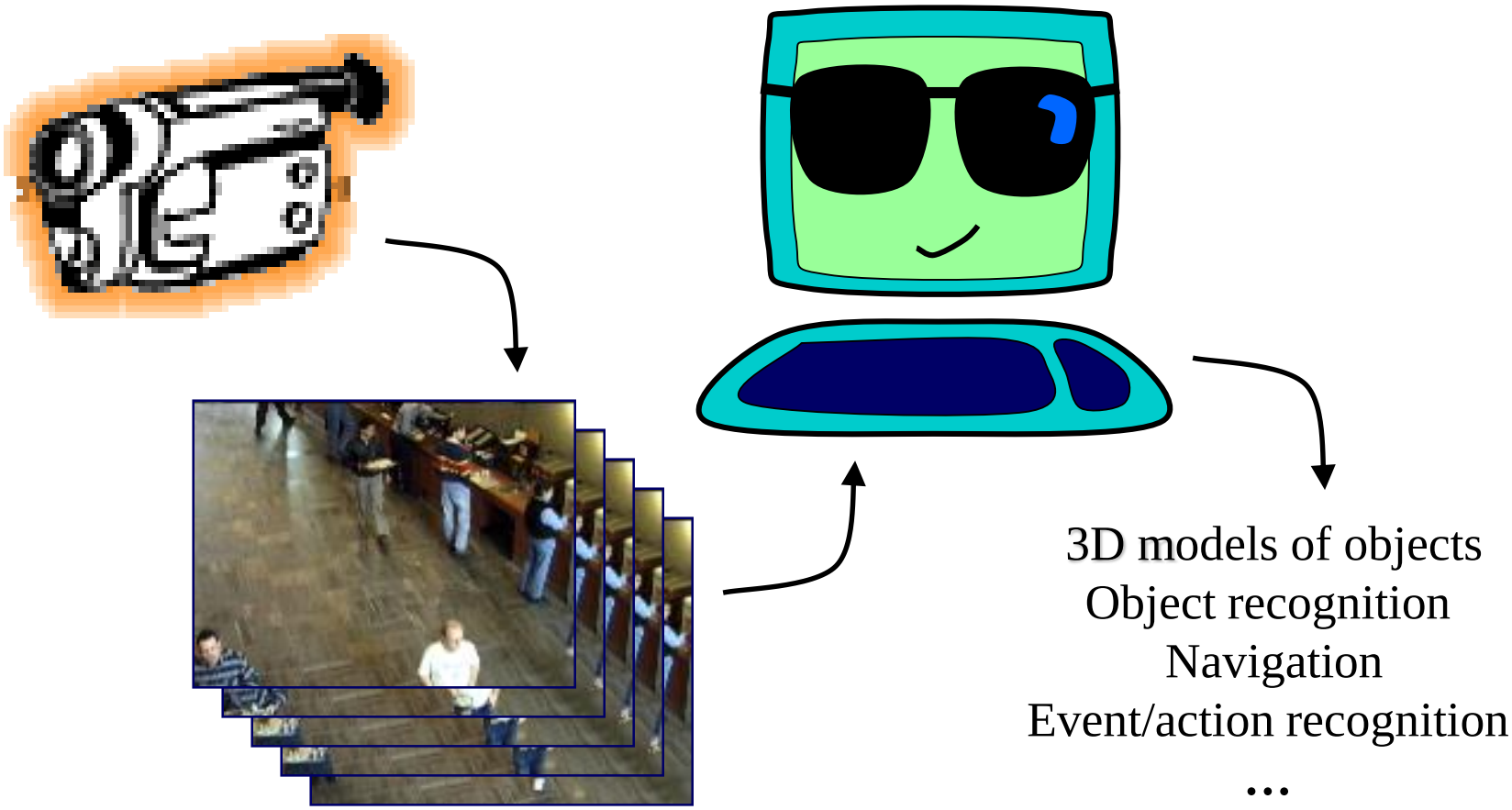
Aims of Computer Vision

- Automate visual perception
- Construct scene descriptions from images
- Make useful decisions about real physical objects and scenes based on sensed images
- Produce symbolic (perhaps task-dependent) descriptions from images
- Produce from images of the external world a useful description that is not cluttered with irrelevant information
- Support tasks that require visual information

Some applications of computer vision

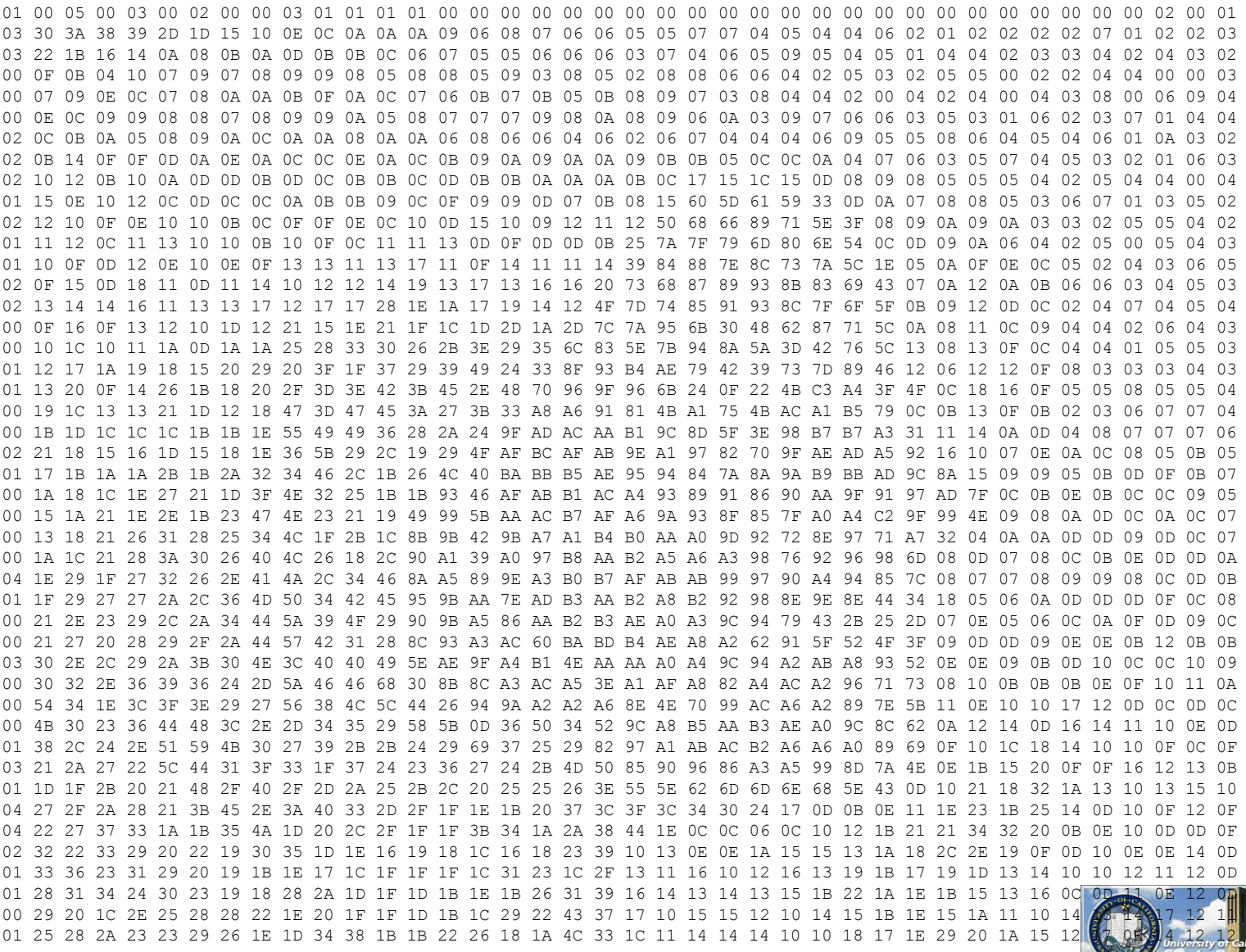
- Photogrammetry, GIS
  - Commercial, military, government
- Robotics
  - Industrial, military, medical, space, entertainment
- Inspection, measurement
- Medical imaging
  - Automatic detection, outlining, measurement
- Graphics and animation, special effects
- Surveillance and security
- Multimedia database indexing and retrieval, compression
- Human-computer interaction (HCI)

What Does Computer Vision Do?



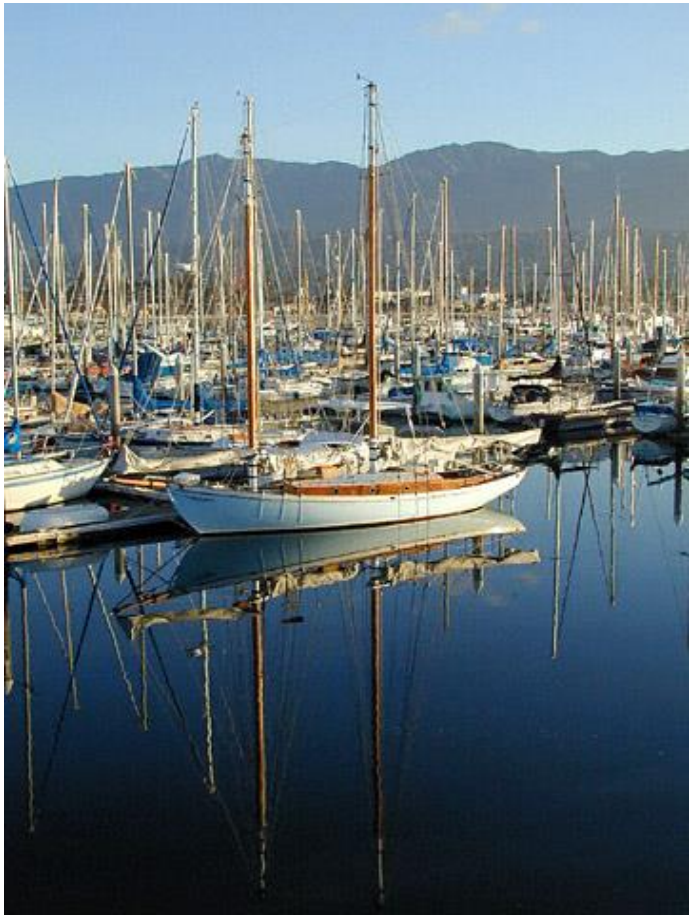
3D models of objects
Object recognition
Navigation
Event/action recognition
…

Vision Transforms From This…

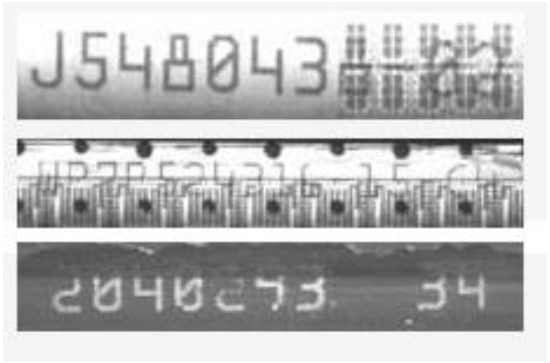


To this…

- Objects
  - Cat, chair, window, star, bush, water, a shoe, my mother…
- Properties
  - Big, bright, yellow, fast, graspable, moving…
- Relations
  - In front, behind, on top, next to, larger, closer, identical…
- Shapes
  - Round, rectangular, star-shaped, symmetric…
- Textures
  - Rough, smooth, irregular…
- Movement
  - Turning, looming, rolling…

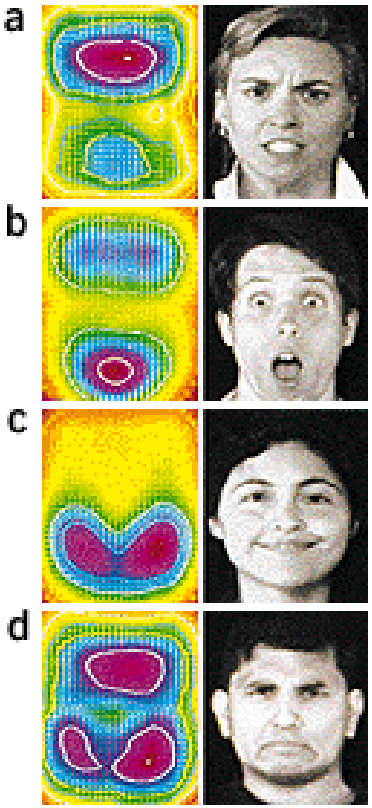


- “A harbor with many dozens of boats; water is calm and glassy; masts are all vertical; mountains in background, blue sky with a touch of clouds…”

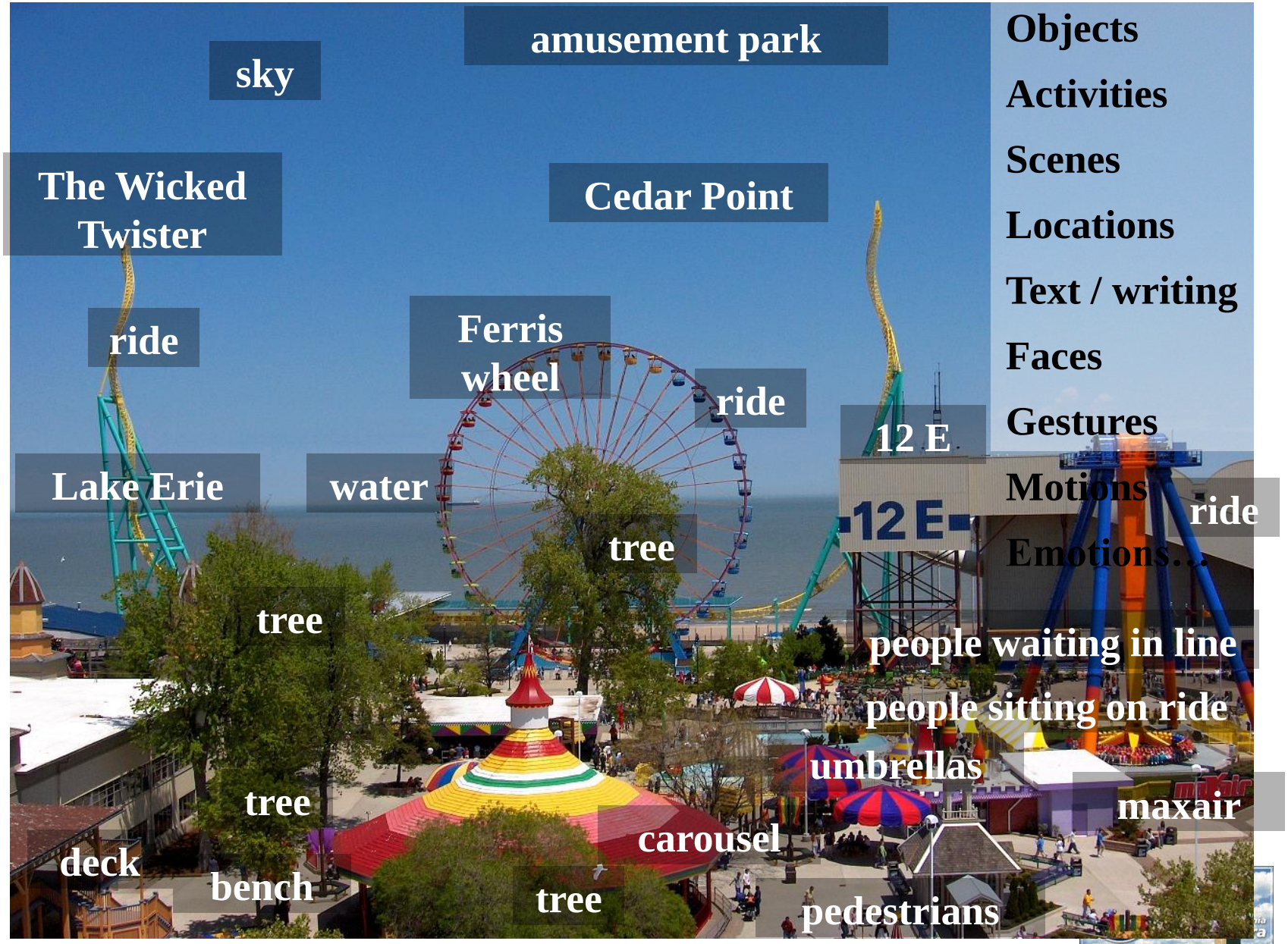


- “J548043”

- “Hallway straight ahead”



- "Angry
  Surprised
  Happy
  Disgusted"


amusement park
sky
The Wicked Twister
Cedar Point
ride
Ferris wheel
ride
12 E
Lake Erie
water
tree
ride
tree
people waiting in line
people sitting on ride
umbrellas
tree
maxair
carousel
deck
bench
tree
pedestrians
Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions…

Why is Vision Difficult?

Consider the input...

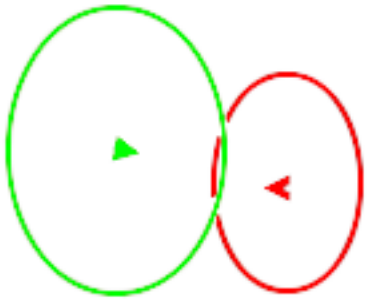
Not this

But this…

```
01 00 05 00 03 00 02 00 00 03 01 01 01 01 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 02 00 01
03 30 3A 38 39 2D 1D 15 10 0E 0C 0A 0A 0A 09 06 08 07 06 06 05 05 07 07 04 05 04 04 06 02 01 02 02 02 02 07 01 02 02 03
03 22 1B 16 14 0A 08 0B 0A 0D 0B 0B 0C 06 07 05 05 06 06 06 03 07 04 06 05 09 05 04 05 01 04 04 02 03 03 04 02 04 03 02
00 0F 0B 04 10 07 09 07 08 09 09 08 05 08 08 05 09 03 08 05 02 08 08 06 06 04 02 05 03 02 05 05 00 02 02 04 04 00 00 03
00 07 09 0E 0C 07 08 0A 0A 0B 0F 0A 0C 07 06 0B 07 0B 05 0B 08 09 07 03 08 04 04 02 00 04 02 04 00 04 03 08 00 06 09 04
00 0E 0C 09 09 08 08 07 08 09 09 0A 05 08 07 07 07 09 08 0A 08 09 06 0A 03 09 07 06 06 03 05 03 01 06 02 03 07 01 04 04
02 0C 0B 0A 05 08 09 0A 0C 0A 0A 08 0A 0A 06 08 06 06 04 06 02 06 07 04 04 04 06 09 05 05 08 06 04 05 04 06 01 0A 03 02
02 0B 14 0F 0F 0D 0A 0E 0A 0C 0C 0E 0A 0C 0B 09 0A 09 0A 0A 09 0B 0B 05 0C 0C 0A 04 07 06 03 05 07 04 05 03 02 01 06 03
02 10 12 0B 10 0A 0D 0D 0B 0D 0C 0B 0B 0C 0D 0B 0B 0A 0A 0A 0B 0C 17 15 1C 15 0D 08 09 08 05 05 05 04 02 05 04 04 00 04
01 15 0E 10 12 0C 0D 0C 0C 0A 0B 0B 09 0C 0F 09 09 0D 07 0B 08 15 60 5D 61 59 33 0D 0A 07 08 08 05 03 06 07 01 03 05 02
02 12 10 0F 0E 10 10 0B 0C 0F 0F 0E 0C 10 0D 15 10 09 12 11 12 50 68 66 89 71 5E 3F 08 09 0A 09 0A 03 03 02 05 05 04 02
01 11 12 0C 11 13 10 10 0B 10 0F 0C 11 11 13 0D 0F 0D 0D 0B 25 7A 7F 79 6D 80 6E 54 0C 0D 09 0A 06 04 02 05 00 05 04 03
01 10 0F 0D 12 0E 10 0E 0F 13 13 11 13 17 11 0F 14 11 11 14 39 84 88 7E 8C 73 7A 5C 1E 05 0A 0F 0E 0C 05 02 04 03 06 05
02 0F 15 0D 18 11 0D 11 14 10 12 12 14 19 13 17 13 16 16 20 73 68 87 89 93 8B 83 69 43 07 0A 12 0A 0B 06 06 03 04 05 03
02 13 14 14 16 11 13 13 17 12 17 17 28 1E 1A 17 19 14 12 4F 7D 74 85 91 93 8C 7F 6F 5F 0B 09 12 0D 0C 02 04 07 04 05 04
00 0F 16 0F 13 12 10 1D 12 21 15 1E 21 1F 1C 1D 2D 1A 2D 7C 7A 95 6B 30 48 62 87 71 5C 0A 08 11 0C 09 04 04 02 06 04 03
00 10 1C 10 11 1A 0D 1A 1A 25 28 33 30 26 2B 3E 29 35 6C 83 5E 7B 94 8A 5A 3D 42 76 5C 13 08 13 0F 0C 04 04 01 05 05 03
01 12 17 1A 19 18 15 20 29 20 3F 1F 37 29 39 49 24 33 8F 93 B4 AE 79 42 39 73 7D 89 46 12 06 12 12 0F 08 03 03 03 04 03
01 13 20 0F 14 26 1B 18 20 2F 3D 3E 42 3B 45 2E 48 70 96 9F 96 6B 24 0F 22 4B C3 A4 3F 4F 0C 18 16 0F 05 05 08 05 05 04
00 19 1C 13 13 21 1D 12 18 47 3D 47 45 3A 27 3B 33 A8 A6 91 81 4B A1 75 4B AC A1 B5 79 0C 0B 13 0F 0B 02 03 06 07 07 04
00 1B 1D 1C 1C 1C 1B 1B 1E 55 49 49 36 28 2A 24 9F AD AC AA B1 9C 8D 5F 3E 98 B7 B7 A3 31 11 14 0A 0D 04 08 07 07 07 06
02 21 18 15 16 1D 15 18 1E 36 5B 29 2C 19 29 4F AF BC AF AB 9E A1 97 82 70 9F AE AD A5 92 16 10 07 0E 0A 0C 08 05 0B 05
01 17 1B 1A 1A 2B 1B 2A 32 34 46 2C 1B 26 4C 40 BA BB B5 AE 95 94 84 7A 8A 9A B9 BB AD 9C 8A 15 09 09 05 0B 0D 0F 0B 07
00 1A 18 1C 1E 27 21 1D 3F 4E 32 25 1B 1B 93 46 AF AB B1 AC A4 93 89 91 86 90 AA 9F 91 97 AD 7F 0C 0B 0E 0B 0C 0C 09 05
00 15 1A 21 1E 2E 1B 23 47 4E 23 21 19 49 99 5B AA AC B7 AF A6 9A 93 8F 85 7F A0 A4 C2 9F 99 4E 09 08 0A 0D 0C 0A 0C 07
00 13 18 21 26 31 28 25 34 4C 1F 2B 1C 8B 9B 42 9B A7 A1 B4 B0 AA A0 9D 92 72 8E 97 71 A7 32 04 0A 0A 0D 0D 09 0D 0C 07
00 1A 1C 21 28 3A 30 26 40 4C 26 18 2C 90 A1 39 A0 97 B8 AA B2 A5 A6 A3 98 76 92 96 98 6D 08 0D 07 08 0C 0B 0E 0D 0D 0A
04 1E 29 1F 27 32 26 2E 41 4A 2C 34 46 8A A5 89 9E A3 B0 B7 AF AB AB 99 97 90 A4 94 85 7C 08 07 07 08 09 09 08 0C 0D 0B
01 1F 29 27 27 2A 2C 36 4D 50 34 42 45 95 9B AA 7E AD B3 AA B2 A8 B2 92 98 8E 9E 8E 44 34 18 05 06 0A 0D 0D 0D 0F 0C 08
00 21 2E 23 29 2C 2A 34 44 5A 39 4F 29 90 9B A5 86 AA B2 B3 AE A0 A3 9C 94 79 43 2B 25 2D 07 0E 05 06 0C 0A 0F 0D 09 0C
00 21 27 20 28 29 2F 2A 44 57 42 31 28 8C 93 A3 AC 60 BA BD B4 AE A8 A2 62 91 5F 52 4F 3F 09 0D 0D 09 0E 0E 0B 12 0B 0B
03 30 2E 2C 29 2A 3B 30 4E 3C 40 40 49 5E AE 9F A4 B1 4E AA AA A0 A4 9C 94 A2 AB A8 93 52 0E 0E 09 0B 0D 10 0C 0C 10 09
00 30 32 2E 36 39 36 24 2D 5A 46 46 68 30 8B 8C A3 AC A5 3E A1 AF A8 82 A4 AC A2 96 71 73 08 10 0B 0B 0B 0E 0F 10 11 0A
00 54 34 1E 3C 3F 3E 29 27 56 38 4C 5C 44 26 94 9A A2 A2 A6 8E 4E 70 99 AC A6 A2 89 7E 5B 11 0E 10 10 17 12 0D 0C 0D 0C
00 4B 30 23 36 44 48 3C 2E 2D 34 35 29 58 5B 0D 36 50 34 52 9C A8 B5 AA B3 AE A0 9C 8C 62 0A 12 14 0D 16 14 11 10 0E 0D
01 38 2C 24 2E 51 59 4B 30 27 39 2B 2B 24 29 69 37 25 29 82 97 A1 AB AC B2 A6 A6 A0 89 69 0F 10 1C 18 14 10 10 0F 0C 0F
03 21 2A 27 22 5C 44 31 3F 33 1F 37 24 23 36 27 24 2B 4D 50 85 90 96 86 A3 A5 99 8D 7A 4E 0E 1B 15 20 0F 0F 16 12 13 0B
01 1D 1F 2B 20 21 48 2F 40 2F 2D 2A 25 2B 2C 20 25 25 26 3E 55 5E 62 6D 6D 6E 68 5E 43 0D 10 21 18 32 1A 13 10 13 15 10
04 27 2F 2A 28 21 3B 45 2E 3A 40 33 2D 2F 1F 1E 1B 20 37 3C 3F 3C 34 30 24 17 0D 0B 0E 11 1E 23 1B 25 14 0D 10 0F 12 0F
04 22 27 37 33 1A 1B 35 4A 1D 20 2C 2F 1F 1F 3B 34 1A 2A 38 44 1E 0C 0C 06 0C 10 12 1B 21 21 34 32 20 0B 0E 10 0D 0D 0F
02 32 22 33 29 20 22 19 30 35 1D 1E 16 19 18 1C 16 18 23 39 10 13 0E 0E 1A 15 15 13 1A 18 2C 2E 19 0F 0D 10 0E 0E 14 0D
01 33 36 23 31 29 20 19 1B 1E 17 1C 1F 1F 1F 1C 31 23 1C 2F 13 11 16 10 12 16 13 19 1B 17 19 1D 13 14 10 10 12 11 12 0D
01 28 31 34 24 30 23 19 18 28 2A 1D 1F 1D 1B 1E 1B 26 31 39 16 14 13 14 13 15 1B 22 1A 1E 1B 15 13 16 0C 0D 11 0E 12 0D
00 29 20 1C 2E 25 28 28 22 1E 20 1F 1F 1D 1B 1C 29 22 43 37 17 10 15 15 12 10 14 15 1B 1E 15 1A 11 10 14 13 14 17 12 11
01 25 28 2A 23 23 29 26 1E 1D 34 38 1B 1B 22 26 18 1A 4C 33 1C 11 14 14 14 10 10 18 17 1E 29 20 1A 15 12 17 0E 14 12 12
02 25 23 21 21 24 27 28 22 1E 2D 2D 23 1D 25 28 27 2A 5F 24 22 15 14 13 19 15 16 15 17 1A 1B 34 29 1B 16 1[illegible] 16 17 1[illegible]
00 24 1F 20 28 22 1B 22 27 20 17 1E 1B 20 22 21 1C 5E 72 23 18 25 16 15 11 0F 17 15 14 14 18 1F 21 1B 16 18 10 13 16 10
02 24 23 25 21 24 21 22 24 28 2F 26 23 1A 1D 16 21 B0 2C 26 22 2C 22 1D 1A 10 1A 1D 1A 13 14 1C 21 1B 17 [illegible] 13 14
```

Some Possible Outputs

depth
or
segmentation



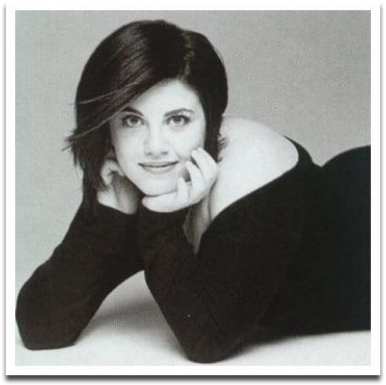
object pose
(facing away,
facing forward)

object
recognition

action
understanding

What Your Brain Does

Variation in Appearance



What makes computer vision hard?

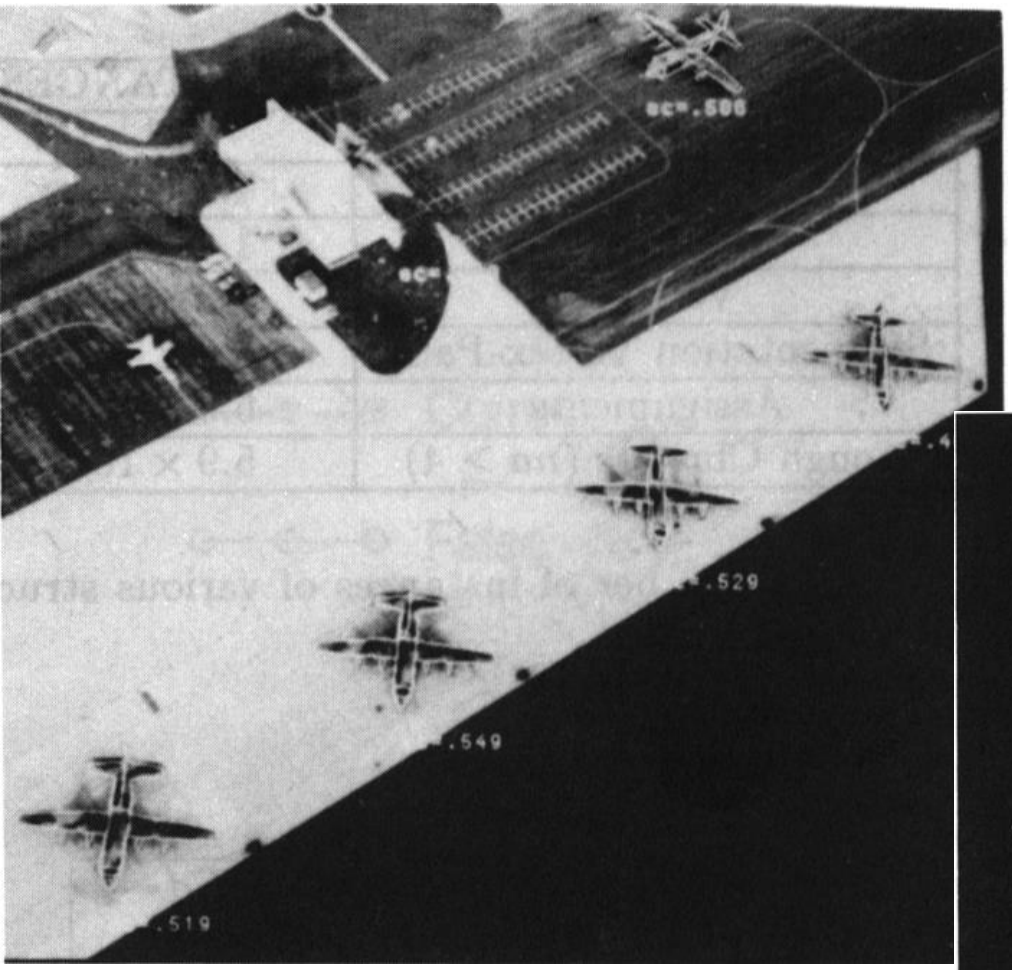
- Underspecified, ill-posed problem!
  - 3D world projected onto 2D sensor(s)
- Environment
  - Lighting, background, movement, camera
- Varying appearance of objects
- Calibration, FOV, camera control, image quality
- Computational complexity (speed of processing)
- Etc., etc….

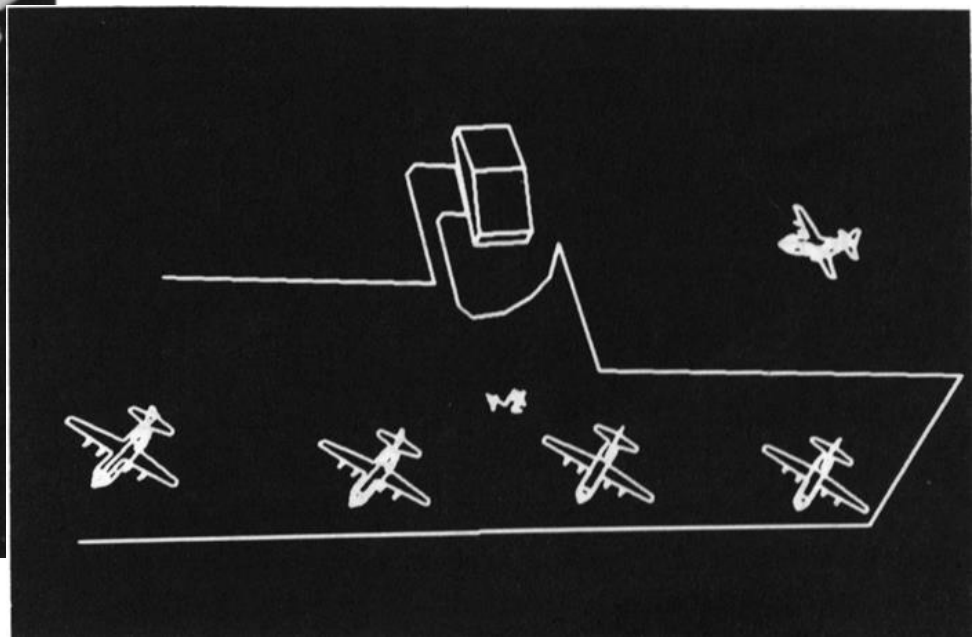
***Robustness* is the primary challenge to computer vision!**

Progress in Computer Vision

- First generation: Military/Early Research
  - Few systems, each custom-built, cost $Ms
  - "Users" have PhDs
  - 1 hour per frame
- Second generation: Industrial/Medical
  - Numerous systems, 1-1000 of each, cost $10Ks
  - Users have college degree
  - RT with special hardware
- Third generation: Consumer
  - 100000(00) systems, cost $100s
  - Users have little or no training
  - RT in software

Early CV research





University of California Santa Barbara

Examples

CMU NavLab

Cognex

MIT Media Lab

More Examples

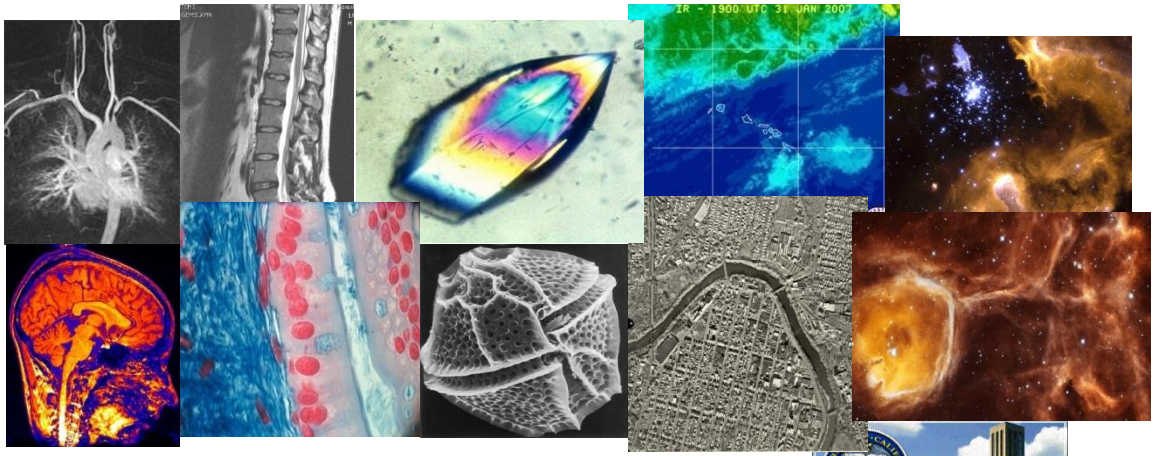
- Images and video are everywhere!

Personal photo albums

Movies, news, sports

Google Image Search™ Picasa™ flickr™ webshots™ picsearch™ YouTube Broadcast Yourself™

Surveillance and security



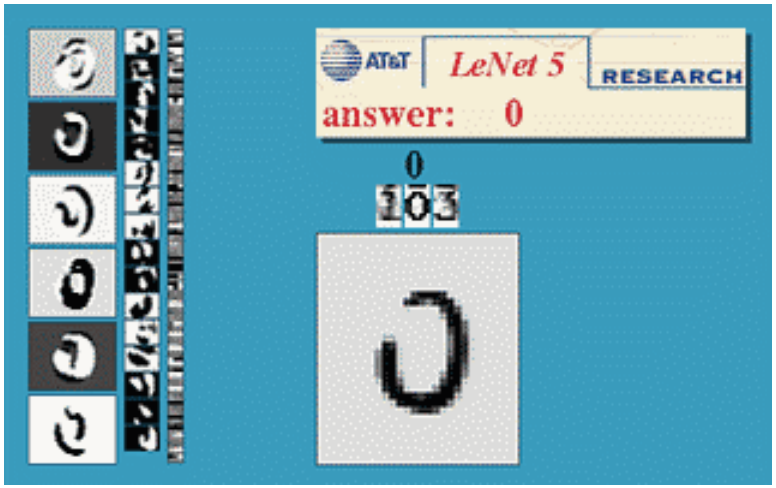
Medical and scientific images

Slide credit, L. Lazebnik

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs

http://www.research.att.com/~yann/

License plate readers

http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

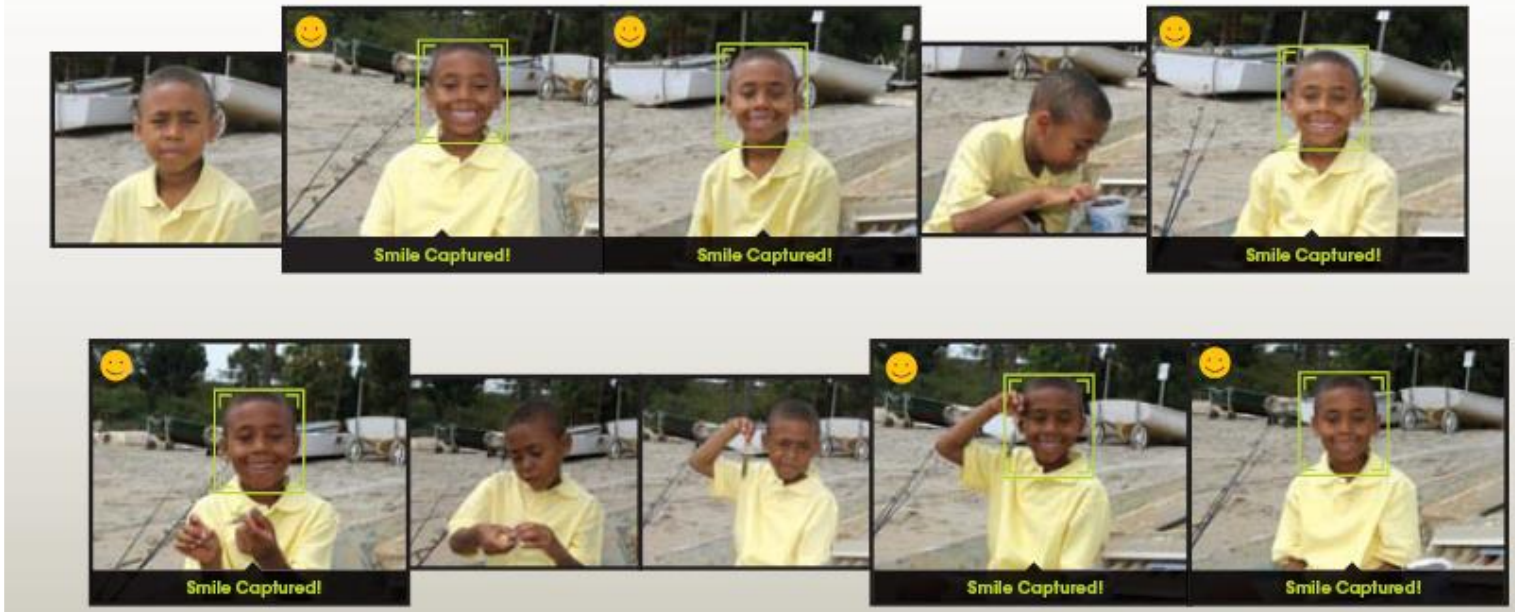
University of California Santa Barbara

Face detection

Smile detection?

The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



Sony Cyber-shot® T70 Digital Still Camera

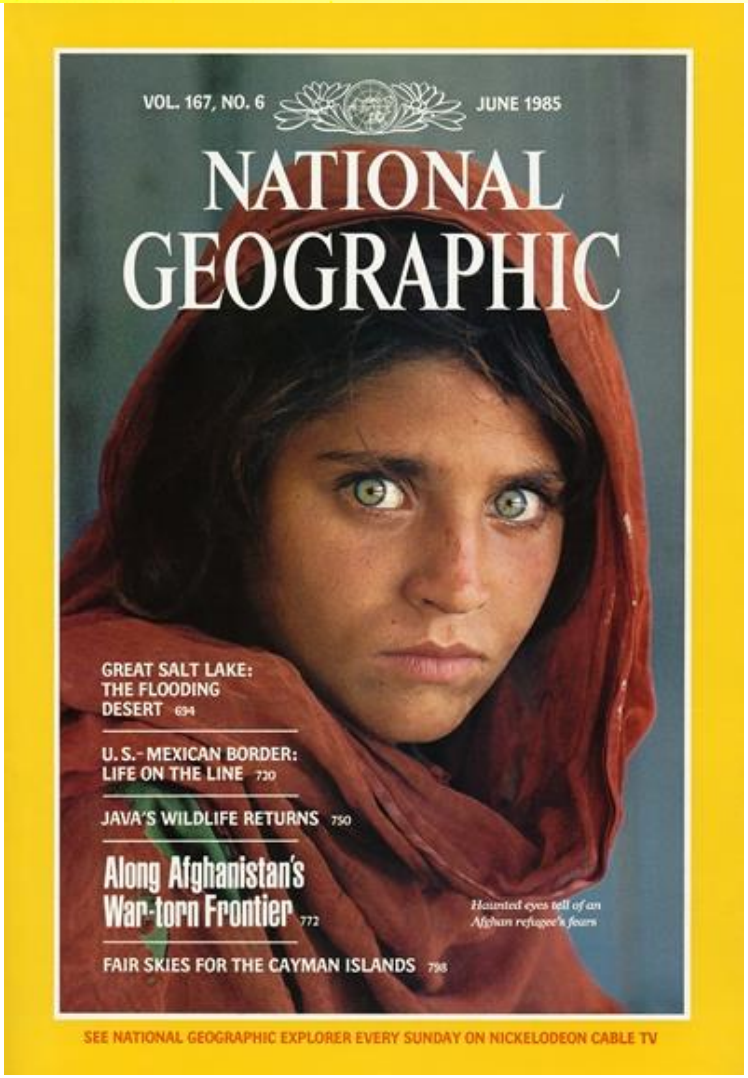
University of California Santa Barbara

Object recognition (in supermarkets)

LaneHawk by EvolutionRobotics

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

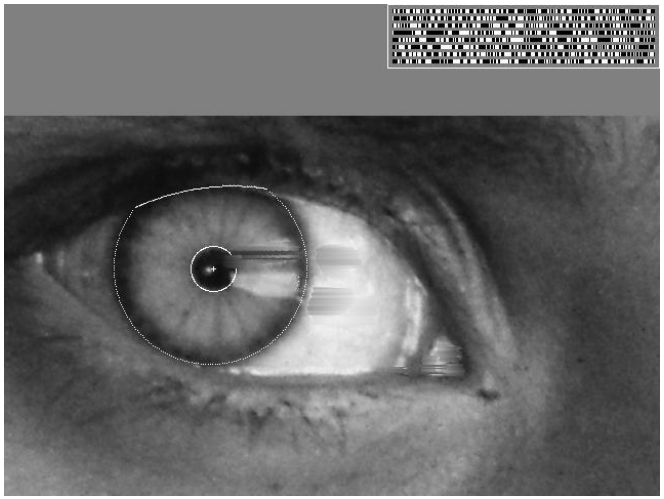
Face recognition

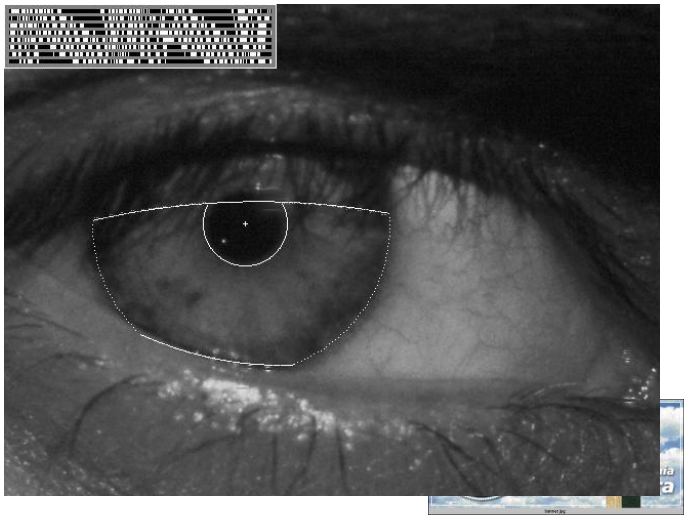


Who is she?

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story





Login without a password…

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

University of California Santa Barbara

Object recognition (in mobile phones)

- This is becoming real:
  - Lincoln Microsoft Research
  - Point & Find, Nokia
  - SnapTell.com (now amazon)

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Click here for interactive demo

Sports

Sportvision first down line

Nice explanation on www.howstuffworks.com

Smart cars

Slide content courtesy of Amnon Shashua

Smart cars

Slide content courtesy of Amnon Shashua

Vision-based interaction (and games)

Nintendo Wii has camera-based IR tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

Digimask: put your face on a 3D avatar.

"Game turns moviegoers into Human Joysticks", CNET

Camera tracking a crowd, based on this work.

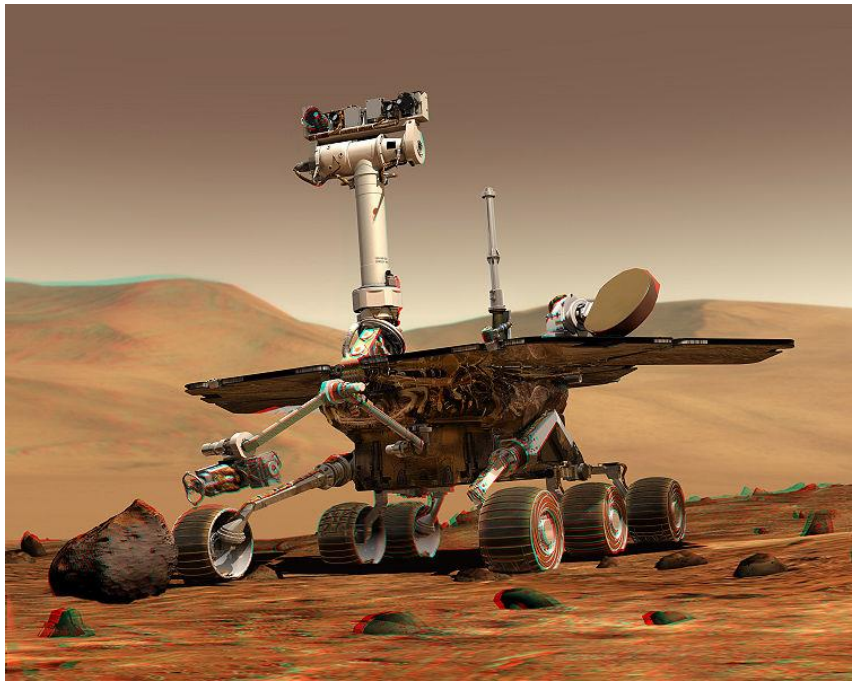
Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

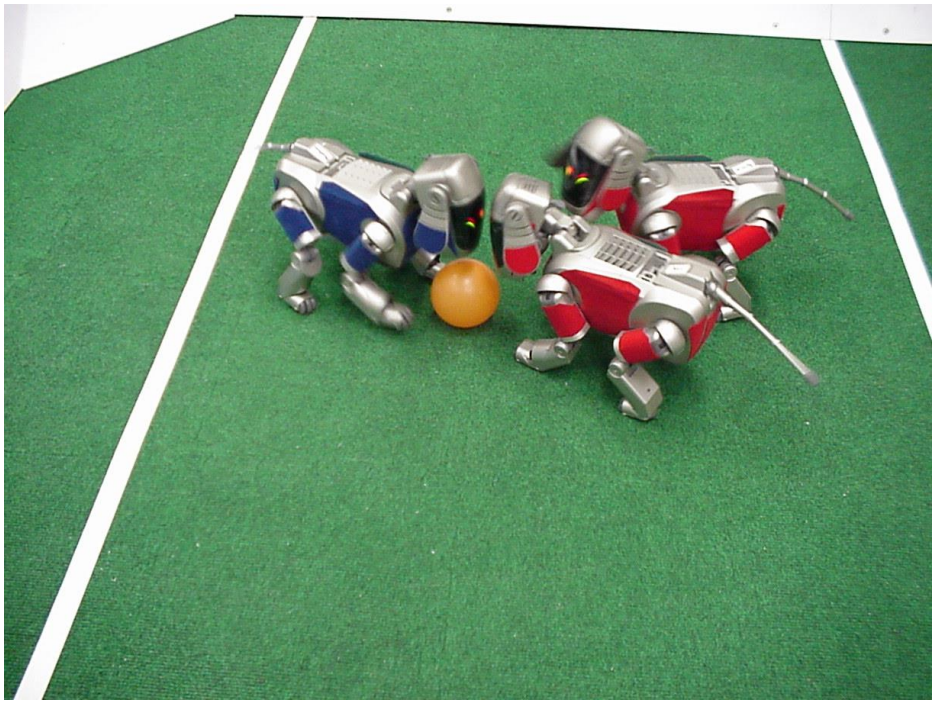
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Robotics



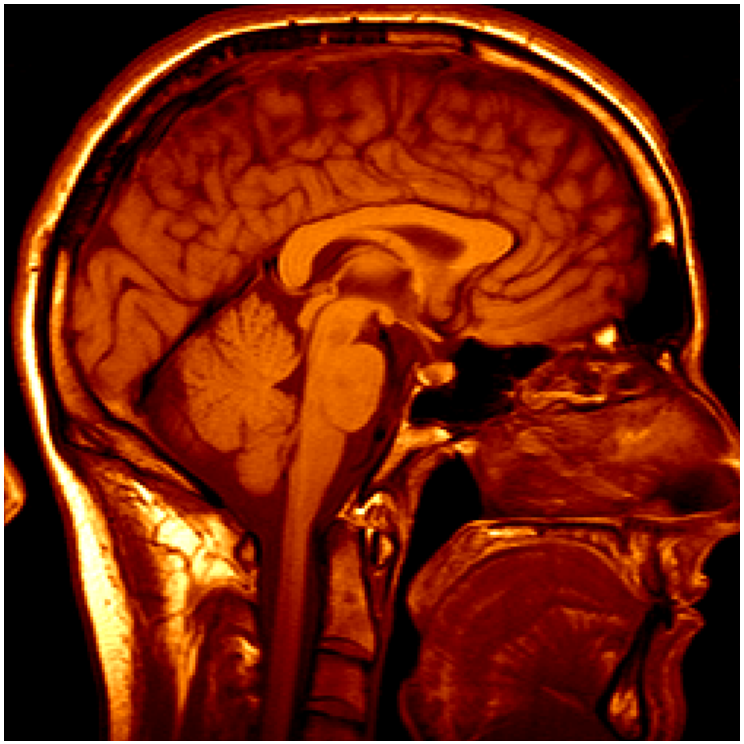
NASA's Mars Spirit Rover

http://en.wikipedia.org/wiki/Spirit_rover



http://www.robocup.org/

Medical imaging



3D imaging

MRI, CT

Image guided surgery

Grimson et al., MIT

Netscape: Blobworld Query Results: Image #108019 (Prefiltered)

File Edit View Go Communicator Help

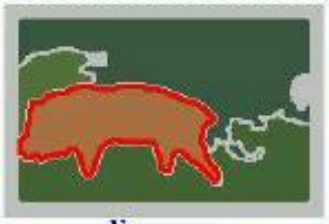
Query image: 108019

Query blobs

	feature importance:				
	overall	color	texture	location	shape
blob	very	very	somewhat	not	not
background	somewhat	very	not	not	not

Querying from 35000 images (2000 returned by the filter).

1: 108044 (score = 0.99)



New query

2: 108023 (score = 0.98)

New query

3: 108006 (score = 0.98)

New query

4: 108029 (score = 0.98)

New query

5: 108051 (score = 0.98)

New query

6: 108084 (score = 0.97)

New query

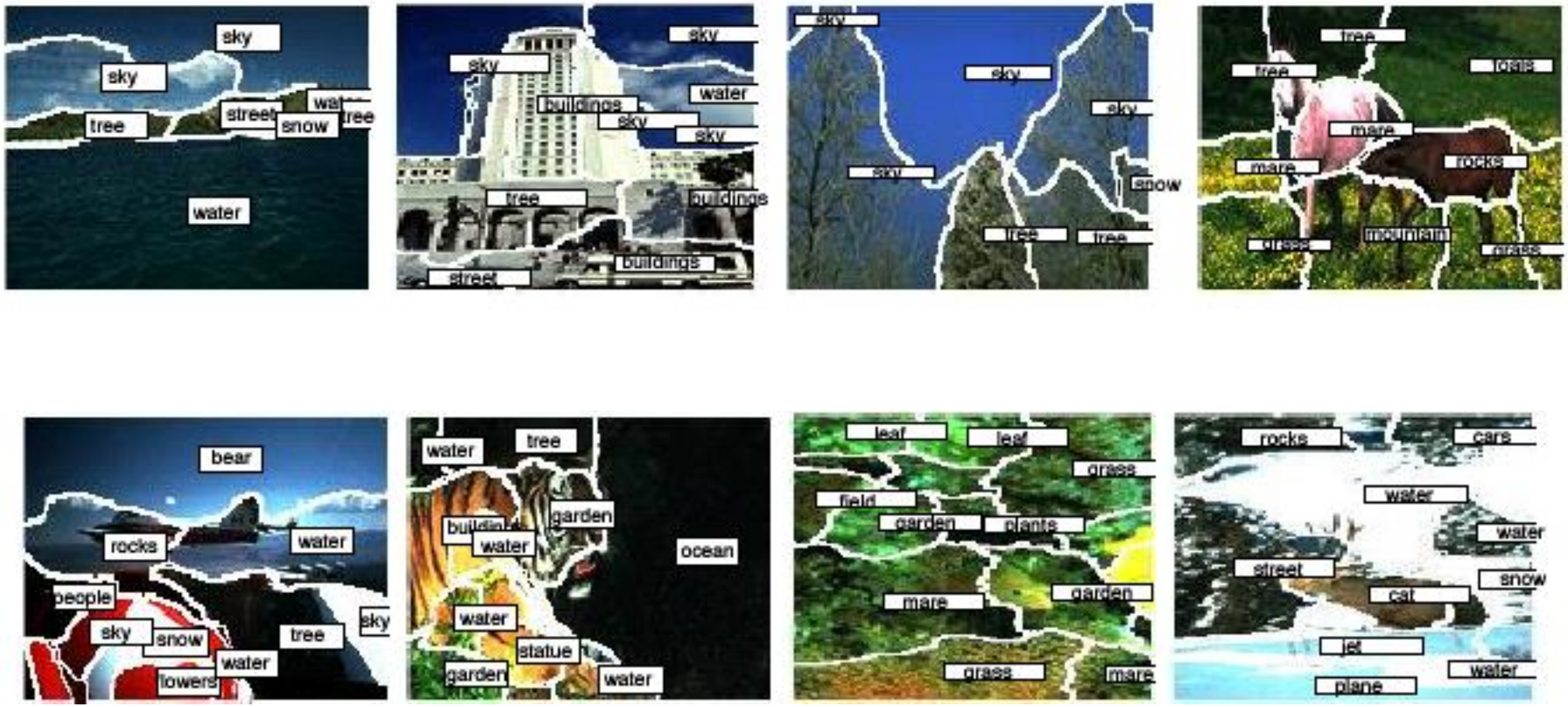
7: 108037 (score = 0.97)

New query

8: 108004 (score = 0.97)

New query

Labeling images



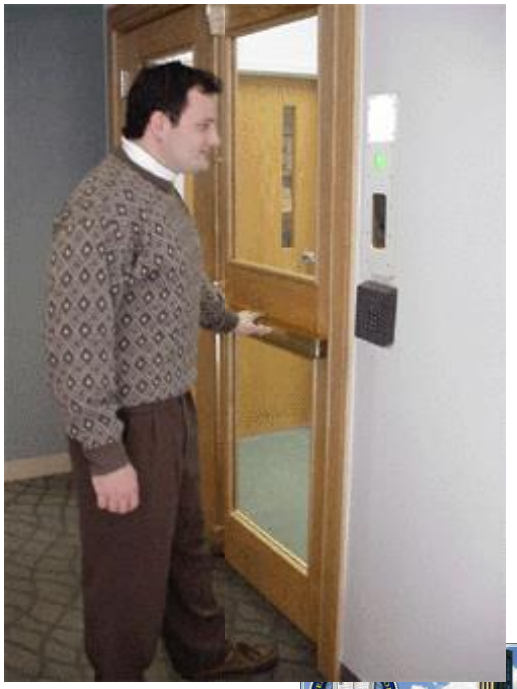
Vision and computer graphics

Eyematic Interfaces

University of California Santa Barbara

Vision for Biometrics

Viisage



Example application: Face recognition

- If we could reliably detect and recognize faces in images…
  - Digital libraries
    - Find me the picture of my brother and Ted
    - Show all the scenes with more than four people
  - Surveillance
    - Alert the operator when a known suspect enters the area
  - HCI
    - Automatically log in as I walk up to the computer
    - Refuse to open my files when someone else is at my machine
  - Military
    - Friend or foe?
  - Personal/entertainment robots
    - Follow the master, look the customer in the eyes

So… why study computer vision?

- "Engineering" reasons
  - Images are become more and more ubiquitous
  - Plenty of useful applications
  - Great mixture of applied mathematics, signal processing, computer science, etc.
  - Solving interesting, underconstrained ("ill-posed") problems
  - Etc.

- "Science" reasons
  - How does the brain do it?
  - Understanding biological vision
  - Deep computational issues
  - Etc.

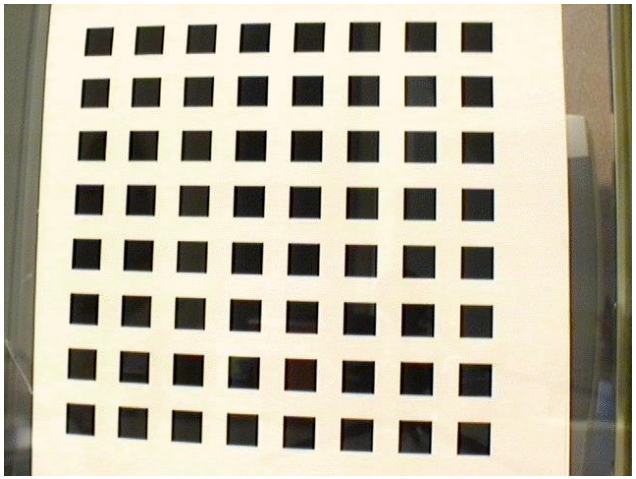
What will we study (not in order)?

- **Part I: The physics of imaging**
  - Camera models and calibration
  - Radiometry
  - Color
- **Part II: "Early" vision**
  - Filtering and edge detection
  - Stereo, optical flow
- **Part III: "Mid-level" vision**
  - Segmentation
  - Classification
  - Tracking
- **Part IV: "High-level" vision and applications**
  - Model-based vision
  - Various applications of computer vision

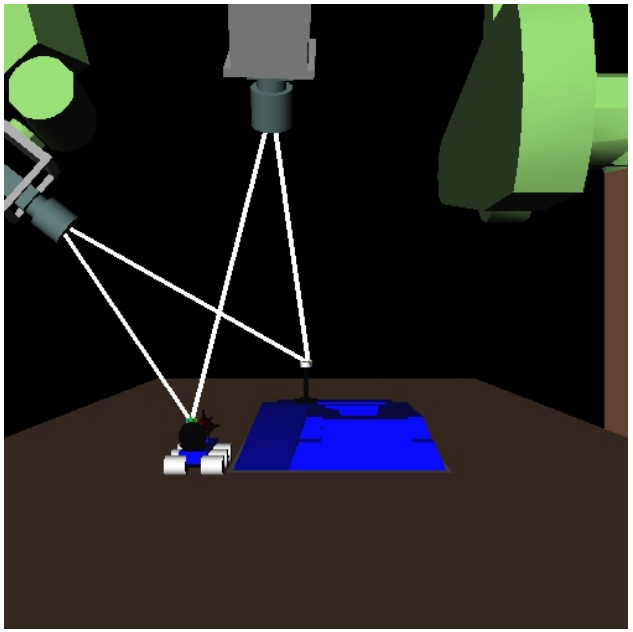
Computer Vision

- Areas of study
  - Camera (sensor) calibration and image formation
  - Binary image analysis
  - Edge detection and low-level filtering
  - Color representation and segmentation
  - Texture description and segmentation
  - Depth/shape from X
  - Stereopsis
  - Optical flow
  - Motion computation
  - Object matching, detection, and recognition
  - 3D sensing and shape description
  - Object and scene tracking

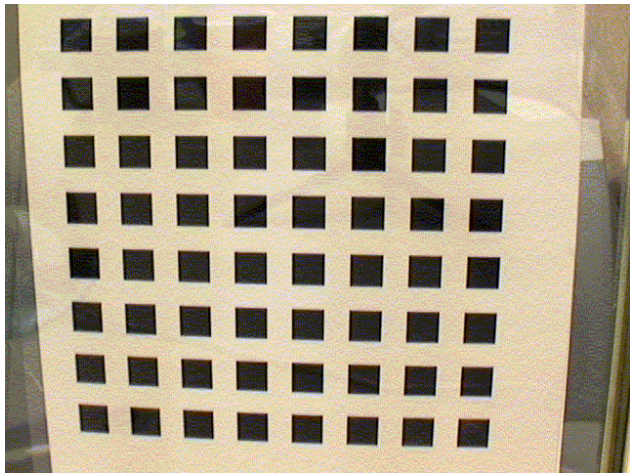
Camera calibration



Original planar pattern

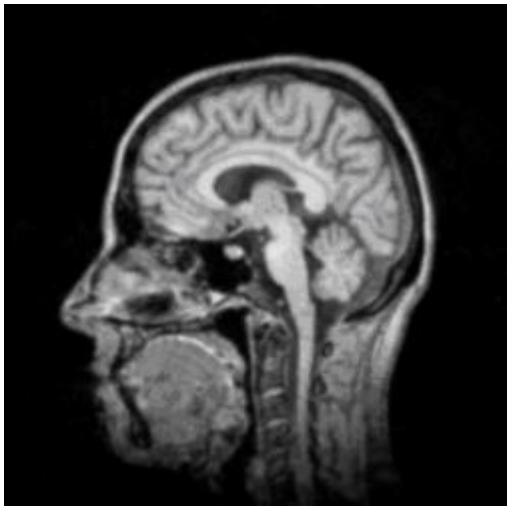


Two calibrated cameras

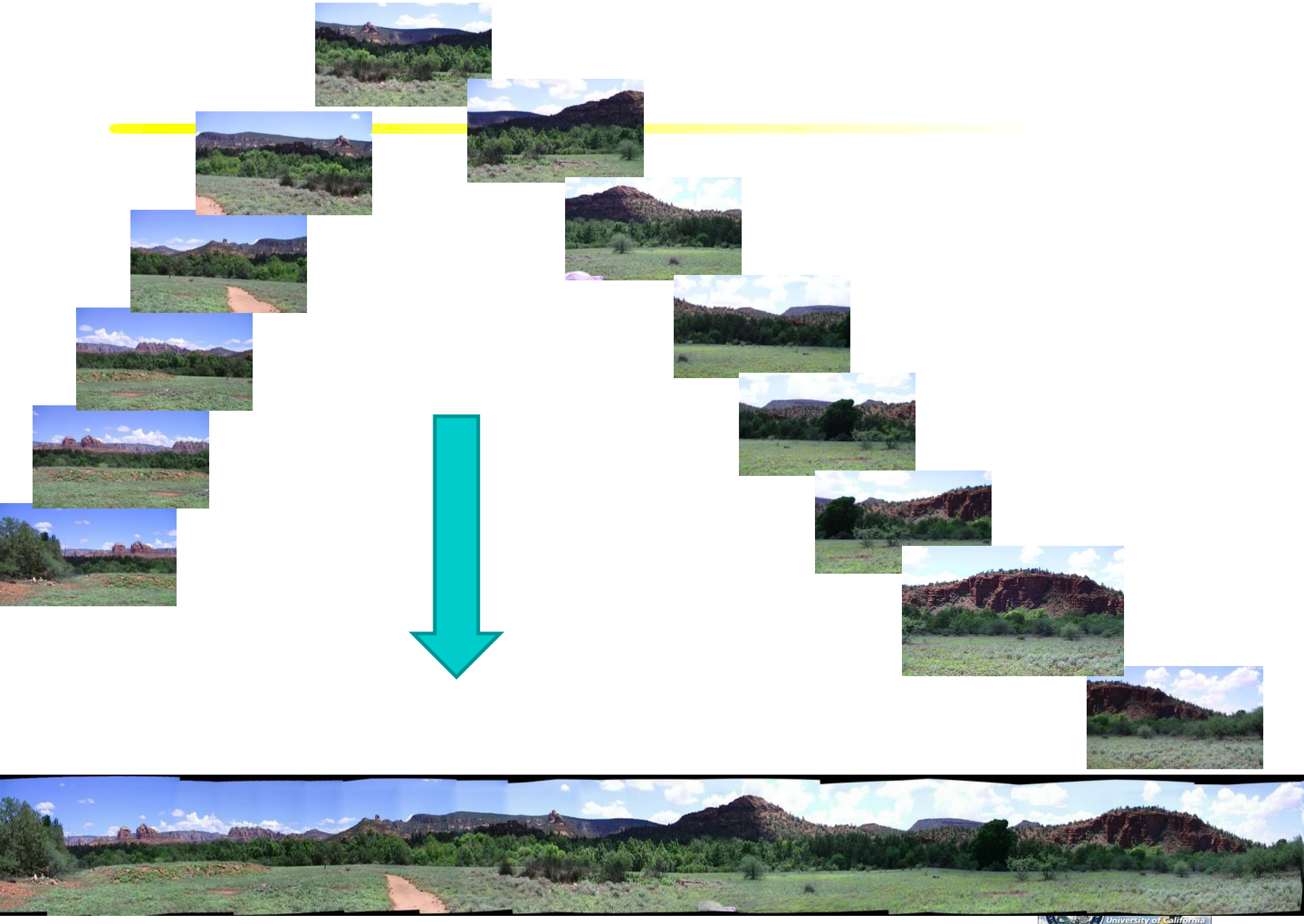


Corrected for radial lens distortion

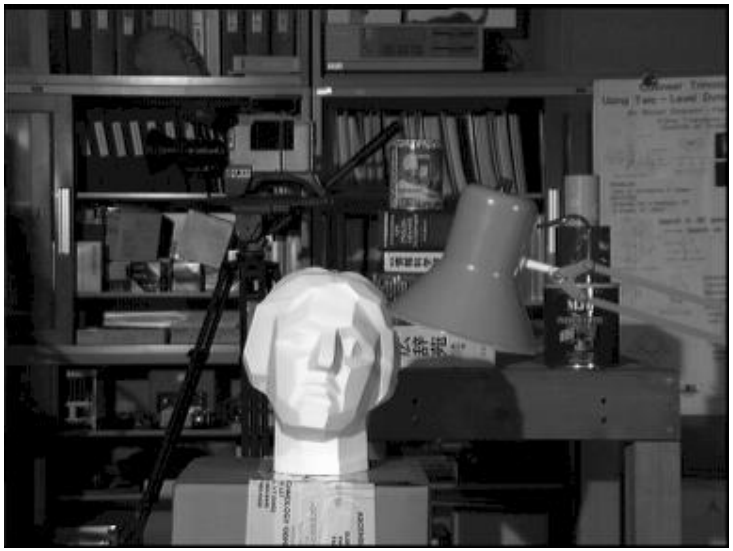
Edge detection

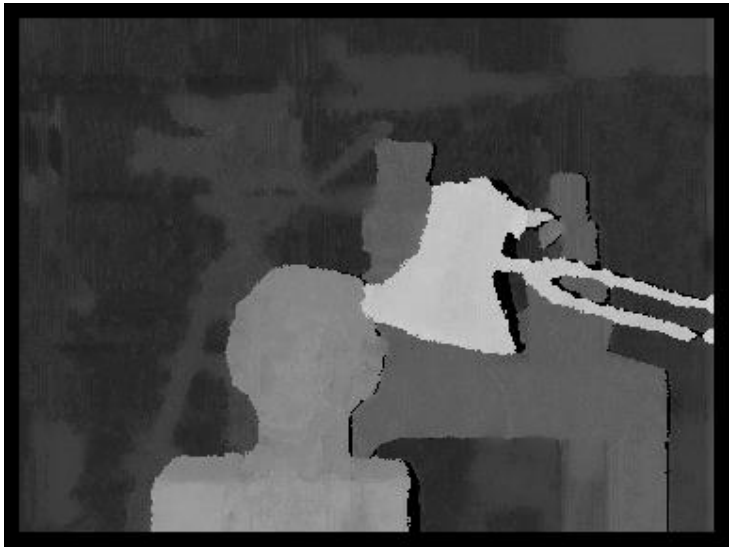


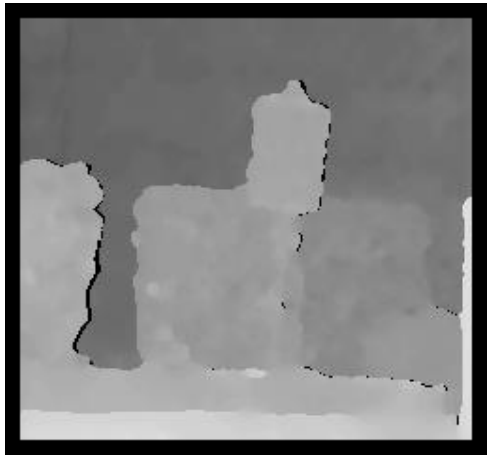

vicomtech
IK4 Research Alliance
University of California
Santa Barbara



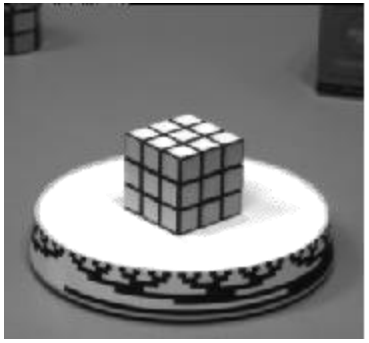
Stereopsis

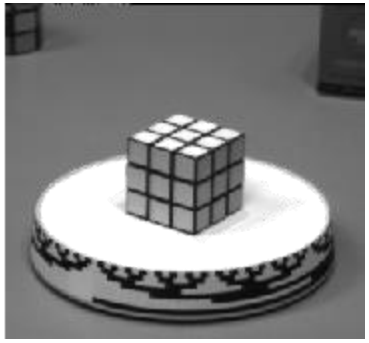


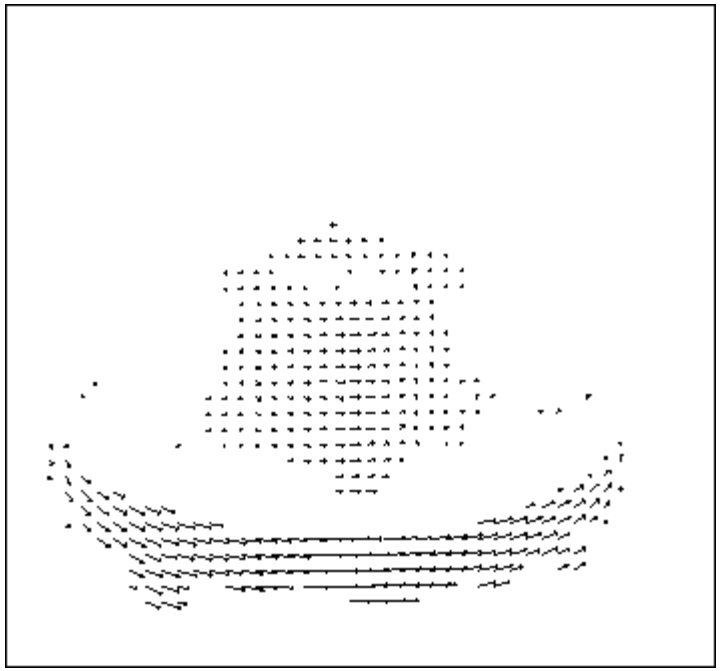




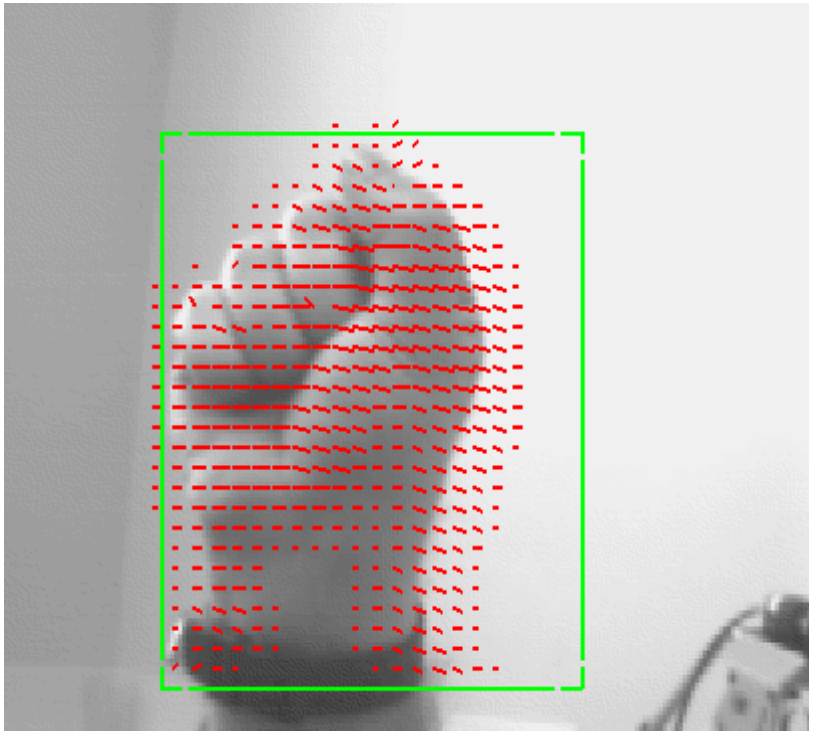
Motion computation



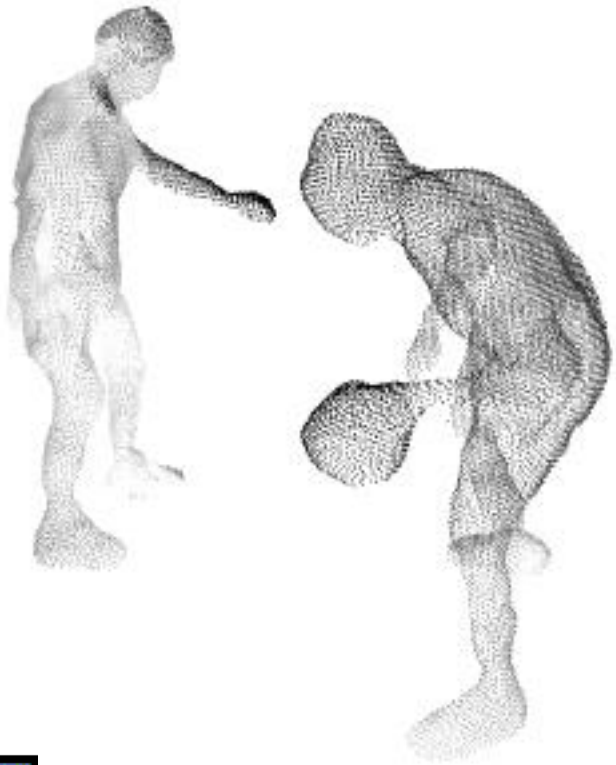




Optical flow

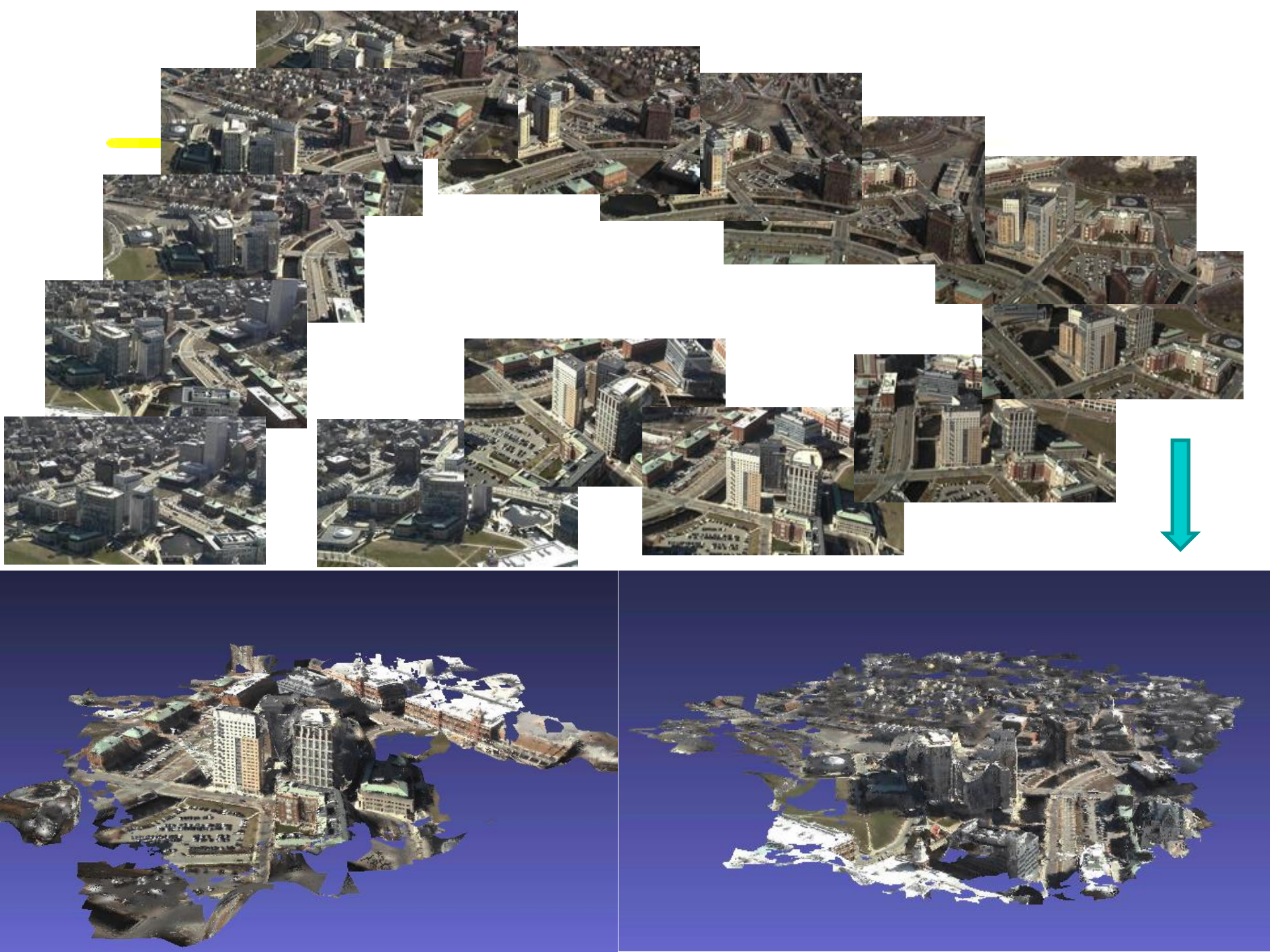


Multi-Camera



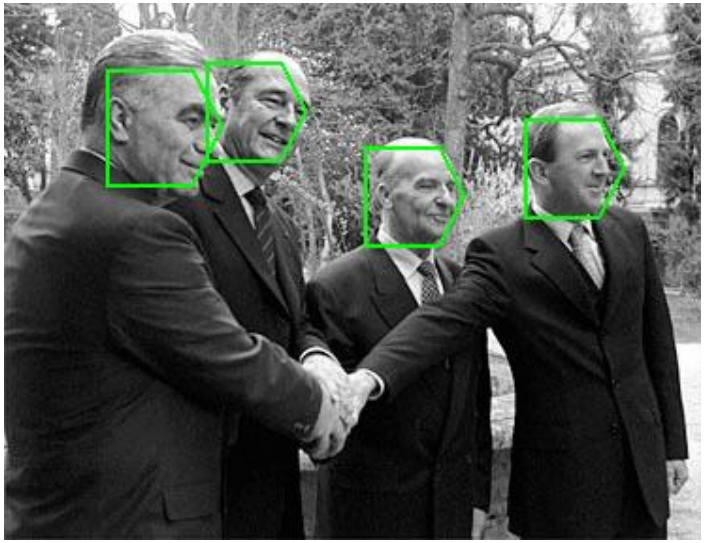
Images

3D reconstruction



Object Detection and Recognition

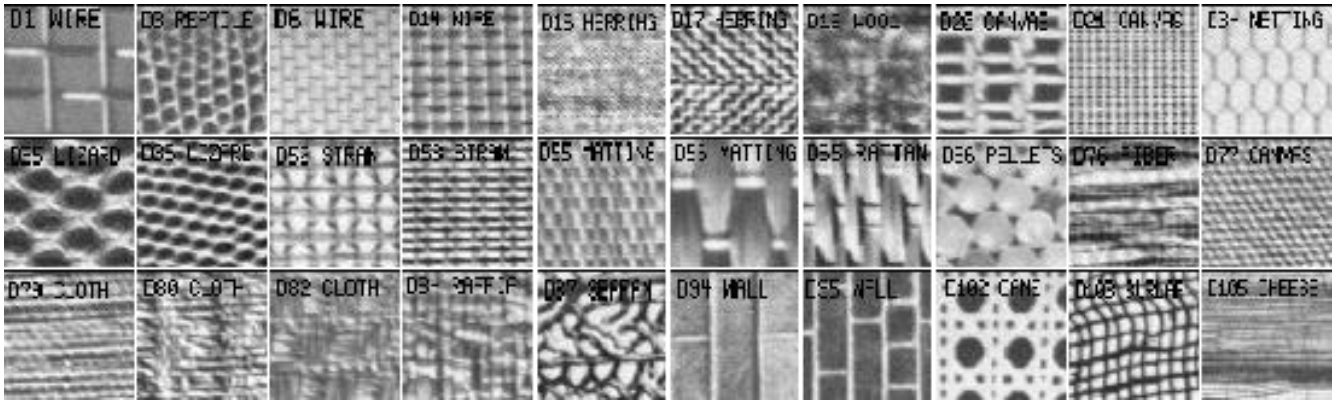
- Model, detect, and recognize objects in images
- Main issues
  - Object representation, feature extraction, matching, learning from data



Faces

Cars

Texture analysis



Course web site

- http://www.cs.ucsb.edu/~cs181b

- Visit it often!