

Introduction to Computer Vision

CS / ECE 181B

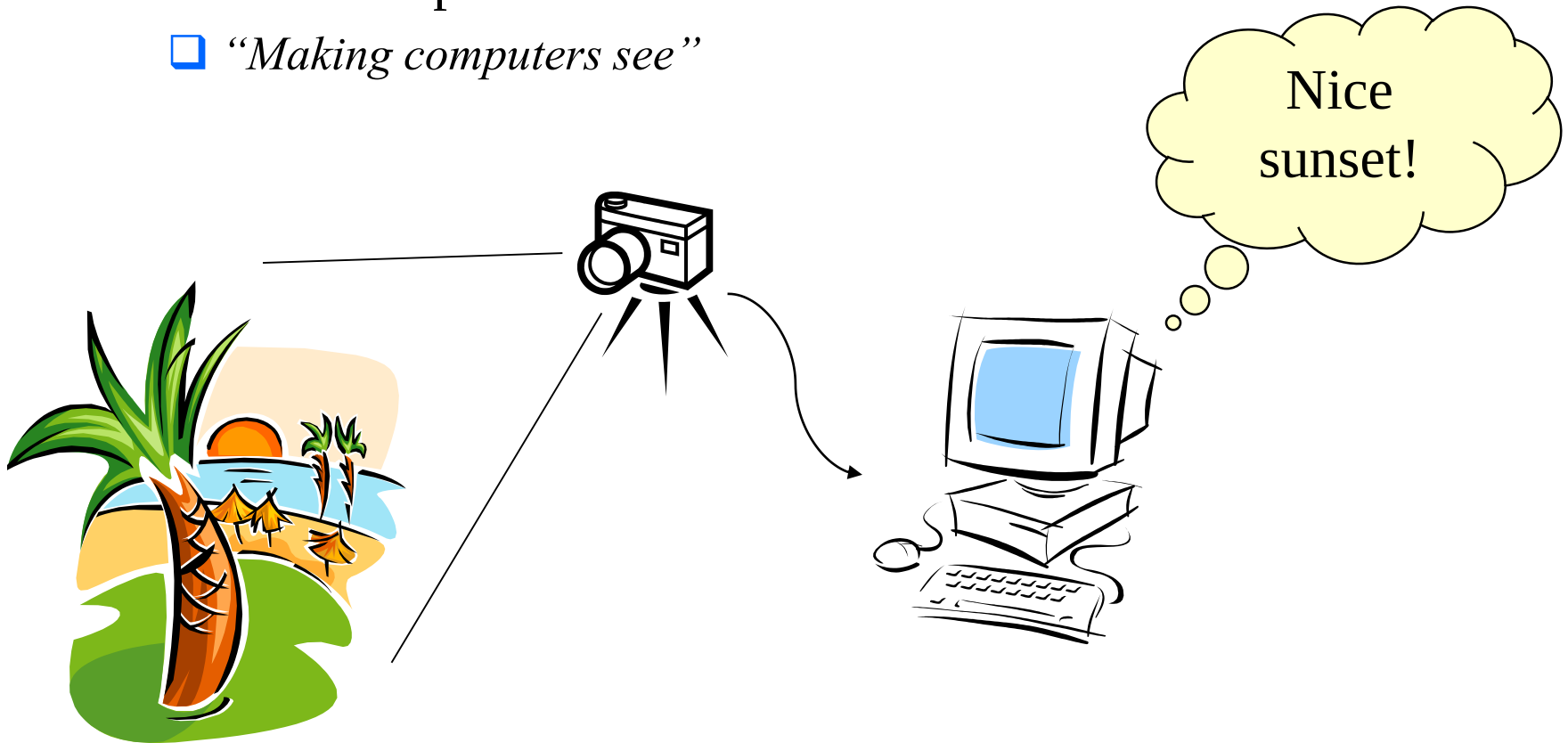
Prof. Yuan-Fang Wang

Computer Science Dept.

Computer Vision

❖ What is computer vision?

❑ *“Making computers see”*



“Extracting descriptions of the world from pictures or sequences of pictures”

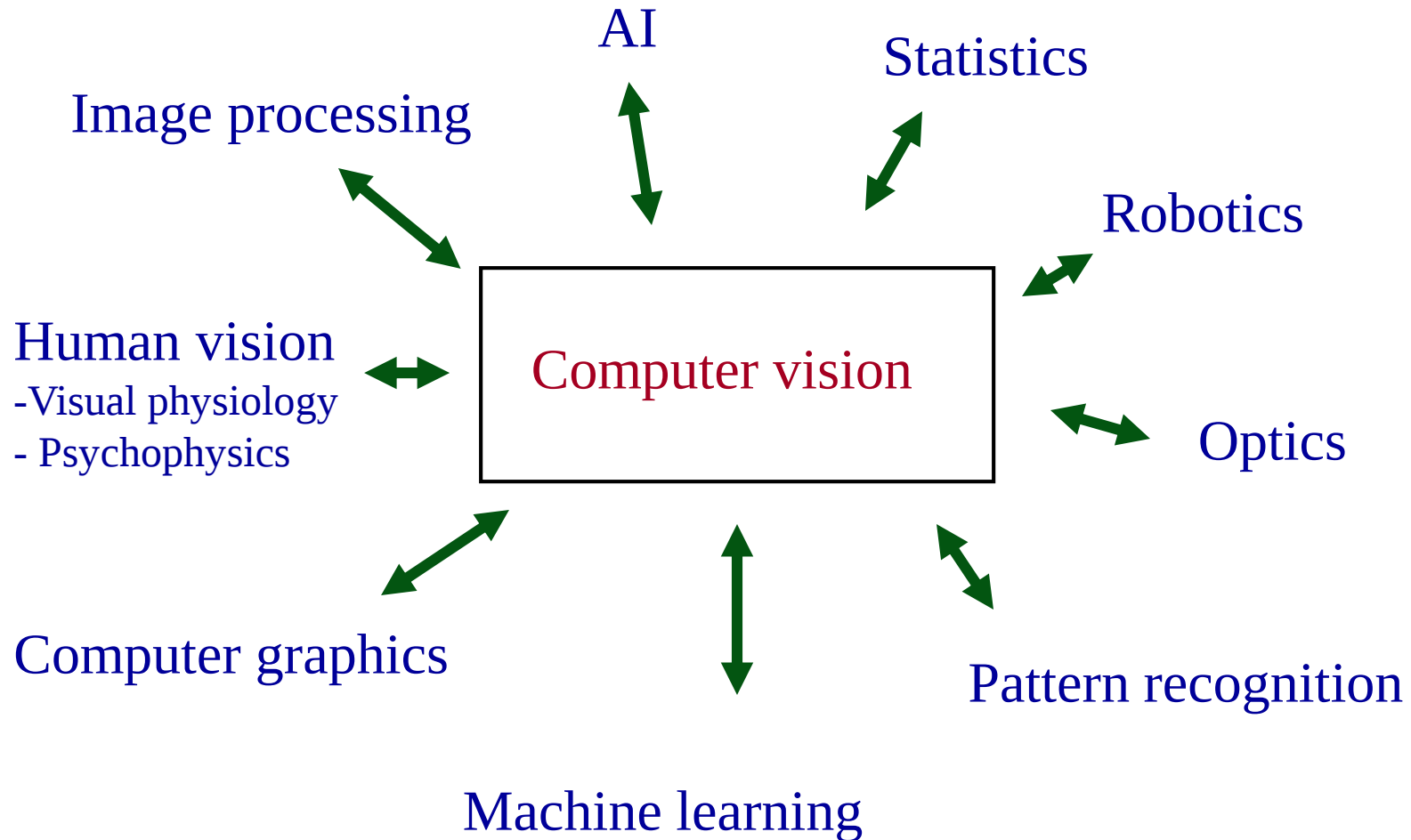
Computer Vision

- ❖ 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

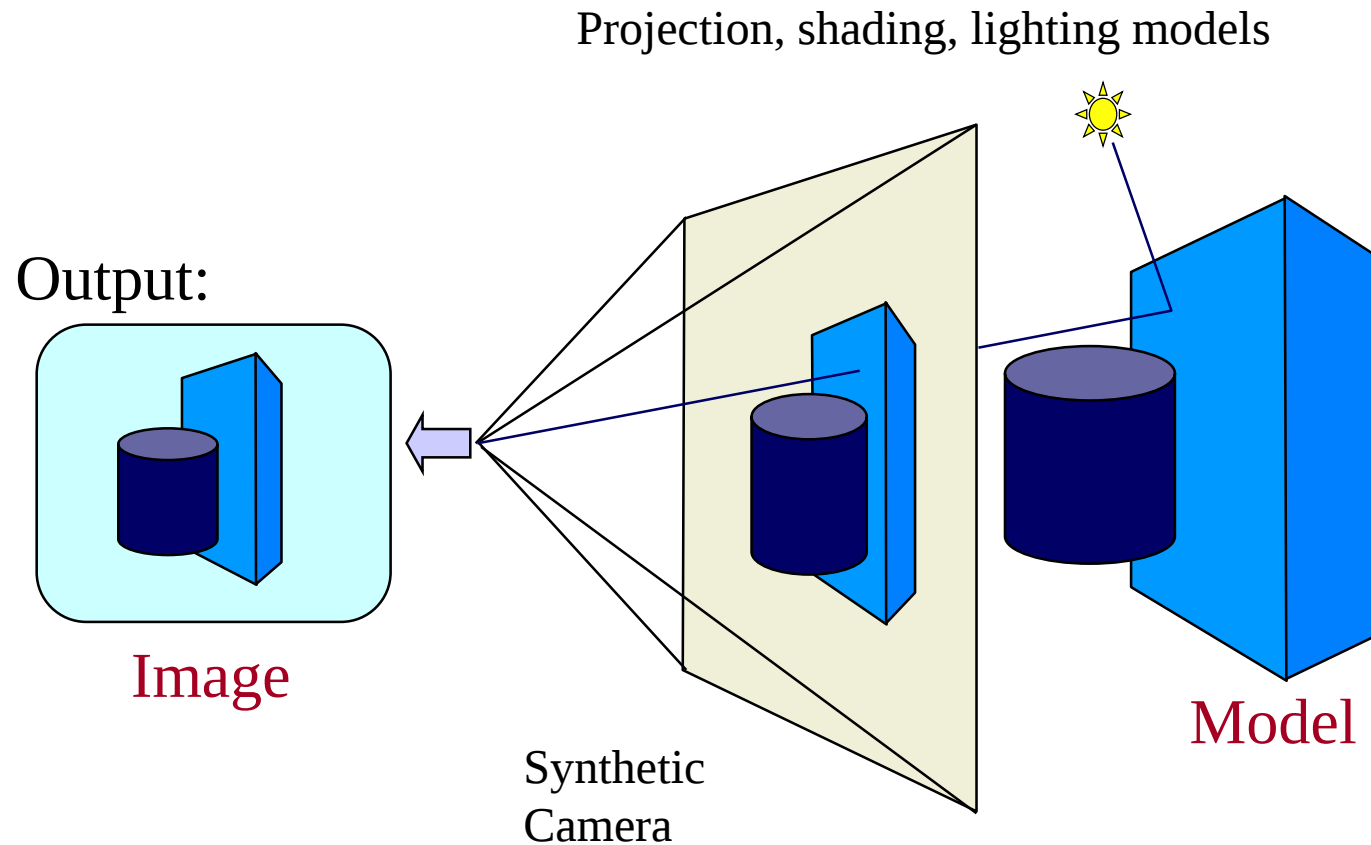
Why Difficult?

- ❖ 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

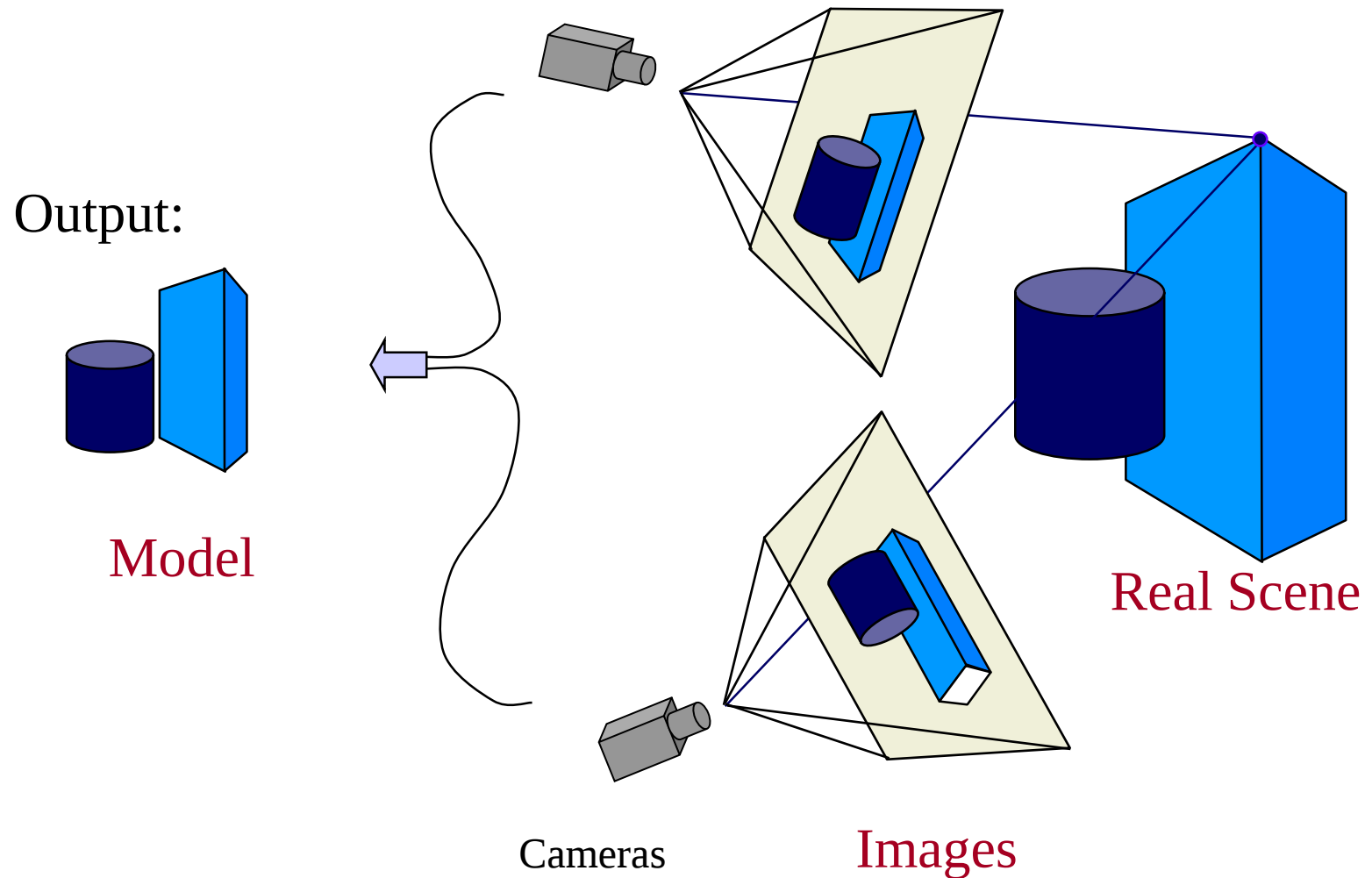
Relation to other fields



Computer Graphics - Synthesis



Computer Vision -Analysis



Computer Vision

- ❖ Imaging process:

$$image = f(world)$$

- ❖ Computer vision

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

including lighting, camera location
and parameters, ...

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?

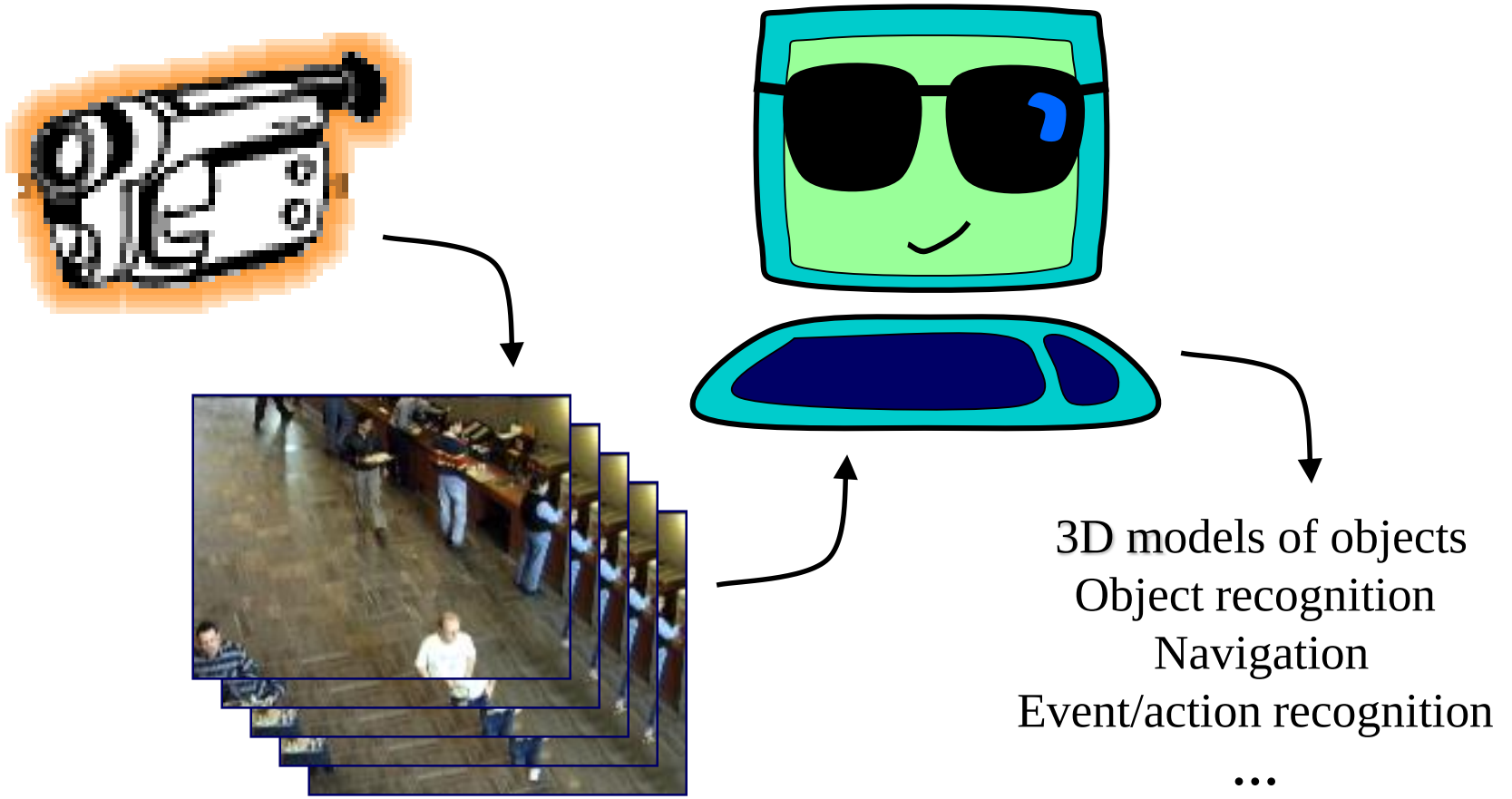
Aims of Computer Vision

- ❖ 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?



Vision Transforms From This...

[illegible]

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...



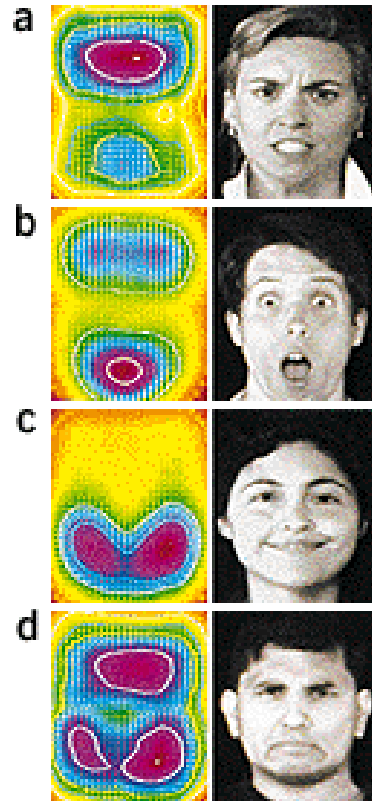
- ❖ “A harbor with many dozens of boats; water is calm and glassy; masts are all vertical; mountains in background, blue sky with a touch of clouds...”



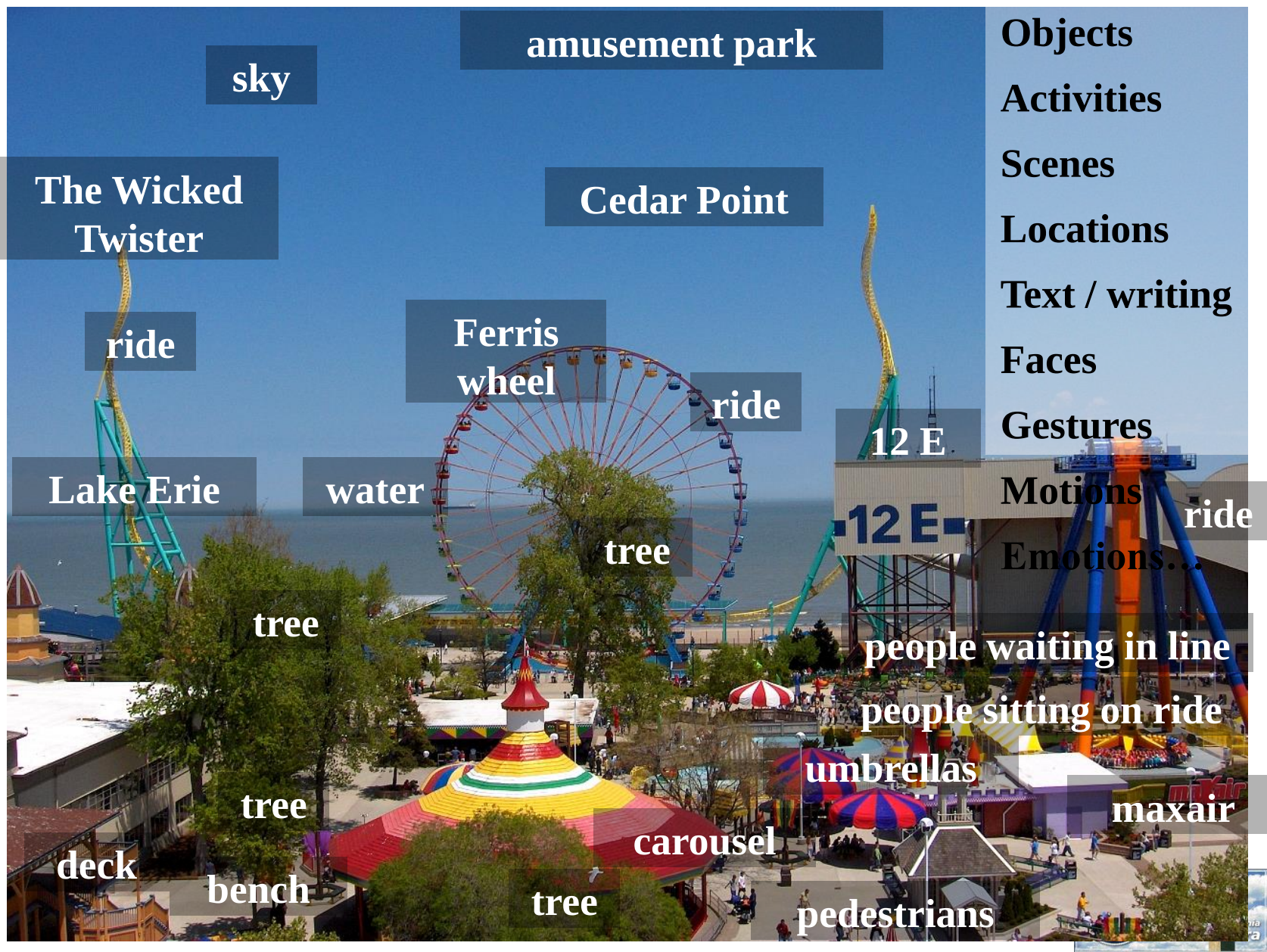
❖ “J548043”



❖ “Hallway straight ahead”



❖ “Angry
Surprised
Happy
Disgusted”



sky

amusement park

Objects

Activities

Scenes

Locations

Text / writing

Faces

Gestures

Motions

Emotions...

The Wicked
Twister

Cedar Point

ride

Ferris
wheel

ride

Lake Erie

water

tree

12 E

12 E

ride

tree

people waiting in line

people sitting on ride

umbrellas

maxair

tree

carousel

deck

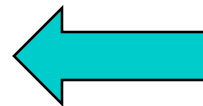
bench

tree

pedestrians

Why is Vision Difficult?

Consider the input...



Not this

But this...

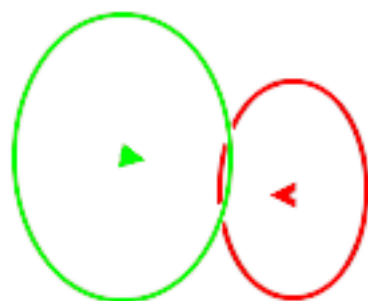
01	00	05	00	03	00	02	00	00	03	01	01	01	01	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00
03	30	3A	38	39	2D	1D	15	10	0E	0C	0A	0A	0A	09	06	08	07	06	06	05	05	07	07	04	05	04	04	06	02	01	02	02	02	02	02	07	01	02	02	02	03					
03	22	1B	16	14	0A	08	0B	0A	0D	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	0A	09	08	05	09	08	08	05	09	03	08	06	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	0C	0B	0C	0B	0A	0A	0A	0B	0C	17	15	1C	15	0D	08	09	08	05	05	05	04	02	05	04	04	00	03								
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	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</																									



Some Possible Outputs



depth
or
segmentation



object pose
(facing away,
facing forward)

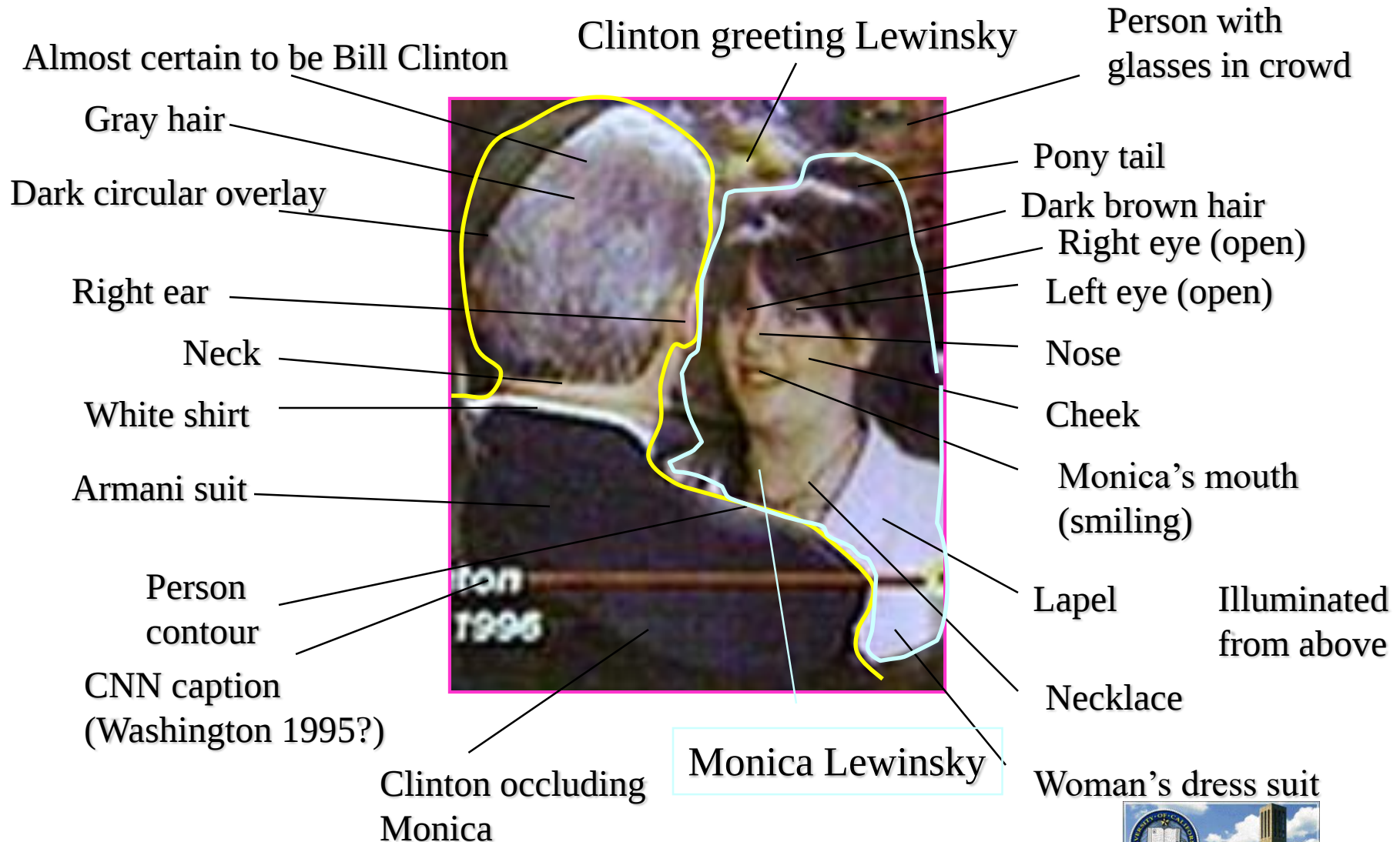


object
recognition

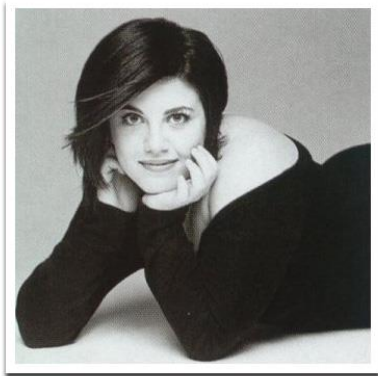


action
understanding

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

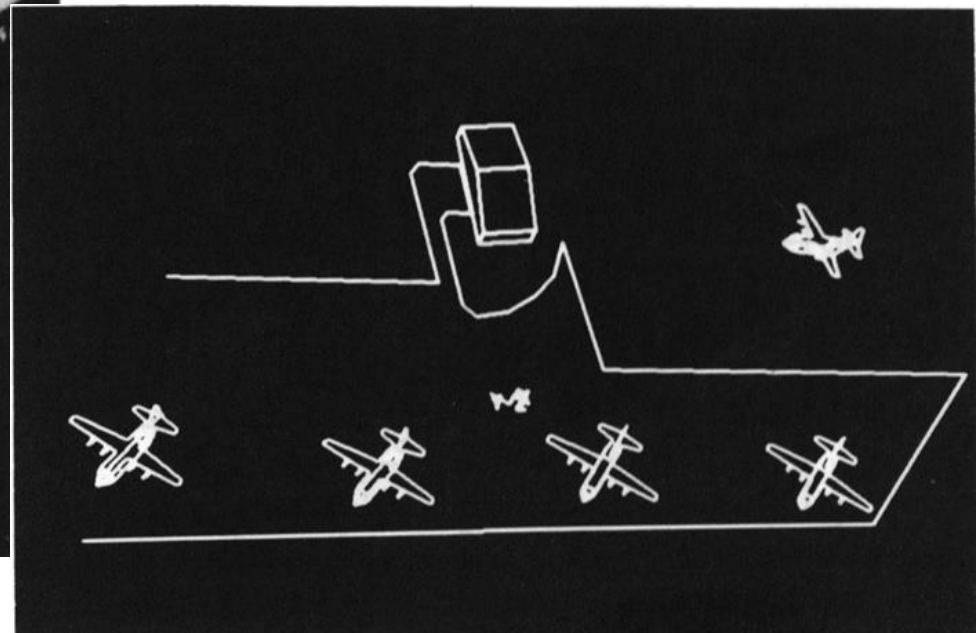
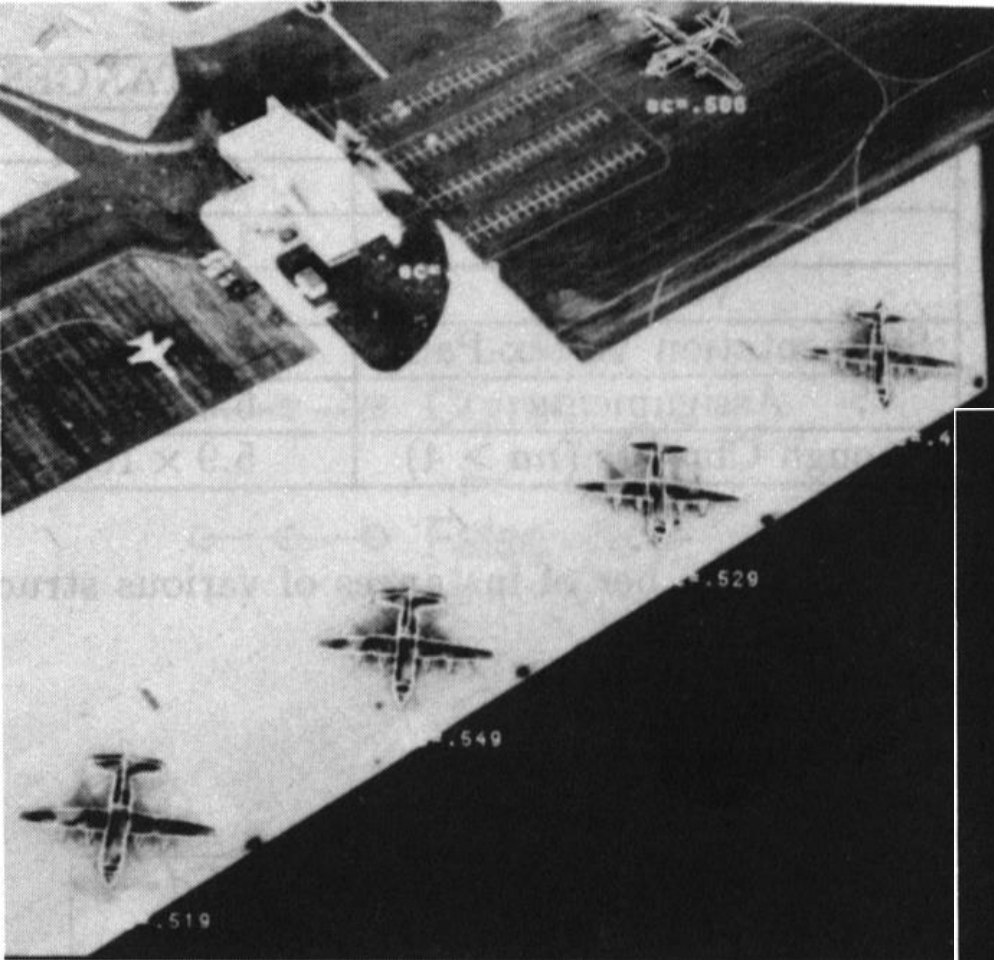
❖ Second generation: Industrial/Medical

- ❑ 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



Examples



CMU NavLab



Cognex

MIT Media Lab



More Examples

❖ Images and video are everywhere!



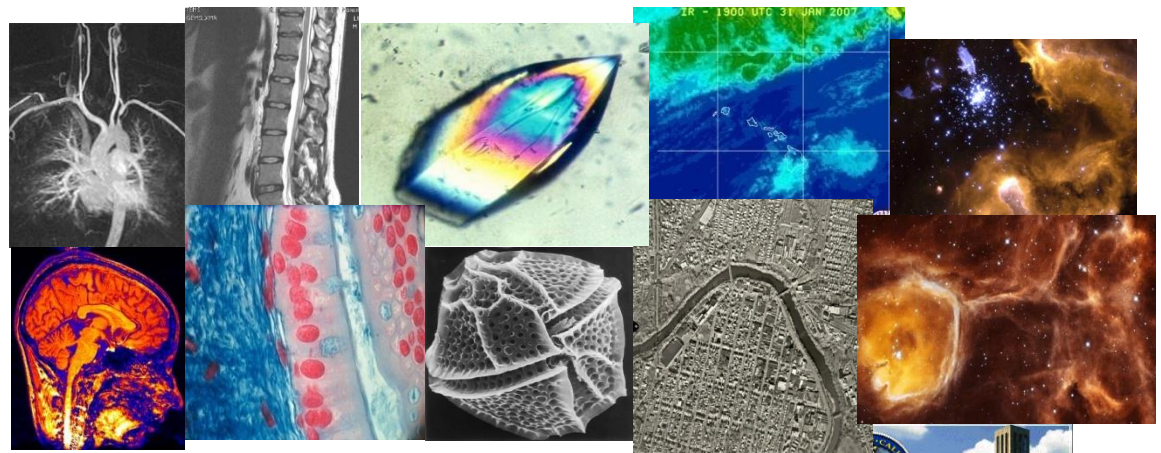
Personal photo albums



Movies, news, sports



Surveillance and security



Medical and scientific images

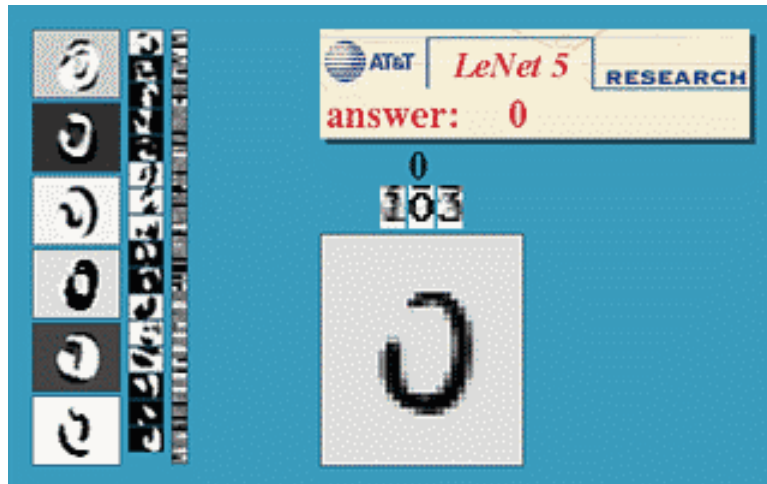


Slide credit, L. Lazebnik

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

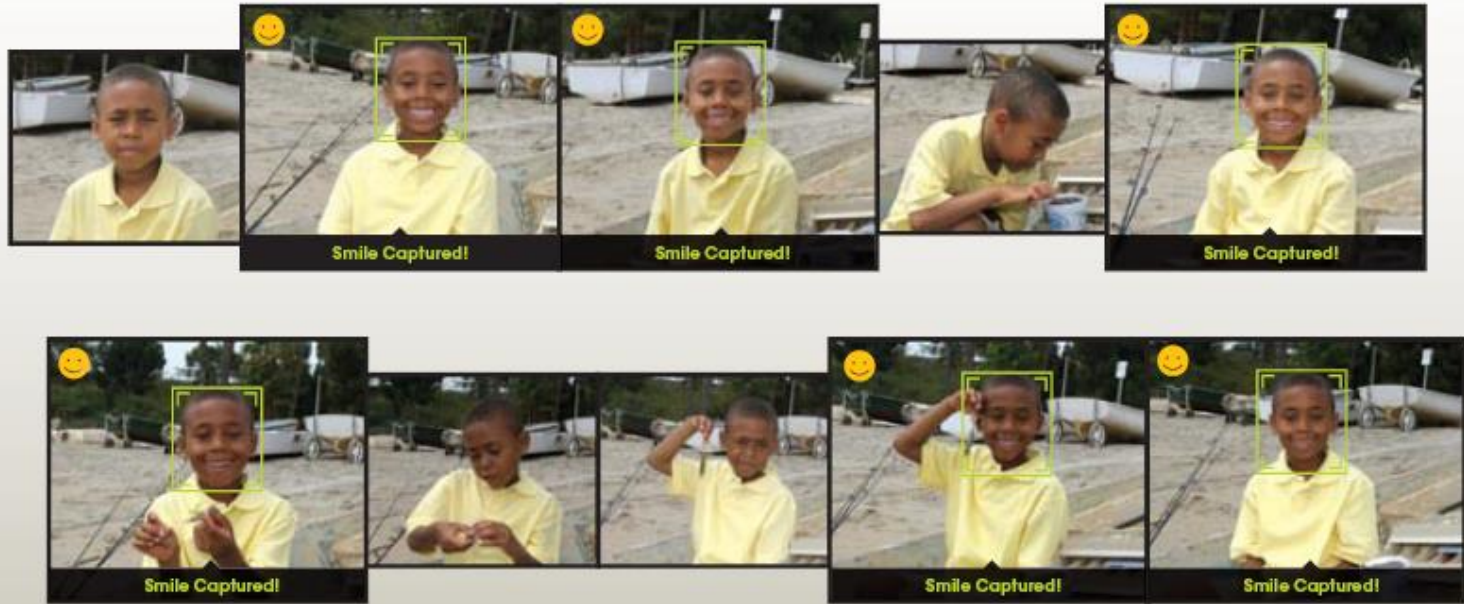
Face detection



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

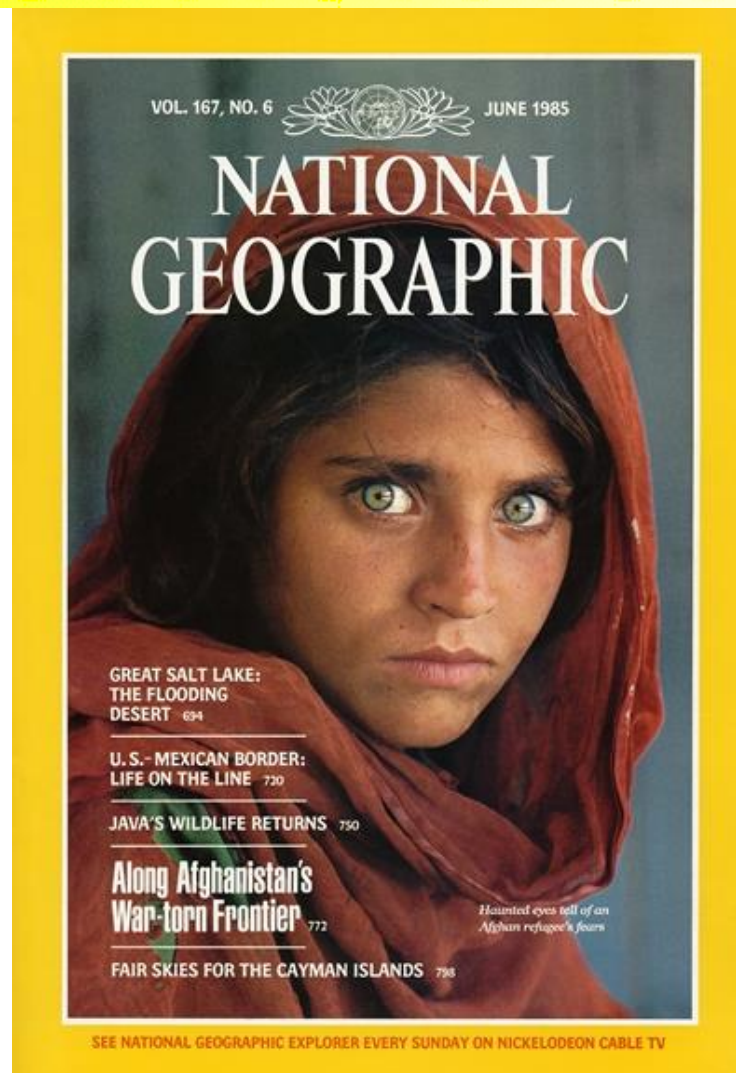
Object recognition (in supermarkets)



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



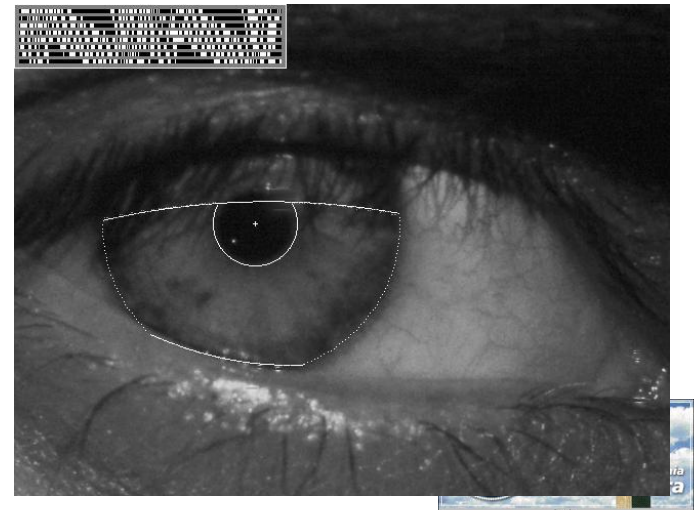
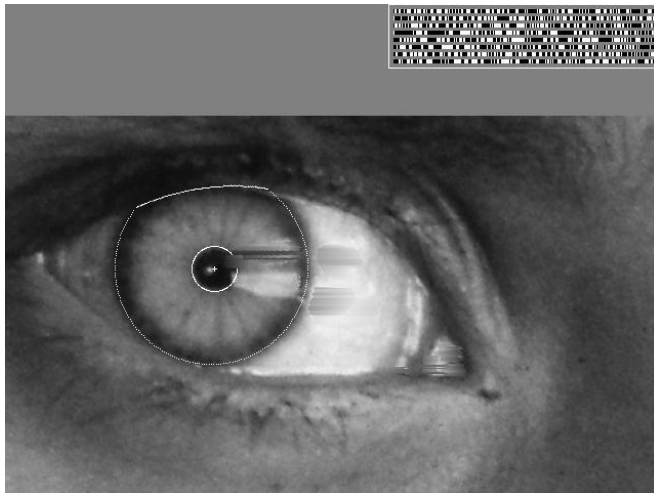
Who is she?



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



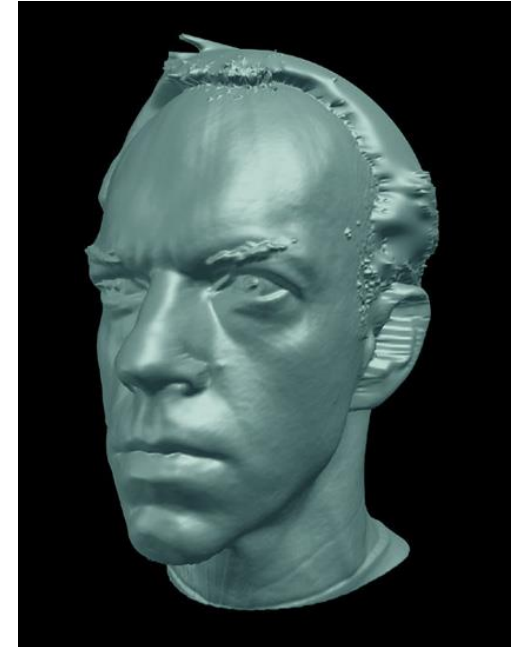
Face recognition systems now
beginning to appear more widely
<http://www.sensiblevision.com/>

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](http://www.howstuffworks.com) on www.howstuffworks.com



Smart cars

[▶▶ manufacturer products](#)[consumer products ◀◀](#)

Our Vision. Your Safety.



rear looking camera

forward looking camera

side looking camera

EyeQ Vision on a Chip

[read more](#)

Vision Applications



Road, Vehicle, Pedestrian Protection and more

[read more](#)

AWS Advance Warning System

[read more](#)

News

- ▶ [Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System](#)
- ▶ [Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end](#)

[all news](#)



Events

- ▶ [Mobileye at Equip Auto, Paris, France](#)
- ▶ [Mobileye at SEMA, Las Vegas, NV](#)

[read more](#)



Smart cars



EyeQ Vision on a Chip



> read more

Vision Applications



Road, Vehicle, Pedestrian Protection and more

> read more

AWS Advance Warning System



> read more

News

- > Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System
- > Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end

> all news

Events

- > Mobileye at Equip Auto, Paris, France
- > Mobileye at SEMA, Las Vegas, NV



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

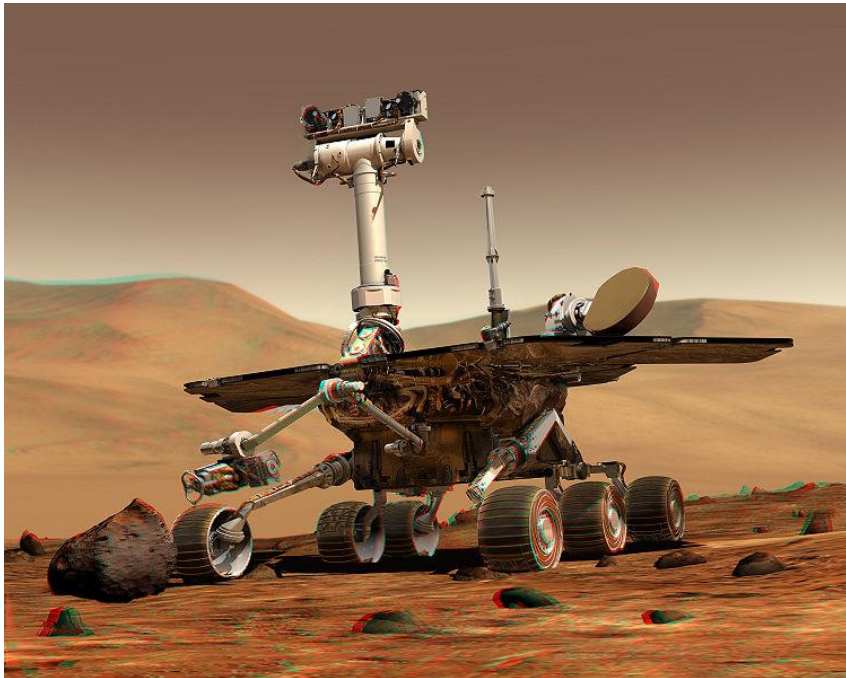


NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

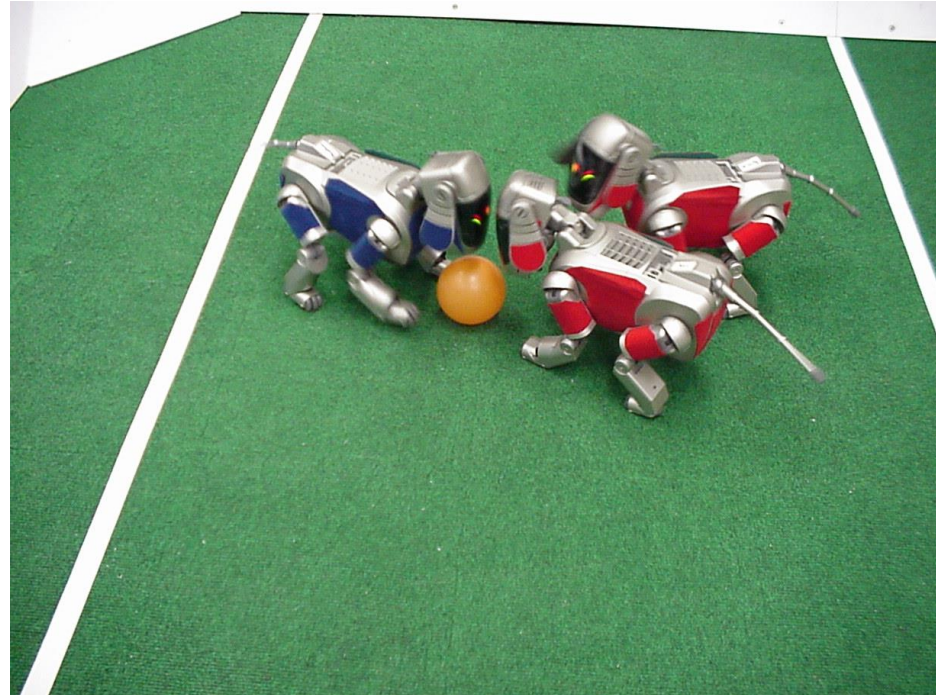
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “[Computer Vision on Mars](#)” by Matthies et al.

Robotics



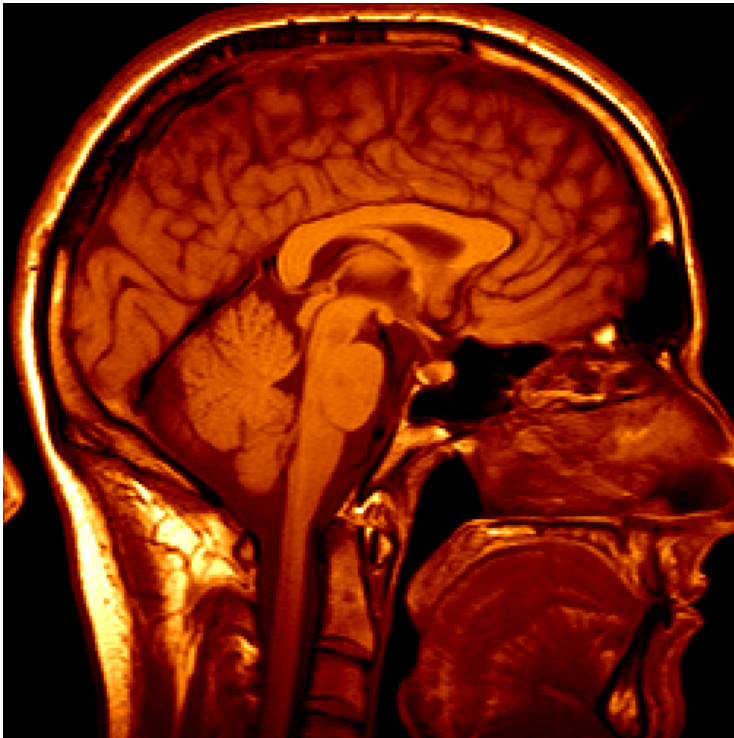
NASA's Mars Spirit Rover

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



<http://www.robocup.org/>

Medical imaging



3D imaging
MRI, CT



Image guided surgery
[Grimson et al., MIT](#)

M

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

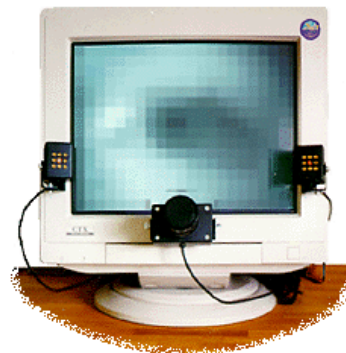


*Eyematic
Interfaces*

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**

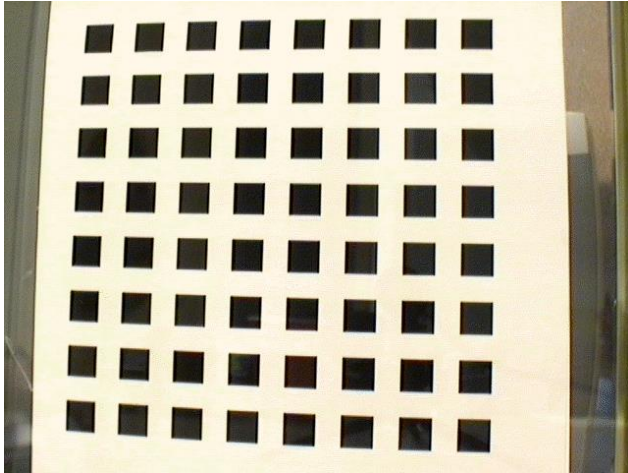
- ☐ Model-based vision
- ☐ Various applications of computer vision

Computer Vision

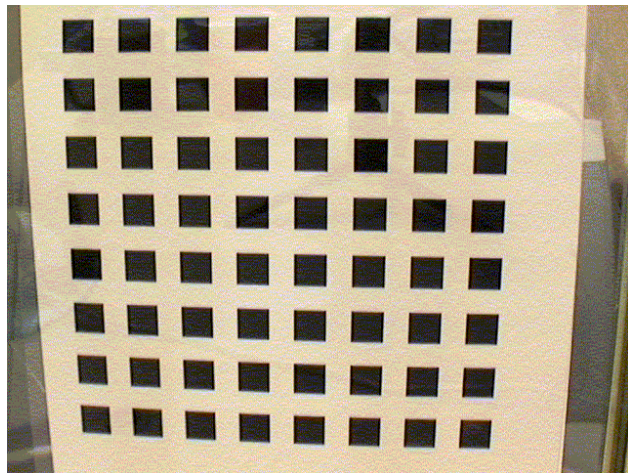
❖ Areas of study

- ☐ 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

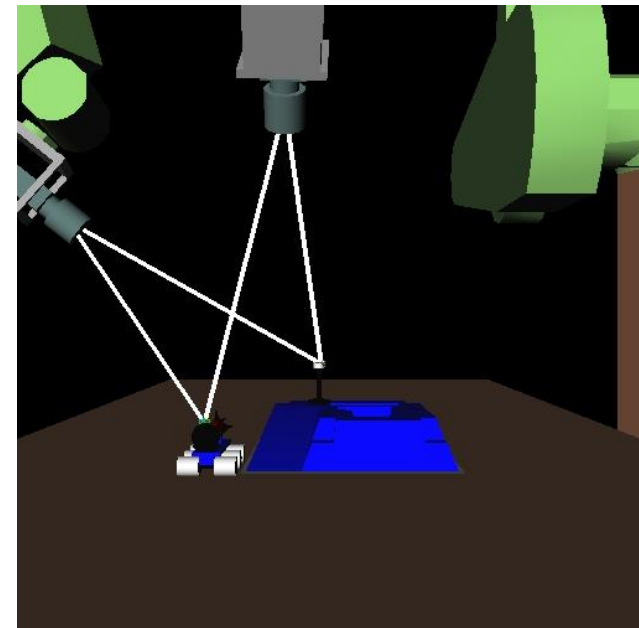
Camera calibration



Original planar
pattern

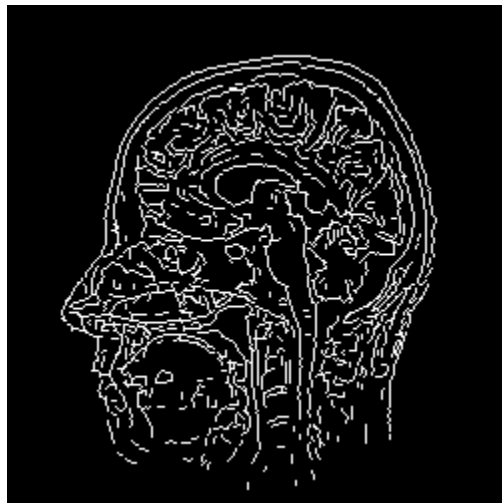
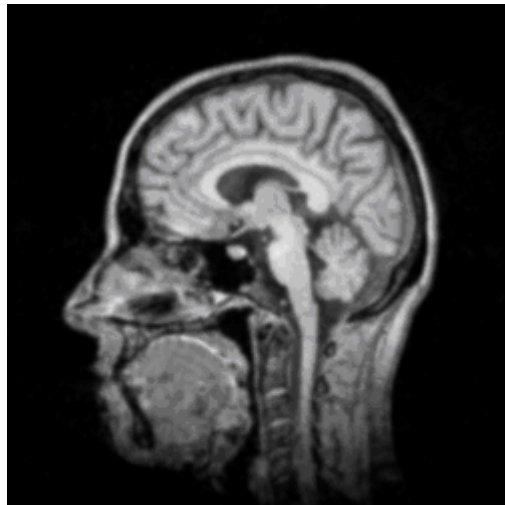


Corrected for radial
lens distortion

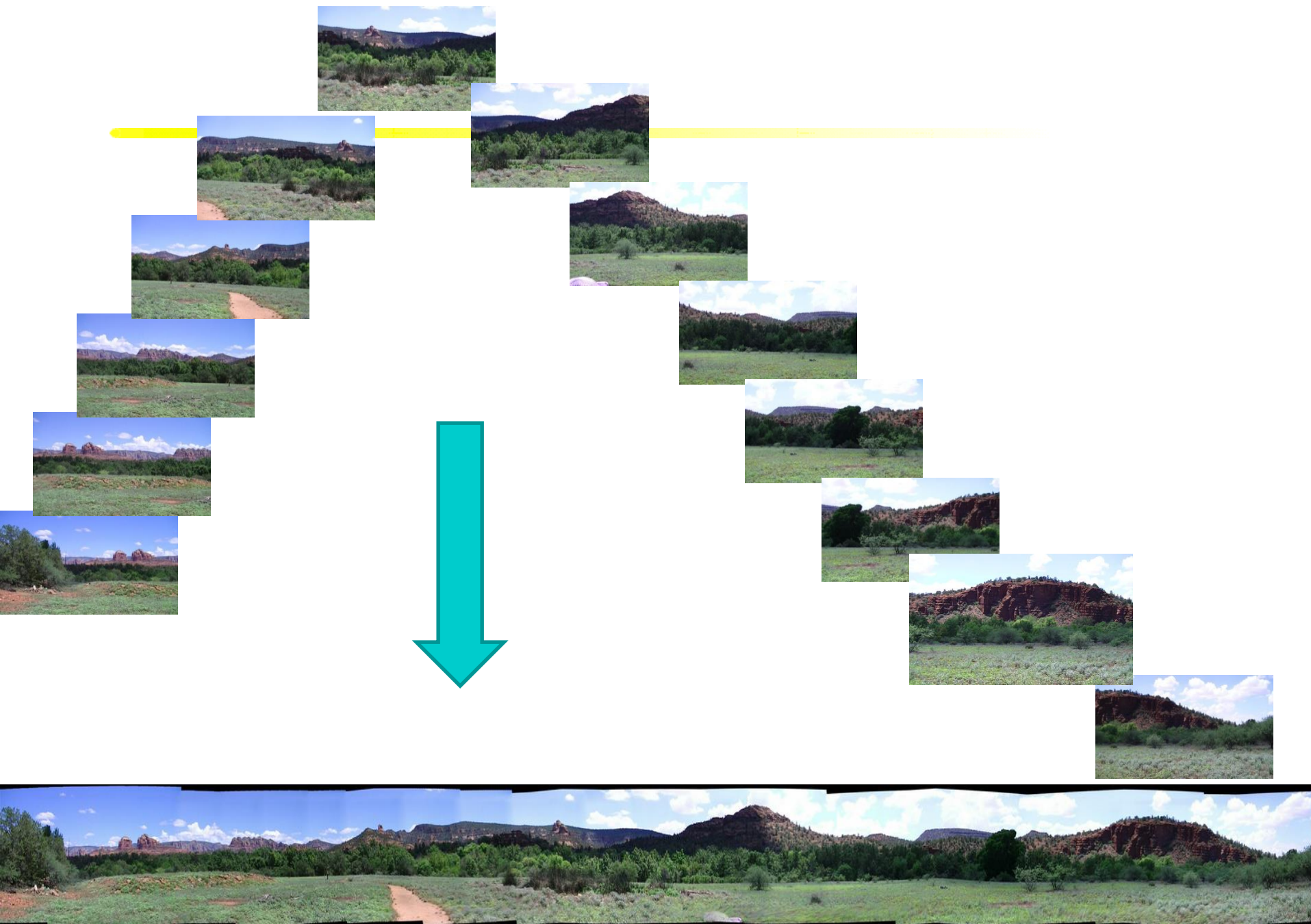


Two calibrated
cameras

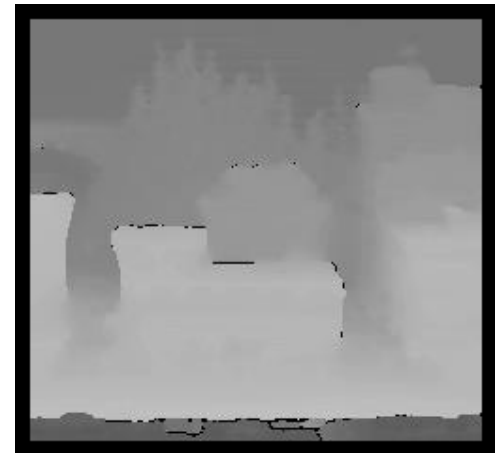
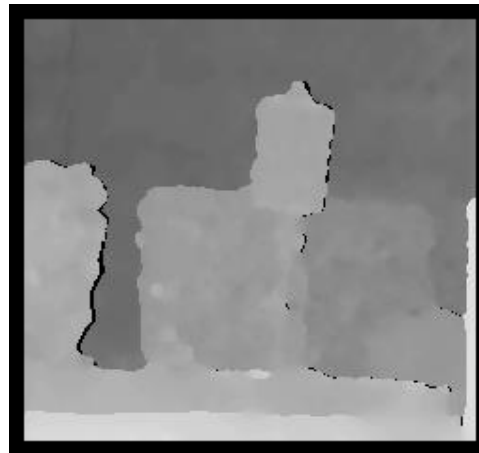
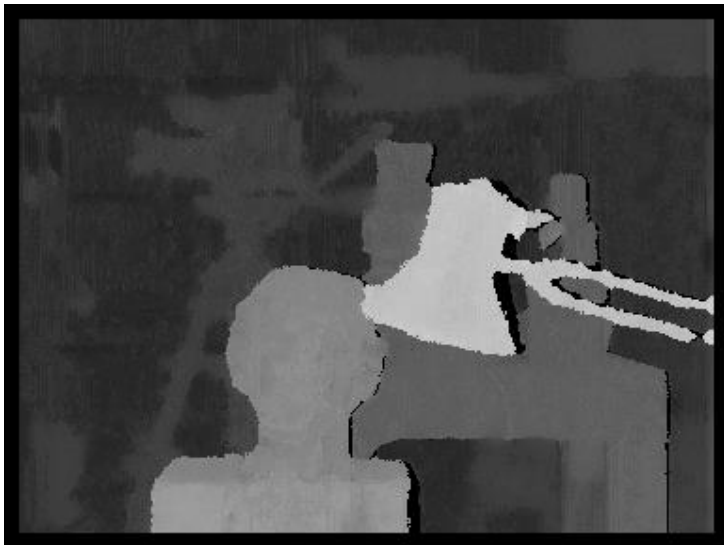
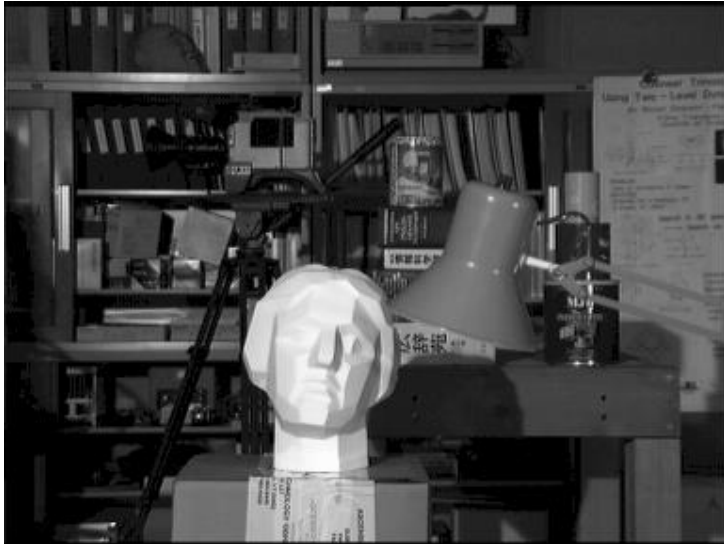
Edge detection



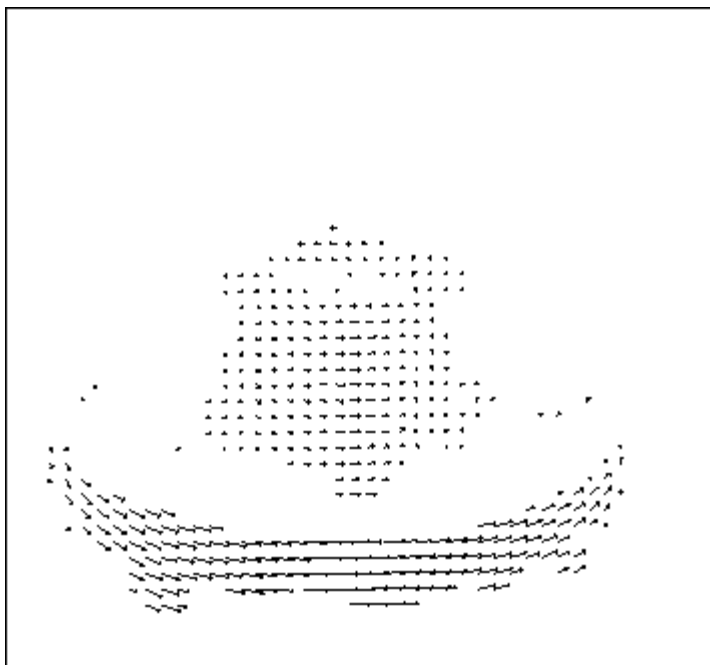
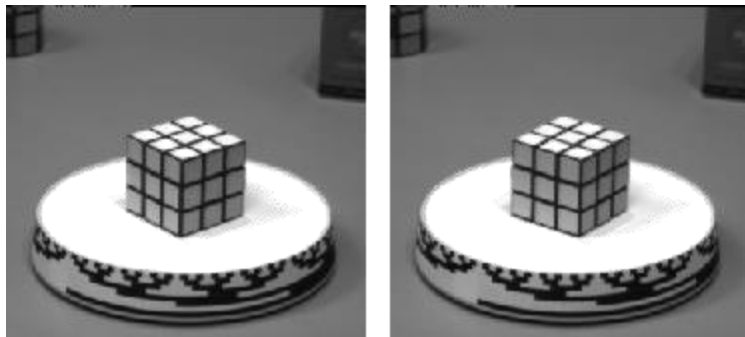




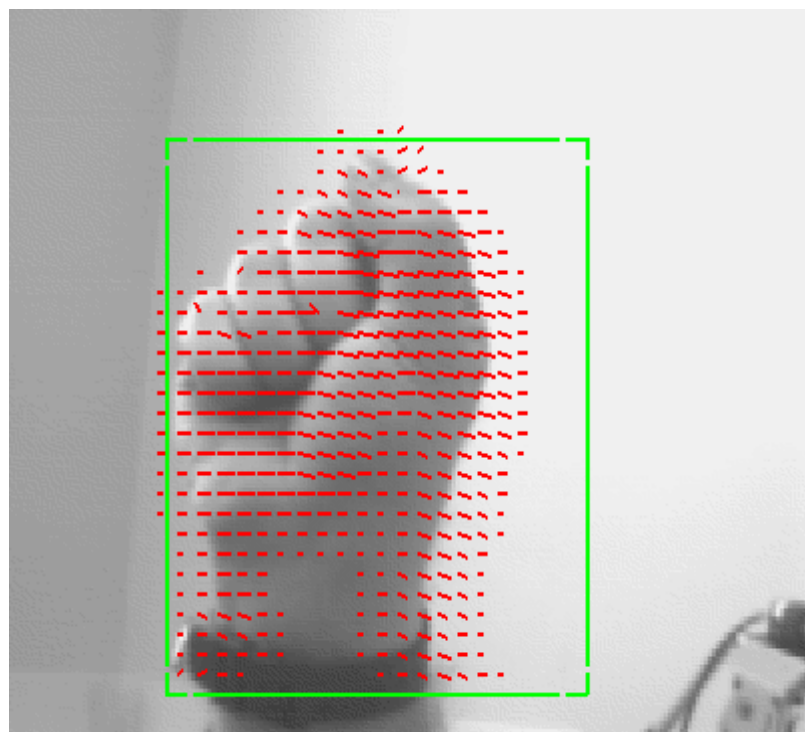
Stereopsis



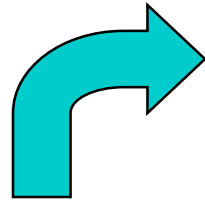
Motion computation



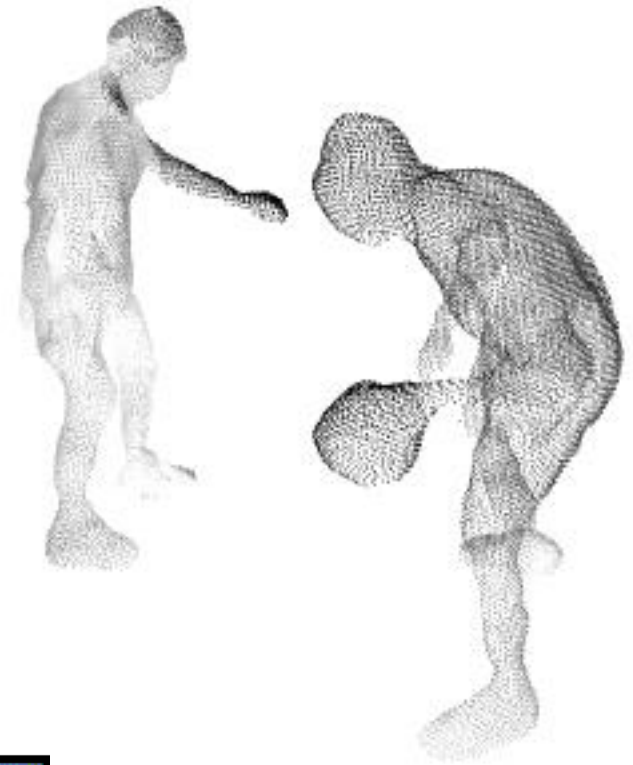
Optical flow



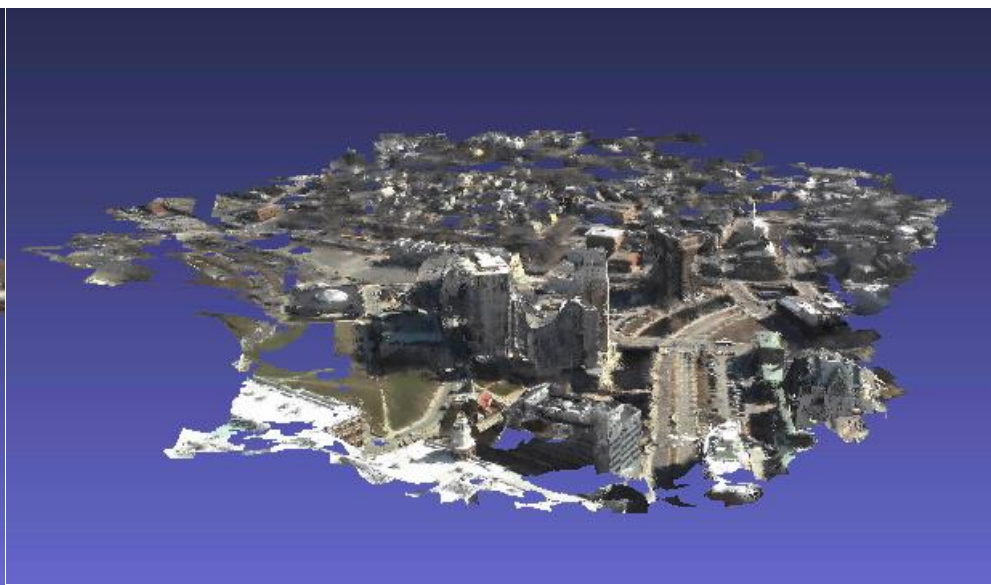
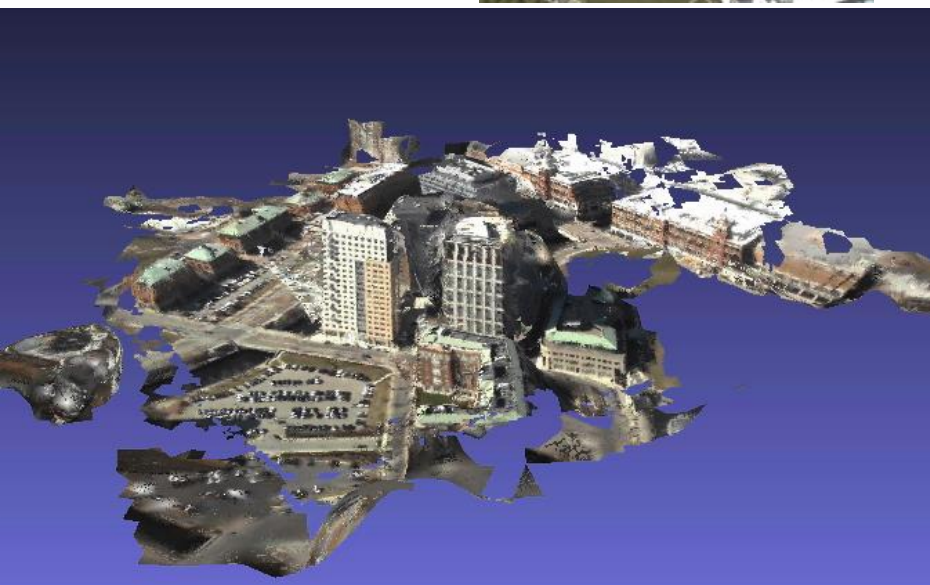
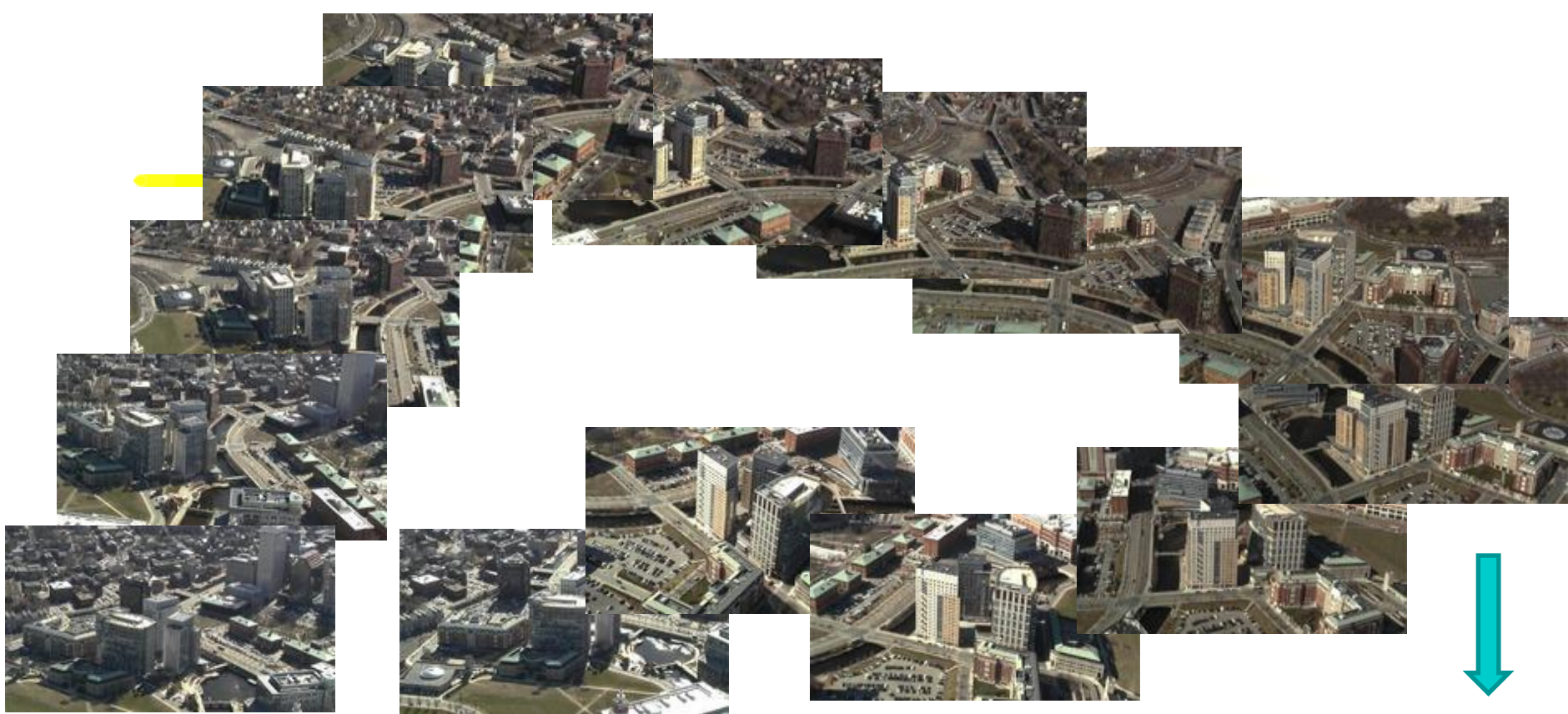
Multi-Camera



Images

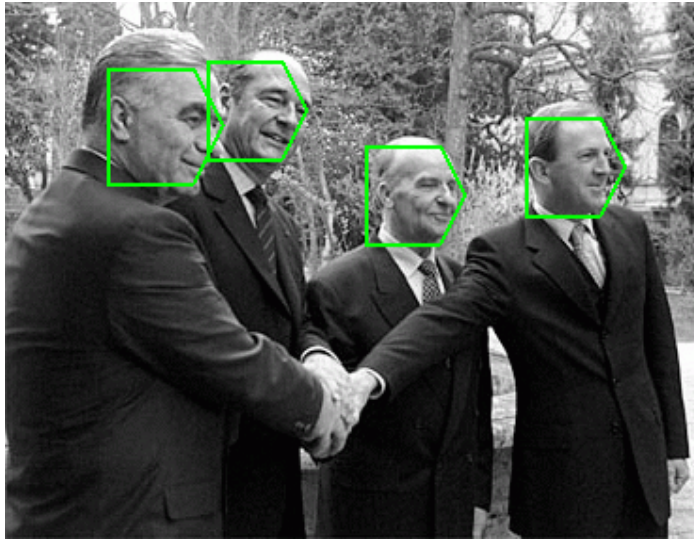


3D reconstruction



Object Detection and Recognition

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

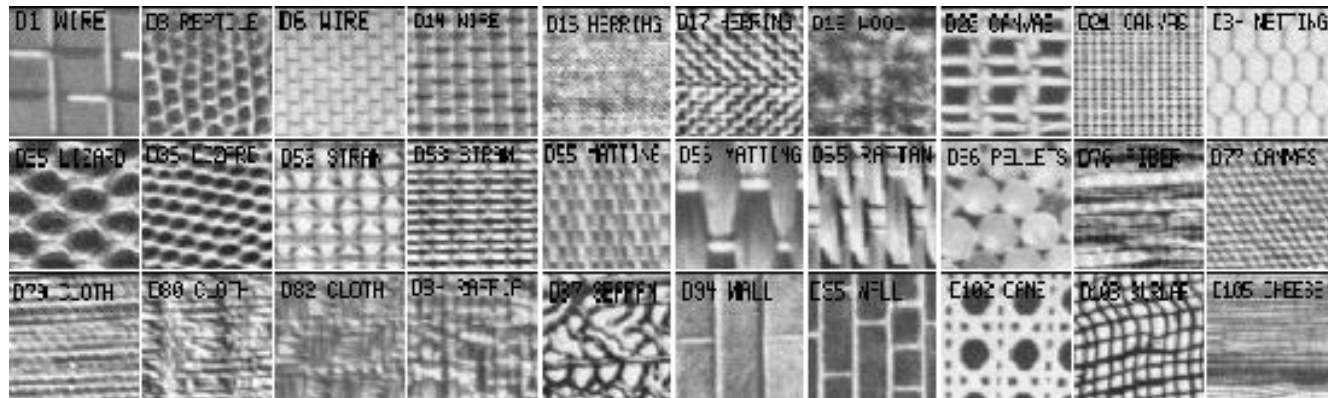


Faces



Cars

Texture analysis



Course web site

❖ <http://www.cs.ucsb.edu/~cs181b>

❖ Visit it often!

“Modern” Computer Vision

- ❖ Modern – learning-based, AI-ish vision is one of the hottest R&D areas in the past five years
- ❖ Recent advances in vision and speech understanding lead to the explosive growth in AI, neural network, and machine learning (topics not covered in ugrad 181b in 2014)
- ❖ Big \$\$\$ and job opportunities
- ❖ New data science focus and initiatives (AI bandwagon ☺)



Central Problems

```
01 00 05 00 03 00 02 00 00 03 01 01 01 01 00
03 30 3A 38 39 2D 1D 15 10 0E 0C 0A 0A 0A 09
03 22 1B 16 14 0A 08 0B 0A 0D 0B 0B 0C 06 07
00 0F 0B 04 10 07 09 07 08 09 09 08 05 08 08
00 07 09 0E 0C 07 08 0A 0A 0B 0F 0A 0C 07 06
00 0E 0C 09 09 08 08 07 08 09 09 0A 05 08 07
02 0C 0B 0A 05 08 09 0A 0C 0A 0A 08 0A 0A 06
02 0B 14 0F 0F 0D 0A 0E 0A 0C 0C 0E 0A 0C 0B
02 10 12 0B 10 0A 0D 0D 0B 0D 0C 0B 0B 0C 0D
01 15 0E 10 12 0C 0D 0C 0C 0A 0B 0B 09 0C 0F
02 12 10 0F 0E 10 10 0B 0C 0F 0F 0E 0C 10 0D
01 11 12 0C 11 13 10 10 0B 10 0F 0C 11 11 13
01 10 0F 0D 12 0E 10 0E 0F 13 13 11 13 17 11
02 0F 15 0D 18 11 0D 11 14 10 12 12 14 19 13
02 13 14 14 16 11 13 13 17 12 17 17 28 1E 1A
00 0F 16 0F 13 12 10 1D 12 21 15 1E 21 1F 1C
```

❖ What are the **essential, invariant** traits?

❖ What is essential?

☐ For human? For animals? For cars?

☐ Why we still recognize an object even if it is partially occluded (a person behind a tree? A bike occluded by a parked car?)

❖ Invariant to what?

☐ Lighting

☐ Movement (translational, rotational, scaling)

☐ Pose changes (both camera and object)

Central Problems

❖ How to represent such **essential, invariant** traits?

☐ Basis

- X, Y, Z
- Fourier, wavelet
- Body fitted basis

❖ Lead to

☐ Hand crafted

☐ Semi-Robust (invariant to certain environmental factor)

☐ Low-level, generic, and flat

☐ Still useful for geometrical problems (calibration, registration, 3D modeling) requiring precise correspondences

❖ Need

☐ Hi-level, domain specific, hierarchical, robust



Central Problems

❖ Hierarchical Basis Expansions

- ❑ Don't need to know what those bases are
- ❑ Don't need to know how many of them are needed
- ❑ Don't need to know how many layers of summarization

❖ Everything can be learned from data!!!

Central Problems

- ❖ What are such representations good for?
- ❖ Recognition
 - ❑ A well studied problem in ML, AI, PR
 - ❑ Nearest neighbor leads to linear decision boundary
 - ❑ Two major extensions:
 - Massage the classifier: neural nets
 - Massage the data: SVM
- ❖ Theoretically
 - ❑ Three layers of neural networks are sufficed

Modern CV

❖ End-to-end optimization

- ❑ General feature learning with
- ❑ General multi-layered neural network classifiers
- ❑ A simple $\text{err} = f(\text{feature})$ optimization with gradient descent

❖ Made possible by

- ❑ Large storage
- ❑ High speed internet (sharing of large amount of data)
- ❑ GPU