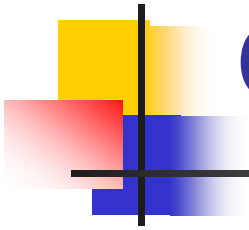
ECSE 4961/6650

Computer Vision

Qiang Ji

jiq@rpi.edu



Computer Vision

- Computer vision is concerned with modeling and replicating human vision using computer software and hardware. It provides machine with the ability of perception.
- It automatically processes and extracts information from images or videos to reconstruct, interpret and understand a 3D scene from its 2D images in terms of the properties of the scene objects.
- Computer vision can augment human vision. It will NEVER replace human vision.

Computer Vision

Make computers understand images and videos

Where is this?
What are there?
Who are there?
What are they doing?
How fast are they moving?
….

Computer vision vs human vision

What we see

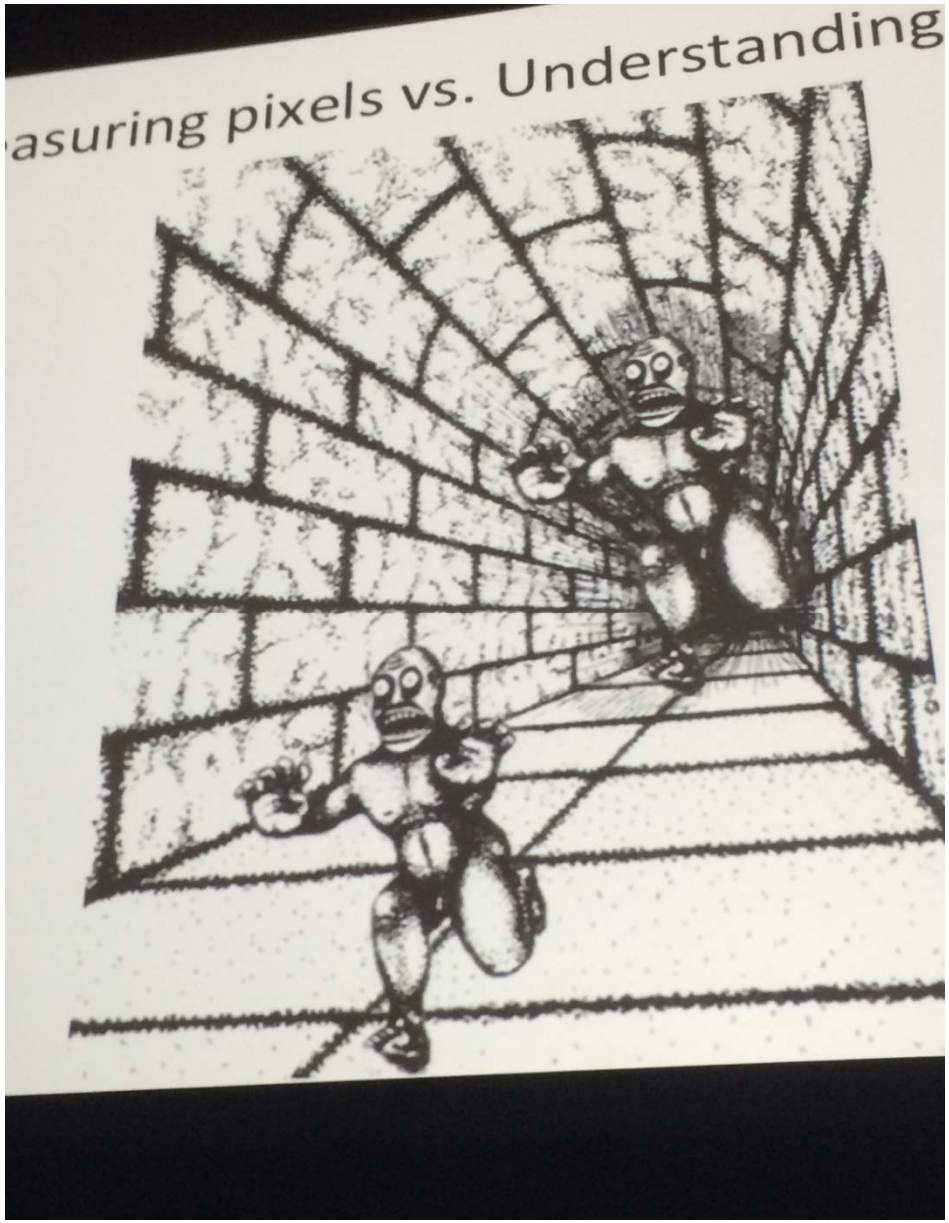
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

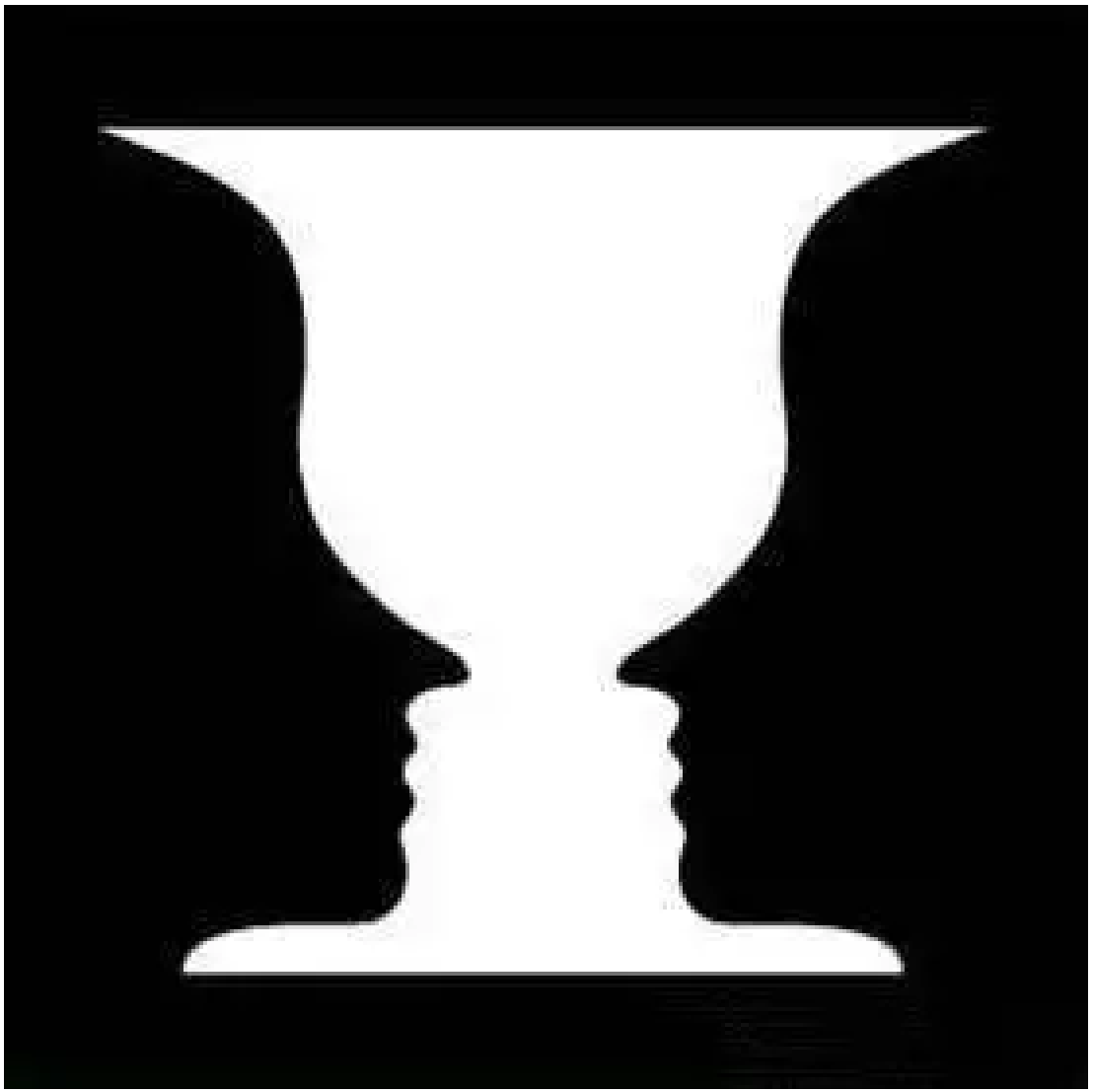
Human Vision v.s. Computer Vision

- Human vision
  - Very powerful
    - Recognize patterns under different illuminations and orientations
    - Visual cortex occupies about 50% of the brain. More human brain devoted to vision than anything else
  - Limitations
    - Limited memory-cannot remember a quickly flashed image
    - Limited to visible spectrum
    - Subjective and inconsistent
    - Optical illusion

Optical Illusion



Optical Illusion



The image patterns are different, depending on if you focus on the white or black part of the image

Human Vision v.s. Computer Vision (cont'd)

- Computer Vision
  - Limited recognition ability, not robust and is sensitive to noise, illumination and orientation variations
  - Often, tasks easy for human are often very hard for computers
  - Can be biased and unfair
  - But consistent and objective

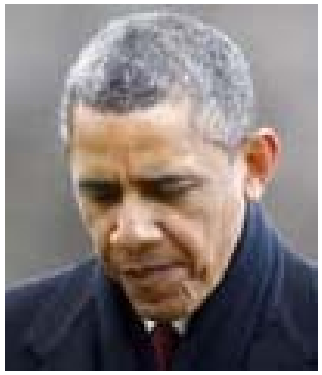
Computer Vision Limitations

- Computer vision often fails to recognize under different variations, including orientation, shape, appearance (illumination), and occlusion

Frontal

expression

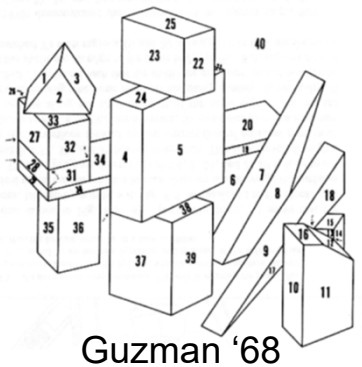
head pose & occlusion



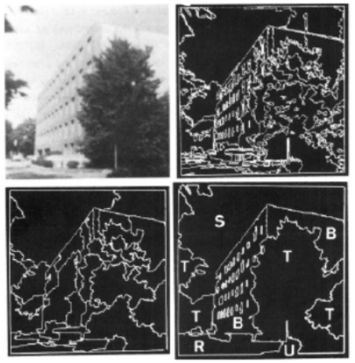
aging & head pose

Brief history of computer vision

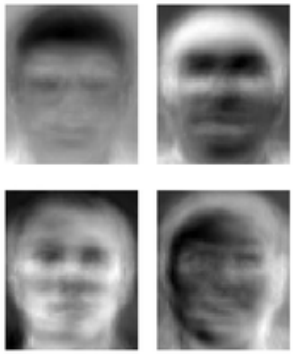
- 1966: Marvin Minsky assigns computer vision as an undergrad summer project (died on Jan. 26, 2016 at age of 88)
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts; vision & graphis; vision for HCI; internet vision, etc.
- 2010's (2012+) : Deep learning (CNN) and large dataset (ImageNet) dominate computer vision



Guzman '68

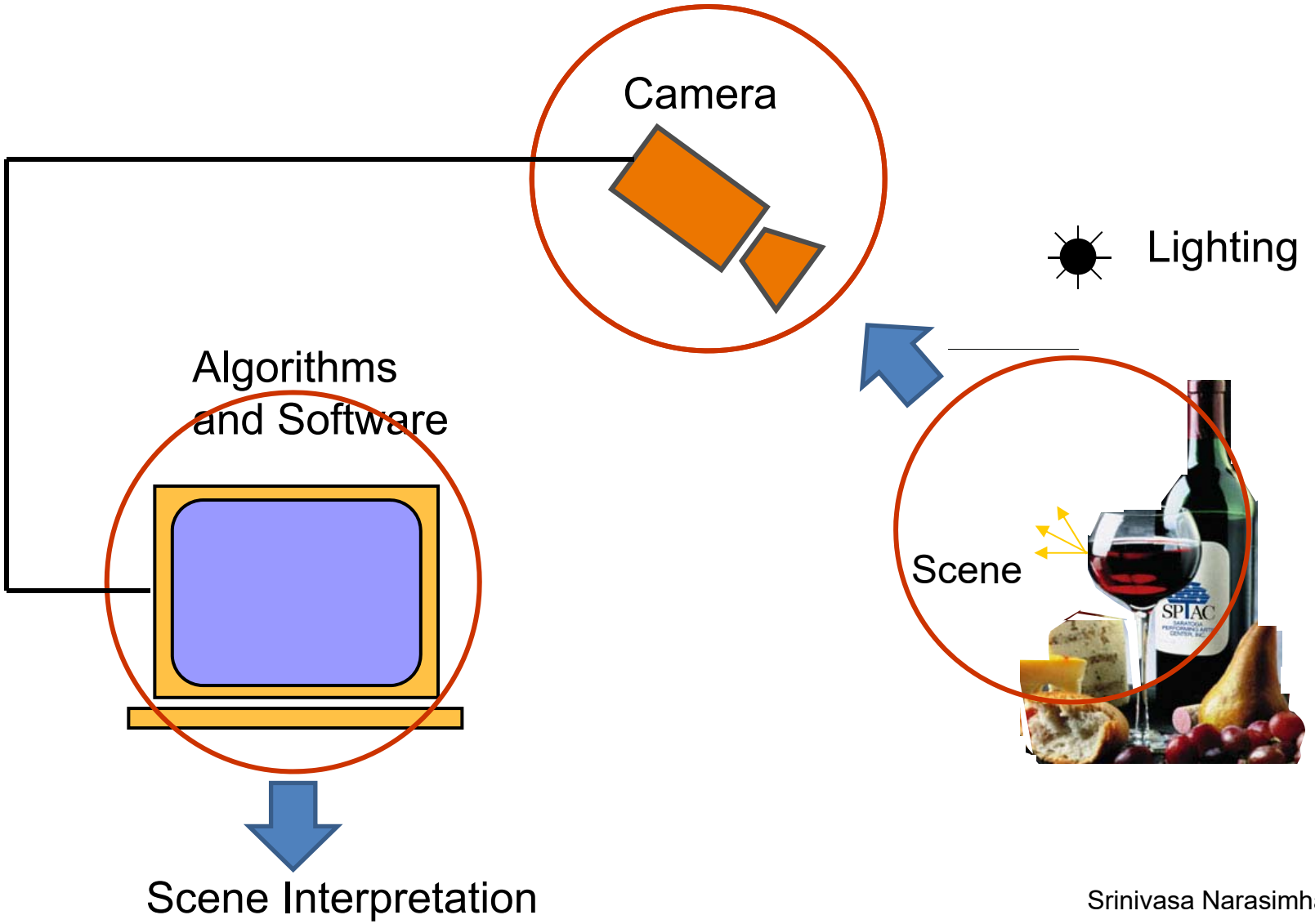


Ohta Kanade '78



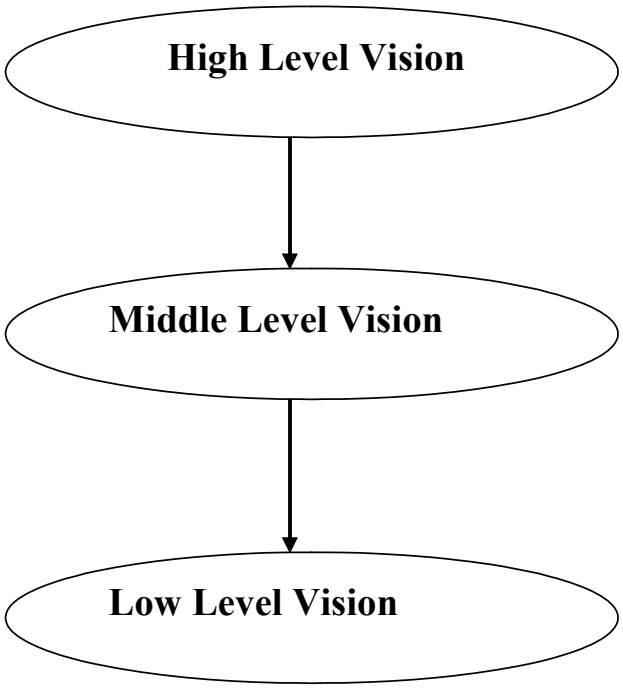
Turk and Pentland '91

Components of a computer vision system



Srinivasa Narasimhan's slide

Computer Vision Hierarchy



Scene content interpretation, and event recognition: human action/activity recognition

Scene object property computation: object detection, tracking, recognition, camera calibration, pose estimation, 3D reconstruction, etc..

Image processing and information extraction: image acquisition, image enhancement, feature extraction, and image segmentation.

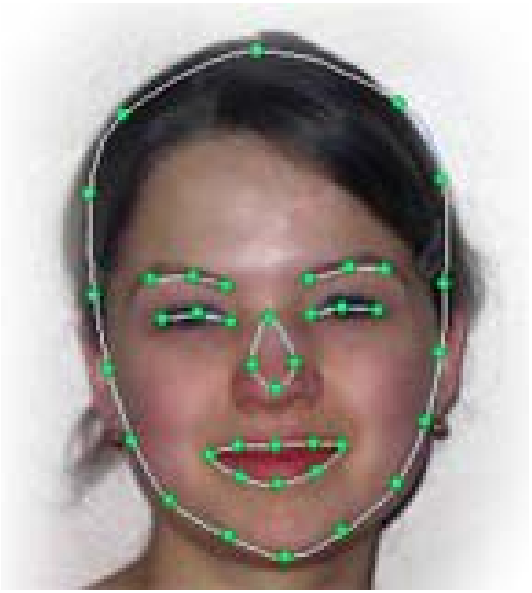
Low Level Computer Vision

Feature extraction: extract image features to represent an image or its content

Hand-crafted features:

- Corner, lines, circle or ellipses detection
- Scale-invariant Feature Transform (SIFT)
  - Invariant to scale, orientation, and illumination variation

Automatic features via deep learning



Facial point detection

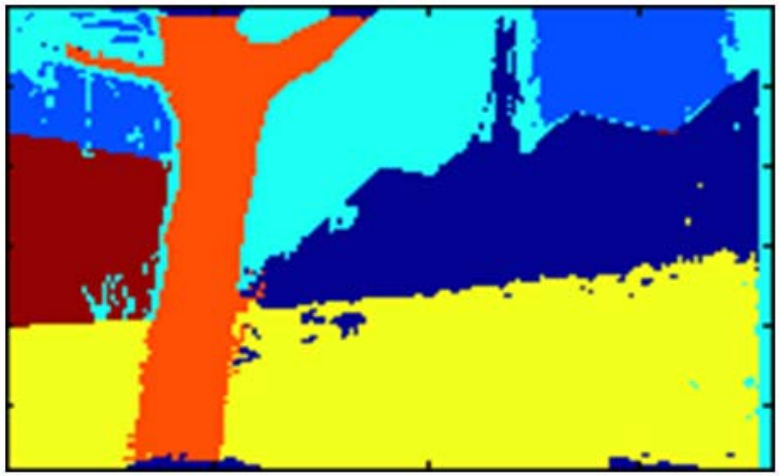
SIFT feature detection

Low level Vision

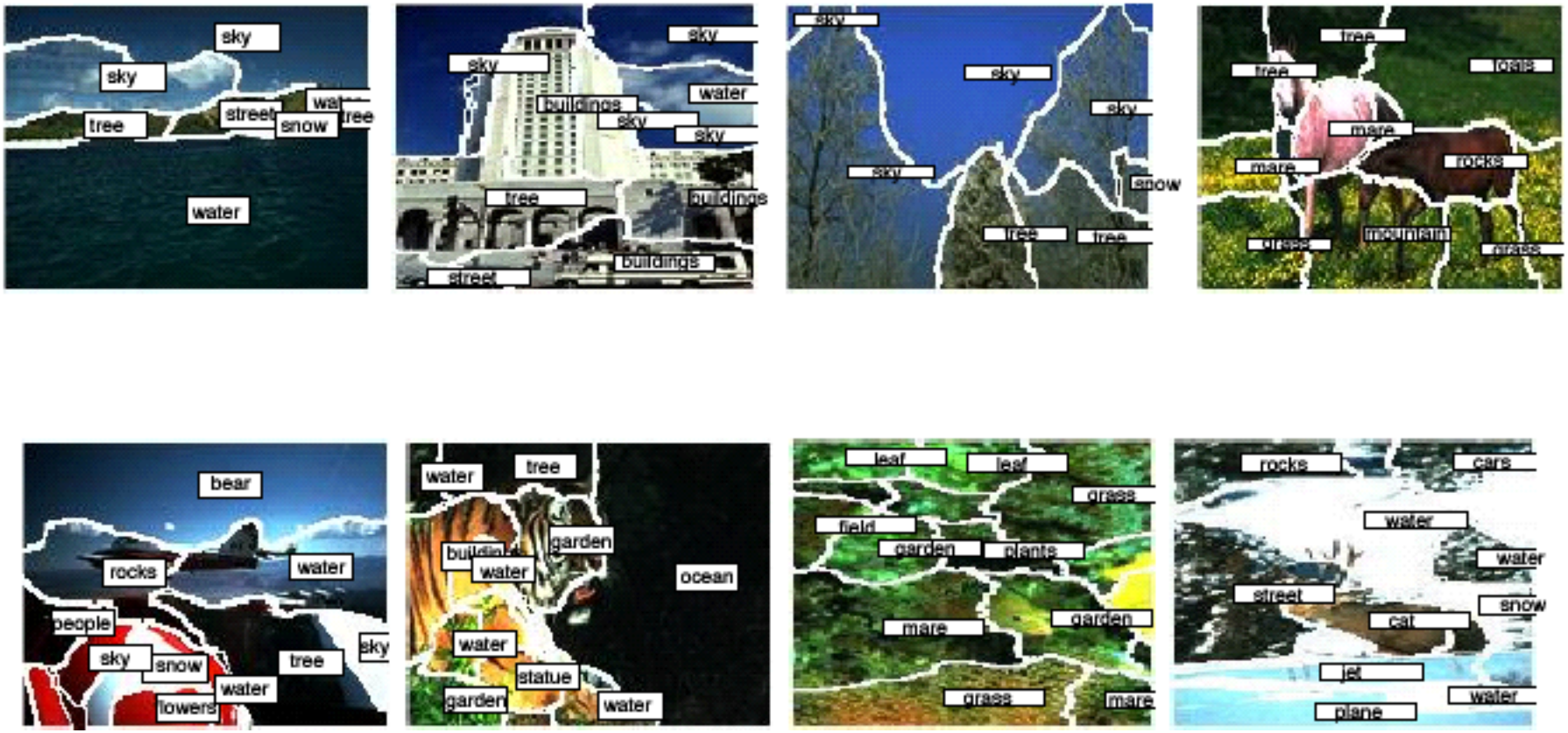
Image segmentation: Divide an image into homogeneous regions with respect to certain property

- Region based, edge based or hybrid methods
- Stochastic (Bayesian) or deterministic (active contour, level set)
- Model-based segmentation
- Interactive image segmentation
- Static images and video

Image Segmentation



Semantic Segmentation

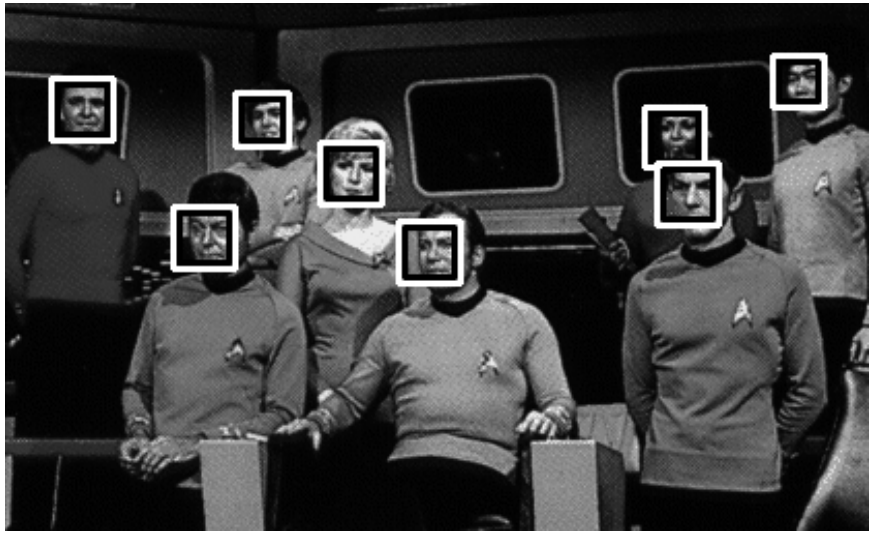


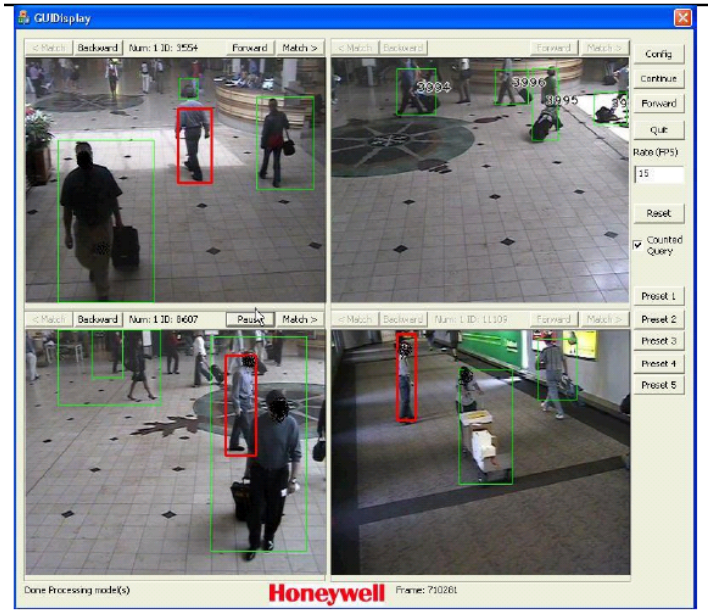
Middle Level Vision

Object detection: localize the objects in an image

- Holistic approach (e.g. template matching)
  - Intensity, PCA, LDA, Gabor features, deep learning features
- Local approach (e.g. the use of SIFT features)
  - Local feature selection
  - Local feature combination
  - Use local features and their spatial relationships
- Objects to detect
  - Face, human body, vehicles, etc..

Object Detection



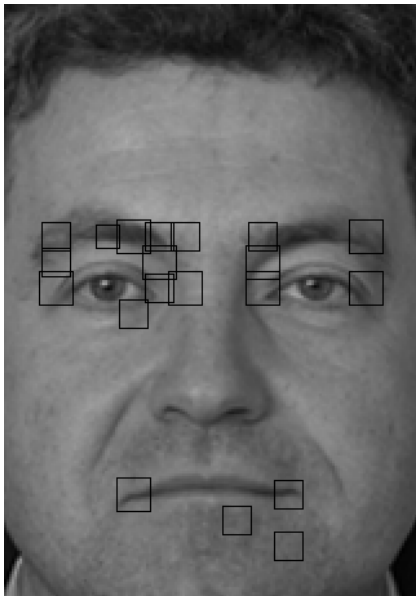


Middle Level Vision (cont'd)

Object recognition: determine the identity of an detected object

- Holistic v.s. local approach
  - Holistic: use entire image of the object
  - Local: use a set sub-images (patches) of the object
  - Features: Intensity, Gabor wavelet, PCA, and ICA, SIFT, DL features
  - Local method is more robust to changes in shape, illumination, pose, and background.
- Hard recognition v.s. soft recognition
  - Hard recognition: determine the identity such as face recognition
    - » Fine-grained recognition – recognize sub-category
  - Soft recognition: determine the object category such as gender, age or race classification.

Object Recognition

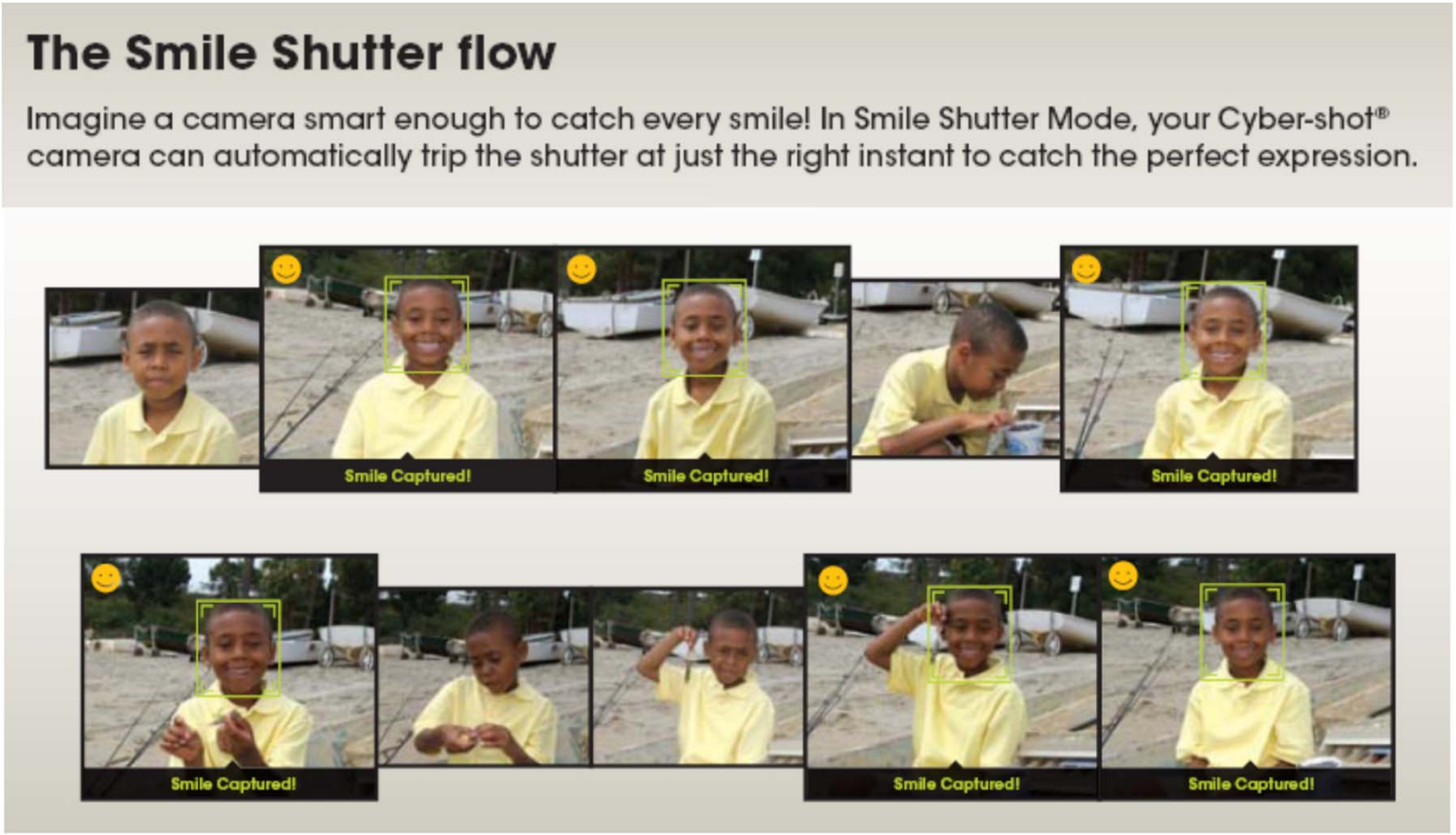


face recognition

object category recognition

Caltech 101

Smile recognition

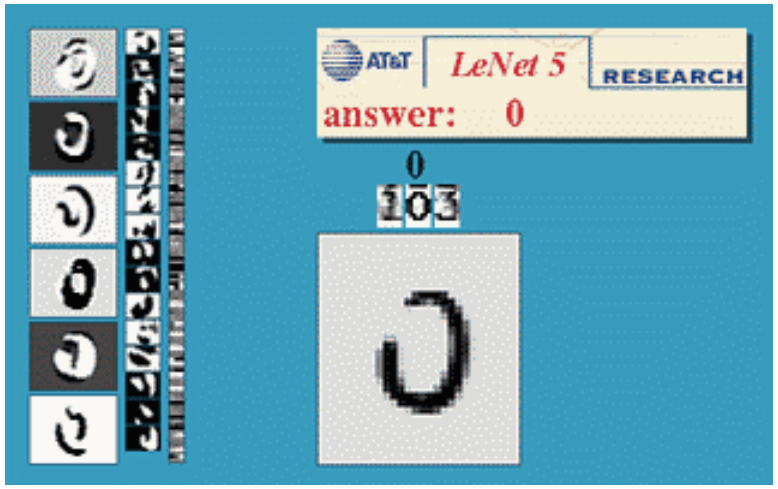


[Sony Cyber-shot® T70 Digital Still Camera]

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

Object recognition (in mobile phones)

Point & Find-Google Goggles (started in 2010 and discontinued on August 20, 2018)

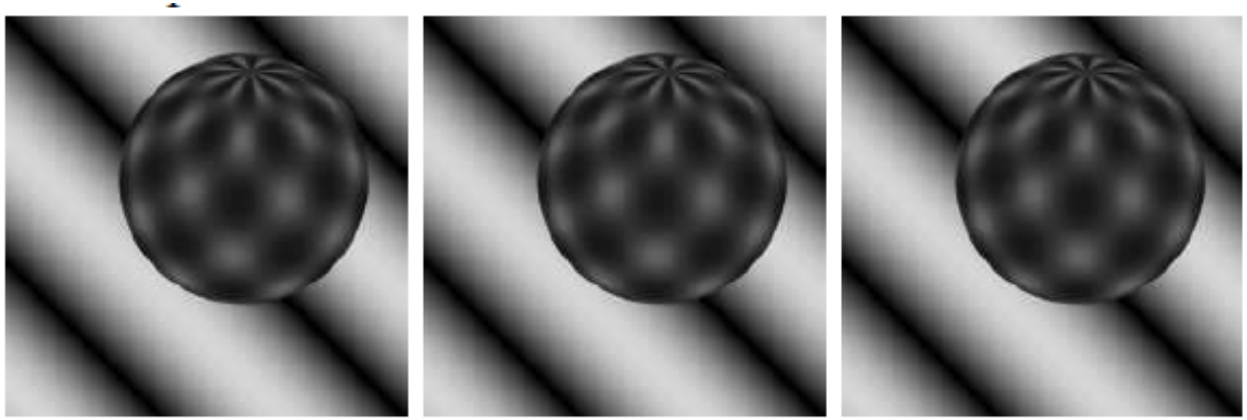
Middle Level Vision (cont'd)

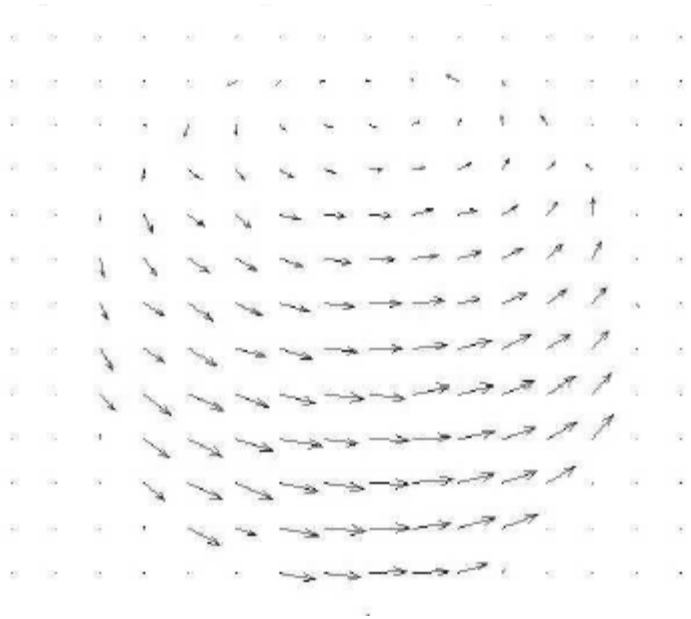
Motion Analysis: estimate the 2D or 3D motion of the object

- Optical flow estimation
  - Estimate the 2D pixel motion

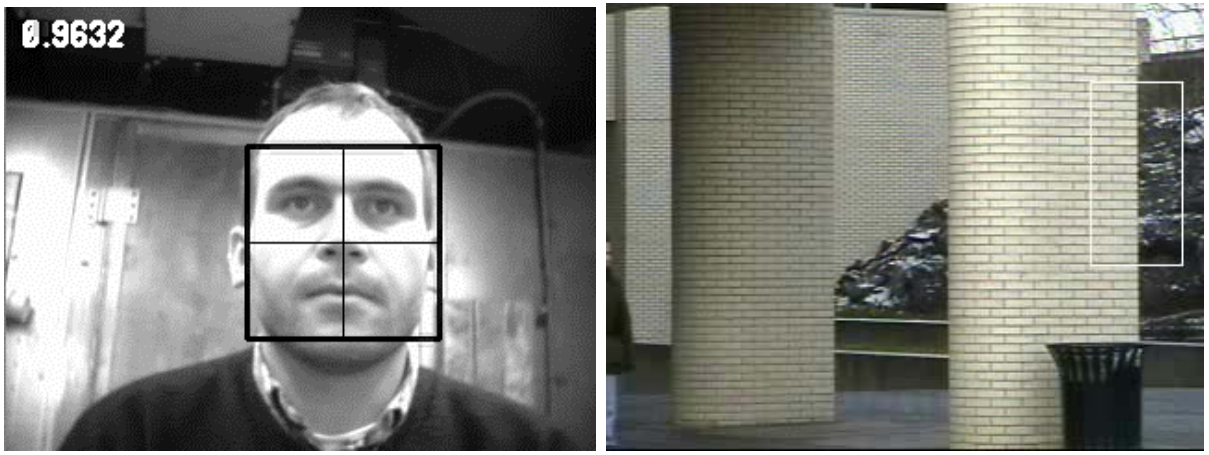
- Tracking: track a target over frames
  - Kalman filtering
    - » Track one object with Gaussian and linear assumptions
  - Particle filtering (condensation)
    - » Track multiple objects, no linear or Gaussian assumptions but computationally more intense.

Optical Flow Estimation

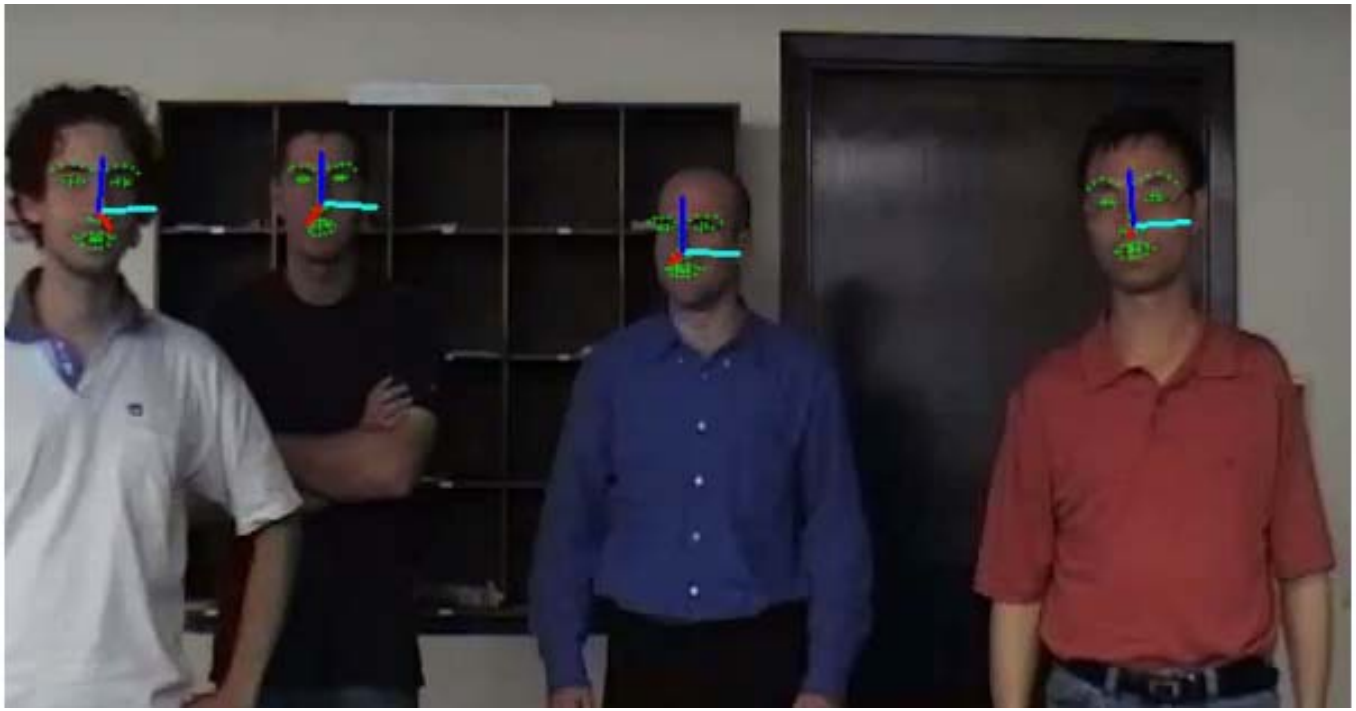




Human Face and Body Tracking

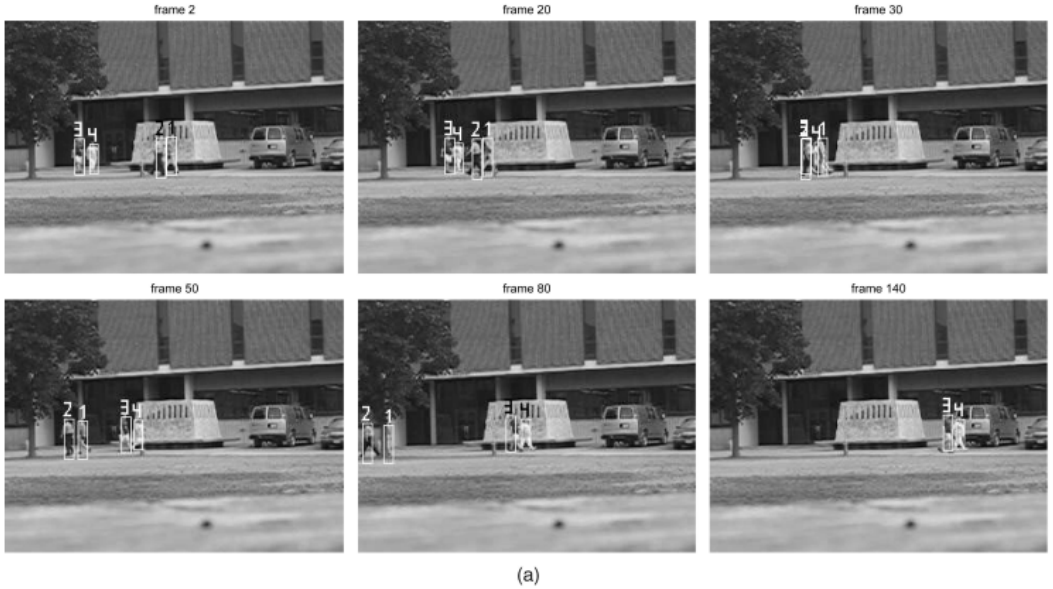


Facial Behavior Tracking



https://www.ecse.rpi.edu/~cvrl/Demo/demo_landmark_tracking_expression_multi.mp4

Multi-people Tracking

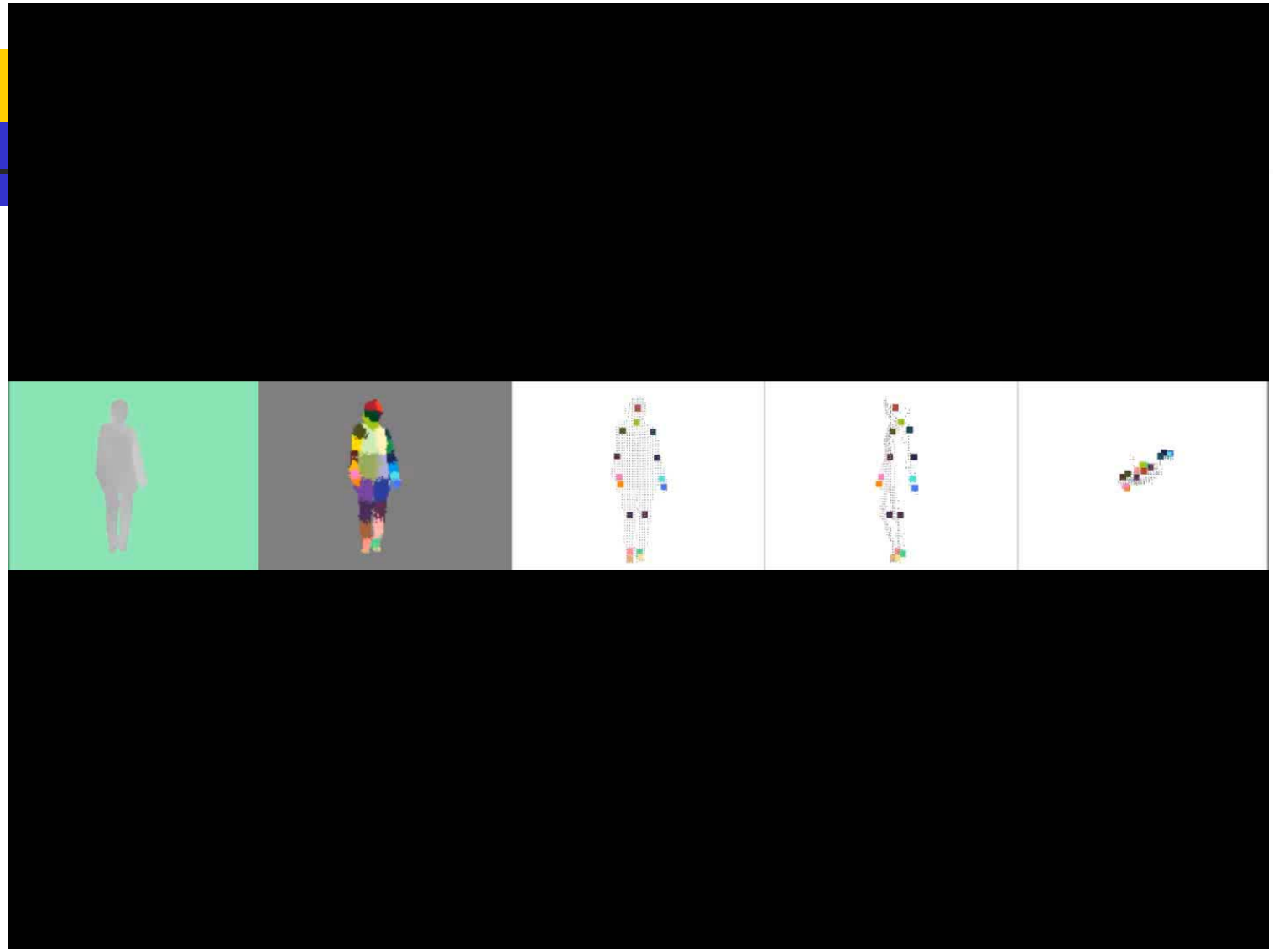


(a)

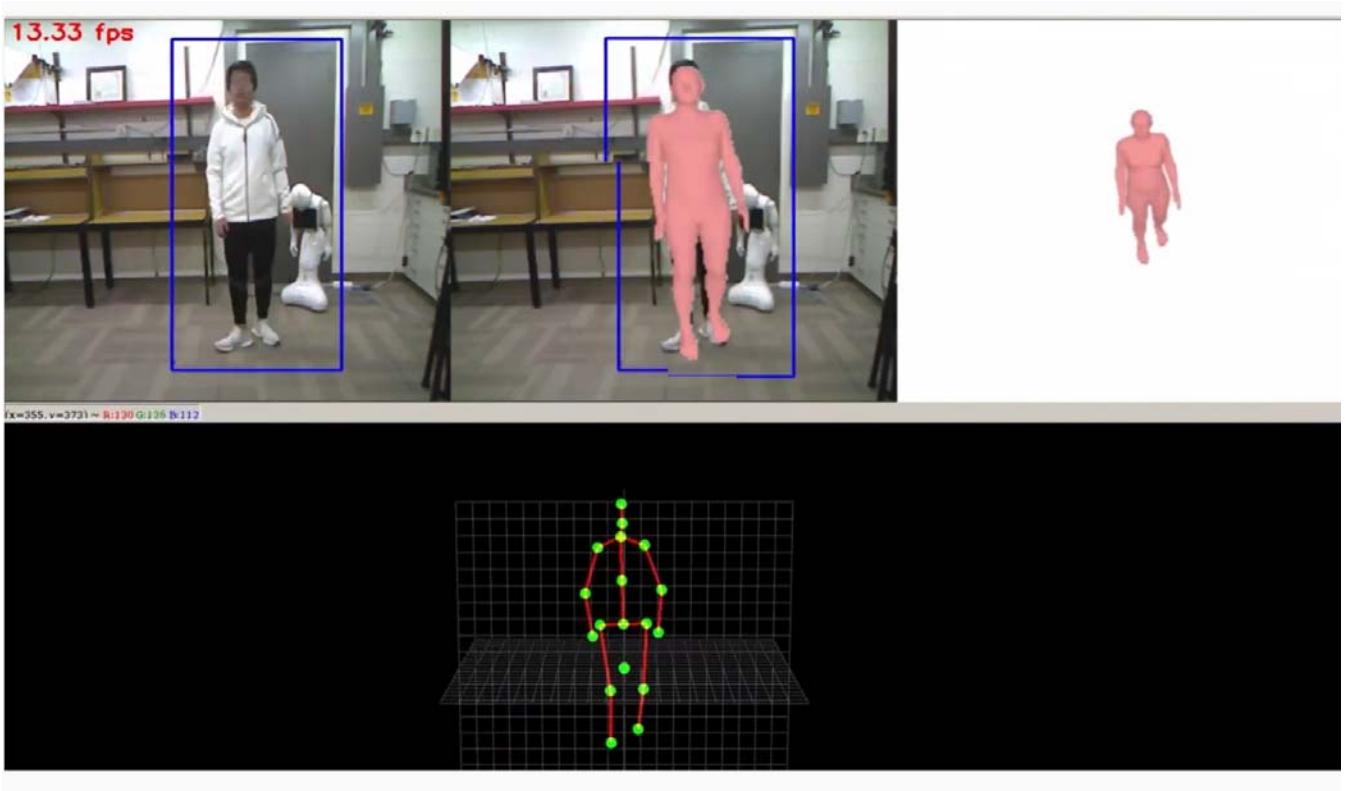
(b)

Proposed algorithm

3D Body Pose Tracking

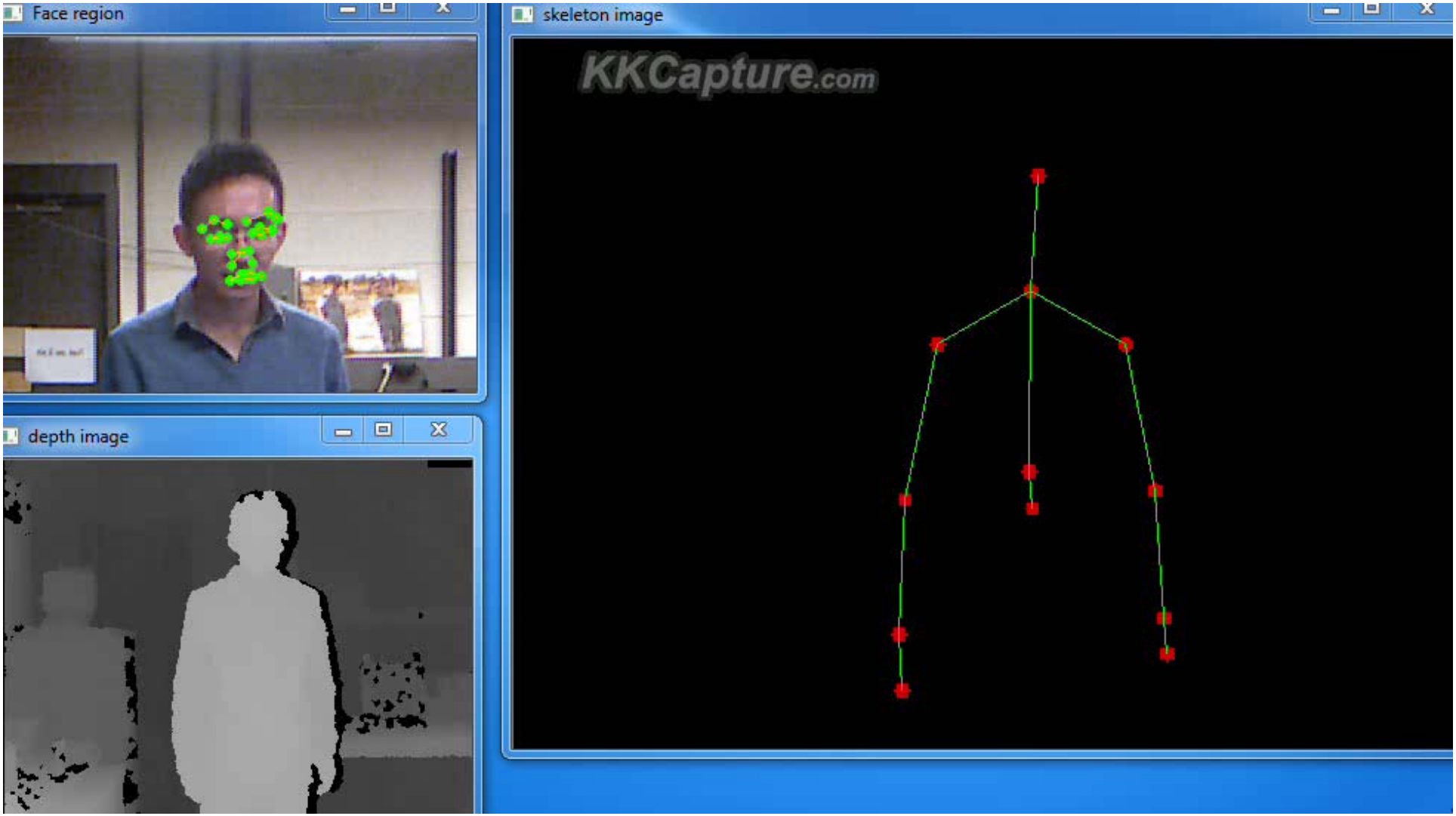


3D Body Mesh Reconstruction



https://youtu.be/UswZElQQCIs

Joint Face and Body Tracking



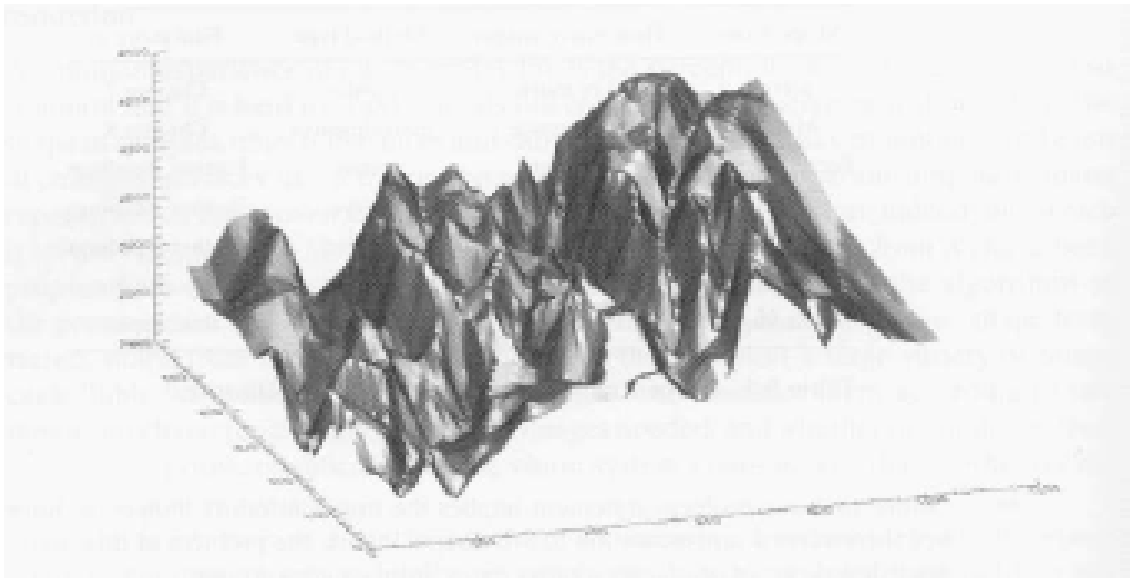
Middle Level Vision (cont'd)

3D Reconstruction: reconstruct the 3D shape or geometry of an object from its images

- 3D reconstruction from single image
  - Shape from X (shape from shading, texture, focus,)
  - Photometric stereo.
- 3D reconstruction from two images
  - Passive stereo-from two images
  - Active stereo-one image
- 3D reconstruction from a sequence of images
  - Structure from motion

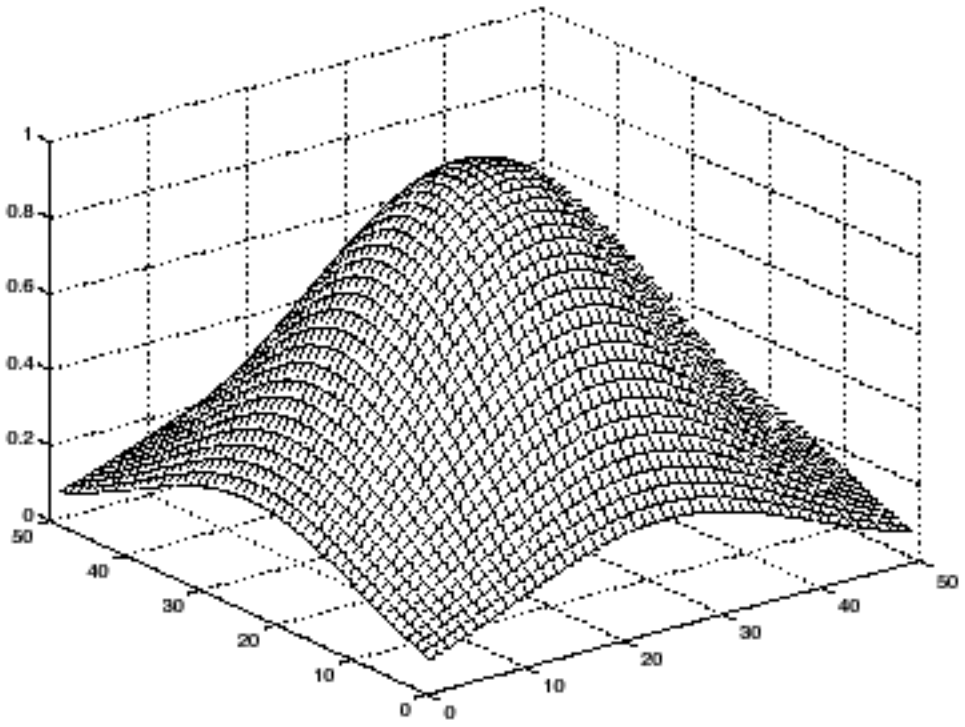
Shape from Shading

- Recover the 3D shape of an object from its image based on the intensity information



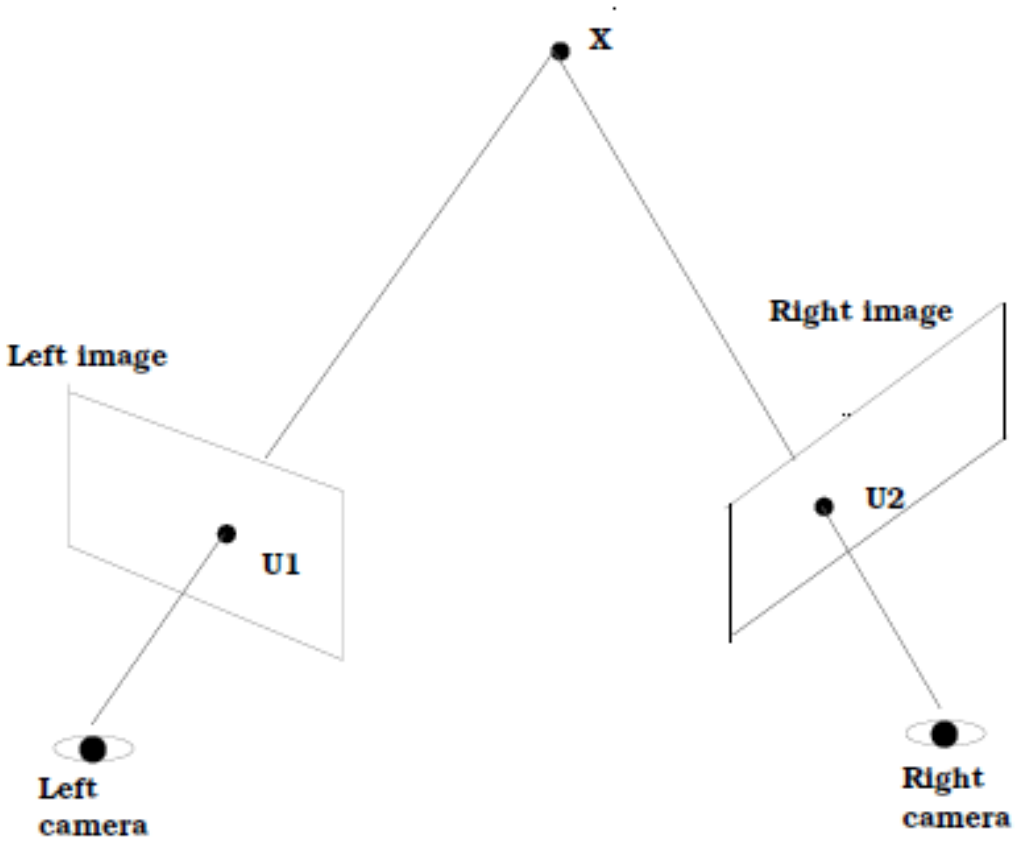
Shape from texture

Recover the 3D shape of an object using the textural features extracted from the image

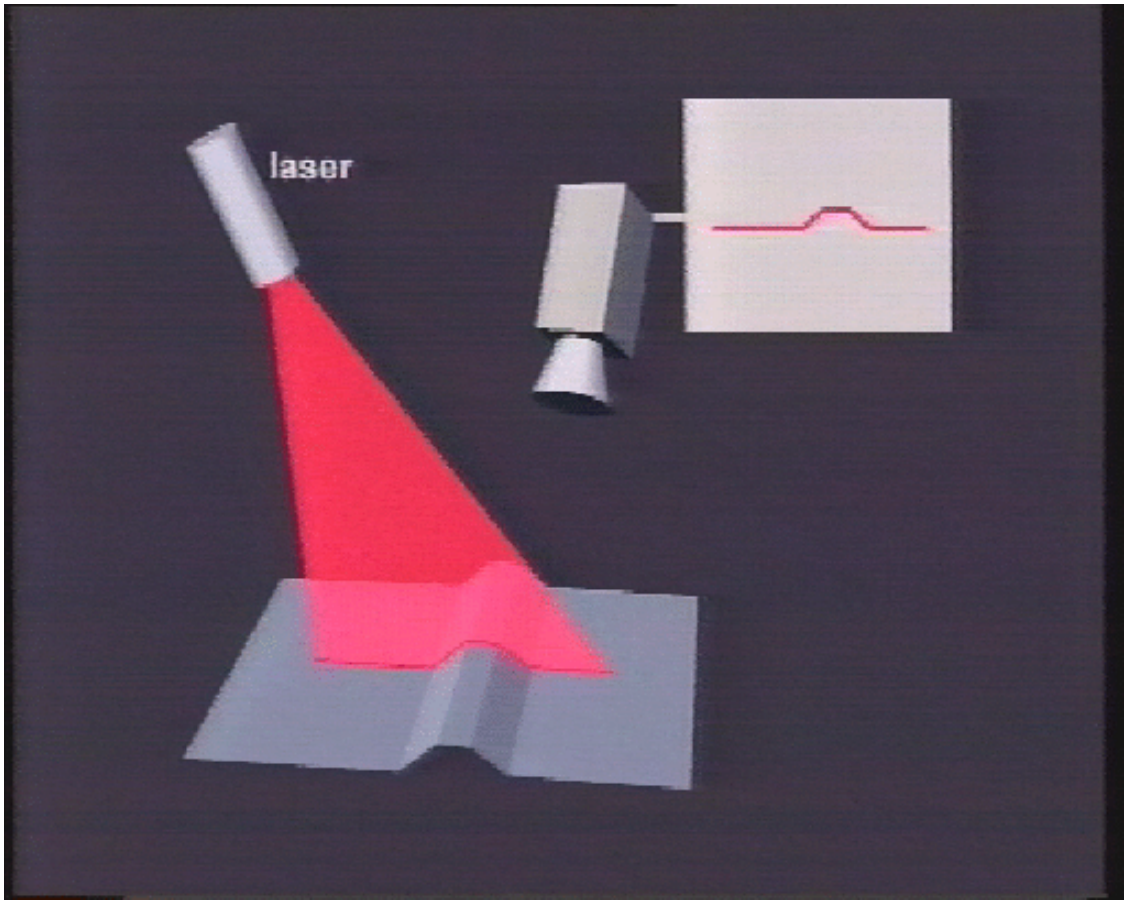


Passive Stereo

- Theoretical basis : triangulation

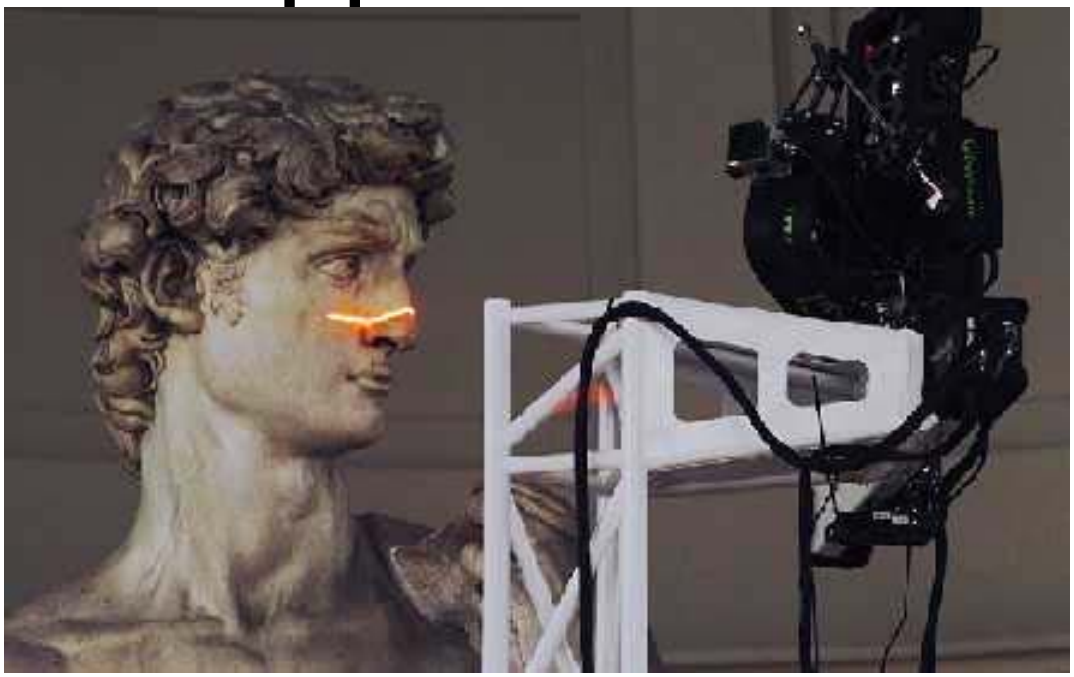


Active Stereo



Active manipulation of scene: Project light pattern on object. Observe geometry of pattern via camera → 3D geometry

Applications: 3D Scanning



Scanning Michelangelo's *"The David"*

- The Digital Michelangelo Project
  - http://graphics.stanford.edu/projects/mich/
- UW Prof. Brian Curless, collaborator
- 2 BILLION polygons, accuracy to .29mm

High Level Vision

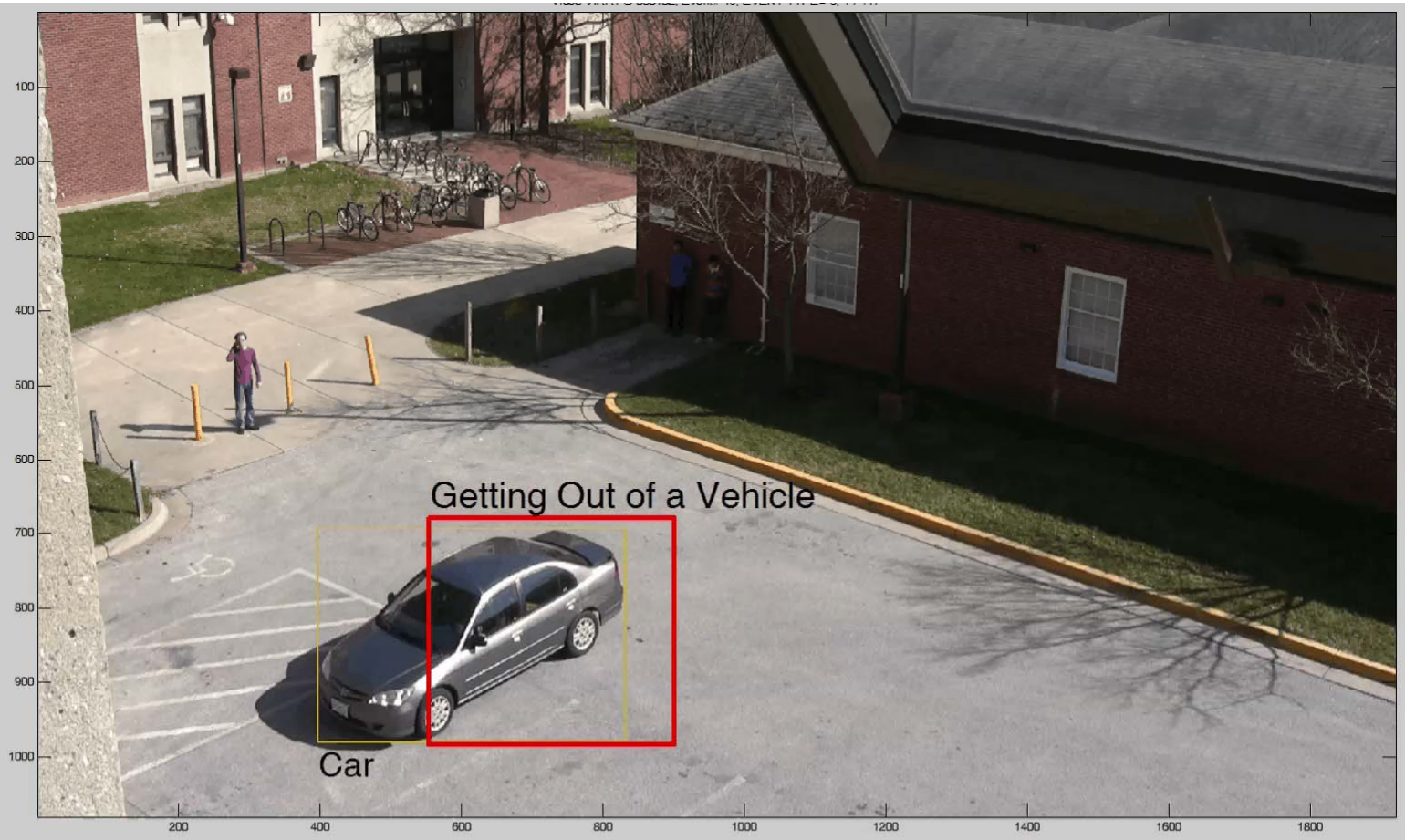
- Action recognition
  - Individual action v.s. group actions
    - Walking, running, crawling, digging, meeting, etc..
  - Probabilistic approach (HMM and variants) v.s. deterministic approach (Context-Free Grammar)
- Activity/event recognition
  - Require to recognize interactions among entities
  - Require to use context to help disambiguate
- Object function and intent recognition
  - Determination of the purpose and function of an object based on its attributes and the attributes of the objects it interacts with
  - Often need dynamic features

Human Action Recognition

(Show px12_20080409_007-activitymosaic.avi

Human Activity Recognition

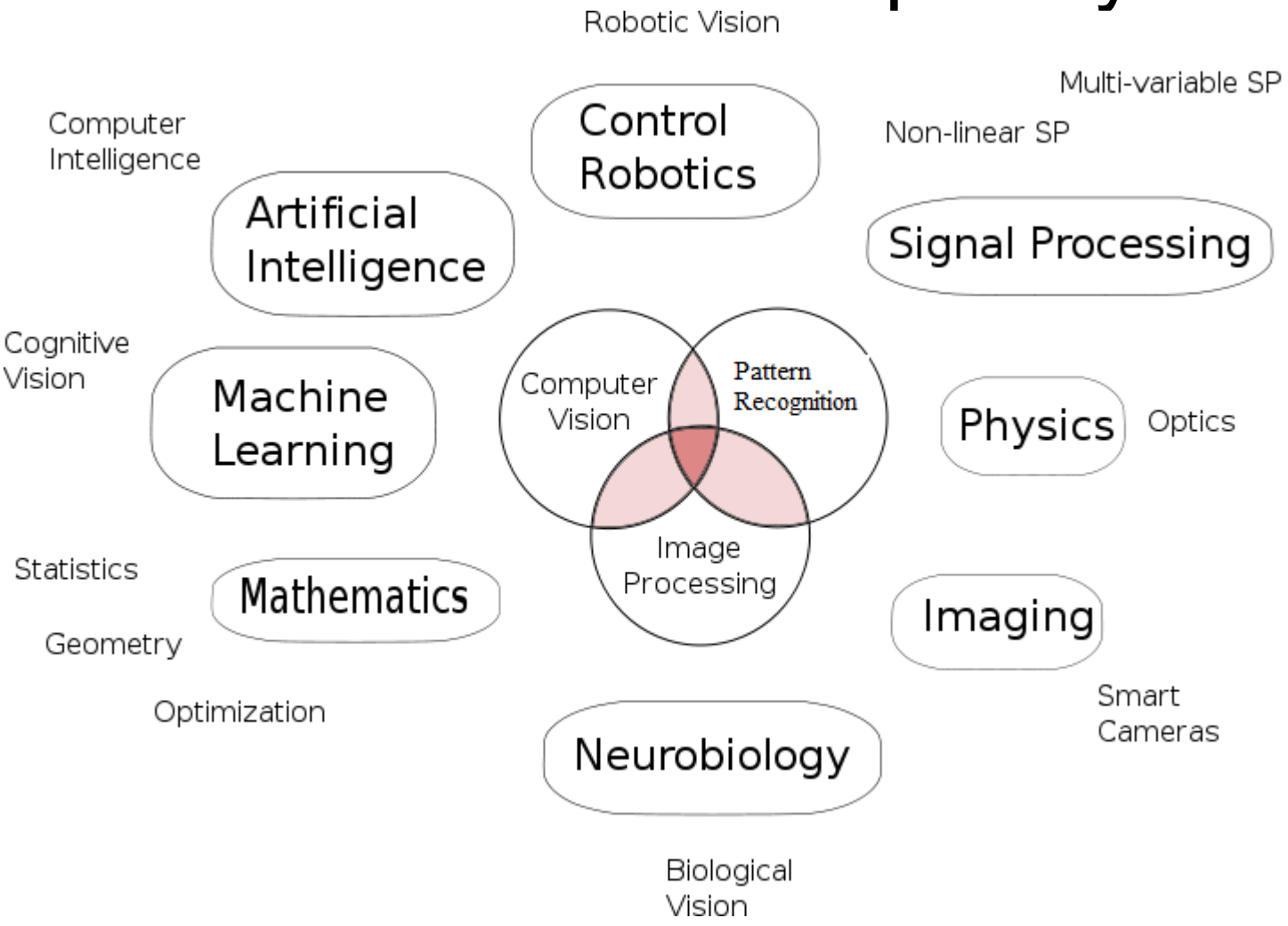
Human Event and Action Recognition



Why is computer vision difficult?

- Inverse problem (2D -> 3D)
- Ill-posed
- High-dimensional data
- Noise
- Large variation due to illumination, pose, shape, and occlusion
- Insufficient data

Vision is multidisciplinary



From wiki

Related Fields

Computer vision overlaps significantly with

- Image processing
- Pattern recognition (machine learning)

Image Processing

- Image processing studies image-to-image transformation. The input and output of image processing are both images. Typical image processing operations include
  - image compression
  - image restoration
  - image enhancement
- Most computer vision algorithms usually assumes a significant amount of image processing has taken place to improve image quality.

Computer Vision

- Computer vision is the construction of explicit, meaningful descriptions of physical objects from their images.

- The output of computer vision are a description or an interpretation or some quantitative measurements of the structures in the 3D scene.

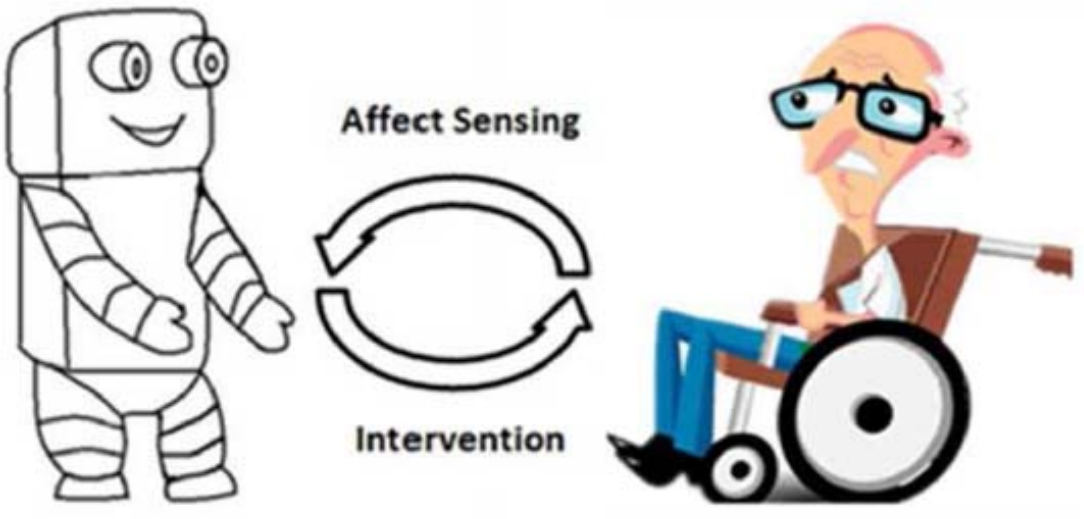
Pattern Recognition

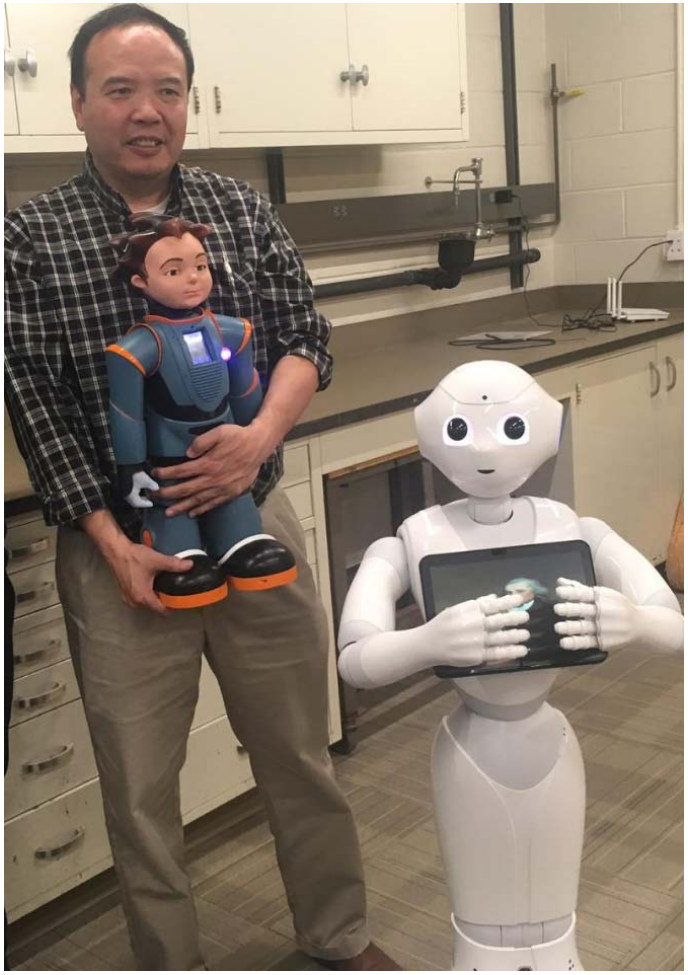
- Pattern recognition (also called machine learning) studies various mathematical techniques (such as statistical techniques, neural network, support vector machine, deep models, etc..) to classify different patterns.
- The input data for pattern recognition can be any data and the output is typically symbolic labels.
- Pattern recognition techniques are widely used in computer vision. Many vision problems can be formulated as classification problem.

Applications

- Robotics
- Human and computer interaction
- Biometrics and security
- Games and entertainment
- Transportation
- Medicine and health
- Image/video databases
- And many more ….

Human Robot Interaction (HRI): companion robot





HRI Demos

Pepper robot
https://youtu.be/DZnSswYGlgo

Facial behavior mirroring

https://youtu.be/H4b8MuT9Ecg

Other Robotics Applications

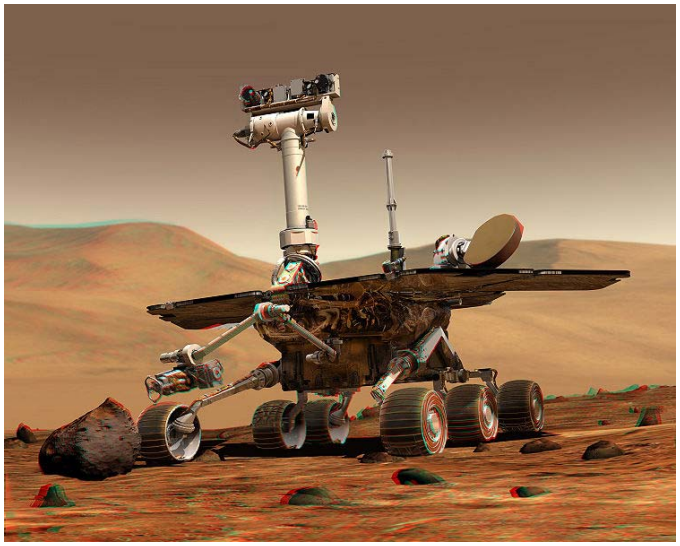
- Localization-determine robot location automatically (e.g. Vision-based GPS)
- Obstacles avoidance
- Navigation and visual servoing
- Assembly (peg-in-hole, welding, painting)
- Manipulation (e.g. PUMA robo manipulator)

Industrial robots

Vision-guided robots position nut runners on wheels

Real time visual servoing for robot grasping

Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

http://www.robocup.org/

Amazon warehouse robots

Human Computer Interaction

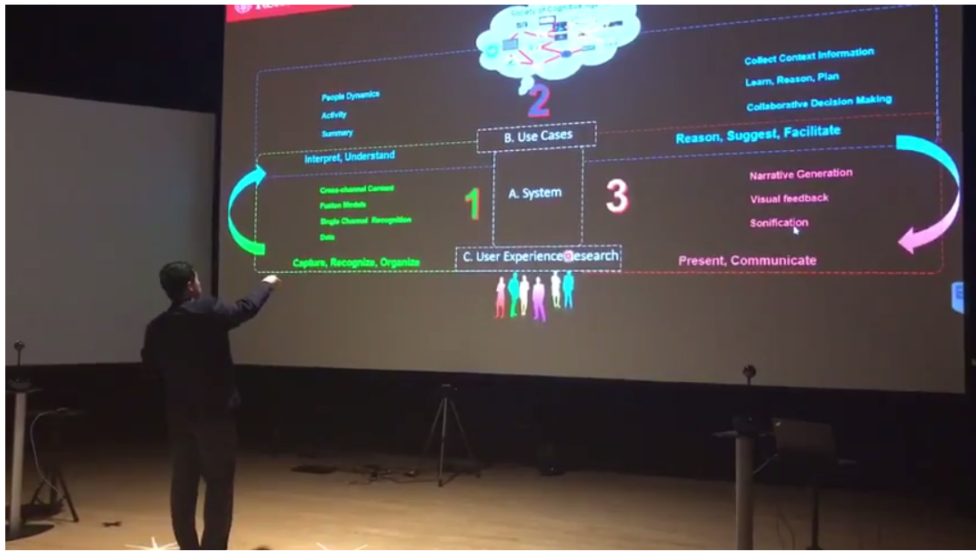
Naturally interact with computer through

- Hand gesture
- Body gesture
- Face movement and facial expression
- Eye movement and gaze

Human computer Interaction

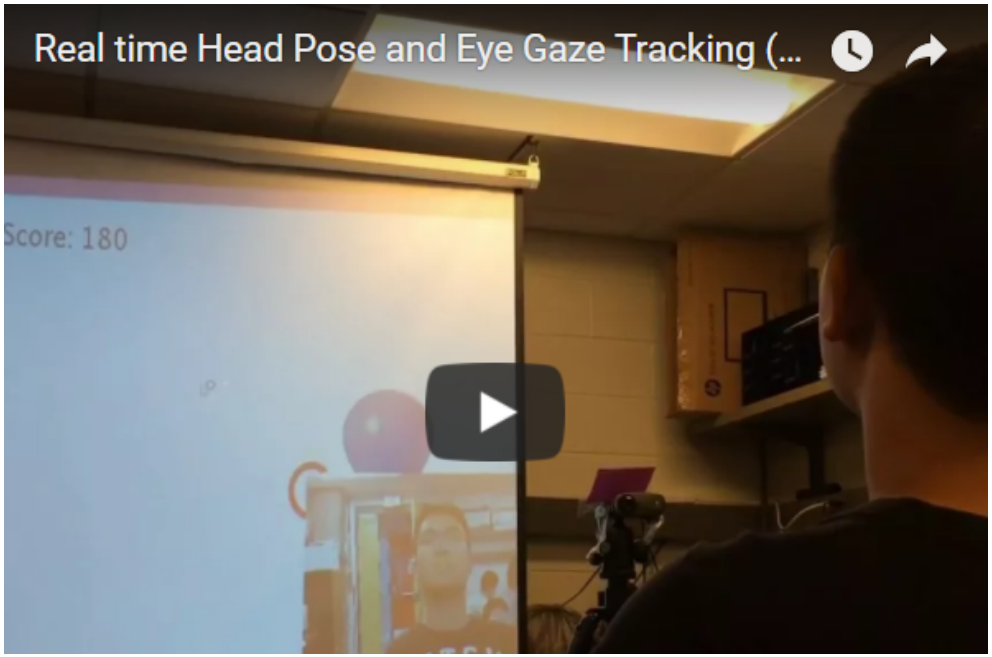
- Interaction with body gesture via Kinect

Interaction with Hand and Body Gestures



https://www.youtube.com/watch?v=m9g7mXaKstw&pbjreload=10

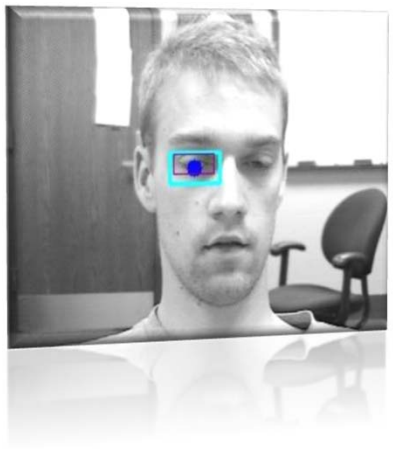
Interactions with Head Movement

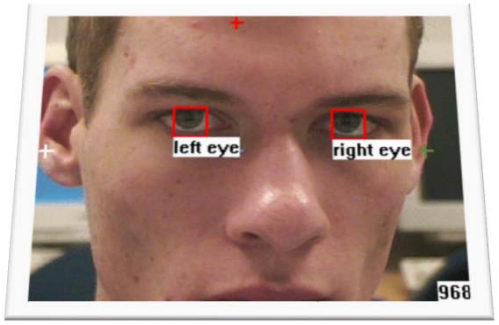


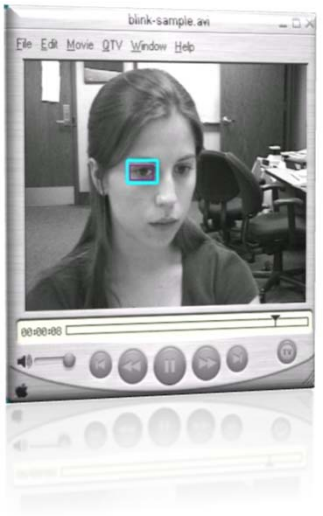
https://youtu.be/a12allAfnK4

Communication with eye movements

- Use eye gaze to point or aim and use eye blink to activate such as eye mouse
- Eye movements can also communicate one's emotion (raising eye brows)



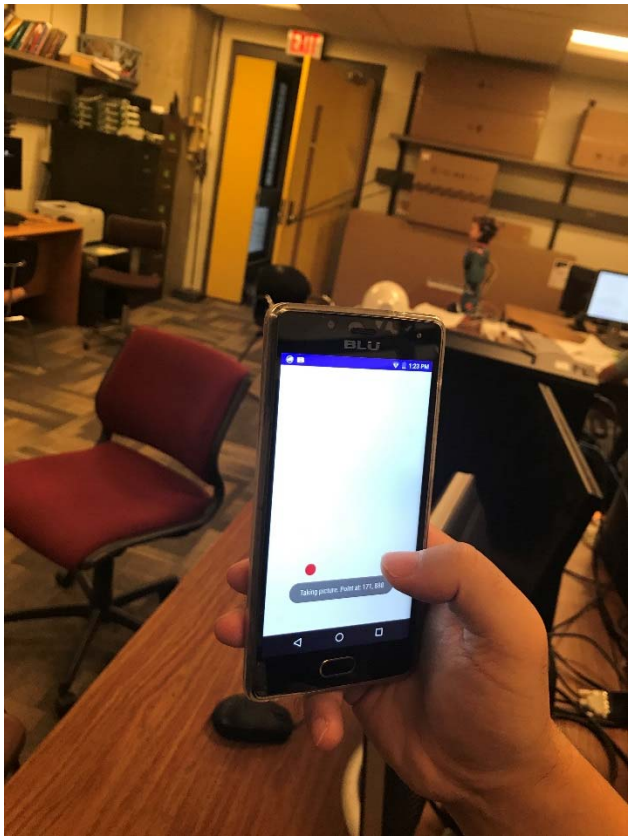


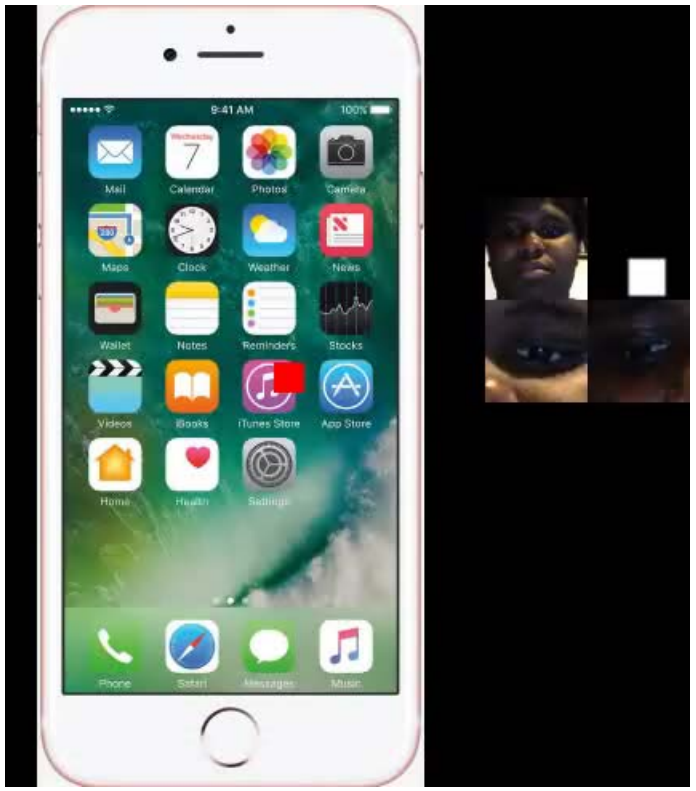


Eye Gaze Tracking

https://www.youtube.com/watch?v=4uH1_2qjbtA

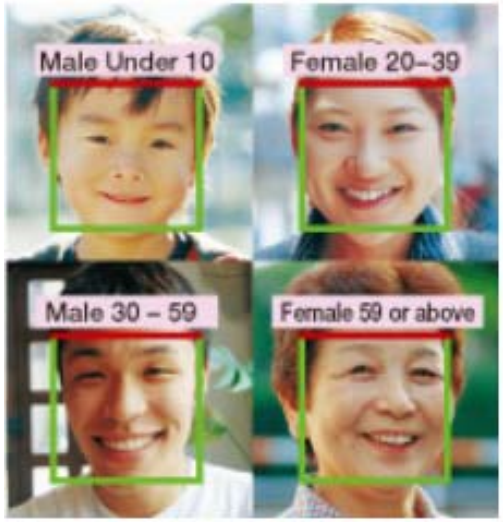
Mobile Eye Gaze Tracking





Biometrics and Security

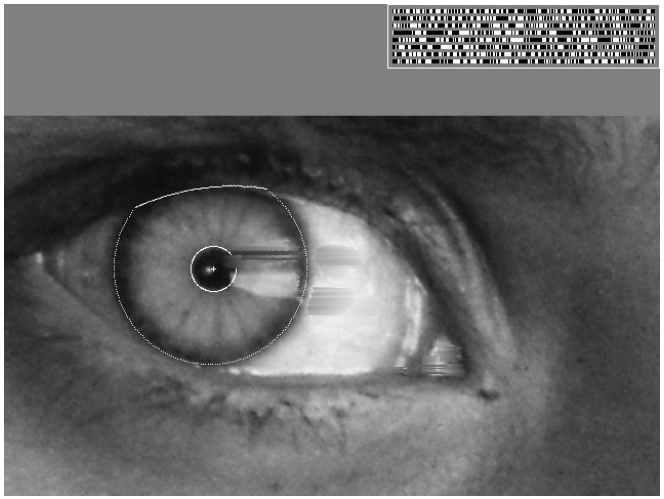
- Facial recognition
- Gender classification
- Age estimation
- Ethnicity classification

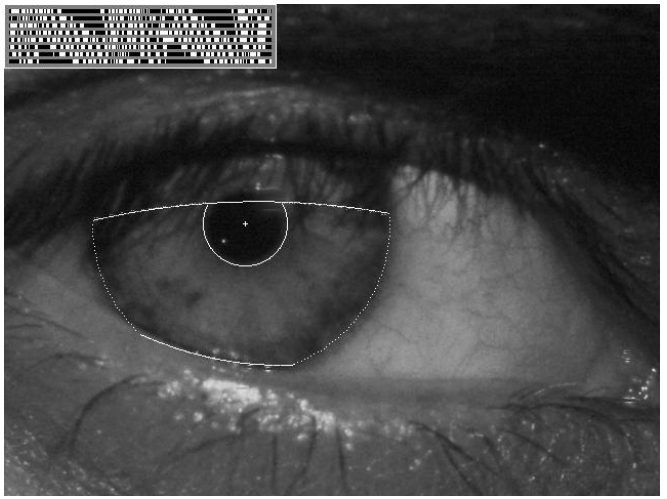


Asian Non-Asian

Vision-based biometrics

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





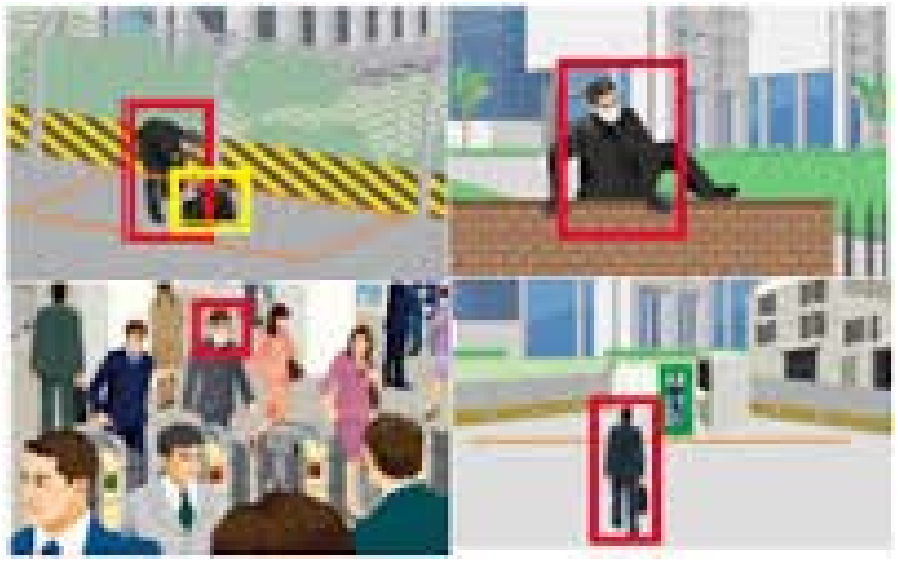
Login without a password…

Fingerprint scanners on many new laptops, other devices

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

Security

- Surveillance-detecting certain suspicious activities or behaviors



Vision for Games and Entertainment

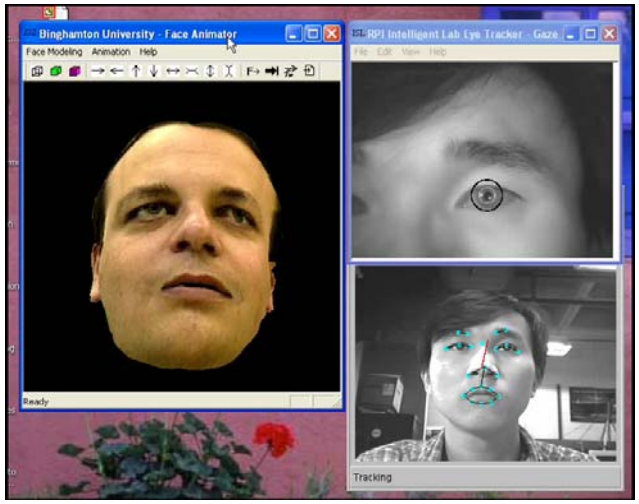
Nintendo Wii has camera-based IR tracking built in. See Lee's work at CMU on clever tricks on using it to create a multi-touch display!

Digimask: put your face on a 3D avatar.

"Game turns moviegoers into Human Joysticks", CNET
Camera tracking a crowd, based on this work.

Facial Motion Capture and Animation

- Facial motion includes eye movement tracking, facial muscle movement tracking, and head movement tracking



Body Motion Capture for Games

Examples of Motion Capture for Movies

Transportation

- Autonomous vehicle (self-driving)
- Driver behavior monitoring
- Augmented driving

Google cars

http://www.nytimes.com/2010/10/10/science/10google.html?ref=artificialintelligence

Self-Driving

Driver Behavior Monitoring

Normal

Eating

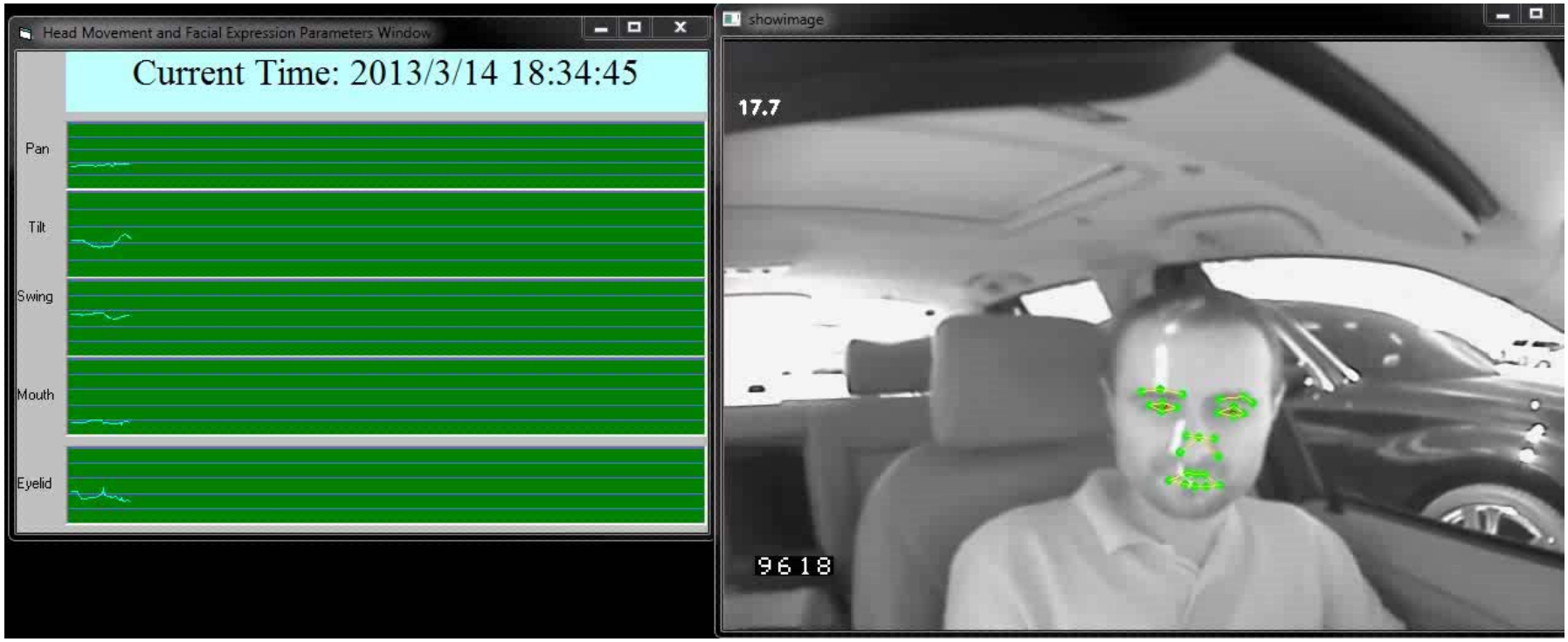
Calling

Makeup

Texting

Adjusting radio

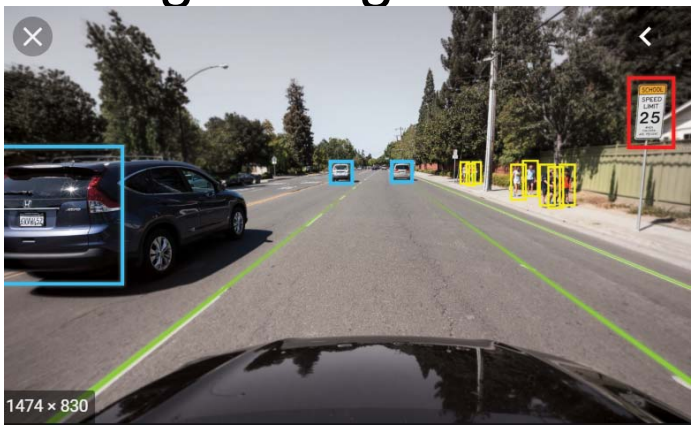
Driving Behavior Tracking



http://www.ecse.rpi.edu/homepages/cvrl/driving_data_tracking_demo_mpeg4.avi

Augmented Driving

- Advanced self-driving combined with driver behavior monitoring
- OpenPilot from Comma.AI
  - Self-driving through outward looking cameras



  - Driver behavior monitoring via inward looking cameras

Medicine Applications

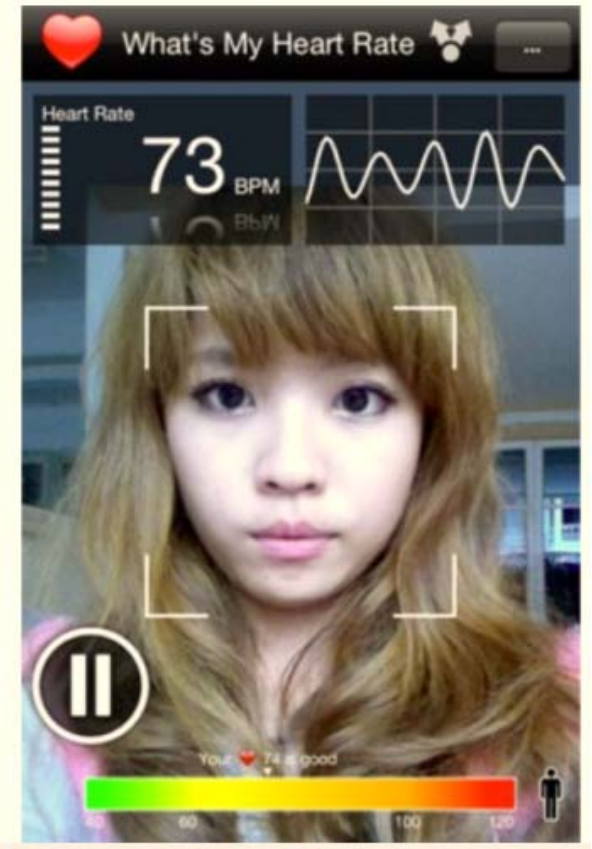
Medical Imaging

- Classification and detection (e.g. lesion or cells classification and tumor detection)
- 2D/3D medical image segmentation
- 3D human organ reconstruction (MRI or ultrasound)
- Vision-guided robotics surgery

Medical Applications (cont'd)

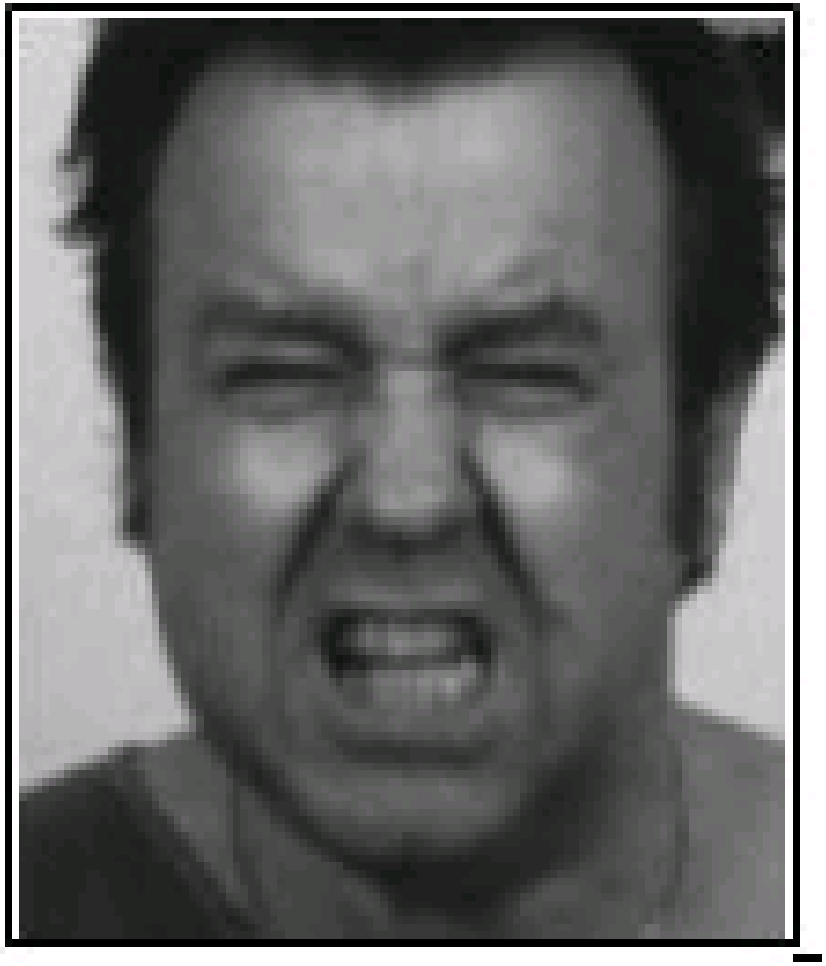
- Body mass index (BMI) prediction
  - Relation between facial features and BMI
- Heart rate estimation
  - measuring the redness of the face as blood flows through the face.

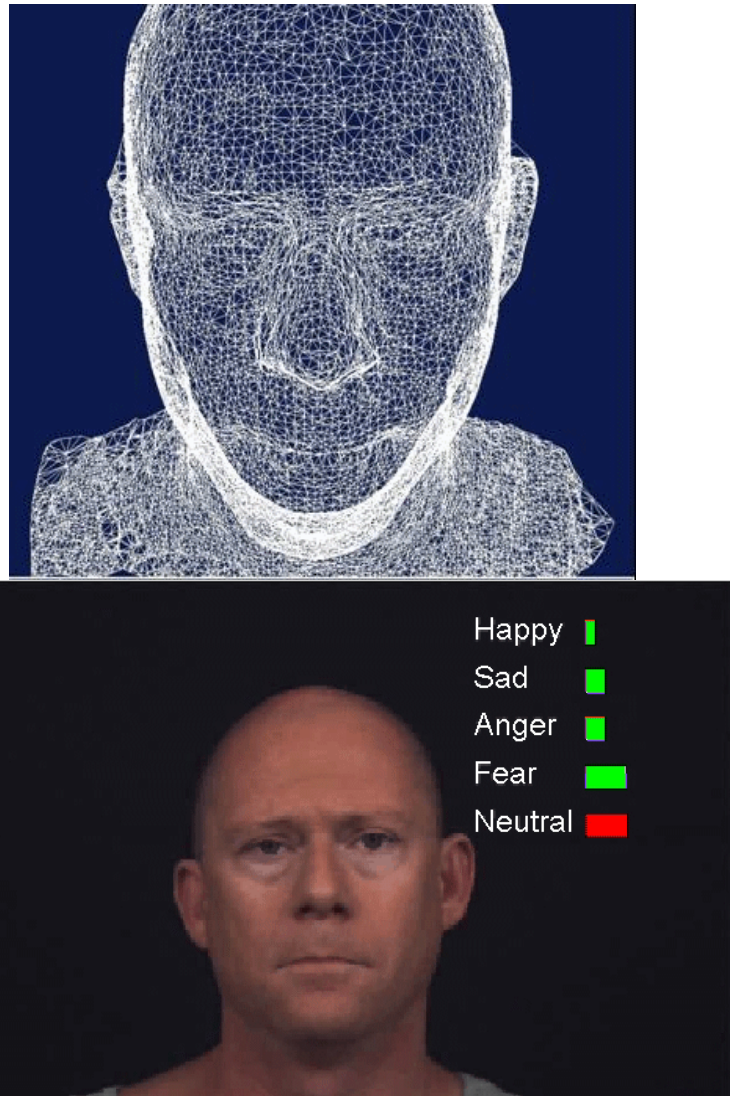
  See iPhone App *What's My Heart Rate*



Medical Applications (cont'd)

Medical diagnosis: facial Expression Analysis of Schizophrenia





Image/Video Database Search/Retrieval

- Image/video retrieval based on image content.

- Affect-based video retrieval

CV Trends

- **From static images to video**
  - Perform various tasks in video since video provides additional motion information that is not available in the static images
  - For example, motion segmentation, face recognition from video, activity recognition from video.

CV Trends (cont'd)

- **From global to local**
  - Use local region features (patch) for object detection tracking, and recognition
  - More robust, more tolerant to illumination, shape, background change and clutters.
  - Issues in local approach
    - Feature identification such as patch or SIFT
    - Different feature selection methods both online (adaptive) and offline
    - Different methods to combine selected features
    - Learn their relationships

CV Trends (Cont'd)

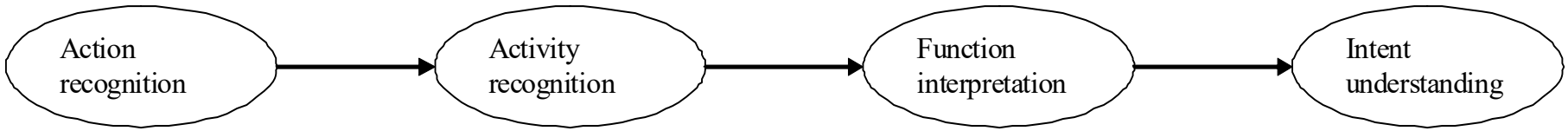
- **Use image context**
  - Use local spatial and temporal context for object recognition and object tracking.
  - Dynamically identify the useful context and learn its relationship to the target

CV Trends (Cont'd)

- **Use of prior knowledge**
  - Systematically capture related prior knowledge about the objects including photometric, geometric and physical constraints, anatomic/ physiological knowledge
  - Combine these knowledge with image measurements for robust visual interpretation and understanding.
  - For example, for body tracking, capture kinematics and biomechanics relationships among body parts. For facial expression recognition, capture the anatomical relationships among the facial muscles.

CV Trends (Cont'd)

- **More high level interpretation and understanding**
  - Research is moving more to high level understanding and interpretation including activity recognition, object function and intent understanding, etc..



CV Trends (Cont'd)

- **More Machine Learning**
  - Machine learning is playing an increasingly more important role.
  - Feature extraction, object representation, and classification.
    - Feature extraction through dimensionality reduction
      - PCA, ICA, LDA, data embedding, other subspace or manifold learning methods
    - Learning methods
      - online or offline learning methods such as Adaboost
  - Deep learning nowadays dominates vision

Computer Vision Literature

1. Journals

- IEEE transactions on Pattern Recognition and Machine Intelligence (PAMI)
- International Journal of Computer Vision
- Computer vision and image understanding
- Image vision and computing
- Machine vision and application
- Pattern recognition

2. Conferences

- IEEE conference on computer vision and pattern recognition (CVPR)
- International conference on computer vision(ICCV)
- International conference on image processing (ICIP)
- International conference on pattern recognition (ICPR)
- IEEE conference on robotics and automation

3. Open-access publications

- ArXiv –an open access repository of papers in physics, mathematics, computer science, etc..

Top Computer Science Conferences

*Ranking is based on **Conference H5-index>=12** provided by Google Scholar Metrics*

Show **Due** only | All Categories | All Countries | Search by keyword

	Hindex	Publisher	Conference Details
1	158	IEEE	**CVPR : IEEE Conference on Computer Vision and Pattern Recognition, CVPR** Jun 18, 2018 - Jun 18, 2018 - Salt Lake City , **United States** http://cvpr2018.thecvf.com/submission/timeline **Deadline : Wed 08 Nov 2017**
2	101	Neural Information Processing Systems Foundation	**NIPS : Neural Information Processing Systems (NIPS)** Dec 4, 2017 - Dec 4, 2017 - Long Beach Convention Center , **United States** https://nips.cc/Conferences/2017/CallForPapers
3	98	Springer	**ECCV : European Conference on Computer Vision** Oct 8, 2016 - Oct 8, 2016 - Amsterdam , **Netherlands** http://www.eccv2016.org/
4	91	IMLS	**ICML : International Conference on Machine Learning (ICML)** Aug 6, 2017 - Aug 11, 2017 - Sydney , **Australia** http://icml.cc/2017
5	89	IEEE	**ICCV : IEEE International Conference on Computer Vision** Oct 22, 2017 - Oct 29, 2017 - Venice , **Italy** http://iccv2017.thecvf.com/
6	85	Association for Computing Machinery	**CHI : Computer Human Interaction (CHI)** Apr 21, 2018 - Apr 21, 2018 - Montréal , **Canada** https://chi2018.acm.org/ **Deadline : Tue 12 Sep 2017**
7	80	IEEE	**INFOCOM : Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)** Apr 15, 2018 - Apr 19, 2018 - Honolulu HI , **United States** http://infocom2018.ieee-infocom.org/content/call-papers-main-conference

Online Computer Vision Resources

- Computer Vision Information
  http://www.visionbib.com
- Computer vision online
  - http://homepages.inf.ed.ac.uk/rbf/CVonline/
- Computer Vision Central
  - http://cvisioncentral.com/
- Vision mailing list
  - http://list.ku.dk/listinfo/sci-diku-imageworld
- Fei-fei Li's Ted Talk at
  - https://www.youtube.com/watch?v=40riCqvRoMs

Topics

- Image Acquisition and Formation
- Perspective Projection Geometry
- Camera Calibration and Pose Estimation
  - Manual and self calibration
- 3D Reconstruction
  - From single images
  - Passive stereo
  - Active stereo
  - From motion
- Motion Estimation and Tracking
  - Optical flow estimation
  - Object tracking with Kalman filtering
  - Structure from motion
- Feature Extraction (Edge, point, line, curve)
- Object detection and recognition (briefly)
- State of the deep learning models for some vision tasks.

Computer Vision Solutions

Solutions to CV tasks lie in four aspects:

- Physics about the 3D objects/scenes (e.g. humans) , that govern their physical and geometric behaviors and properties
- Vision models, including image formation, illumination, optics, and projection models, that relate 3D objects to their images
- Mathematical tools, including linear algebra, statistics, optimization, neural networks, etc.. that help solve the inverse problem
- Engineering that helps build the sensors, the digitizer, and other necessary equipment for image/video acquisition

Background Needed

- Good mathematical background, in particular linear algebra and optimization methods.

- Good programming skills in one of high level programming languages-Python, C++ or Matlab.

Outcomes

- understand the fundamental computer vision theories
- have the ability to design and implement major computer vision techniques
- have the capability of applying computer vision technologies to applications of interest.

Intelligent Systems Lab

Human-centered computer vision and its applications

- Develop computer vision algorithms to automatically analyze and recognize human facial and body behaviors
  - Facial behaviors: facial expression, head movements, and eye gaze
  - Body behaviors: body pose, body gestures, and human body actions and activities recognition
- Apply computer vision to different applications to augment humans perception, cognitive, and physical capabilities
- More at https://www.ecse.rpi.edu/~cvrl/
- ISL Introduction video at

  https://www.ecse.rpi.edu/~cvrl/Demo/Interview.mp4