Computer Vision

- The ability of computers to see.
  - Image Understanding
  - Machine Vision
  - Robot Vision
  - Image Analysis
  - Video Understanding

Image

- 2-D array of numbers (intensity values, gray levels)
- Gray levels 0 (black) to 255 (white)
- Color image is 3 2-D arrays of numbers
  - Red
  - Green
  - Blue
- Resolution (number of rows and columns)
  - 128X128
  - 256X256
  - 512X512
  - 640X480

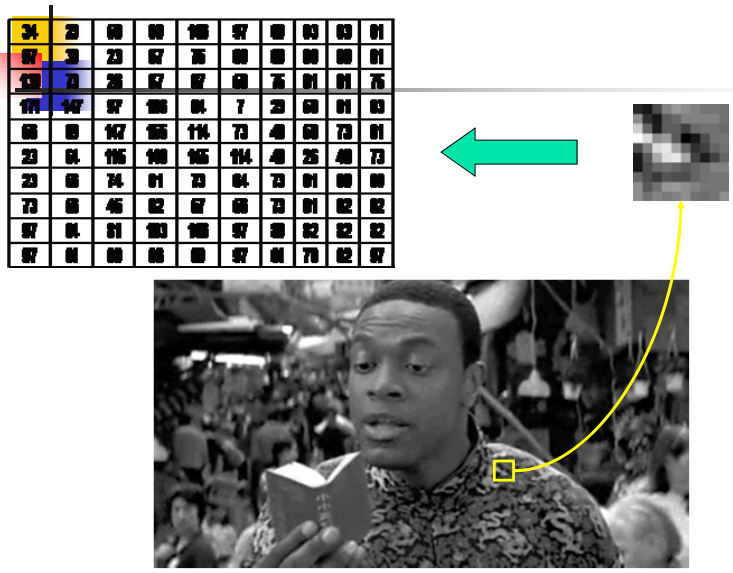


Image Formats

- TIF
- PGM
- PBM
- GIF
- JPEG
- RAW

Video

- Sequence of frames
- 30 frames per second

- Formats
  - AVI
  - MPEG
  - Quick Time

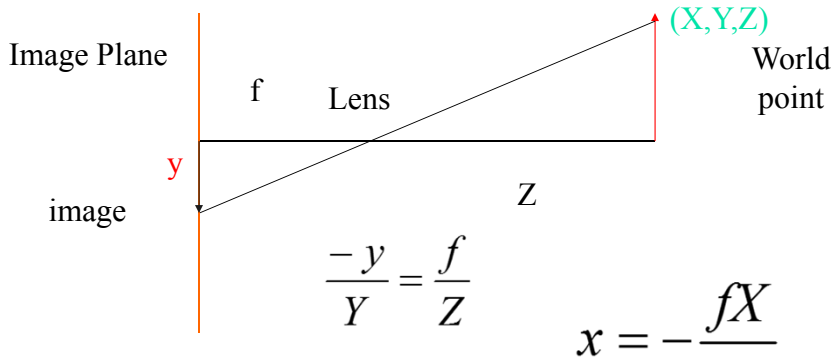
Video Clip

Sequence of Images

Image Formation

- Light Source
- Camera (extrinsic and intrinsic parameters)
- Scene (Surface reflectance, Surface shape)

Perspective Projection (Pin Hole)

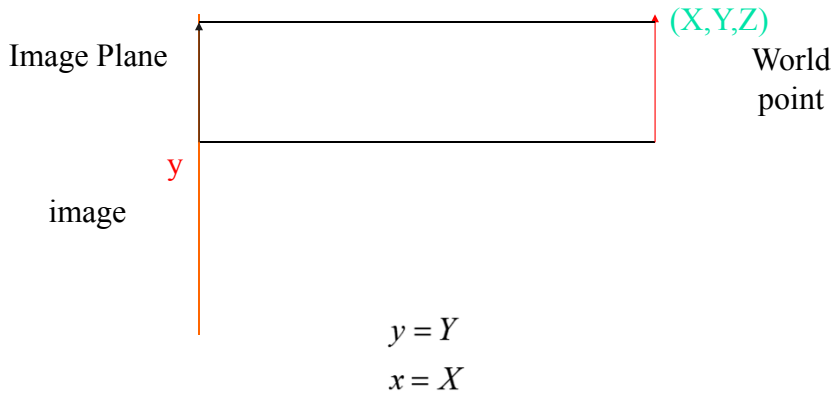


$$\frac{-y}{Y} = \frac{f}{Z}$$

$$y = -\frac{fY}{Z}$$

$$x = -\frac{fX}{Z}$$

Orthographic Projection



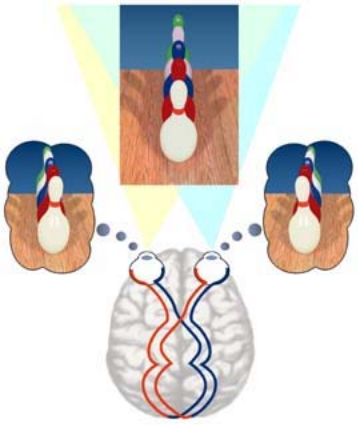
$$y = Y$$
$$x = X$$

Shape from X

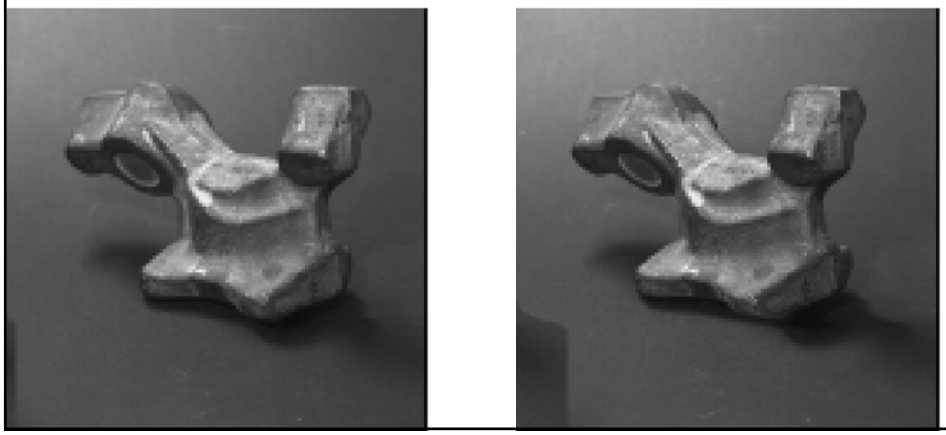
- Recover 3-D shape from 2-D image(s)
  - Stereo
  - Motion
  - Shading
  - Texture
  - Contours

Stereo

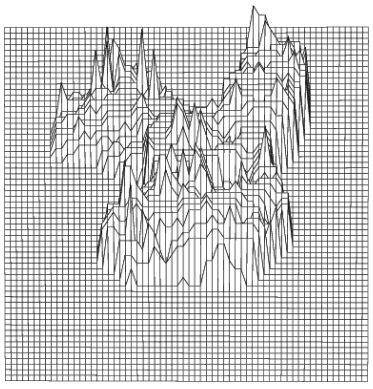
http://www.vision3d.com/stereo.html

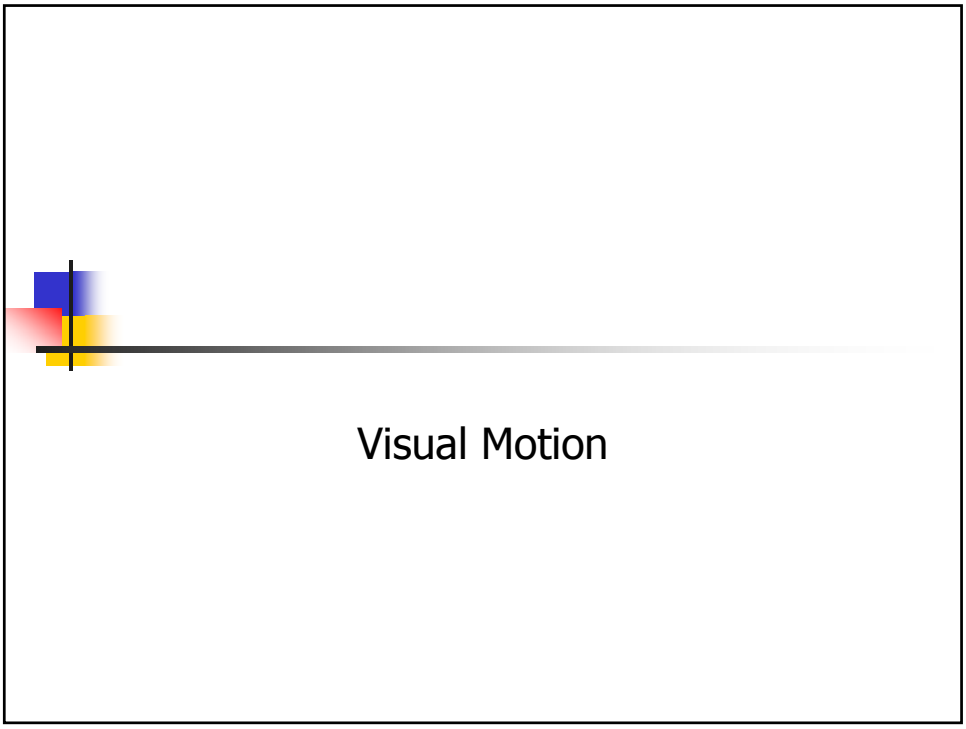


Renault Stereo Pair

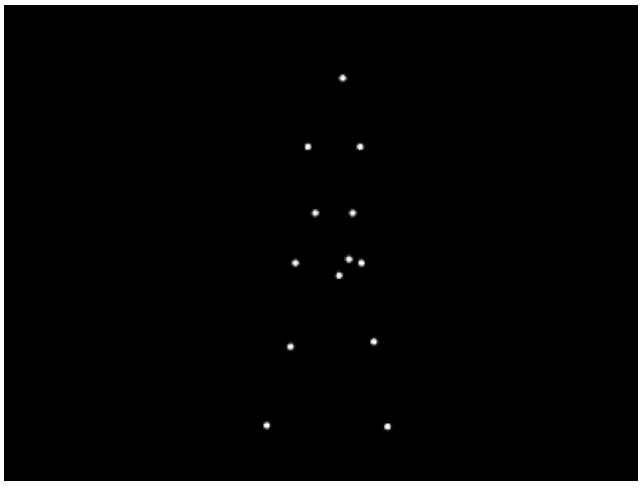


Depth Map

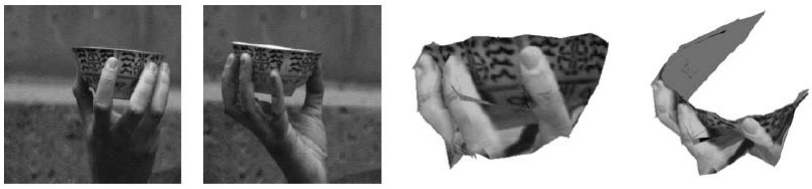




Shape from Motion: Moving Light Display



Shape from Motion



(a) (b) (c) (d)

Applications of Computer Vision

- Face Recognition
- Object Recognition
- Video Surveillance and Monitoring
  - Object detection, tracking and behavior analysis
- Remote Sensing: UAVs
- Robotics
- Computer Graphics

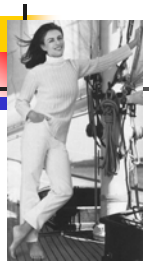
Object Recognition

Finding People in images

Problem 1: Given an image I

Question: Does I contain an image of a person?

"Yes" Instances






"No" Instances





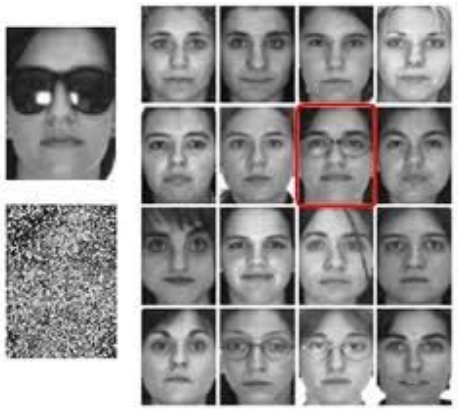

Localize People (Human Detection)

Human Detection

Airplanes

Motor Cycles

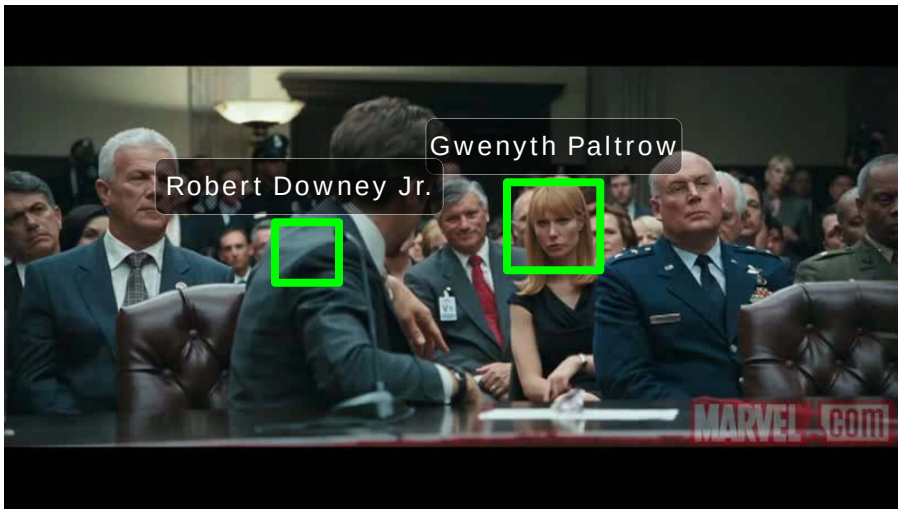
Face Recognition



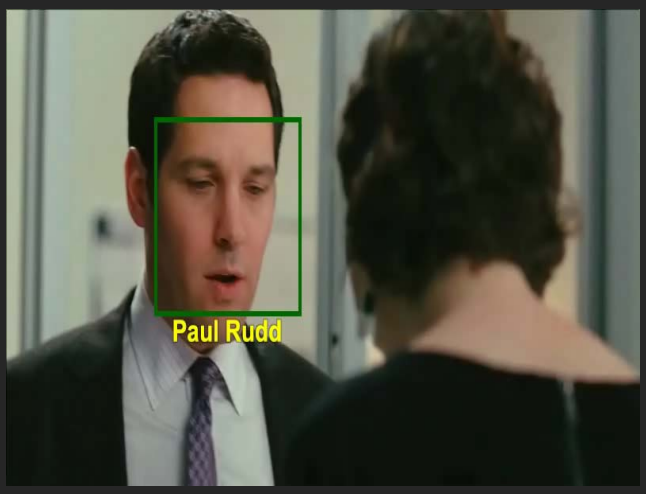
facebook flickr Picasa

YouTube Google play amazon instant video iTunes

Open-Universe Face Identification

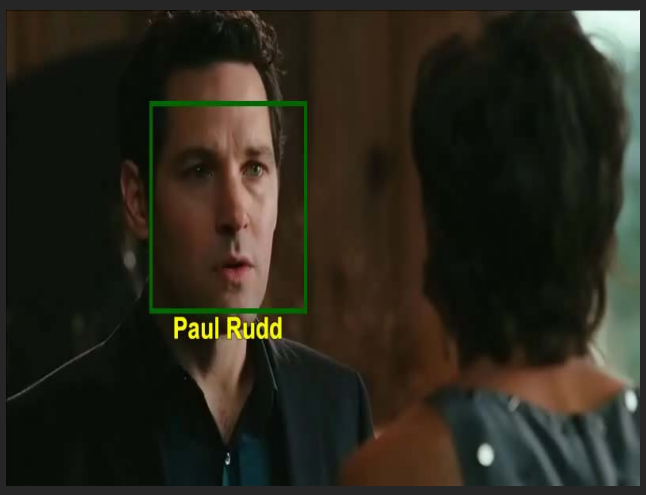


Recognition – Qualitative



Known:
Paul Rudd
Steve Carell

Recognition – Qualitative



Known:
Paul Rudd
Steve Carell

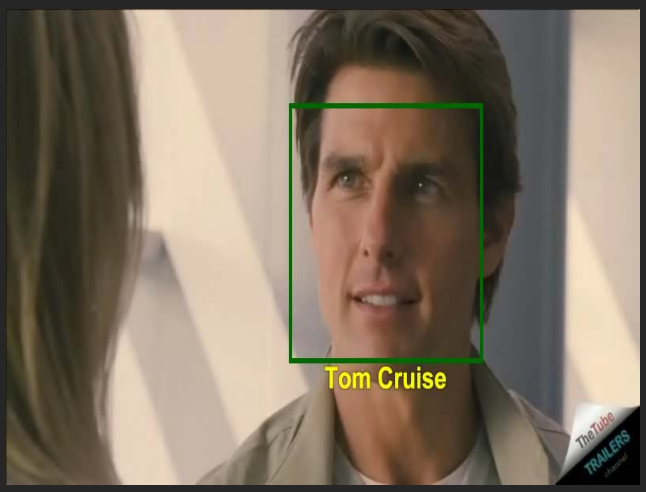
Recognition – Qualitative

Known:
Owen Wilson
Reese Witherspoon
Paul Rudd

Recognition – Qualitative

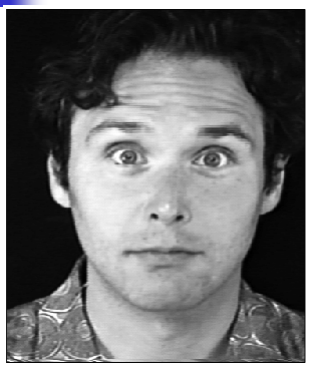
Known:
Bruce Willis
Morgan Freeman

Recognition – Qualitative

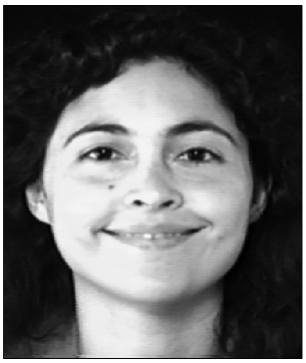


Known:
Tom Cruise
Cameron Diaz

FACIAL EXPRESSIONS



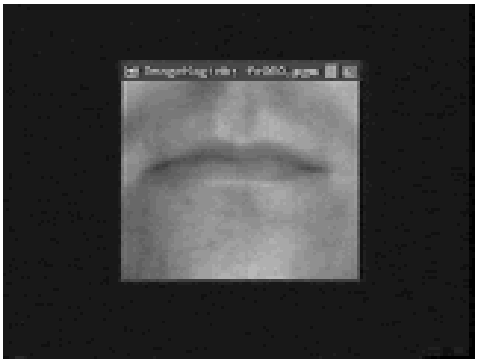
RAISE EYE BROWS



SMILE

Detecting Driver Alertness

Lipreading



Video Surveillance and Monitoring

Object detection → Object tracking → Object categorization and classification → Event or Activities Recognition

- Automated Surveillance System (Detection & Tracking)



A. I Person Tracking


A. II Part Tracking

UAV: Unmanned Aerial Vehicle

UAVs: Unmanned Aerial Vehicles (Drones)

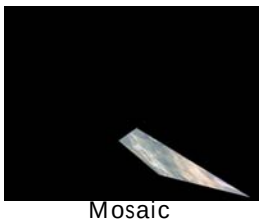
Global Hawk

Predator

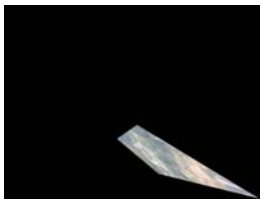
Microdrone

Registration Result - I

Aerial Video - EO



Mosaic

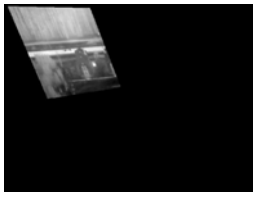


Alignment

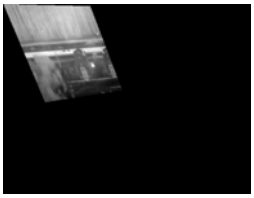
Mask

Registration Result - II

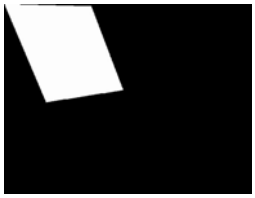
Aerial Video - IR



Mosaic



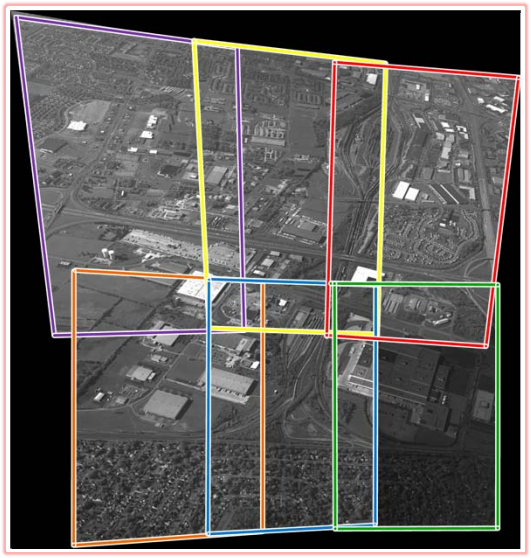
Alignment



Mask

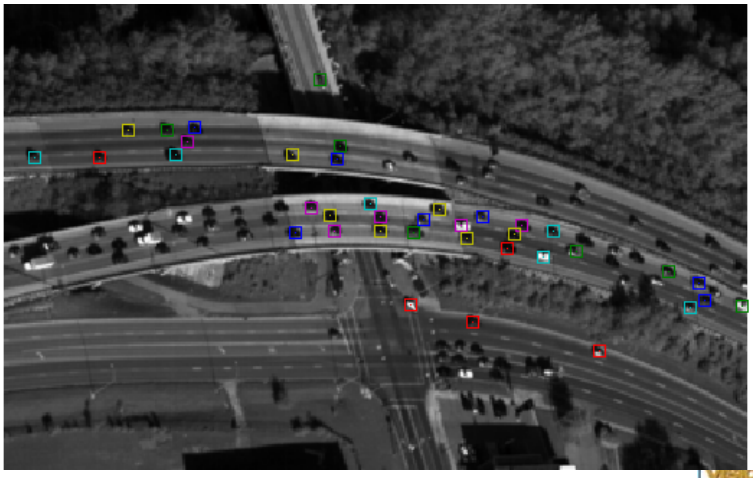
Detection Results

Wide Area Surveillance



Wide Area Surveillance

Tracking Results



Robot Vision (Unmanned Ground Vehicle)

UGV

UGV

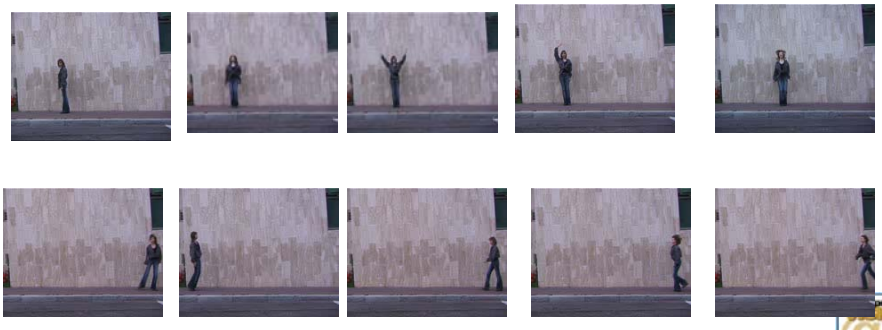
Human Action Recognition

Events, Actions, Activities,

- Action
- Event
- Movement
- Activity
- Interaction
- Verb
- ….

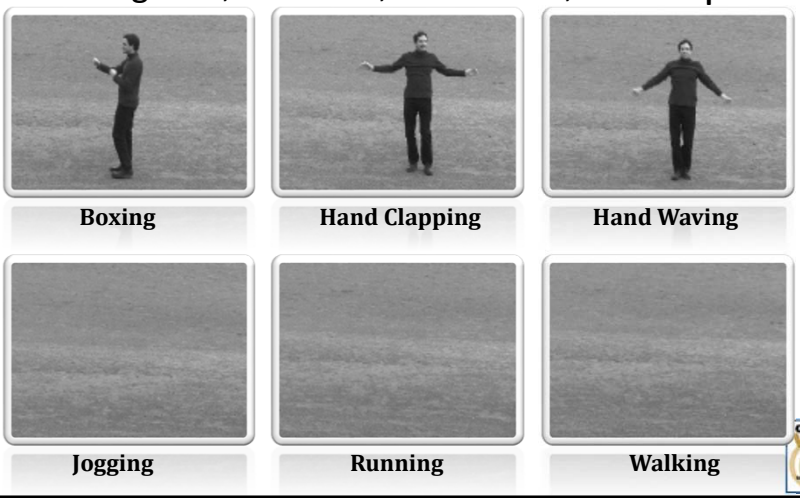
Weizmann Action Dataset

- 10 actions
- 9 actors per action



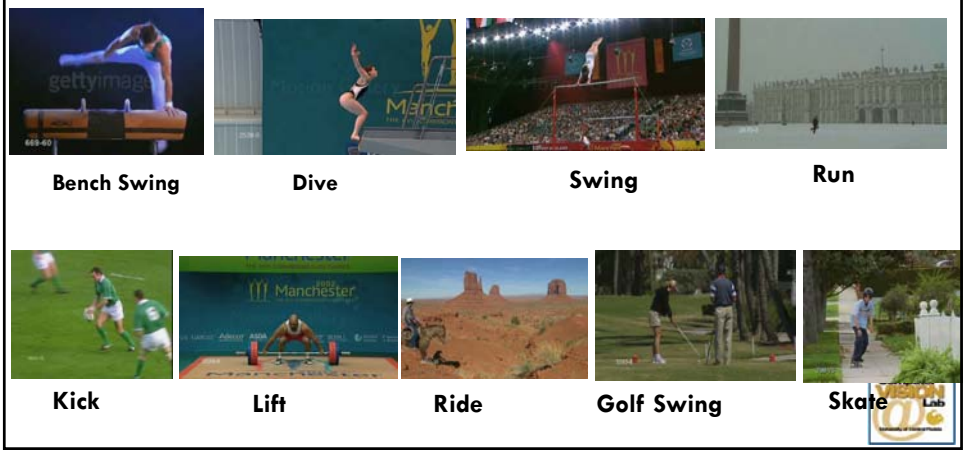
KTH Data Set

- Six Categories, 25 actors, 4 instances, 600. clips



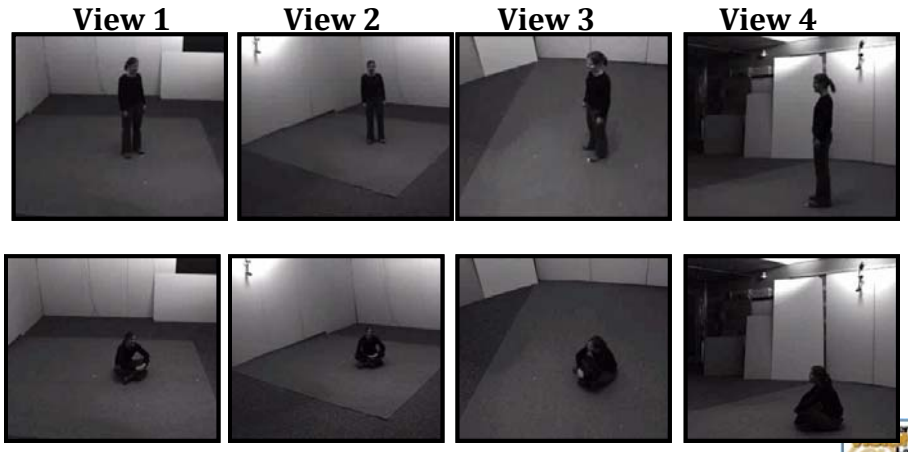
UCF Sports Action Dataset

9 actions, 142 videos.



IXMAS Multi-view Data Set

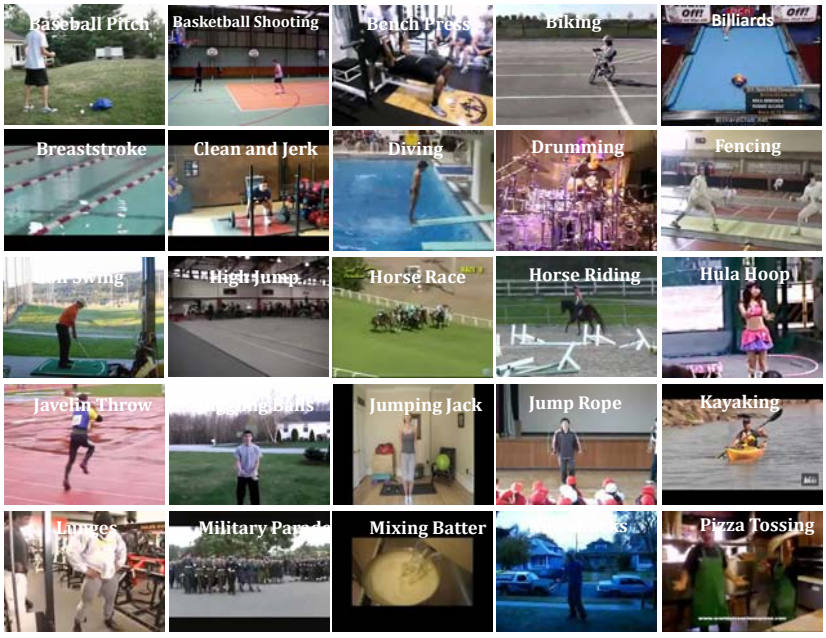
- 13 action categories, 4 camera views, 10 actors, 3 instances.



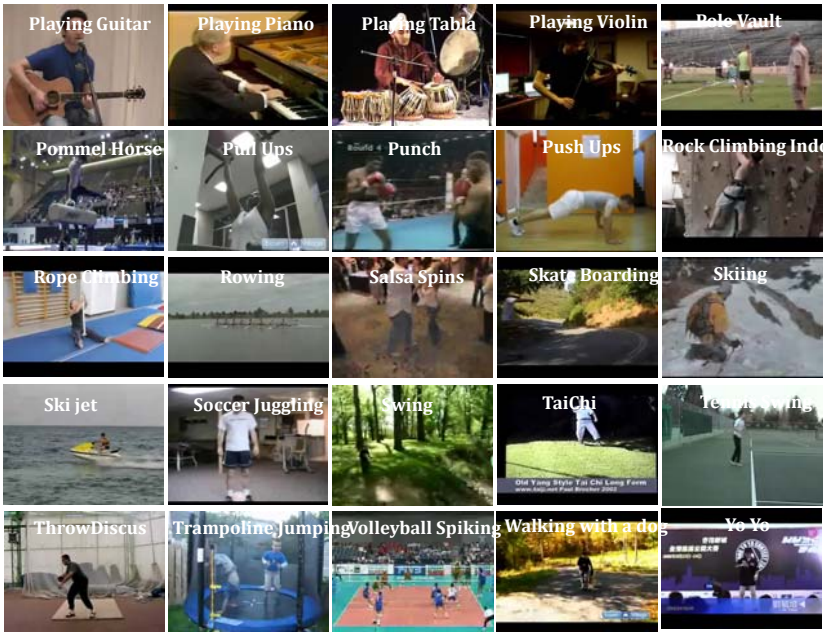
UCF YouTube Action Dataset (UCF-11)

UCF-101

UCF50

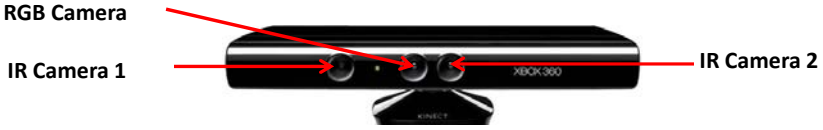


UCF50



Microsoft Kinect sensor

- Data Captured using Microsoft Kinect sensor



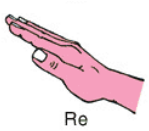
- Approximately 50,000 gesture samples

Gesture Lexicons

Diving Signals

Referee Signals

Nurse Gesture

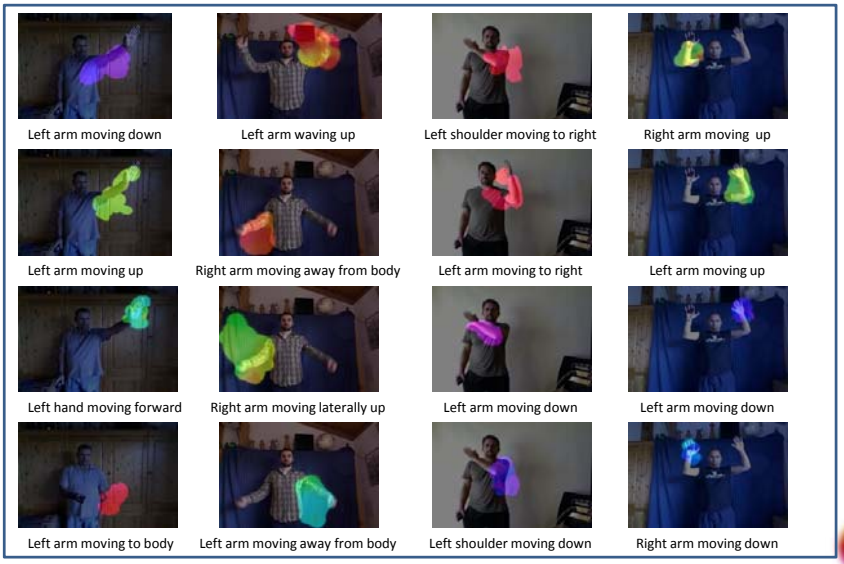


Music Notes

Gestures from Depth camera

Gestures from RGB camera

Discovered Primitives



Representative Motion Primitives (out of 136) for different batches

High Density Crowded Scenes

Political Rallies | Religious Festivals | Marathons | High Density Moving Objects

Tracking in Crowds

- Average chip size 14 x 22 pixels
- 492 Frames
- Selected 199 athletes for tracking
- Successfully tracked 143 athletes

Results

Experiment – 1

Experiment-3

- Average chip size 14 x 17 pixels
- 453 Frames
- Selected 50 athletes for tracking

Experiment – 3

Behaviors in Crowded Scenes

Multi-Source Multi-Scale Counting in Extremely Dense Crowd Images

CVPR 2013

Haroon Idrees, Imran Saleemi, Cody Seibert, Mubarak Shah

Ground truth=634 Proposed Method=640

Ground truth=1567 Proposed Method=1590

Ground truth=1428 Proposed Method=1468

Ground truth=653 Proposed Method=673

Ground truth=2322 Proposed Method=2203

Ground truth=2319 Proposed Method=2496

Where Am I?

"Where Am I?"

Problem:

Accurate Image Localization

Input

Mere Visual Information(Images)

Output

Location in Terms of λ (Lon.) and φ (Lat.)

φ=40.4419, λ=-79.9986

Qualitative Image Localization Results:

68.3%: The best match found:


Query


Match – Error: 9.01 m


Query


Match – Error: 6.4 m


Query


Match – Error: 7.5 m


Query


Match – Error: 62.7 m

Qualitative Image Localization Results:

68.3%: The best match found:
13.2%: A correct match found, but not the best one:


Query


Match – Error: 244.7 m


Query


Match – Error: 157.3 m


Query


Match – Error: 307.7 m

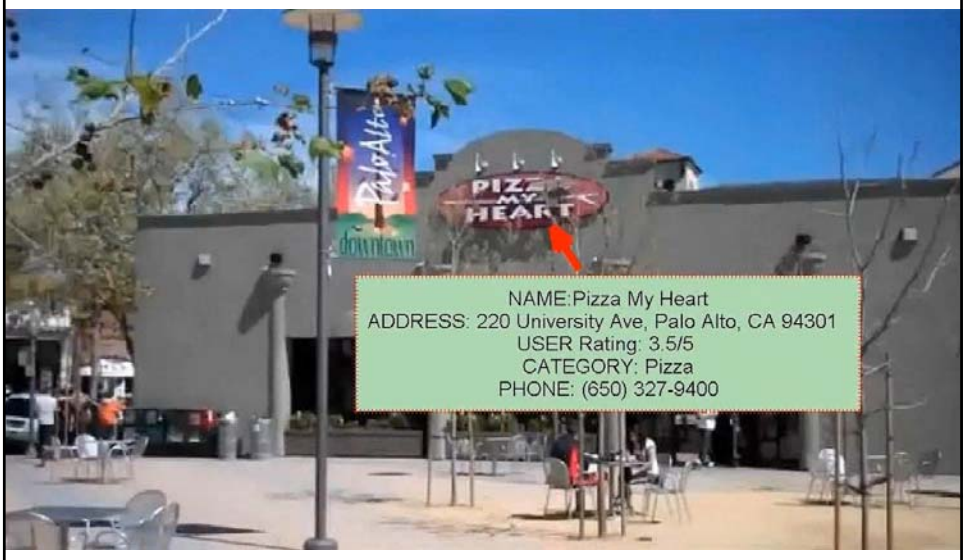

Query


Match – Error: 71.3 m

Geospatial Trajectory Extraction

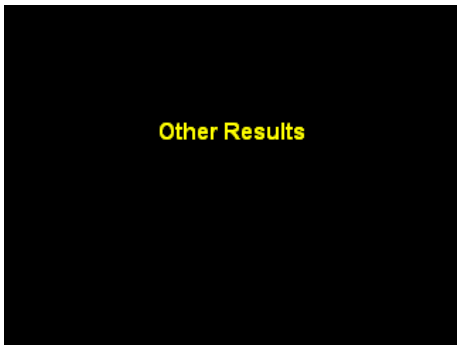
Visual Business Recognition

ACM Multimedia 2013



Computer Vision for Computer Graphics

Video Completion



University of Central Florida VISION

Layer Based Video Composition

University of Central Florida VISION

Results of Doll

University of Central Florida VISION

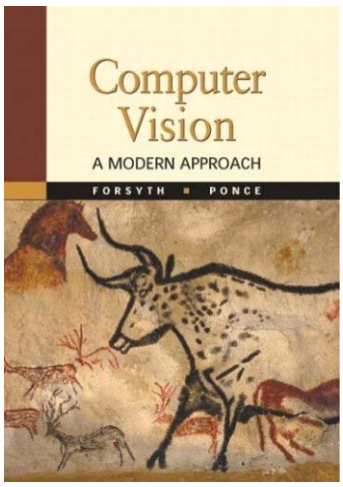
Results of Mom-Daughter

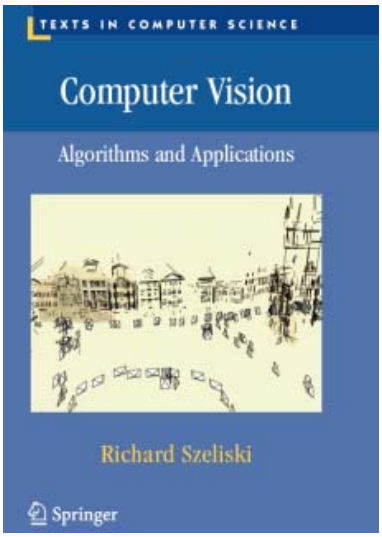
Multimedia: Segmentation of Moving-Sounding Objects

Accepted in IEEE Transactions on Multimedia

120

Computer Vision Text Books





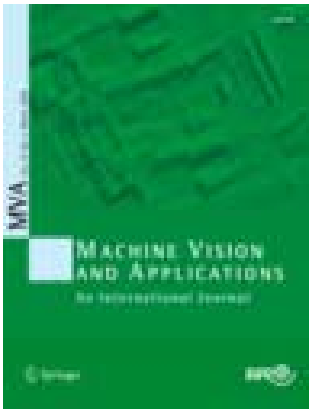
Computer Vision Researchers


IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND MACHINE INTELLIGENCE


Computer Vision
and Image
Understanding


INTERNATIONAL JOURNAL OF
COMPUTER VISION
Now Accepting
Video
Submissions


image
AND
vision
COMPUTING



Computer Vision Conferences

IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

- June 24-27, 2014
- Greater Columbus Convention Center in Columbus, Ohio.

European Conference on Computer Vision (ECCV)

European Conference on Computer Vision

International Conference on Pattern Recognition (ICPR)

Asian Conference on Computer Vision (ACCV)

International Conference on Image Processing