Frontiers of
Human Activity Analysis

J. K. Aggarwal
Michael S. Ryoo
Kris M. Kitani

THE UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

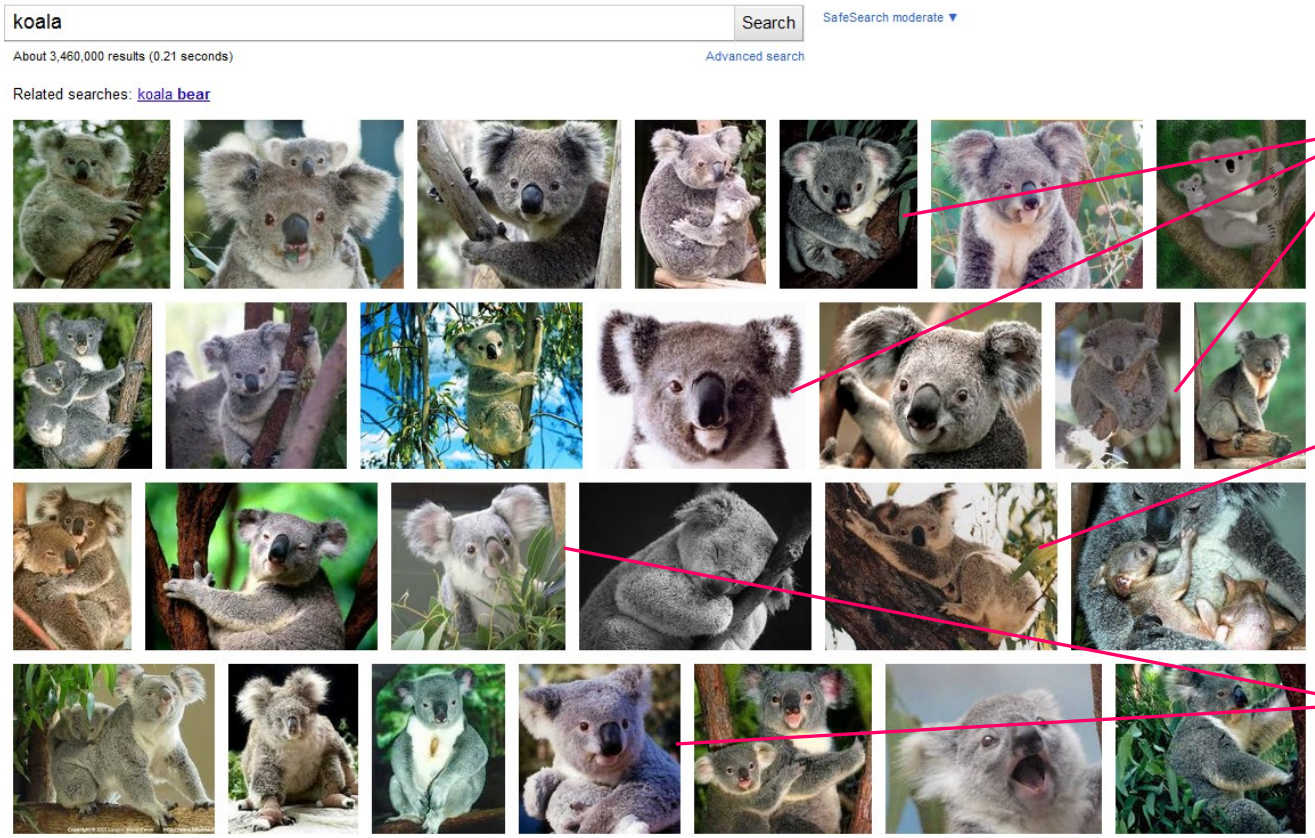
ETRI

Carnegie Mellon
THE ROBOTICS INSTITUTE

Applications and challenges

Object recognition applications

- Applications in practice
  - Google image search



koala.jpg

This is a **koala** in the Australian zoo, …

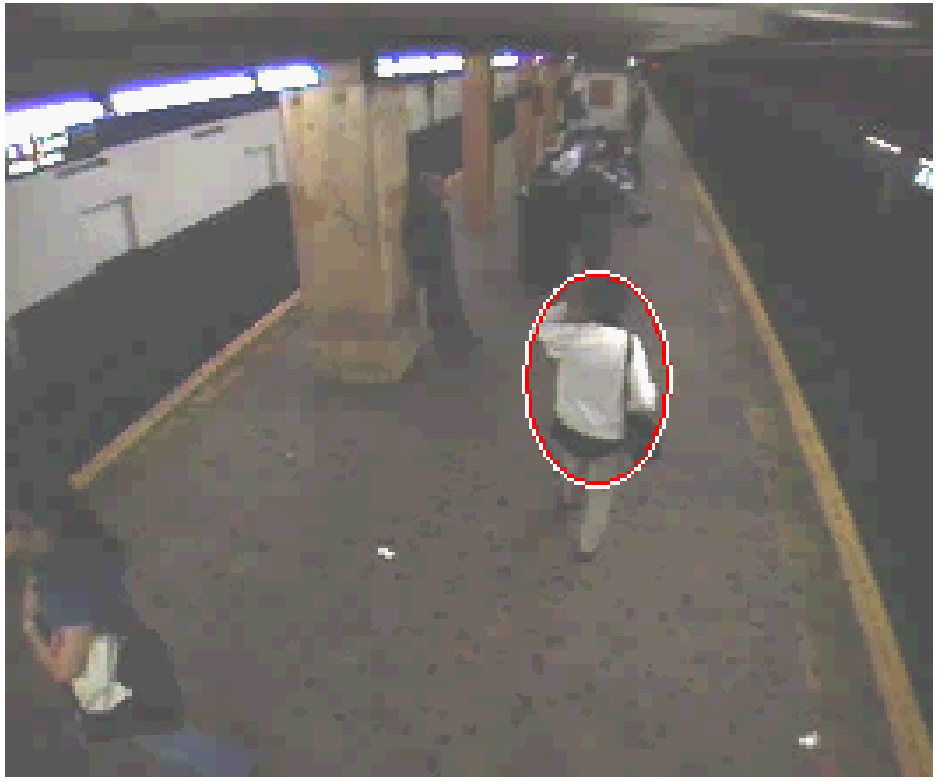
No label at all.

Object recognition applications

- Object recognition application in practice
  - Pedestrian (human) detection
    - Vehicle safety – Volvo automatic brake

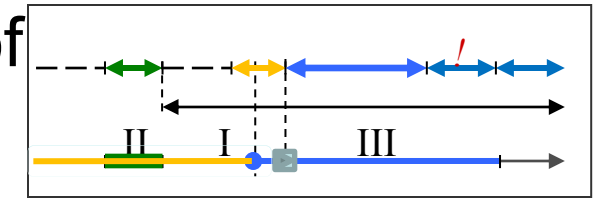
Video analysis applications

- We are living in an age where computer vision applications are working in practice.
  - Surveillance applications
  - Example> Siemens pedestrian tracking



Detection of Object Abandonment

- Activity recognition via reasoning of temporal logic
  - Description-based



Bhargava, Chen, Ryoo, Aggarwal, AVSS 2007

Illegally parked car detection

Foregrounds

Moving vehicles

Staying vehicles

Text Illegally parked vehicles

Lee, Ryoo, Riley, Aggarwal, AVSS 2007

Human-vehicle interactions

- Retrieval of videos involving humans and vehicles
  - Event-based analysis
  - Surveillance
  - Military systems
- Scene state analysis
  - Detailed description of the scene sequence

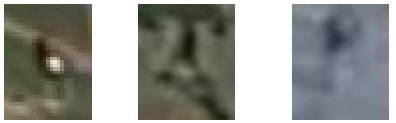
Four persons are going into a car and coming out:

Person locations, door opening, seating, …

Ryoo+, Lee+, and Aggarwal, ACM CIVR 2010

Aerial vehicles

- Unmanned aerial vehicles (UAVs)
  - Automated understanding of aerial images
    - Recognition of military activities
    - Example> carrying, digging, …

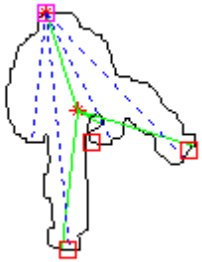


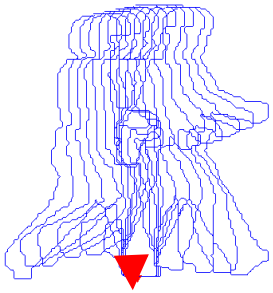
  - Challenge: visual cues are vague because of low-resolution

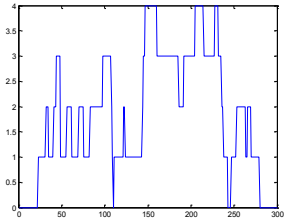
Chen and Aggarwal, WMVC 2009

Detecting Persons Climbing

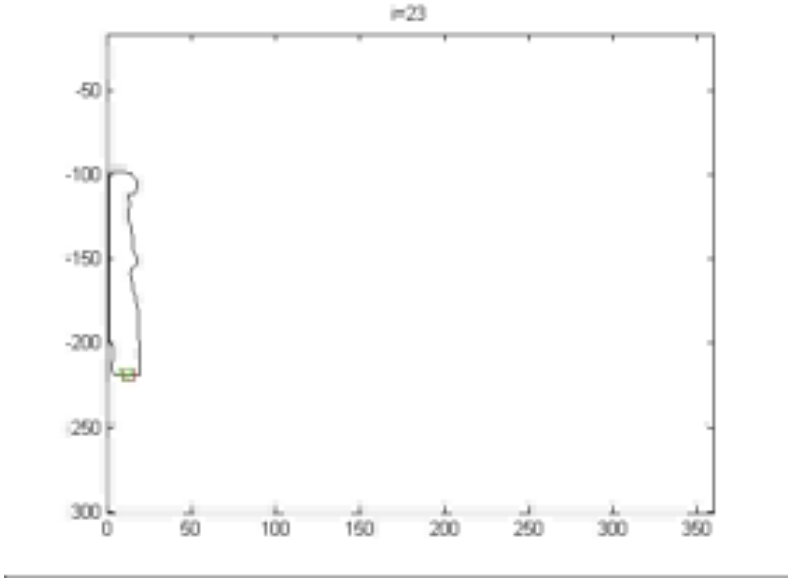
Human contour segmentation → Extracting extreme points → Detecting stable contacts → Constructing primitive intervals







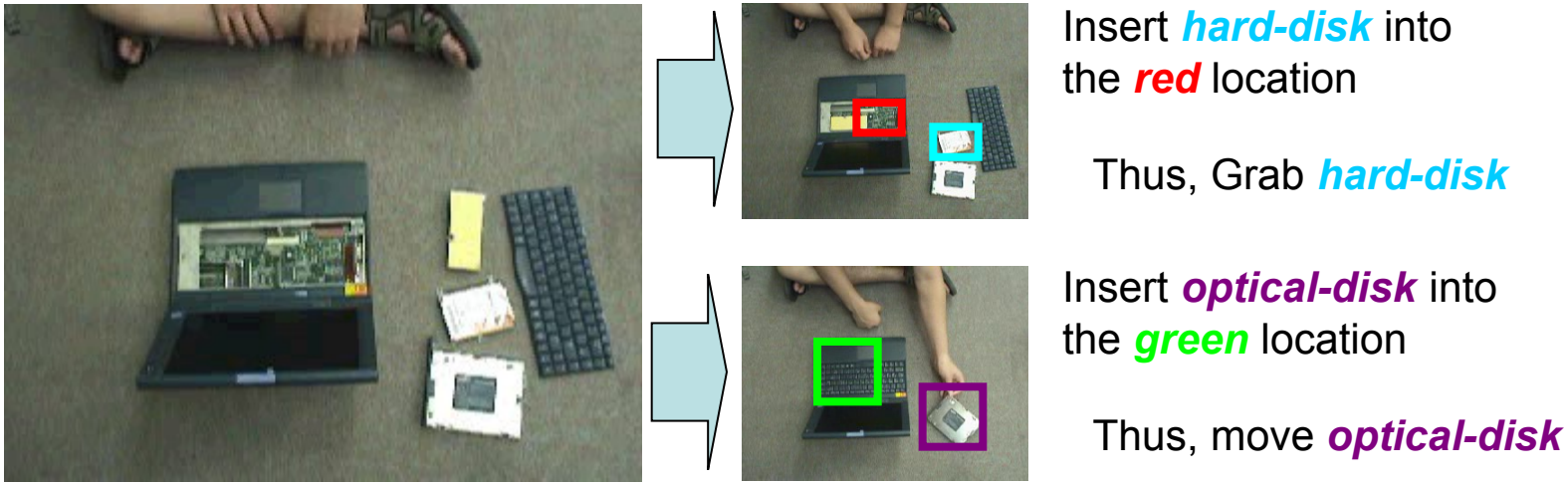
An climbing sequence is decomposed into segments based on primitive intervals formed fr
om stable contacts; the recognition is achieved from searching with HMM



Yu and Aggarwal, 2007

Human Computer Interactions

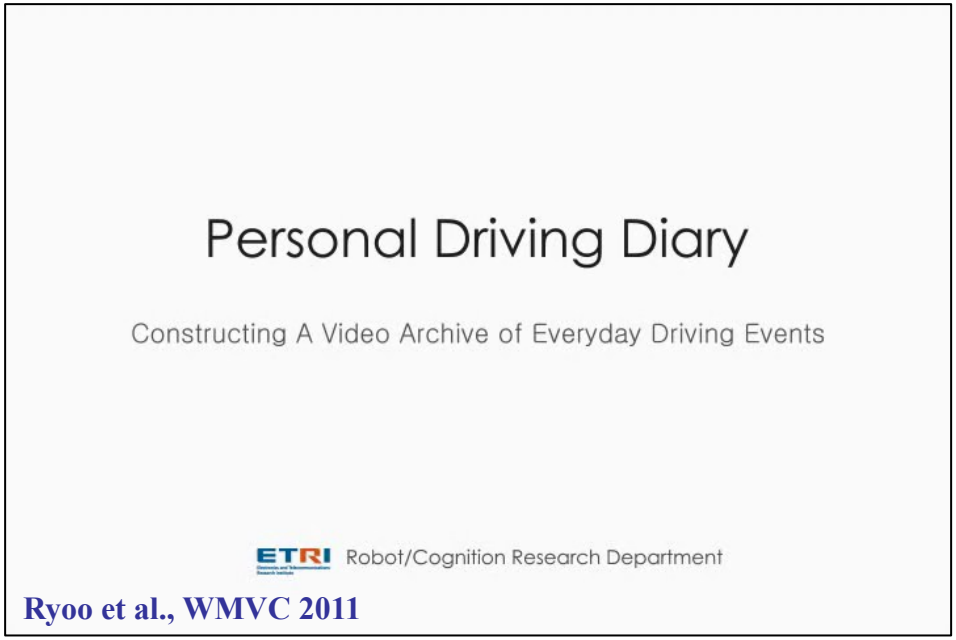
- Intelligent HCI system to help users' physical tasks
  - Our HCI system *observes* the task, and *guides* the user to accomplish it by providing **feedback**.
  - Ex> Assembly tasks



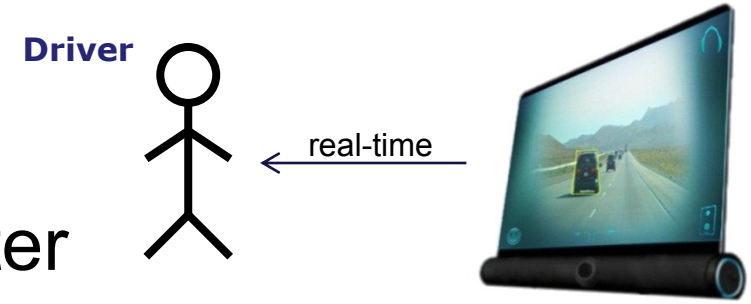
Ryoo, Grauman, and Aggarwal, CVIU 10

Intelligent driving

- Human activity recognition
  - Personal driving diary
    - Passive recording



- Human intention recognition
  - Real-time driving supporter



Future directions

- Future directions of activity recognition research are driven by ***applications***
  - Surveillance
    - Real-time
    - Multiple cameras, continuous streams
  - Video search
    - YouTube – 20 hours of new videos per minute
    - Large-scale database
    - Very noisy – camera viewpoints, lighting, …

Future directions (cont'd)

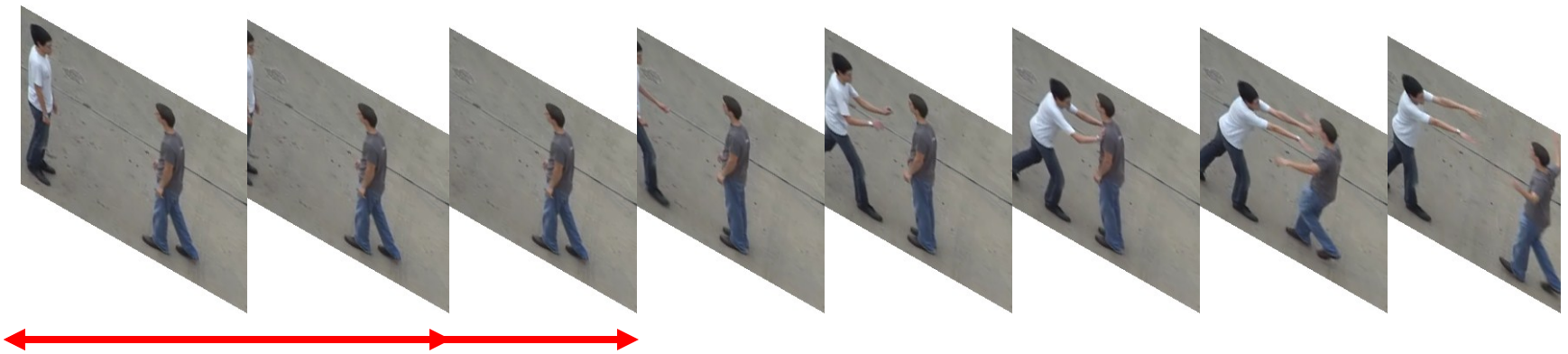
- Longer sequences (story, plot, characters)
- 3D representation and reasoning

- Complex temporal and logical constraints
  - Allen's temporal logic, context
- Incorporating prior knowledge (Ontologies)
  - Data driven high-level learning is difficult

Challenges – real-time

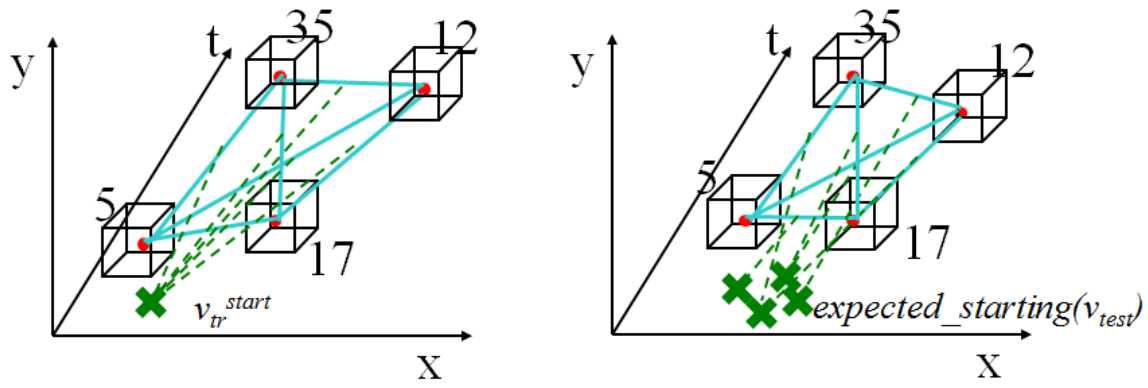
- Real-time implementation is required
  - Surveillance systems
  - Robots and autonomous systems
  - Content-based video retrieval
- Multiple activity detection
  - Continuous video streams

Computations



Challenges – real-time

- Goal
  - Process 20 hours of videos every minute
- Problem
  - Sliding window is slow.
  - Voting?

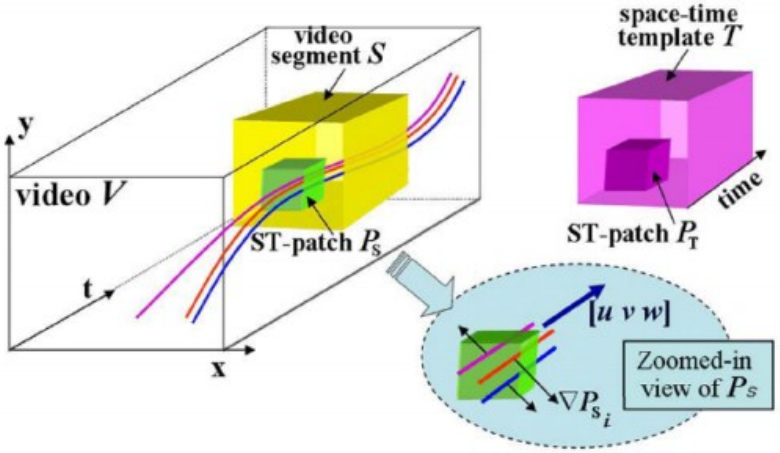


  - Feature extraction itself?

Real-time using GP-GPUs

- General purpose graphics processing units (GP-GPUs)
  - Multiple cores running ***thousands*** of threads
  - Parallel processing

20 times speed-up of video comparison method in Shechtman and Irani 2005:



[Rofouei, M., Moazeni, M., and Sarrafzadeh, M., Fast GPU-based space-time correlation for activity recognition in video sequences, 2008]

Challenges – activity context

- An activity involves interactions among
  - Humans, objects, and scenes

Running?

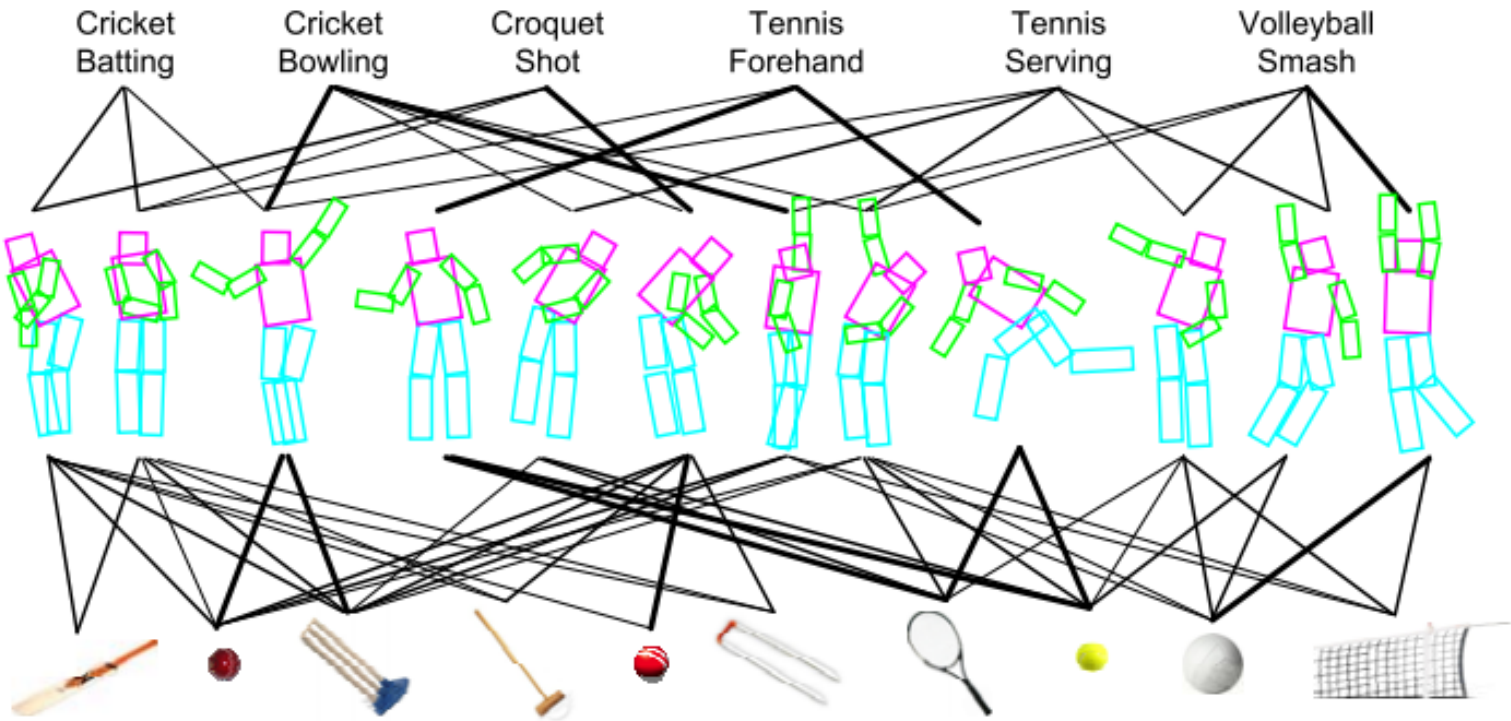
Kicking?

  - Activity recognition must consider motion-object context!

[Gupta, A., Kembhavi, A., and Davis, L., Observing Human-Object Interactions: Using Spatial and Functional Compatibility for Recognition, IEEE T PAMI 2009]

Activity context – pose

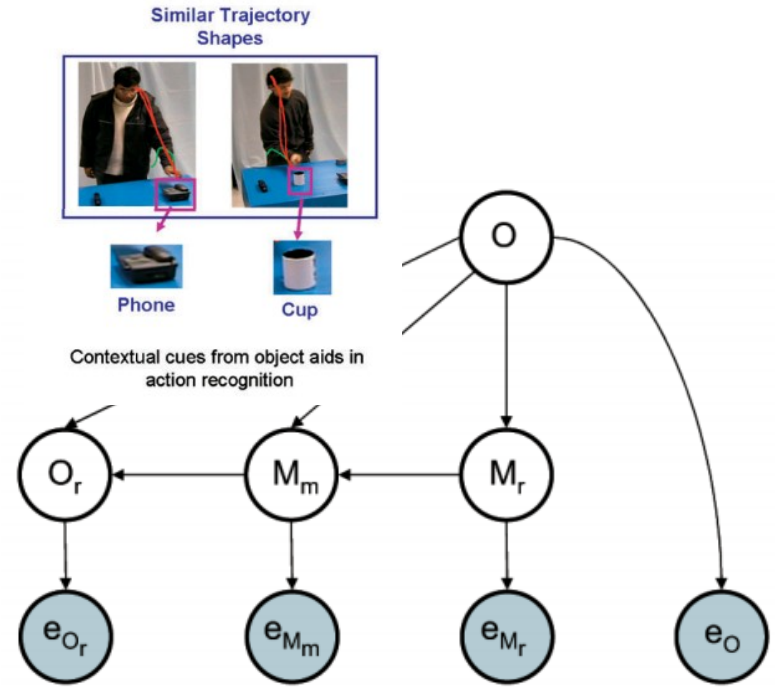
- Object-pose-motion

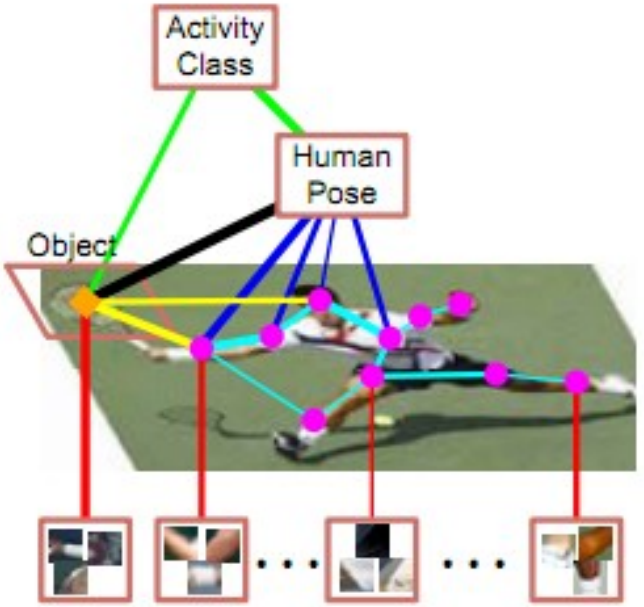


[Yao, B. and Fei-Fei, L., Modeling Mutual Context of Object and Human Pose in Human-Object Interaction Activities , CVPR 2010]

Activity context (cont'd)

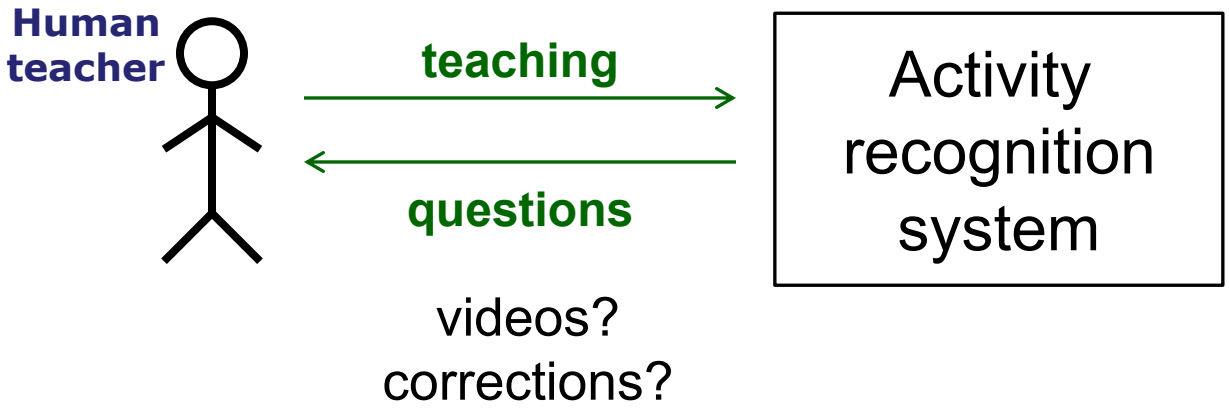
- Probabilistic modeling of dependencies
  - Graphical models





Challenges – interactive learning

- Interactive learning approach
  - Learning by generating questions



  - Interactive learning
    - Human-in-the-loop
    - Explore decision boundaries actively

Active video composition

- A new learning paradigm
  - *Composed videos* from a real video
  - Automatically create **necessary** training videos
  - Structural variations
    - Who stretches his hand first?

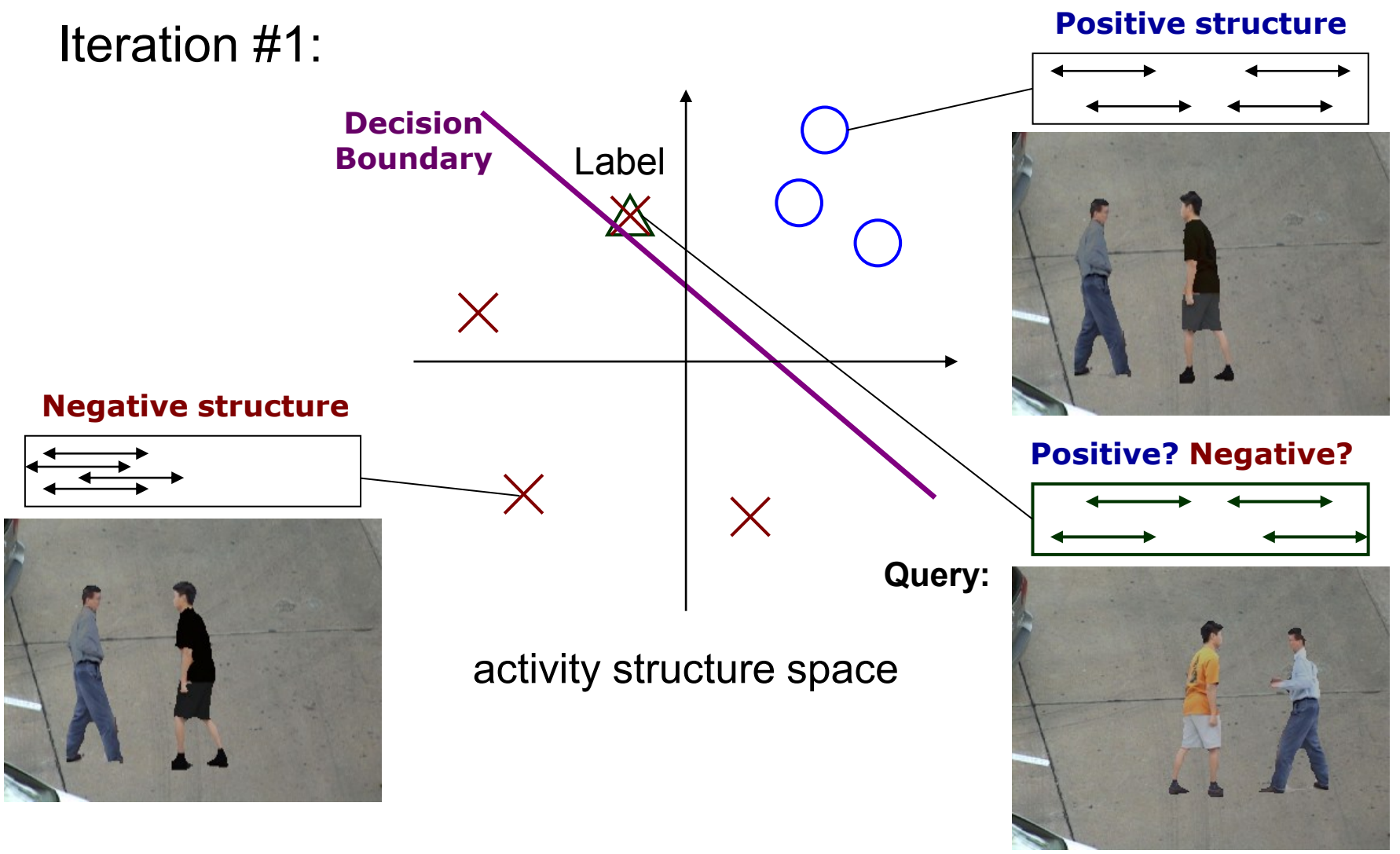
Original video:

Composed videos:

Ryoo and Yu, WMVC 2011

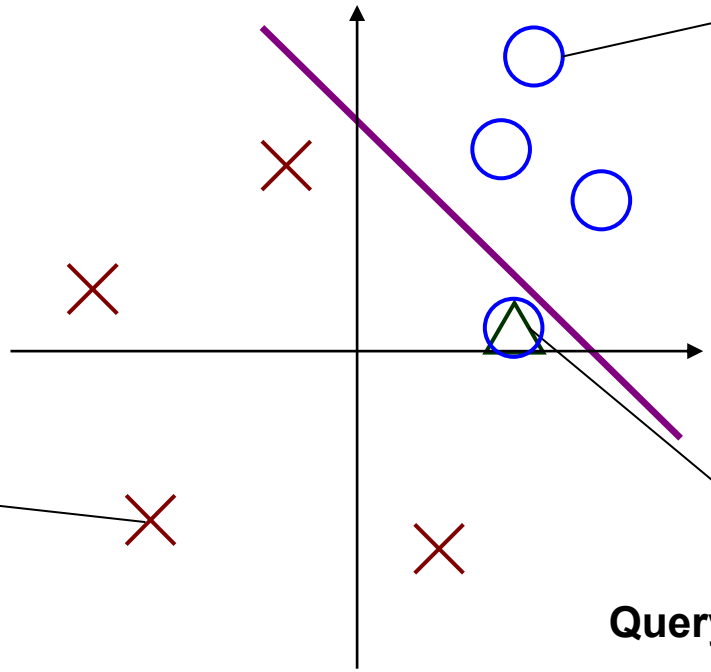
Active structure learning (cont'd)

Iteration #1:



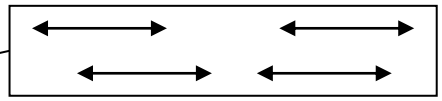
Active structure learning (cont'd)

Iteration #2:

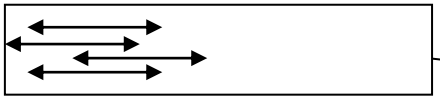


activity structure space

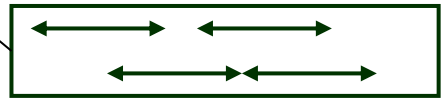
Positive structure



Negative structure



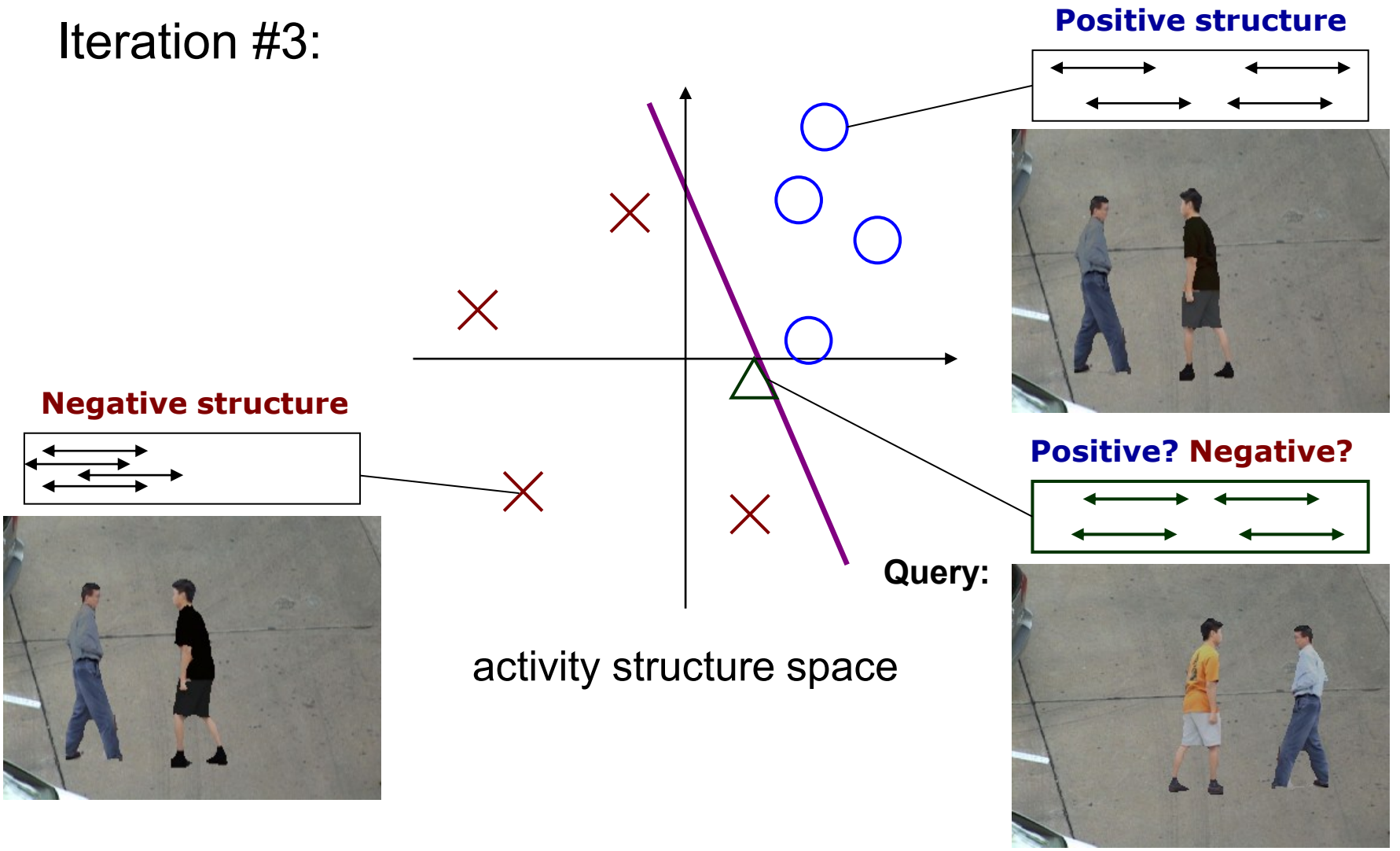
Positive? Negative?



Query:

Active structure learning (cont'd)

Iteration #3:



Recent work on ego-centric activity analysis

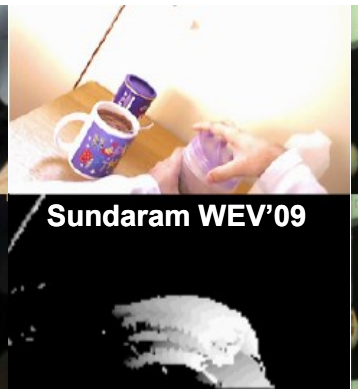
Understanding first-person activities with wearable cameras

(not a new idea in the wearable computing community)


Starner ISWC'98


Torralba ICCV'03

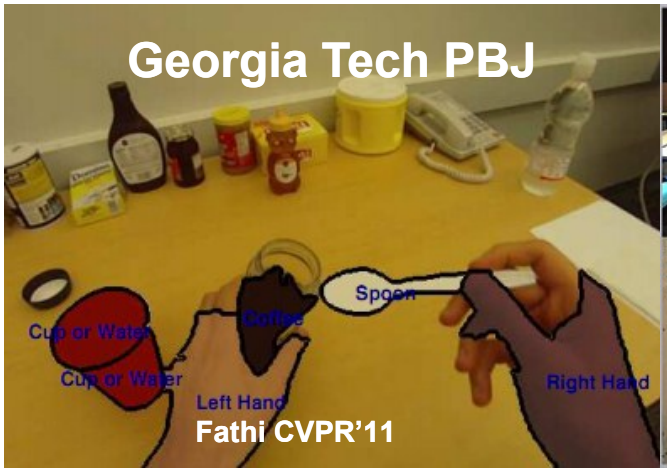

Ren CVPR'10


Sundaram WEV'09

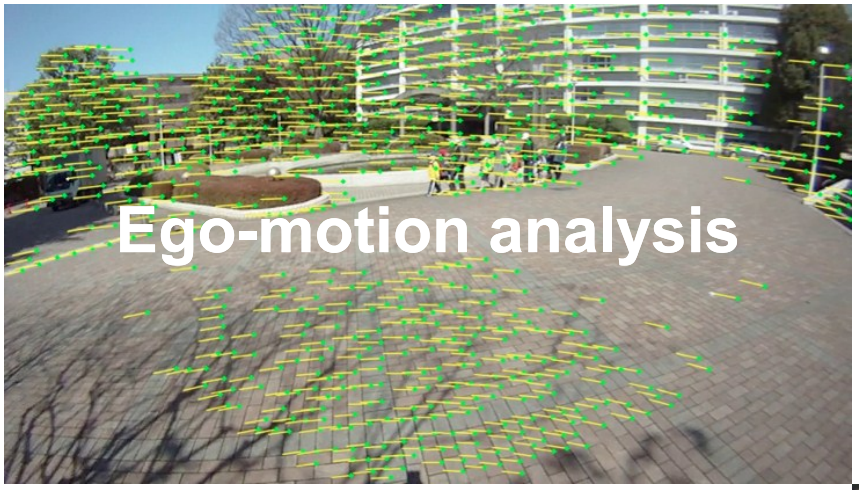

Sun WEV'09


Fathi ICCV'11

Datasets

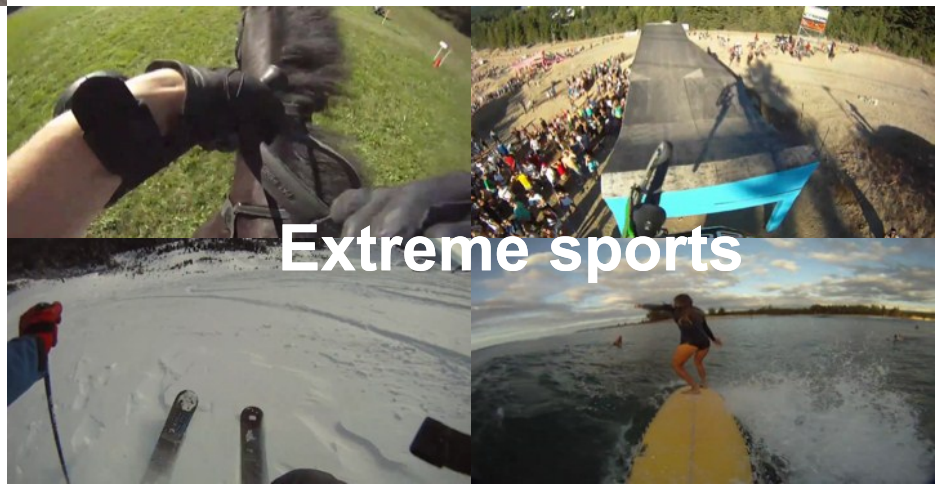


Future of ego-centric vision



Inside-out cameras

2nd person activity analysis



Conclusion

- Single-layered approaches for actions
  - Sequential approaches
  - Space-time approaches
- Hierarchical approaches for activities
  - Syntactic approaches
  - Description-based approaches
- Applications and challenges
  - Real-time applications
  - Context and interactive learning

Thank you

- We thank you for your attendance.
- Are there any questions?