Emerging Topics in Human Activity Recognition

Michael Ryoo NASA Jet Propulsion Laboratory
Ivan Laptev INRIA
Greg Mori Simon Fraser University
Sangmin Oh Kitware

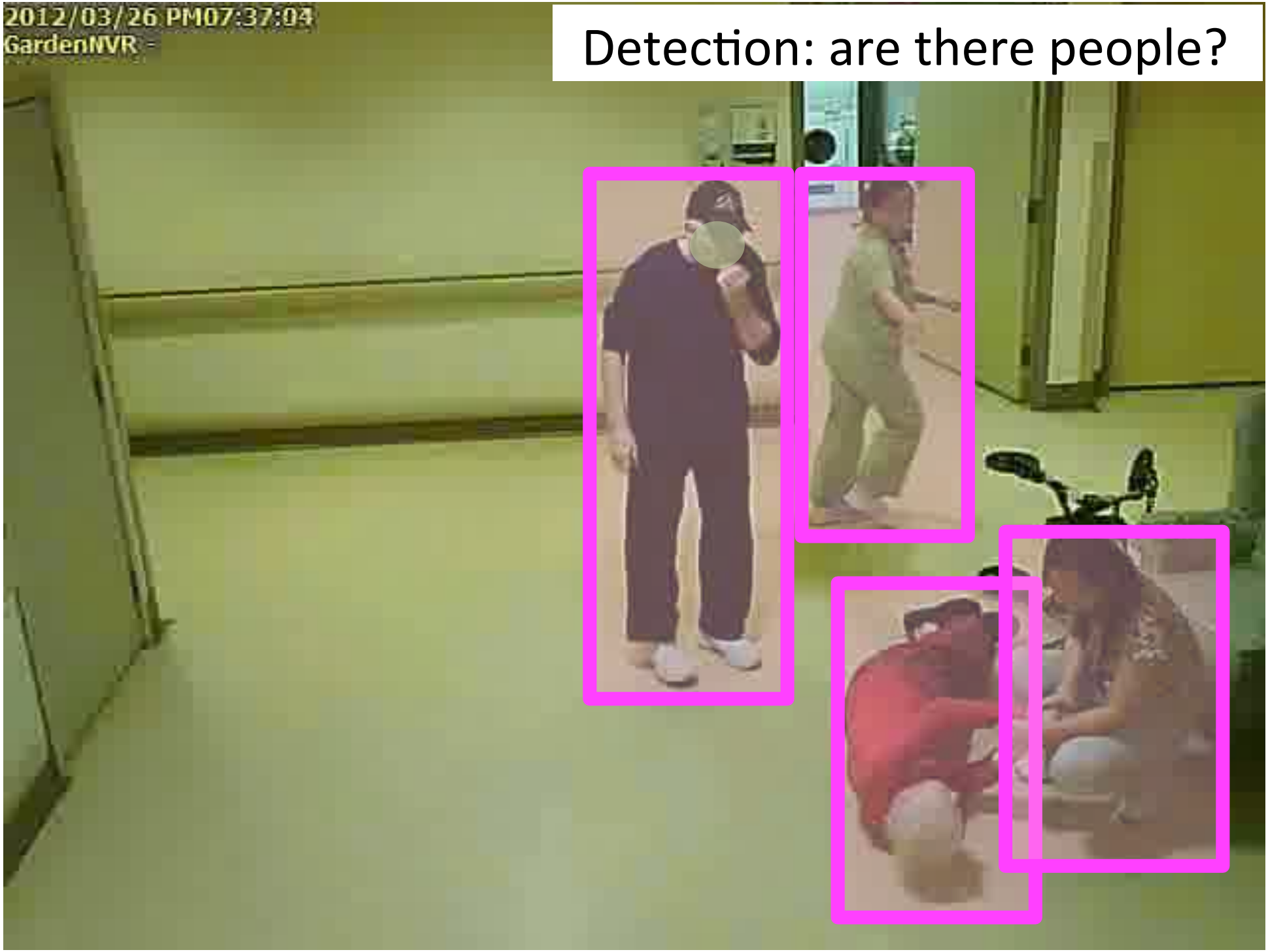
CVPR tutorial on 2014/06/23

Group Activity Recognition

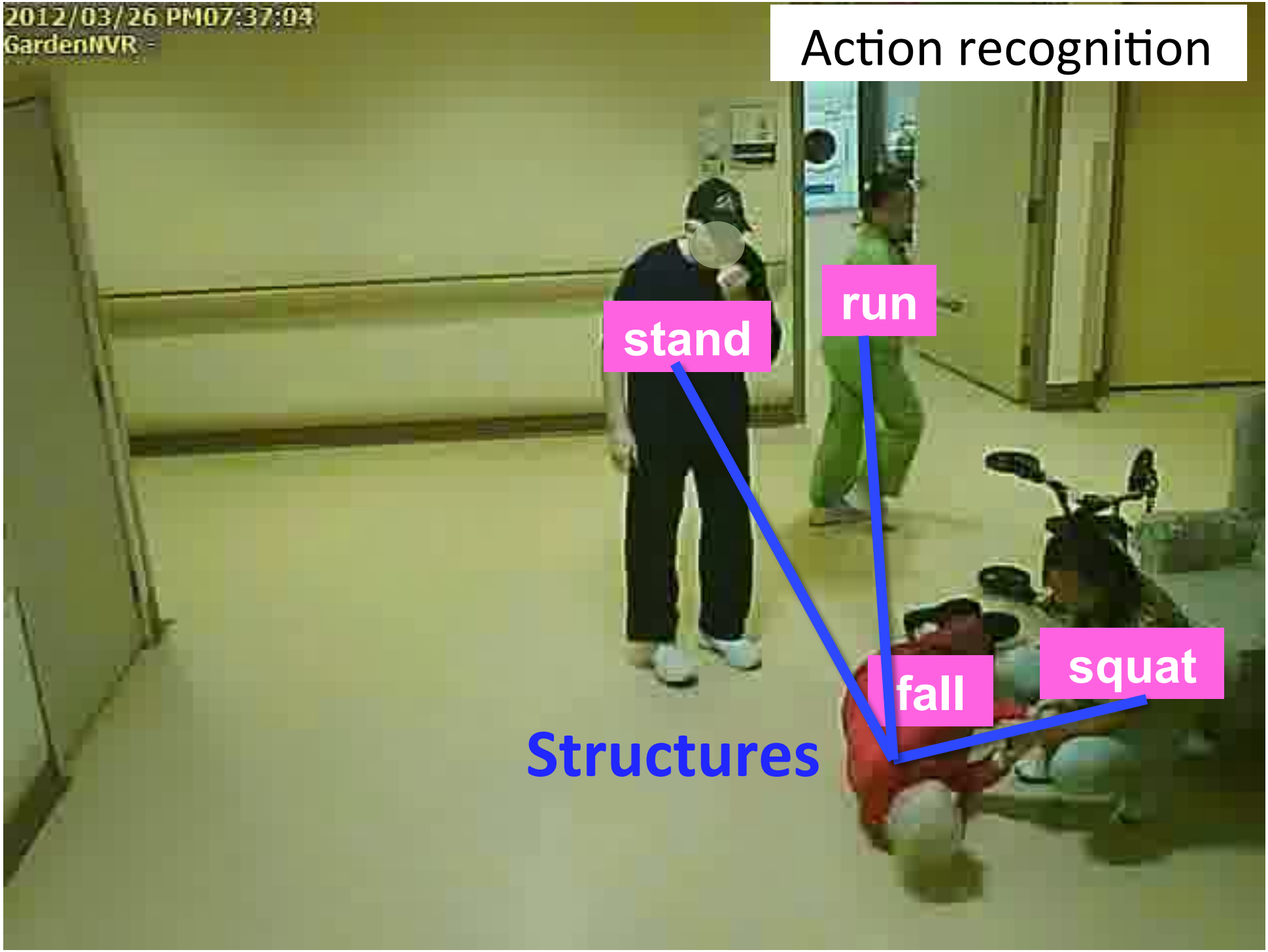
Greg Mori
Simon Fraser University

CVPR tutorial on 2014/06/23

What does activity recognition involve?

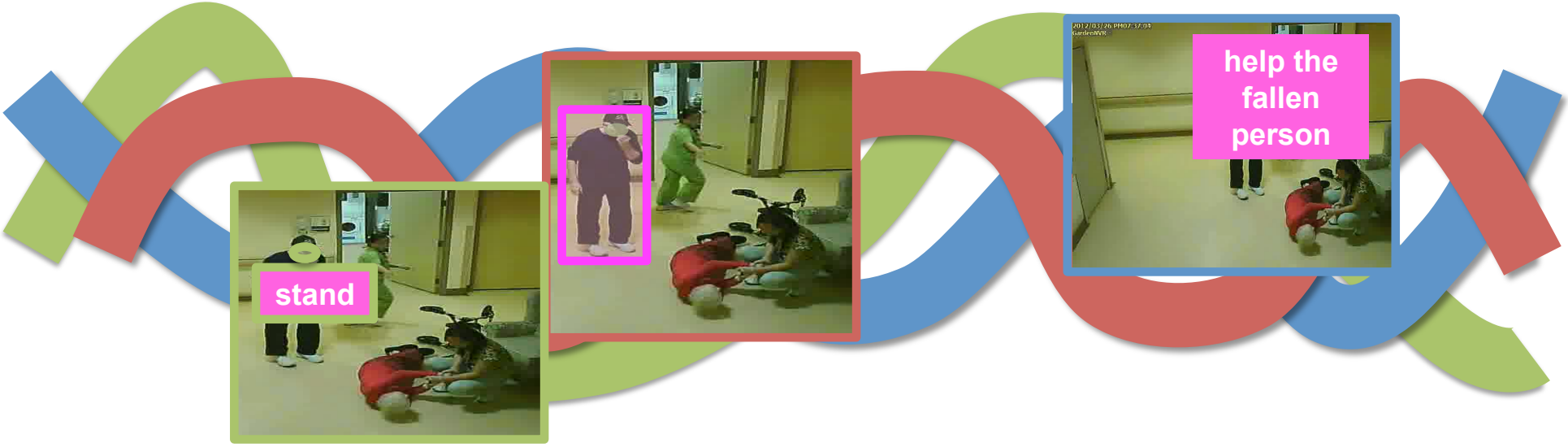


Detection: are there people?


2012/03/26 PM07:37:04
GardenNVR
Action recognition
stand
run
fall
squat
Structures


2012/03/26 PM07:37:04
GardenNVR -
Group activity recognition
help the fallen person


2012/03/26 PM07:37:04
GardenNVR
Intention/social role
get help
watch
comfort



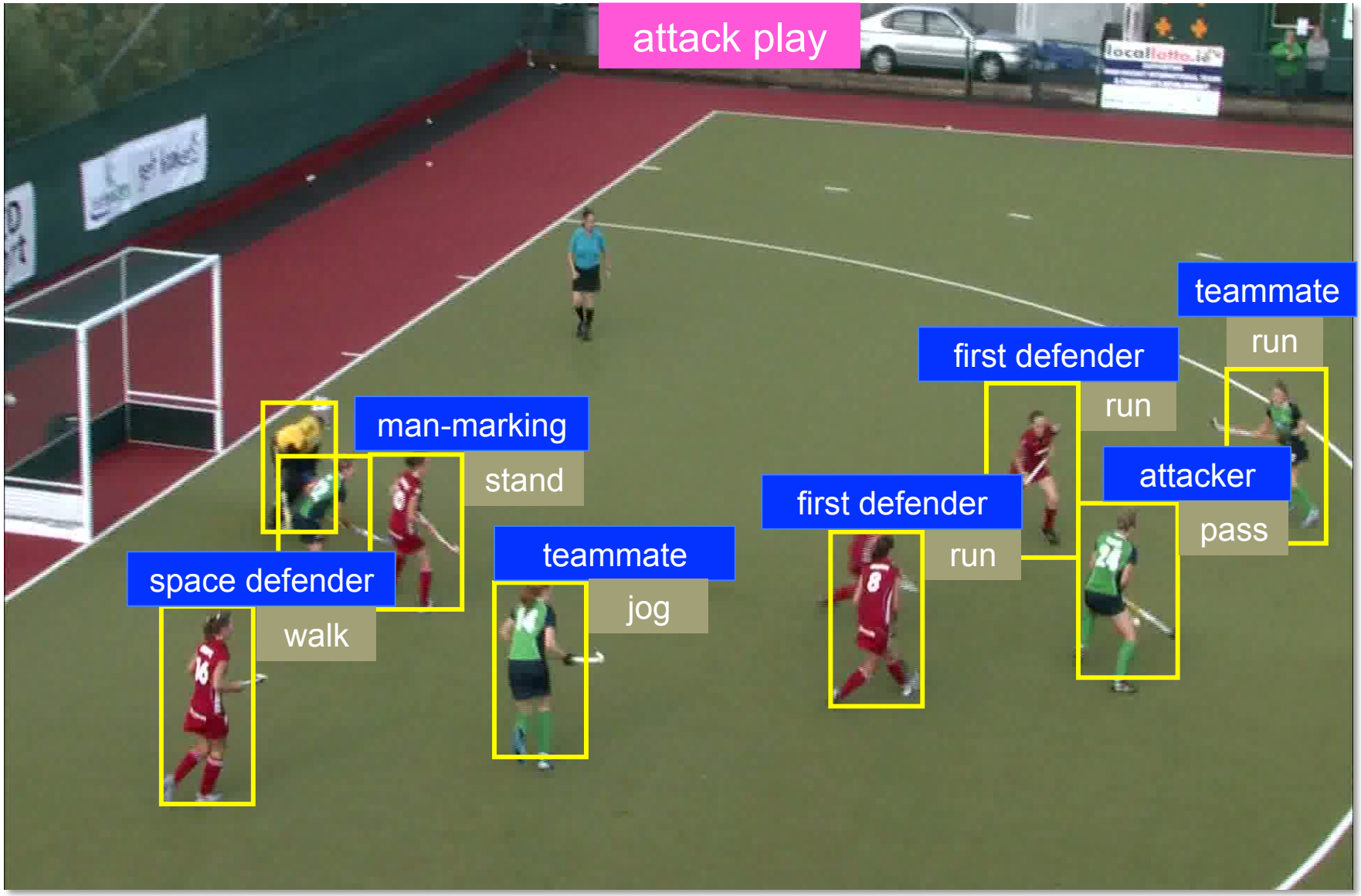
- Intertwined problems
- Build models that jointly solve these problems
  - Ability to focus on sub-problems, leverage information from others
  - Principled learning algorithms

Surveillance

VIRAT dataset: Oh et al. CVPR 2011

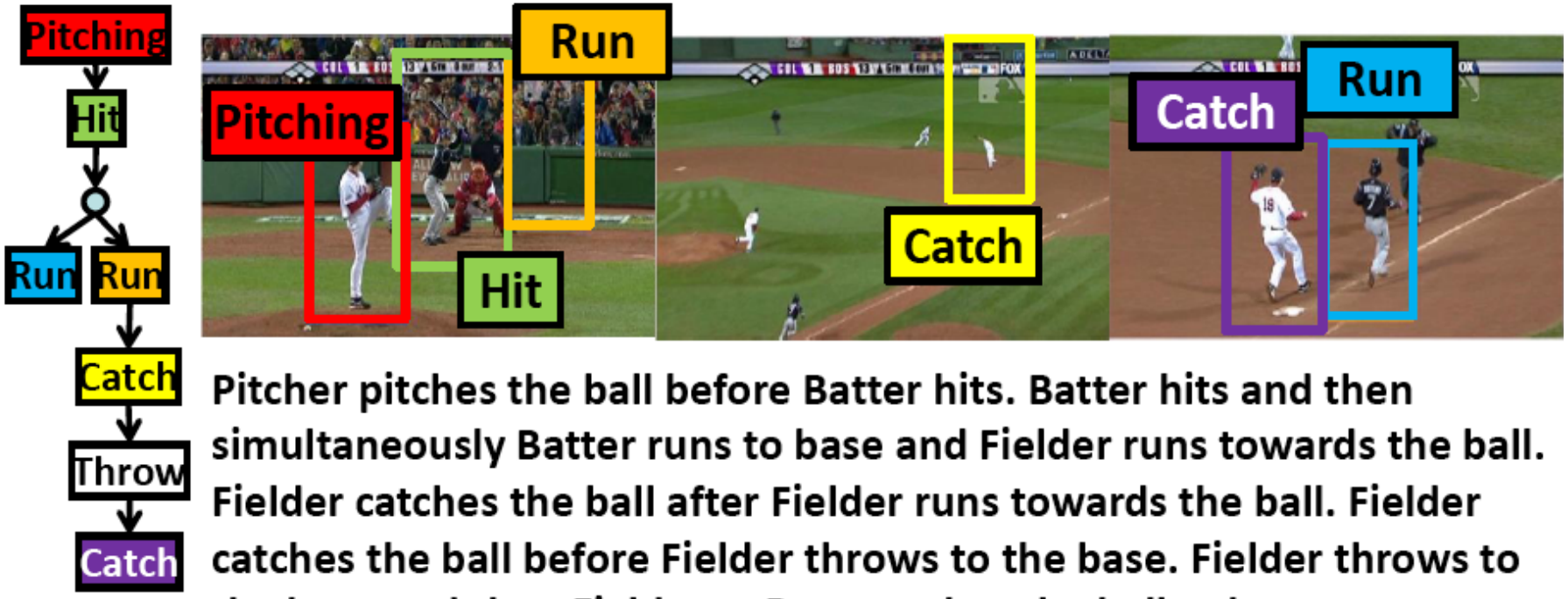
Surveillance – Road Safety

Sports Analysis



Lan, Sigal, Mori CVPR 2012

Text Generation



Pitcher pitches the ball before Batter hits. Batter hits and then simultaneously Batter runs to base and Fielder runs towards the ball. Fielder catches the ball after Fielder runs towards the ball. Fielder catches the ball before Fielder throws to the base. Fielder throws to the base and then Fielder at Base catches the ball at base .

Gupta et al. CVPR 2009

Activity landscape

Actions	Human interactions	Group activities	Events
			
Run	Point	Talk	Hockey
1	2	5-10	~10
• Bobick & Davis, 2001 • Efros et al, 2003 • Schuldt et al, 2004 • Alper & Shah, 2005 • Dollar et al, 2005 • Blank et al, 2005 • Niebles et al, 2006 • Laptev et al, 2008 • Wang & Mori, 2008 • Rodriguez et al, 2008 • Wang & Mori, 2009 • Liu et al, 2009 • Marszalek et al, 2009 •	• Oliver et al, 1998 • Medioni et al, 2001 • Park & Aggarwal, 2004 • Ryoo & Aggarwal, 2006 • Ryoo & Aggarwal, 2009 • Yuan et al, 2010 • Vahdat et al, 2011 • Patron-Perez et al., 2012 • Raptis and Sigal, 2013 • Zhu et al., 2013	• Cupillard et al, 2002 • Moore & Essa, 2002 • Vaswani et al, 2003 • Khan & Shah, 2003 • Zhang et al, 2006 • Mehran et al, 2009 • Gupta et al, 2009 • Choi & Savarese, 2009 • Lan et al, 2010 • Ryoo & Aggarwal, 2011 • Choi & Savarese, 2011 • Amer & Todorovic, 2011 •	*Number of People* • Intille & Bobick, 2001 • Loy et al, 2010 • Lan et al, 2012 • Amer et al, 2012 • Shao et al., 2014

Lan, Sigal, and Mori CVPR 2012

Role of Context in Actions

Is this a fallen person?

Role of Context in Actions

Who has the puck?

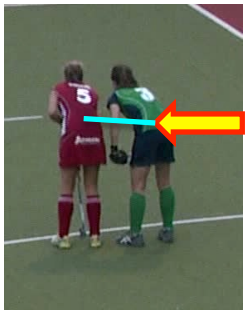
Semantic Descriptions of Videos

actions	social roles	event
walk run jog bend shoot dribble pass	attacker first defenders man-marking defend-space teammate	corner hit free hit attack play

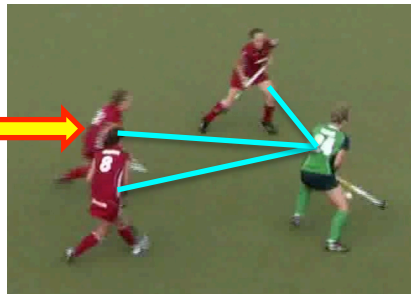
Social Roles

- Mid-level semantics that describe individual/group behaviors in the context of social interactions.

man-marking

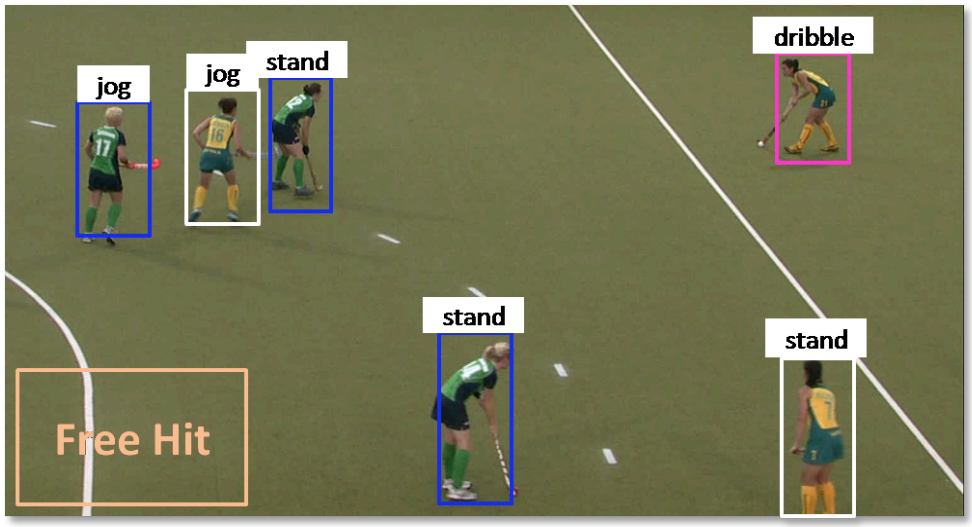


first defenders



Goal

- Label all individuals' actions, social roles and the scene-level events.

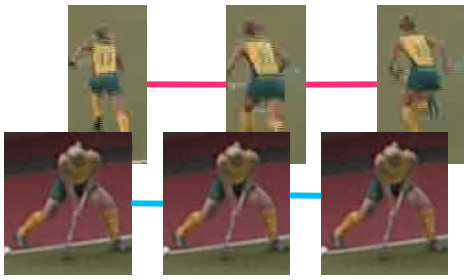


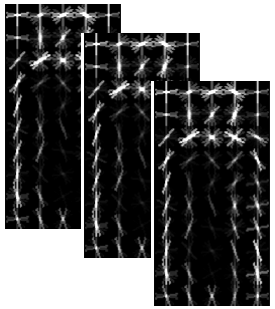
Social Roles

- attacker
- man-marking
- teammate

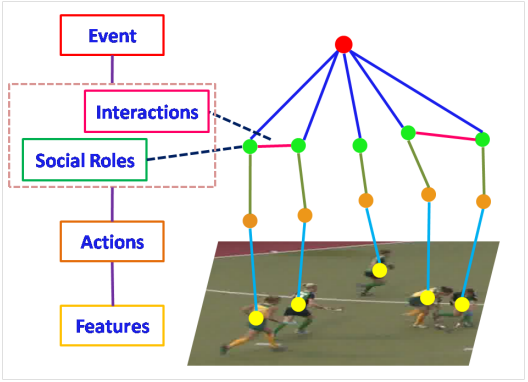
- Search for event/social role/action of interest
  - Who is the attacker? What's the overall game situation?

System Overview





Model



Activity Hierarchy Model Representation

Y :

Corner hit

Attack play

. . .

r :

Attacker

Man-marking

. . .

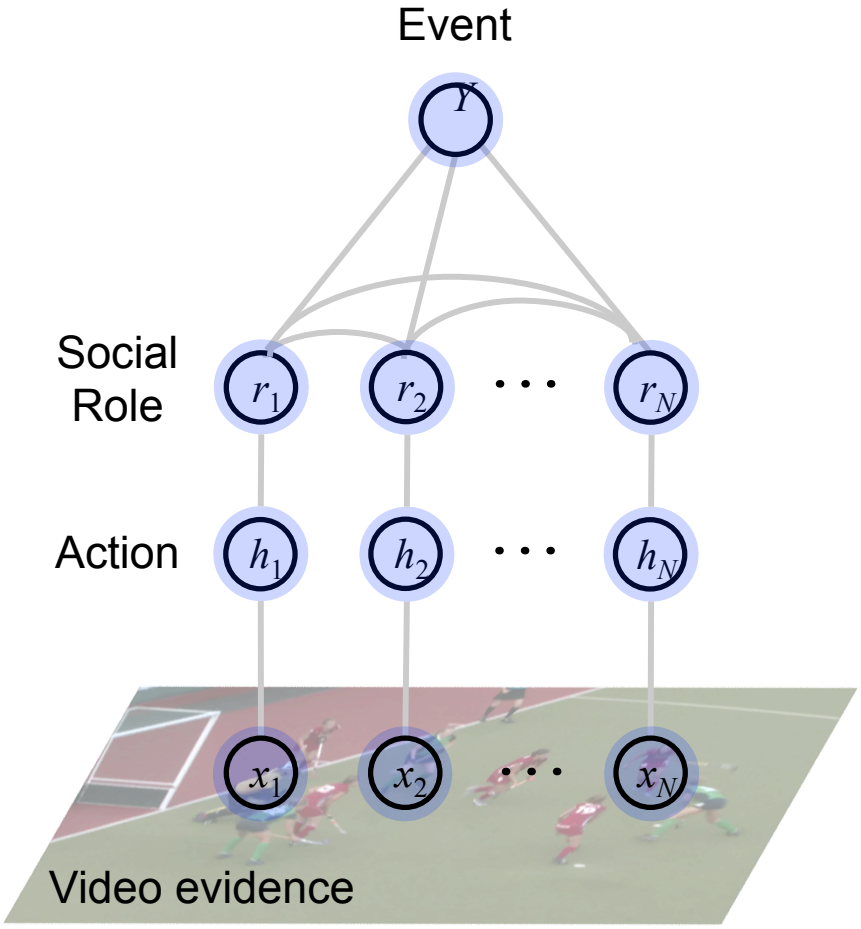
h :

Pass

jog

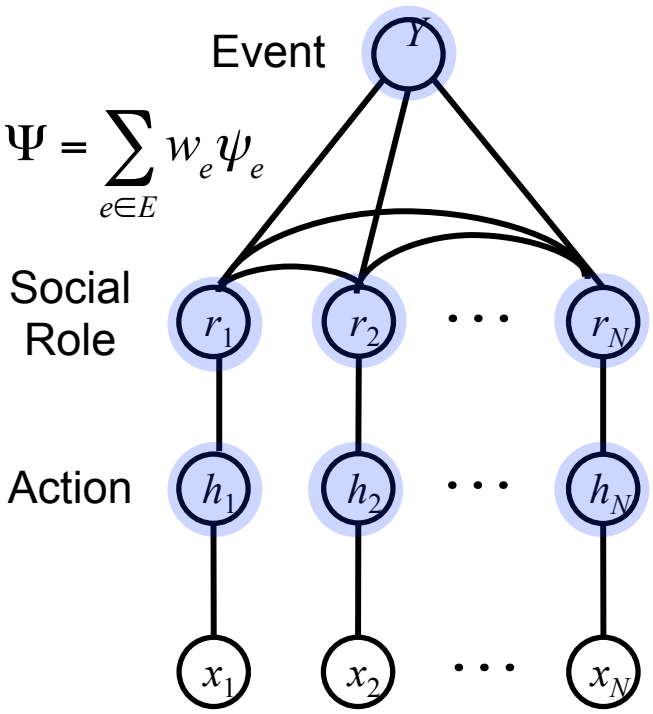
. . .

x: Concatenated HOG [Dalal & Triggs, 2005]



Lan, Sigal, and Mori CVPR 2012

Model Learning



Query for event: $loss = \Delta(y, y_i)$

$$\Delta(y, y_i) = \begin{cases} 1 & \text{if } y \neq y_i \\ 0 & \text{otherwise} \end{cases}$$

Query for social roles: $loss = \Delta(r, r_i)$

Query for actions: $loss = \Delta(h, h_i)$

Scene labeling: $loss = \Delta(y, y_i) + \Delta(r, r_i) + \Delta(h, h_i)$

$$\min_{\mathbf{w},\xi} \frac{1}{2}\|\mathbf{w}\|_2^2 + \beta \sum_i \xi_i$$

$$\text{s.t. } \forall i, y, r, h$$

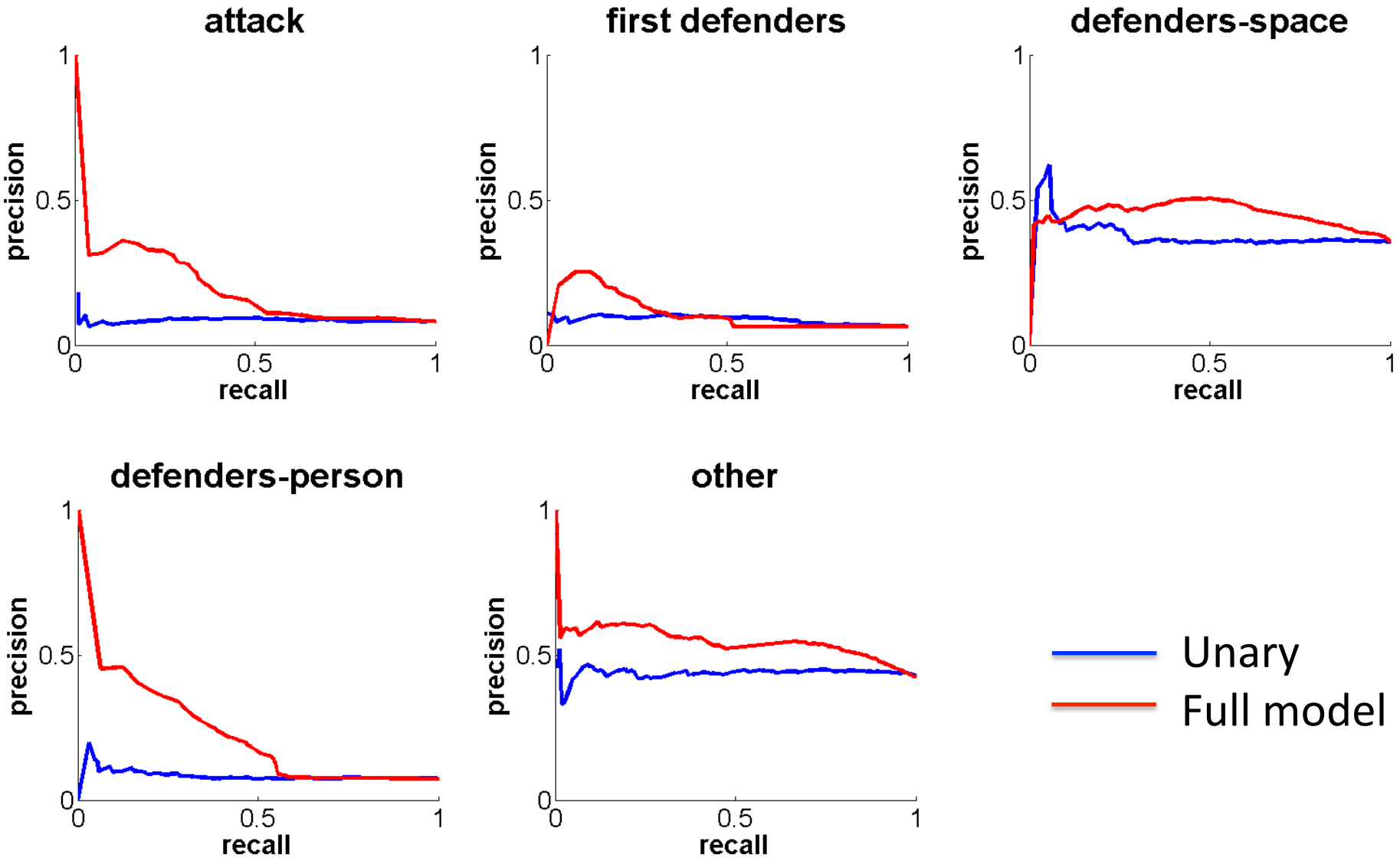
$$\mathbf{w}_{y_i r_i h_i} \cdot \psi_i - \mathbf{w}_{yrh} \cdot \psi_i \geq loss - \xi_i$$

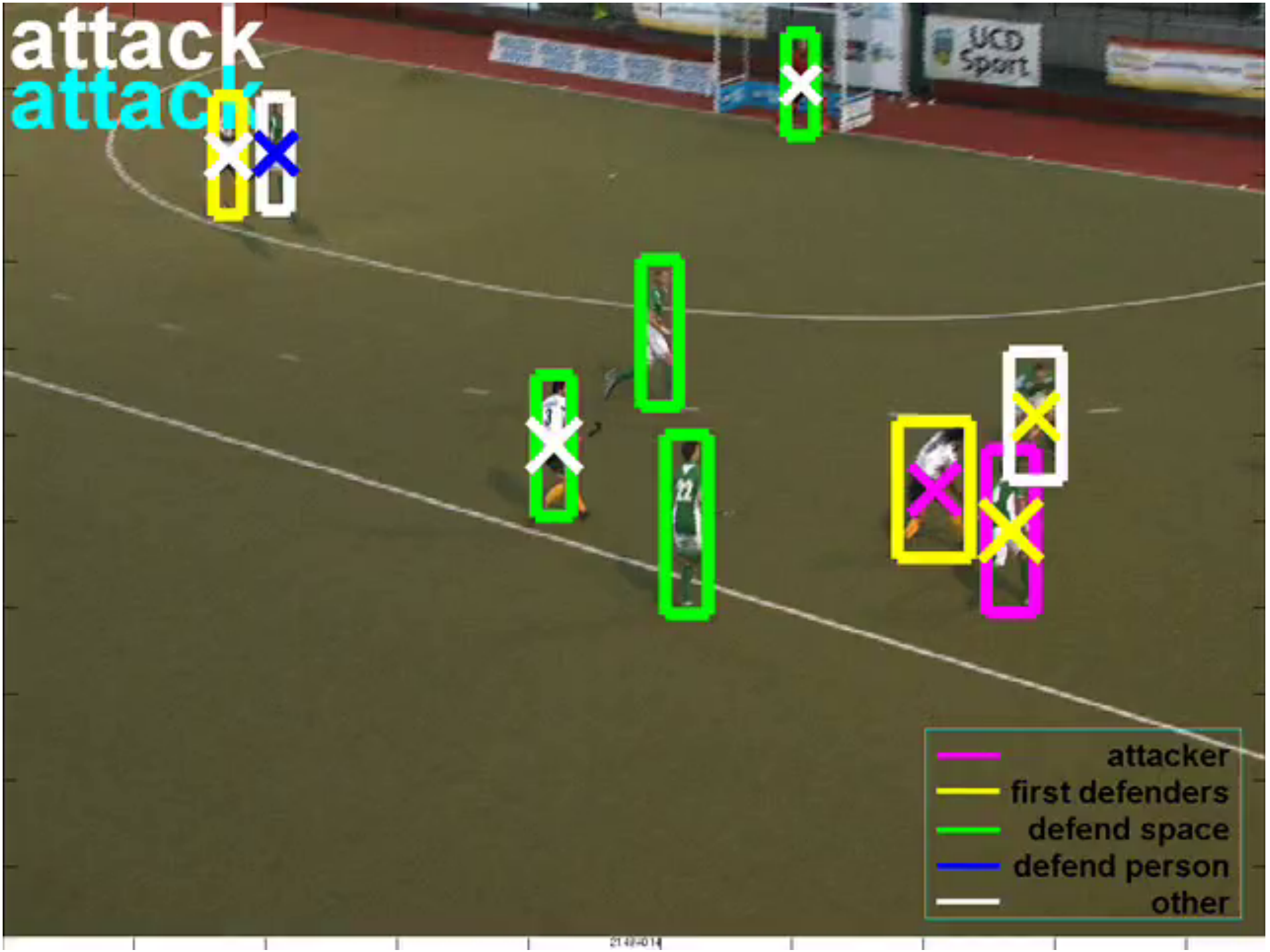
$$\forall i, \xi_i \geq 0$$

ESPN Broadcast Field Hockey Data

- 58 videos, 11 actions, 5 social roles, 3 scene-level events

Results – Query for Social Roles




attack
attack
attacker
first defenders
defend space
defend person
other

Activity landscape

Actions

Run

1

- Bobick & Davis, 2001
- Efros et al, 2003
- Schuldt et al, 2004
- Alper & Shah, 2005
- Dollar et al, 2005
- Blank et al, 2005
- Niebles et al, 2006
- Laptev et al, 2008
- Wang & Mori, 2008
- Rodriguez et al, 2008
- Wang & Mori, 2009
- Liu et al, 2009
- Marszalek et al, 2009
- ……

Human interactions

Point

2

- Oliver et al, 1998
- Medioni et al, 2001
- Park & Aggarwal, 2004
- Ryoo & Aggarwal, 2006
- Ryoo & Aggarwal, 2009
- Yuan et al, 2010
- Vahdat et al, 2011
- Patron-Perez et al., 2012
- Raptis and Sigal, 2013
- Zhu et al., 2013

Group activities

Talk

5-10

- Cupillard et al, 2002
- Moore & Essa, 2002
- Vaswani et al, 2003
- Khan & Shah, 2003
- Zhang et al, 2006
- Mehran et al, 2009
- Gupta et al, 2009
- Choi & Savarese, 2009
- Lan et al, 2010
- Ryoo & Aggarwal, 2011
- Choi & Savarese, 2011
- Amer & Todorovic, 2011
- ……

Events

Hockey

~10

Number of People

- Intille & Bobick, 2001
- Loy et al, 2010
- Lan et al, 2012
- Amer et al, 2012
- Shao et al., 2014

Interaction Models I: Rule-based System

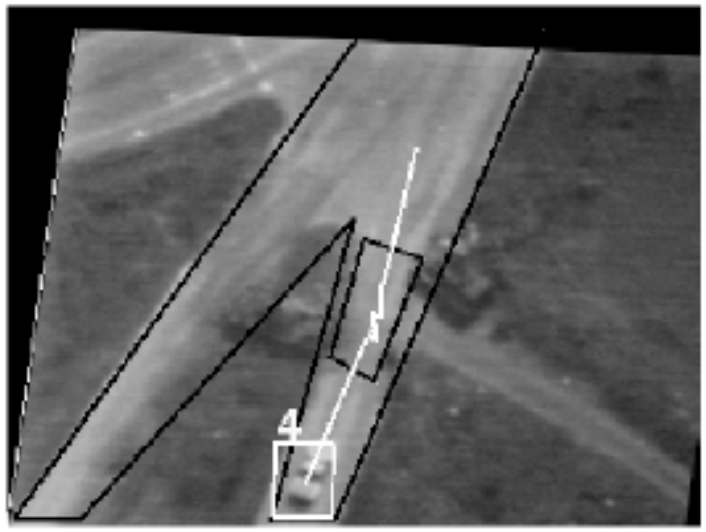
Detect and track moving objects

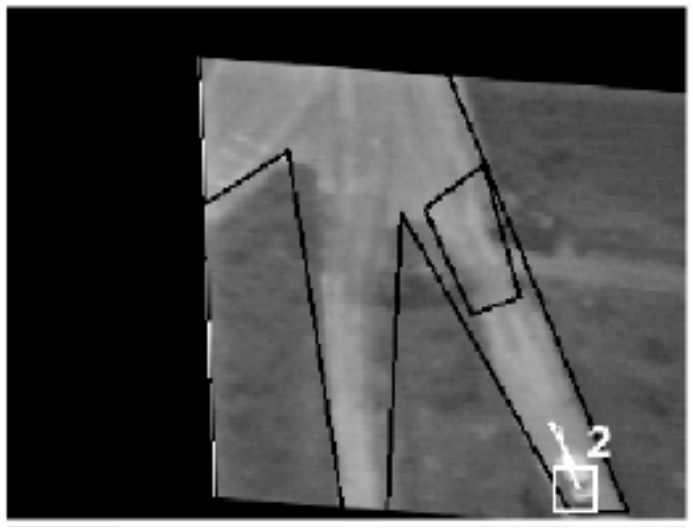
Manually identify key regions in scene

- E.g. road, checkpoint

Scenarios **describe relative arrangements of objects in scene**

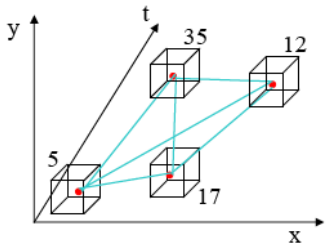
- E.g. proximity of car to checkpoint
- Notions of scene **context**

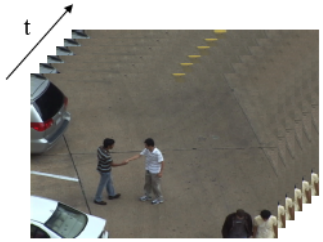


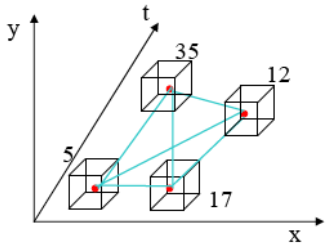


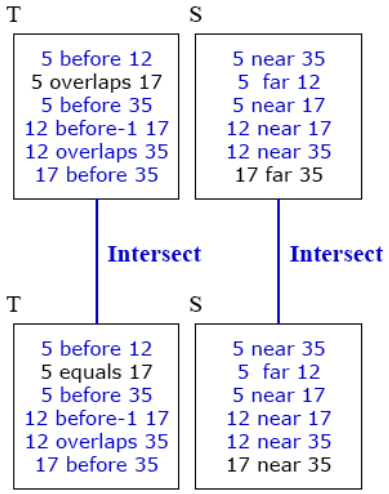
Medioni, Cohen, Bremond, Hongeng, Nevatia PAMI 2001

Interaction Models II: Local Feature Relationships





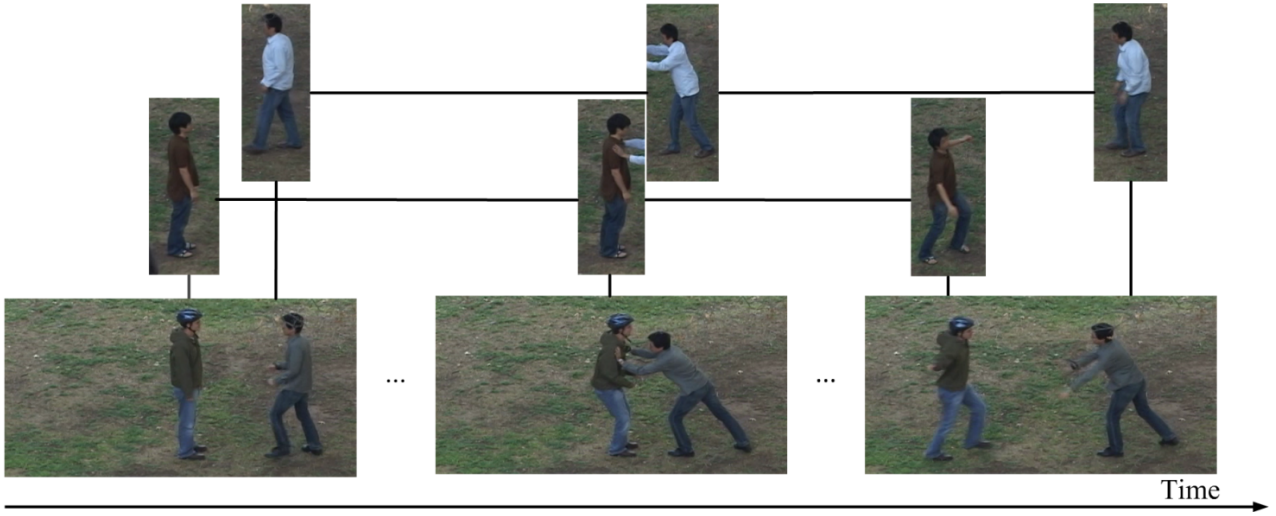




- UT Interaction dataset
- Focus on interactions between people
- Local feature approach
  - Define spatio-temporal relationships between points
  - Novel kernel for comparing sets

Ryoo and Aggarwal ICCV 2009

Interaction Models III: Key poses

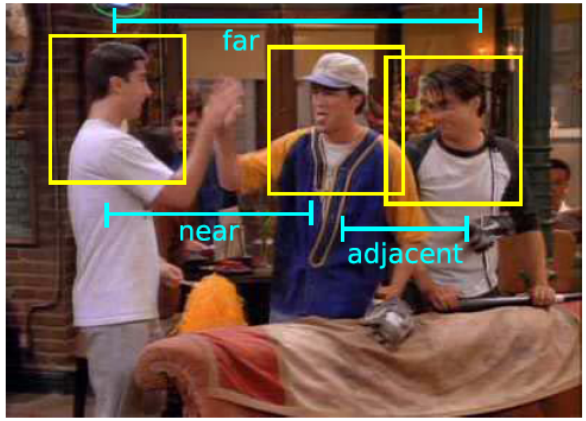


- Identify key points in an interaction
  - Poses of people
  - Relative positions of people
- Latent SVM formulation

Vahdat, Gao, Ranjbar, Mori VS 2011

Interaction Models IV: Upper-body Position / App.

- Human-human interactions in TV shows (dataset)
  - Upper-body detector
  - Estimate head orientation
  - Model combination of relative positions, facing, appearance



Patron-Perez, Marszalek, Zisserman, Reid, BMVC 2010

Interaction Models V: Human-object interactions

opening trunk getting out of vehicle entering a facility exiting a facility

- Human-object interactions in surveillance video (VIRAT)
  - Detect moving objects in video
  - Features: velocity, object type, proximity to vehicle trunk, person is carrying object, person is in parking lot, …
  - Structural learning for parameters

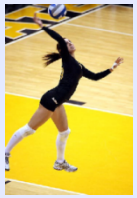
Zhu, Nayak, Roy-Chowdhury CVPR 2013

Interaction Models VI: Objects and Human Pose

A:


Tennis forehand


Croquet shot


Volleyball smash

. . .

O:

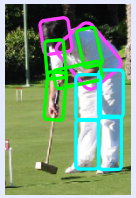

Tennis racket

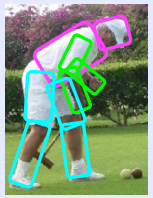

Croquet mallet

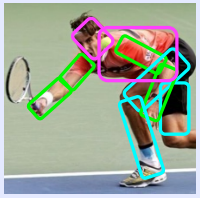
Volleyball

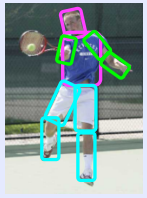
. . .

H:







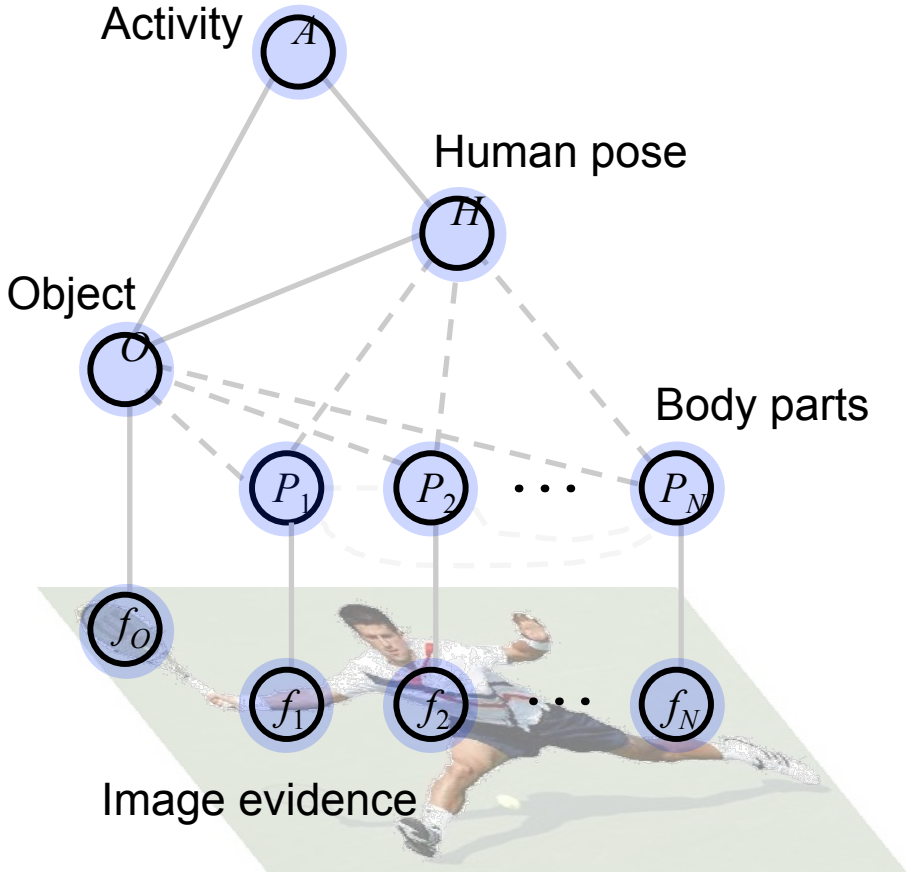


Intra-class variations

- More than one H for each A;
- **Unobserved** during training.

P: l_P: location; θ_P: orientation; s_P: scale.

f: Shape context. [Belongie et al, 2002]



Yao and Fei-Fei, CVPR 2010

Activity landscape

Actions

Run

Human interactions

Point

Group activities

Talk

Events

Hockey

1

- Bobick & Davis, 2001
- Efros et al, 2003
- Schuldt et al, 2004
- Alper & Shah, 2005
- Dollar et al, 2005
- Blank et al, 2005
- Niebles et al, 2006
- Laptev et al, 2008
- Wang & Mori, 2008
- Rodriguez et al, 2008
- Wang & Mori, 2009
- Liu et al, 2009
- Marszalek et al, 2009
- ……

2

- Oliver et al, 1998
- Medioni et al, 2001
- Park & Aggarwal, 2004
- Ryoo & Aggarwal, 2006
- Ryoo & Aggarwal, 2009
- Yuan et al, 2010
- Vahdat et al, 2011
- Patron-Perez et al., 2012
- Raptis and Sigal, 2013
- Zhu et al., 2013

5-10

- Cupillard et al, 2002
- Moore & Essa, 2002
- Vaswani et al, 2003
- Khan & Shah, 2003
- Zhang et al, 2006
- Mehran et al, 2009
- Gupta et al, 2009
- Choi & Savarese, 2009
- Lan et al, 2010
- Ryoo & Aggarwal, 2011
- Choi & Savarese, 2011
- Amer & Todorovic, 2011
- ……

~10

Number of People

- Intille & Bobick, 2001
- Loy et al, 2010
- Lan et al, 2012
- Amer et al, 2012
- Shao et al., 2014

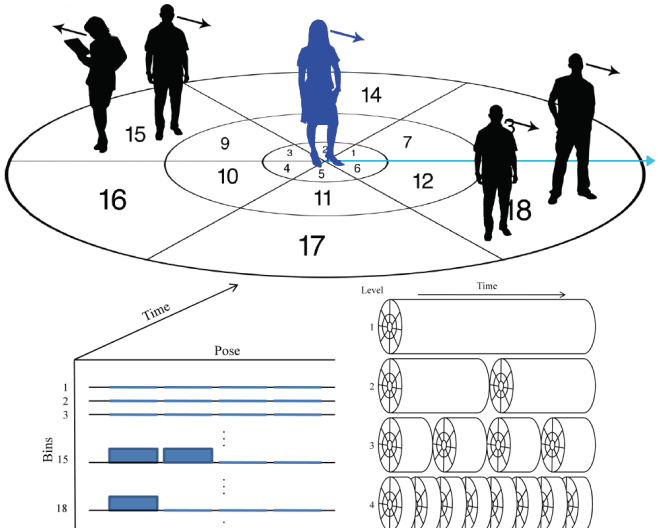
Group Models I: Stochastic Grammars

- Probabilistic grammar for describing domain
- Hand tracking, model object interactions, interactions between players in game
- Detect player strategies, roles, low-level actions

Moore and Essa, AAAI 2002

Group Models II: Person context

- Collective activity dataset
  - Pedestrian behaviours
  - Activities are context-dependent
    - Actions of others in the scene
  - Spatio-temporal context containing pose of other people in scene



Choi, Shahid, Savarese VS 2009

Group Models III: Adaptive Structures

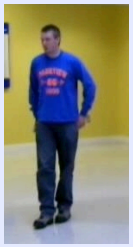
Y:

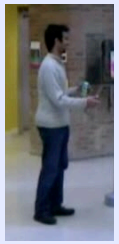
Talk

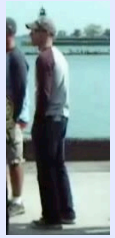
Queue

. . .

h:

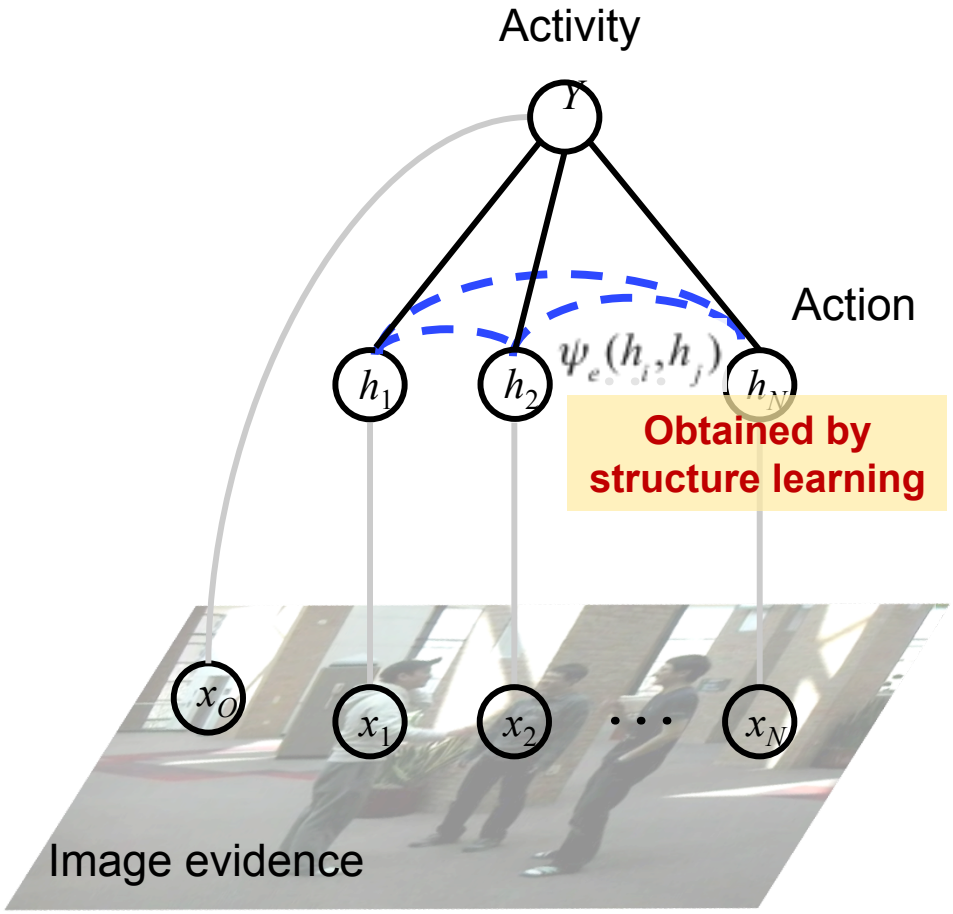






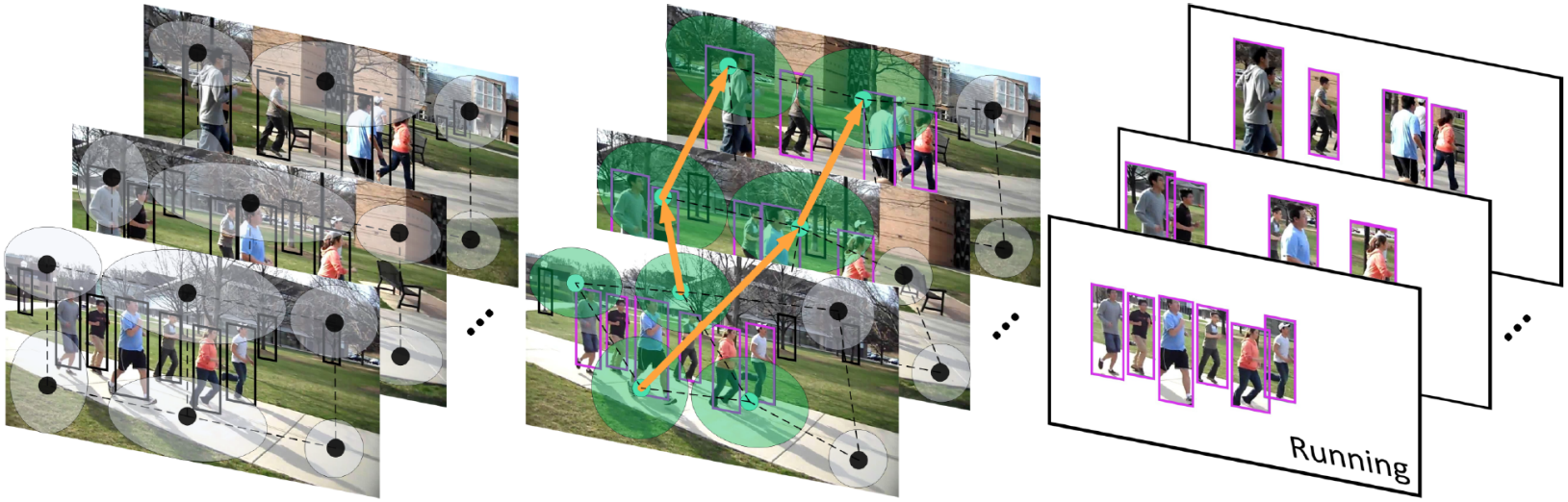
. . .

x: HOG [Dalal & Triggs, 2005]



Lan, Wang, Yang, Mori, NIPS 2010

Group Models IV: Temporal extent

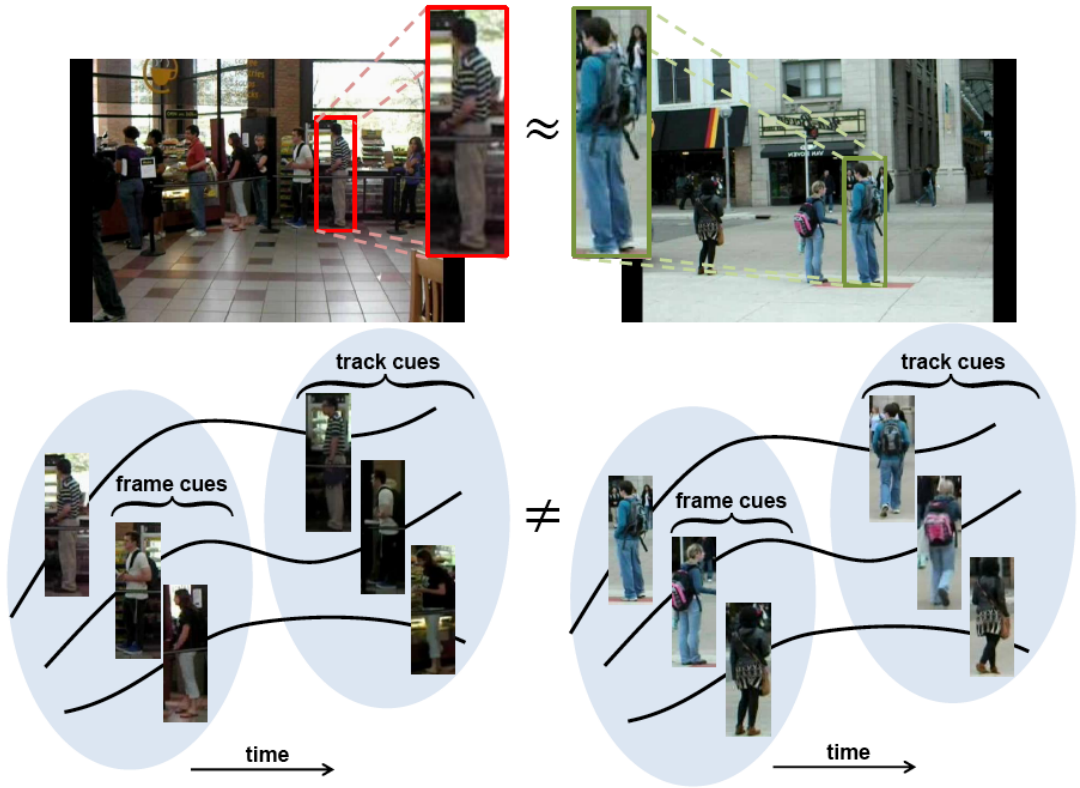


- "Chains model"
  - Find people whose actions are relevant to the group activity
  - Reason about temporal links, data association between noisy detections, temporal extent of a group activity

Amer and Todorovic, ICCV 2011

Group Models V: Linking tracklets

- Collective activity dataset
  - Features on individual detections
  - Features on tracks
    - Short tracklets
  - Data association terms
- Efficient inference



Khamis, Morariu, and Davis, ECCV 2012

Group Models VI: Storyline Model

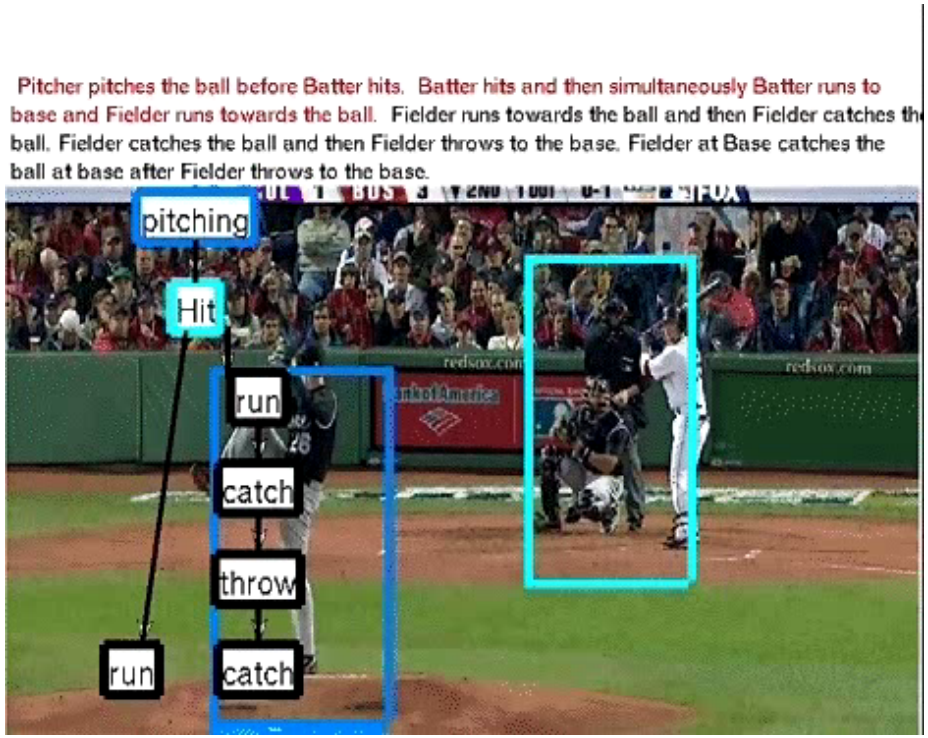
Captioned baseball videos in training

Build AND-OR graph representation of activities

- AND specifies elements of an activity that must occur
- OR allows variation in how an element appears

Describe low-level tracks using STIPs

Match tracks to actions in AND-OR graph



Gupta, Srinivasan, Shi, Davis CVPR 2009

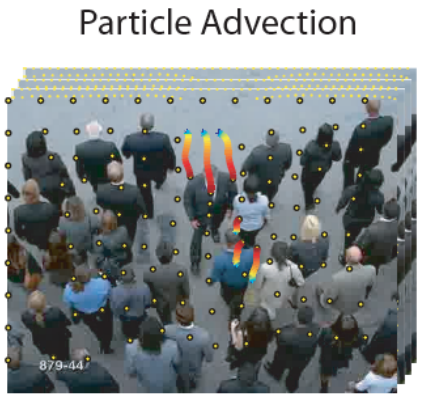
Activity landscape

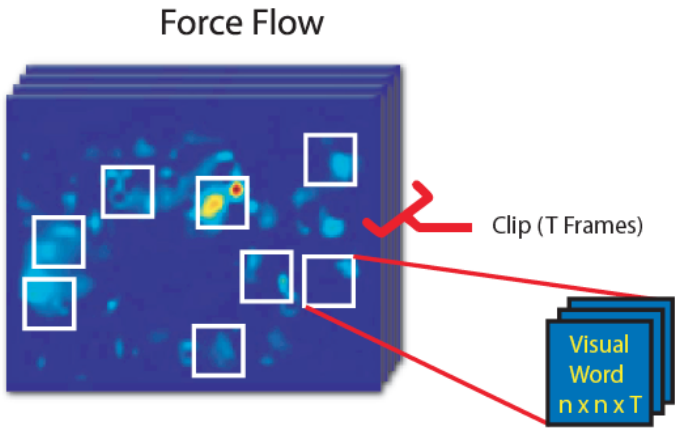
Actions	Human interactions	Group activities	Events
			
Run	Point	Talk	Hockey
1	2	5-10	~10
• Bobick & Davis, 2001 • Efros et al, 2003 • Schuldt et al, 2004 • Alper & Shah, 2005 • Dollar et al, 2005 • Blank et al, 2005 • Niebles et al, 2006 • Laptev et al, 2008 • Wang & Mori, 2008 • Rodriguez et al, 2008 • Wang & Mori, 2009 • Liu et al, 2009 • Marszalek et al, 2009 •	• Oliver et al, 1998 • Medioni et al, 2001 • Park & Aggarwal, 2004 • Ryoo & Aggarwal, 2006 • Ryoo & Aggarwal, 2009 • Yuan et al, 2010 • Vahdat et al, 2011 • Patron-Perez et al., 2012 • Raptis and Sigal, 2013 • Zhu et al., 2013	• Cupillard et al, 2002 • Moore & Essa, 2002 • Vaswani et al, 2003 • Khan & Shah, 2003 • Zhang et al, 2006 • Mehran et al, 2009 • Gupta et al, 2009 • Choi & Savarese, 2009 • Lan et al, 2010 • Ryoo & Aggarwal, 2011 • Choi & Savarese, 2011 • Amer & Todorovic, 2011 •	• Intille & Bobick, 2001 • Loy et al, 2010 • Lan et al, 2012 • Amer et al, 2012 • Shao et al., 2014

Number of People

Event Models I: Holistic flow models

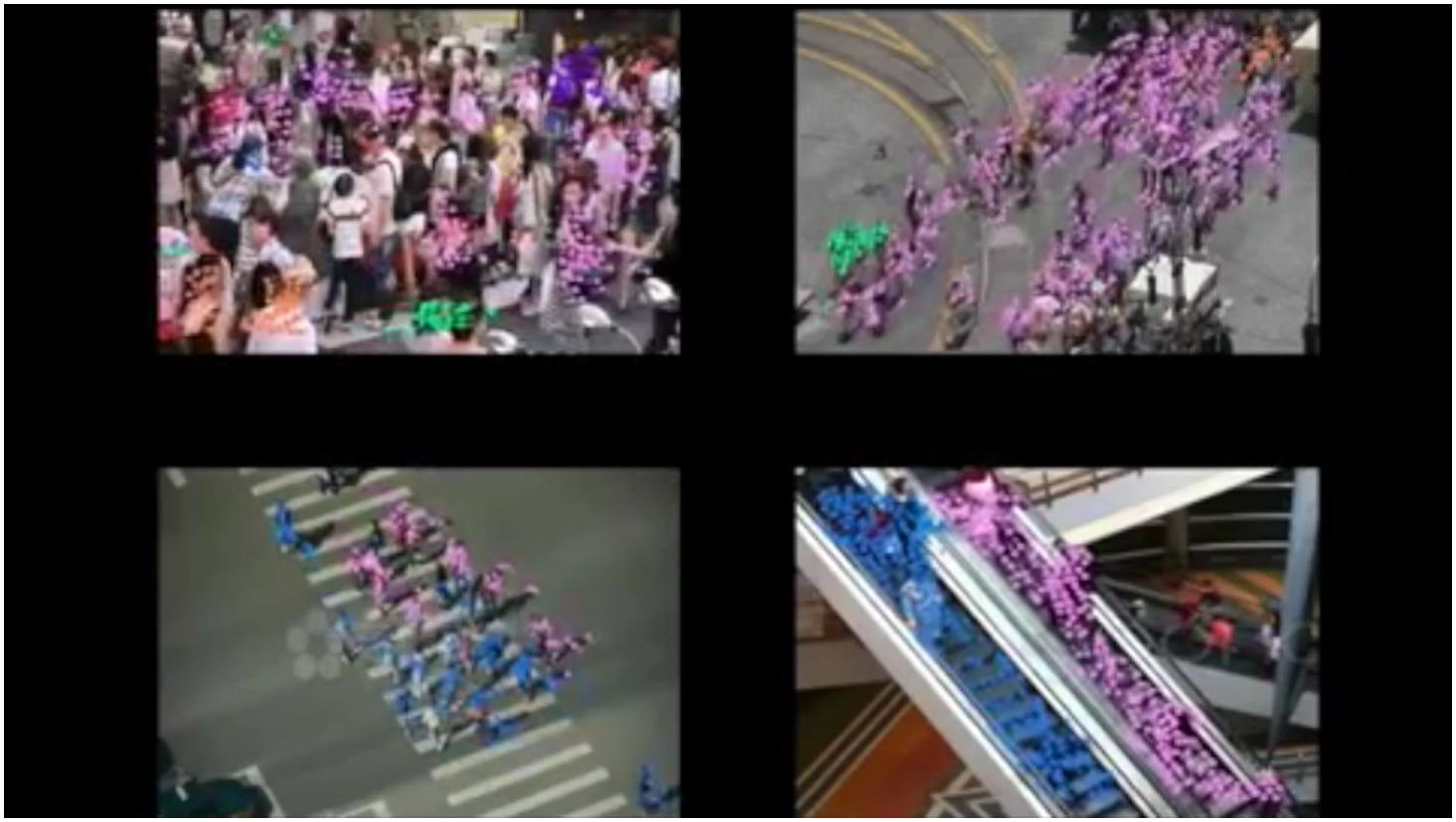
- Global model of crowd
  - No tracking of individuals
- Compute flow on dense grid
- Interaction forces between particles estimated in local regions
- Quantized into bag of words representation over the video





Mehran, Oyama, Shah, CVPR 2009

Event Models II: Group Flow



- Model group motion using Markov chain parameters – probabilistic spatial transition model (from tracklets of individuals)
- Define intra-group and inter-group measurements
  - Collectiveness: fit of individual members to group parameters
  - Stability: maintenance of nearest neighbours within a group, …
  - …

Shao, Loy, Wang, CVPR 2014

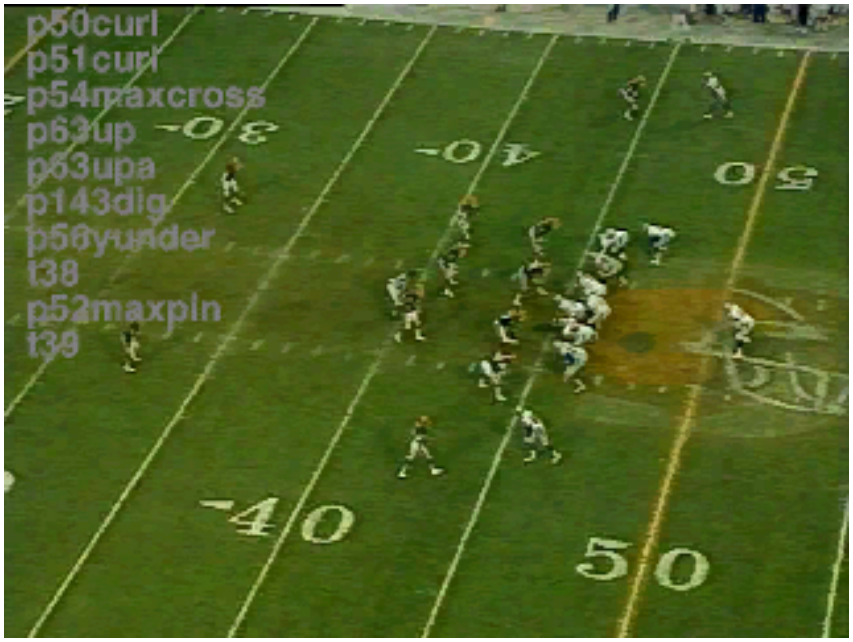
Event Models III: Bayes Nets

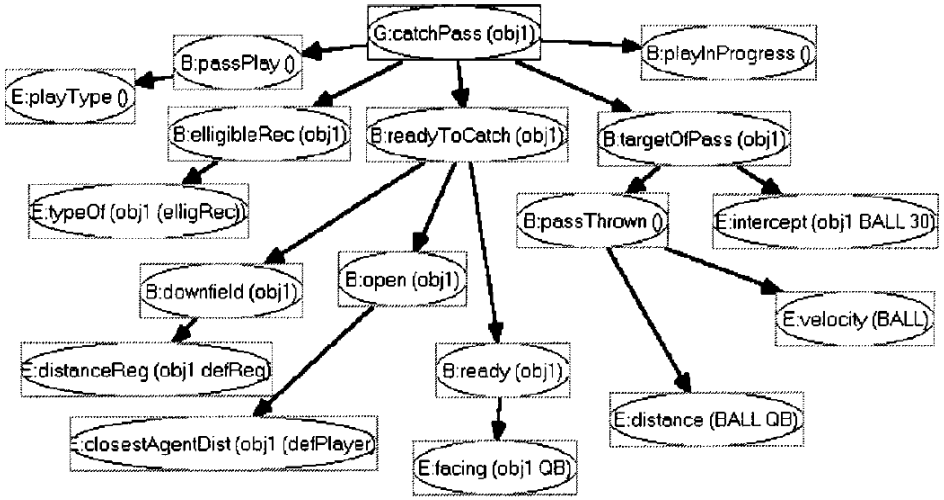
Detect and track players, ball

Low-level action detectors for individual players

Hand-constructed Bayes net for each activity

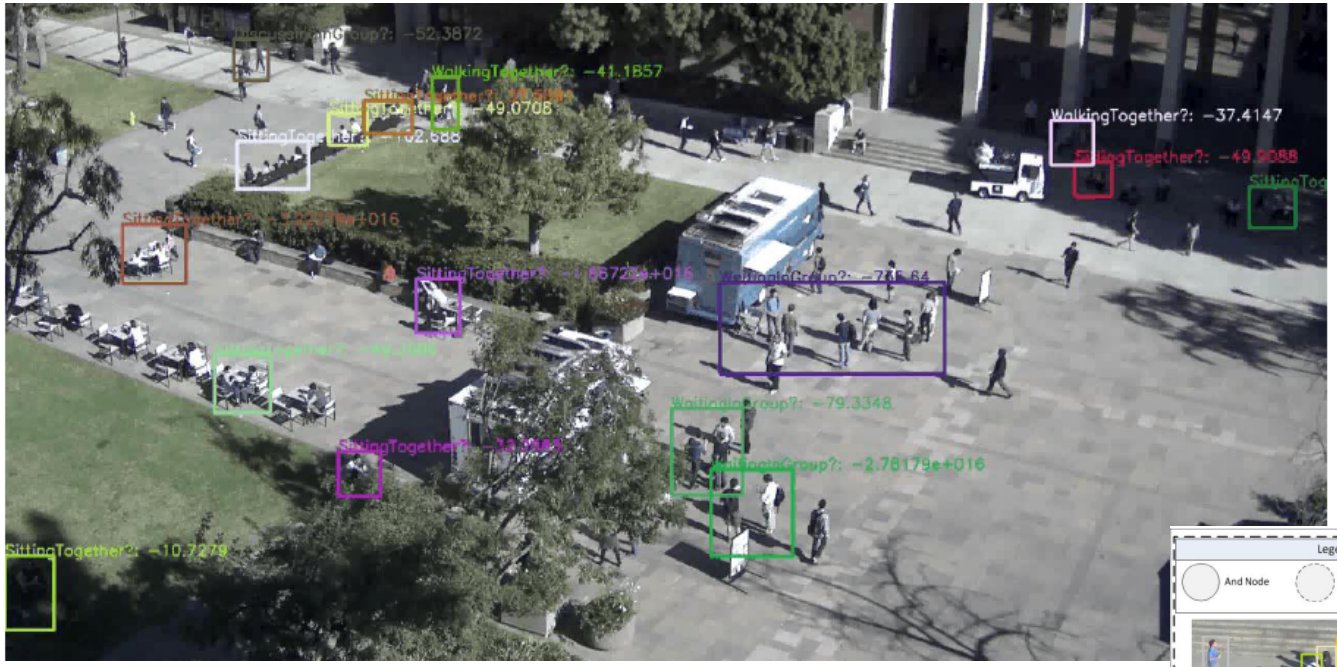
- Spatial and temporal relations between low-level actions

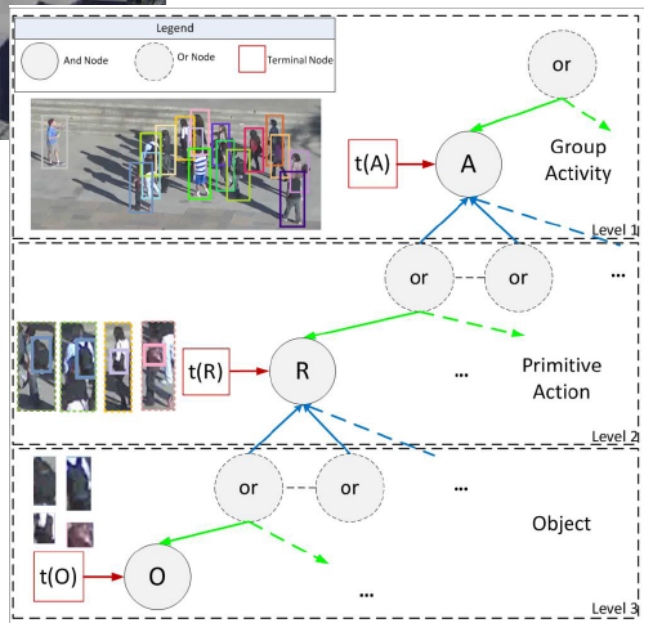




Intille & Bobick CVPR 1999 / CVIU 2001

Event Models IV: Hierarchical models





- UCLA Courtyard dataset: actions, interactions, objects
- AND-OR graph to model relationships between group activity, primitive actions, objects
    - DPM detectors for people (poses) and objects
- Efficient inference

Amer, Xie, Zhao, Todorovic, Zhu, ECCV 2012

Conclusion

Broader scope

- **Pairs of people, groups, levels of interaction**

Types of models

- **Probabilistic models**
- **Grammars**
- **Discriminative models (LSVM)**

Level of detail

- **Individual people: actions, social roles, interactions**
- **Overall crowd flow**