Human Pose Estimation from Static Video Images

Henry Medeiros
Marquette University
henry.medeiros@marquette.edu

Motivation

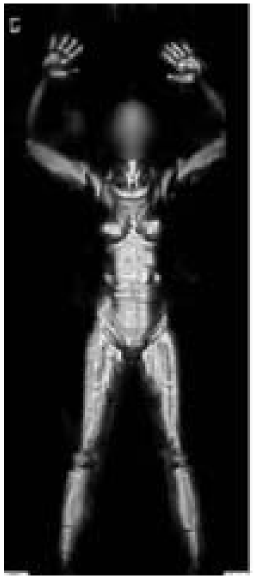
- **Objective:**
  - Measure the positions and orientations of the body parts of individuals in an image
  - A particular arrangement of image patches must correspond to a reasonable configuration of body parts

- **Approach:**
  - Use articulated body model

- **Who cares?**
  - Activities consist of combinations of poses
  - Robust human detection and localization
  - Humans have complex articulated shapes – traditional monolithic detectors perform poorly
  - Provide inputs to ATRs for people being screened by body scanners
  - Assist tracking of passengers and divested items at the checkpoint

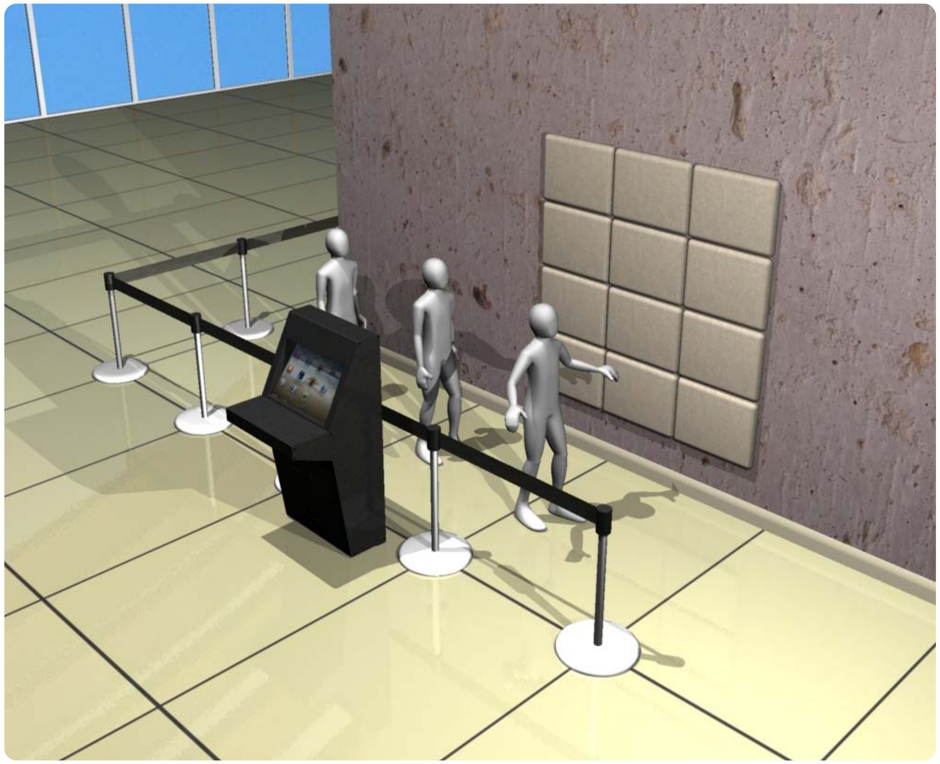
Leeds Sport Pose (LSP) dataset

Body Scanner ATR



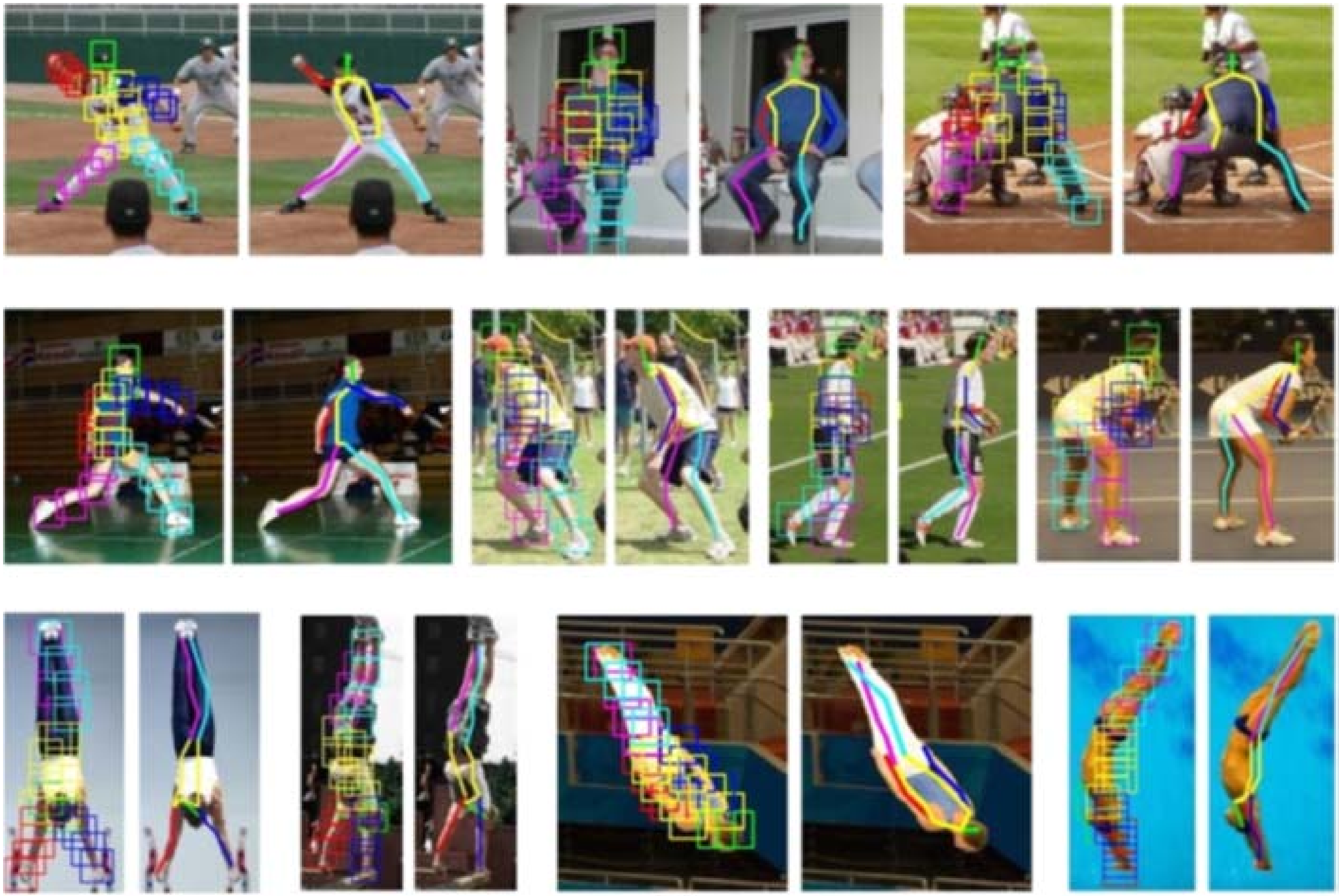
Present: L3 Provision – Fixed Pose

Future: Walkthrough



Jonah Gollub's ADSA12 presentation

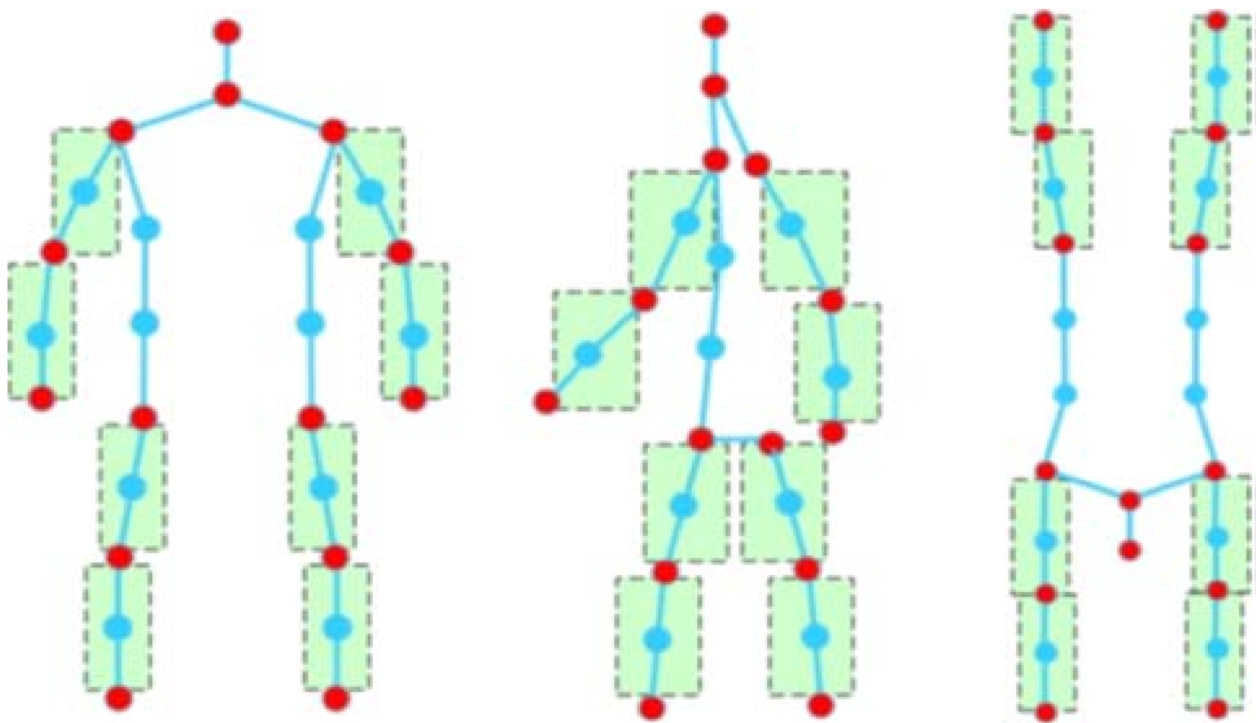
Qualitative Results



Outperforms state-of-the-art by approximately 3% on multiple complex datasets

Zhu et al.'s Pose Estimation*

- **Two stages approach**
  - Detect upper body from three pre-defined categories using simplified model
  - Select pose-specific model according to upper body detection and then perform full body pose estimation
- **Pose-specific Models**
  - Frontal view
  - Side view
  - Handstand



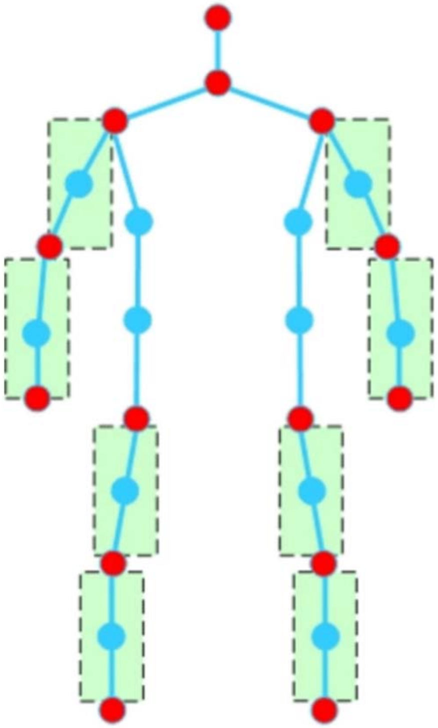
Zu et al.

*Zhu et al. "Human pose estimation with multiple mixture parts model based on upper body categories," *Journal of Electronic Imaging* 24(4), 043021 (Jul/Aug 2015).

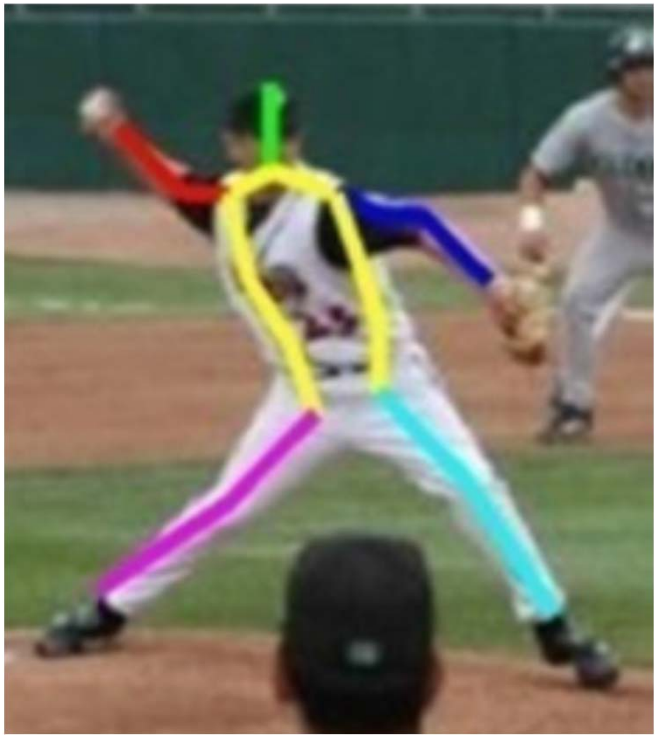
Mixture of Parts Model

- Inspired on Yang and Ramanan's Mixture of Parts Model

- The cost to fit a model to an image is comprised of three terms:
  - Appearance
  - Deformability
  - Compatibility



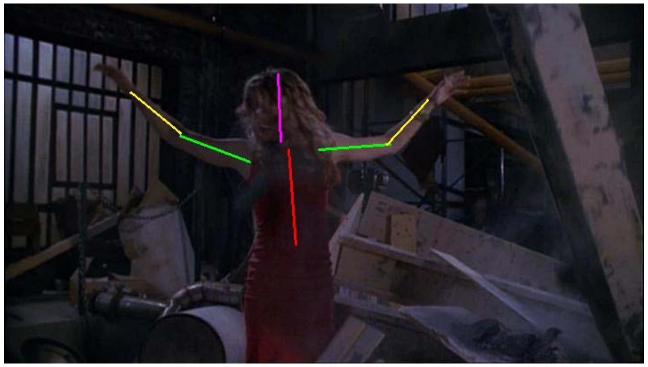
Zu et al.



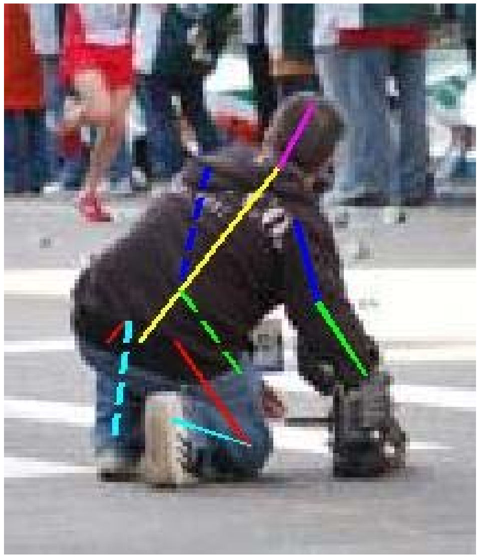
Leeds Sport Pose (LSP) dataset

Quantitative Results

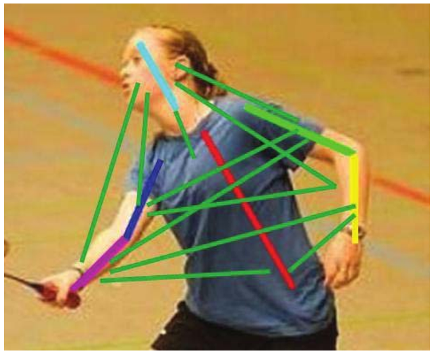
	Zhu et al.	State of the art
Upper body detection	93%	89%
Pose estimation	65% - 71%	63% - 69%



Buffy dataset



Leeds Sport Pose (LSP) dataset



UIUC people dataset

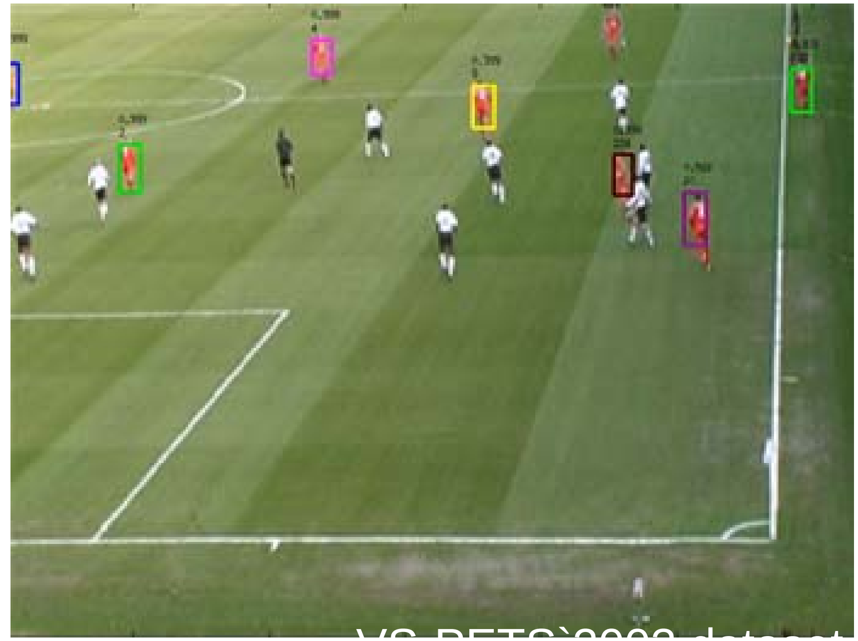
Target Tracking in Videos

Henry Medeiros
Marquette University
henry.medeiros@marquette.edu

Motivation

- **Objective:**
  - Estimate the locations of multiple targets frame-by-frame
  - Information from one frame must influence decision on the following frame

- **Approach:**
  - Recursive stochastic estimation



VS-PETS`2003 dataset

- **Who cares?**
  - Temporal information is also important
  - Improve robustness of detectors
  - Associate detections between frames
    – opens up the possibility of making conclusions about sequences of events
  - Assist tracking of passengers and divested items at the checkpoint

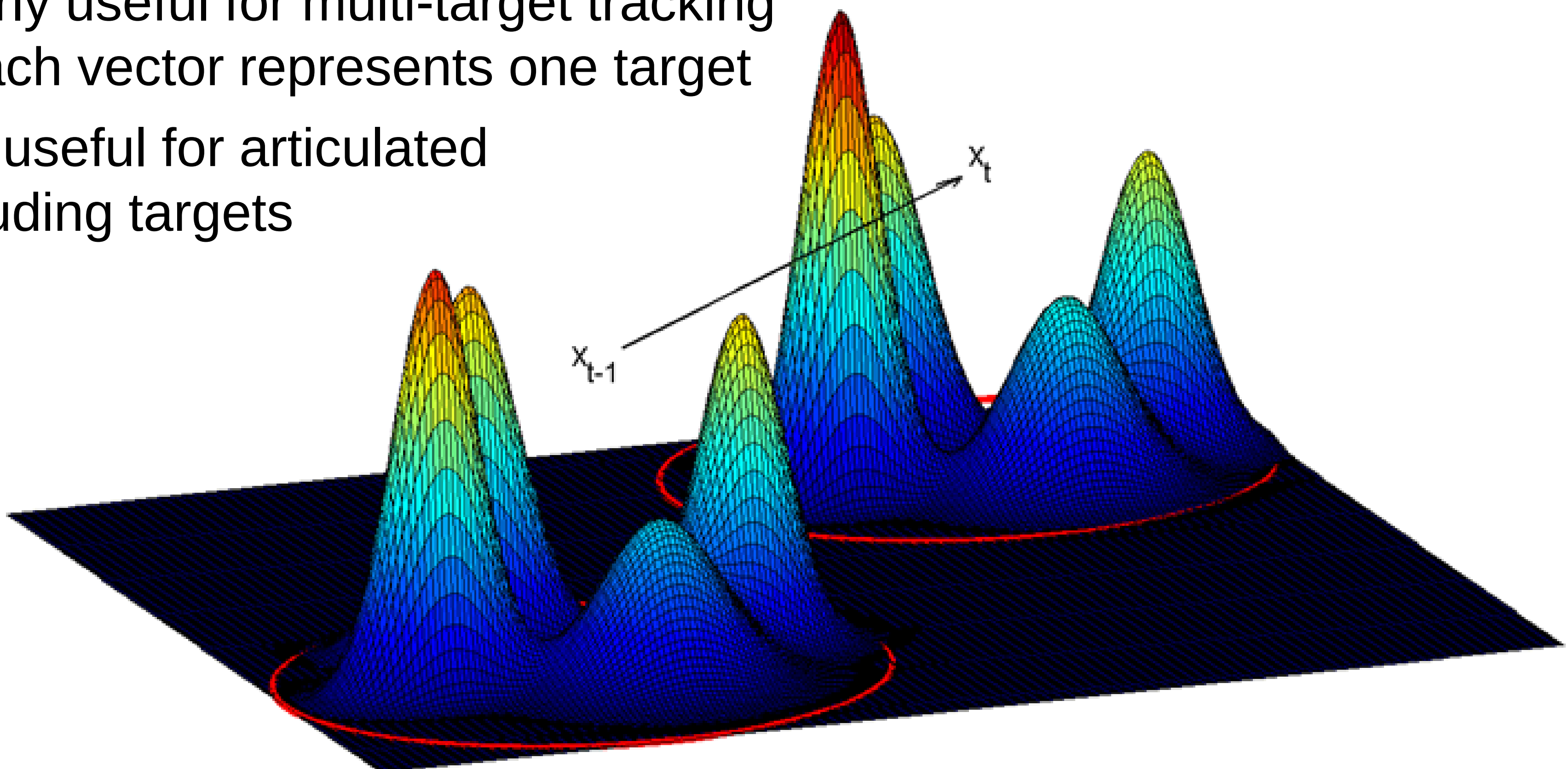
Improving Detection Robustness

Proposed approach

State of the art

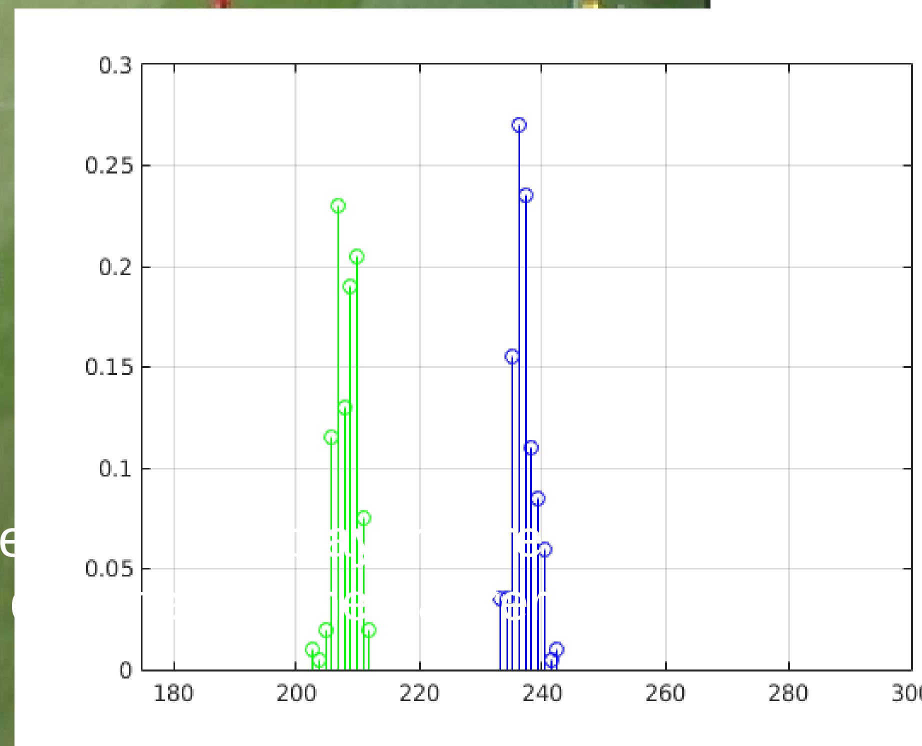
Random Finite Sets Trackers

- Recursive Bayesian framework to estimate the state of multiple targets in the presence of clutter and misdetections
- State is not a single vector but an entire set with random cardinality
- Particularly useful for multi-target tracking where each vector represents one target
- Similarly useful for articulated self-occluding targets



Redrawn from http://randomsets.eps.hw.ac.uk/tutorial.html

Monte Carlo Tracker



Assume
initially

- Image samples are collected (Monte Carlo approach)
- When targets approach, particles get mixed up
- We mitigate this problem by reducing the importance of overlapping samples

- Likelihoods proportional to desired target color
- Easy to estimate target position for well separated targets (e.g., mean)

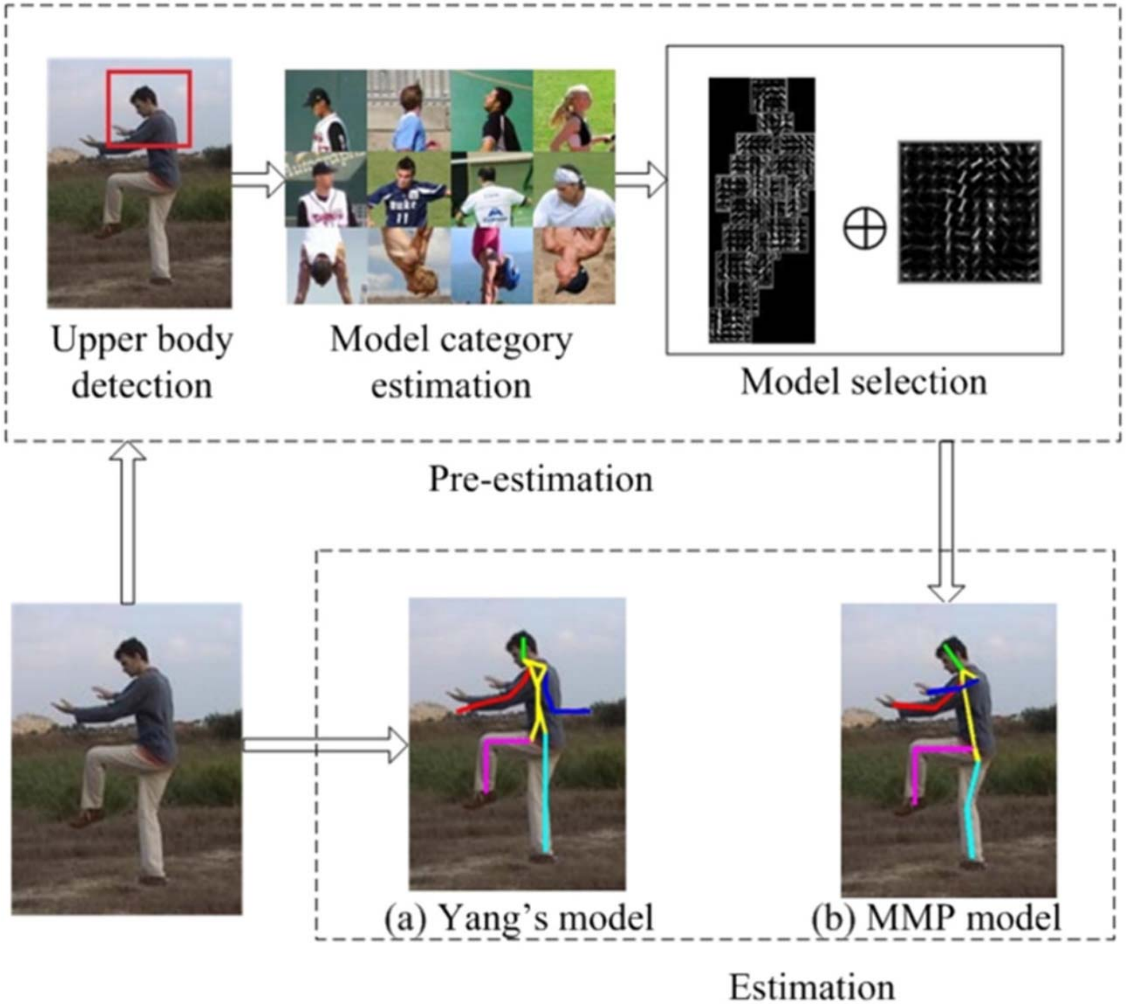
Discussion

- **Main limitations**
  - Likelihood model is the weakest link - Zhu's method could be used as a pose likelihood estimator
  - Proper use of temporal information dependent on good dynamic model – Articulated model can be used to learn pose transitions from videos
  - Computationally expensive – but highly parallelizable
- **How it could be adapted to security?**
  - Integrated with Zhu's method – pose tracker for walkthrough body scanner ATR

Extra Slides

Zhu's Upper Body Detection

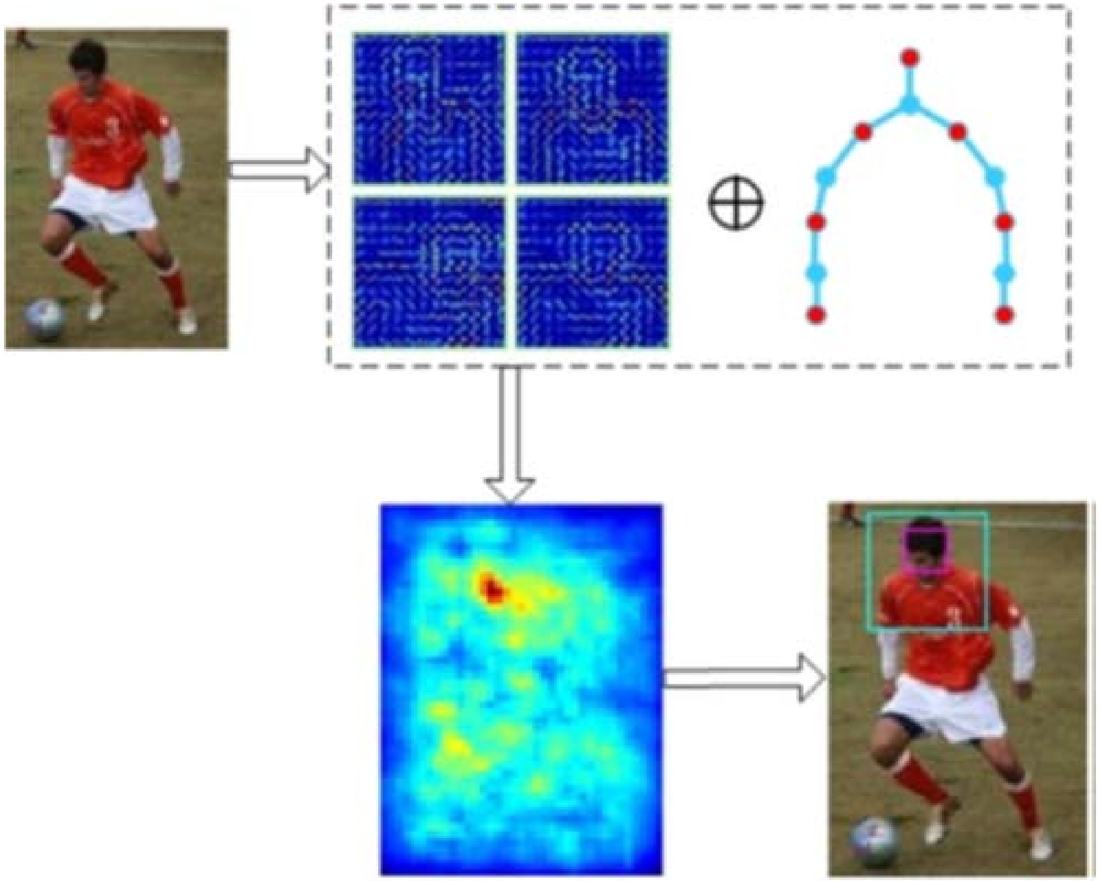
- Hierarchical approach
- Detect upper body from three categories
- Estimate pose for specific category



Zhu et al.

Upper Body Detection

- Three categories
  - Front view
  - Side view
  - Handstand
- Three matching conditions
  - Appearance
  - Deformability
  - Compatibility



Experimental Results - Upper Body

Zhu et al.

Experimental Results - Upper Body

Dataset	Method	Detection rate
Buffy	Eichner et al.[19]	89.01
	Ferrari et al.[6]	88
	Niebles and Fei-Fei[26]	73
	Ours	93.56

Zhu et al.

Experimental Results – Pose Estimation

Dataset	Method	Torso	Head	U. leg	L. leg	U. arm	L. arm	Total
LSP	Yang and Ramanan[4]	92.6	87.4	66.4	57.7	50.0	30.4	58.9
	Johnson and Everingham[8]	78.1	62.9	65.8	58.8	47.4	32.9	55.1
	Tian et al.[5]	**95.8**	**87.8**	69.9	60.0	51.9	32.8	61.3
	Johnson and Everingham[7]	88.1	74.6	**74.5**	66.5	53.7	37.5	62.7
	Pishchulin et al.[14]	88.7	85.1	63.6	58.4	46.0	35.2	58.0
	Fang and Yi[15]	91.9	86.0	74.0	**69.8**	48.9	32.2	62.8
	Ours	94.5	86.9	72.05	62.45	**57.95**	**39.75**	**64.6**
UIUC	Wang et al.[11]	86.6	68.8	56.3	50.2	30.8	20.3	47.0
	Tian et al.[5]	**98.8**	**96.8**	**78.7**	64.2	62.2	39.5	68.5
	Ours	97.57	95.95	78.34	**64.98**	**66.19**	**49.19**	**71.1**

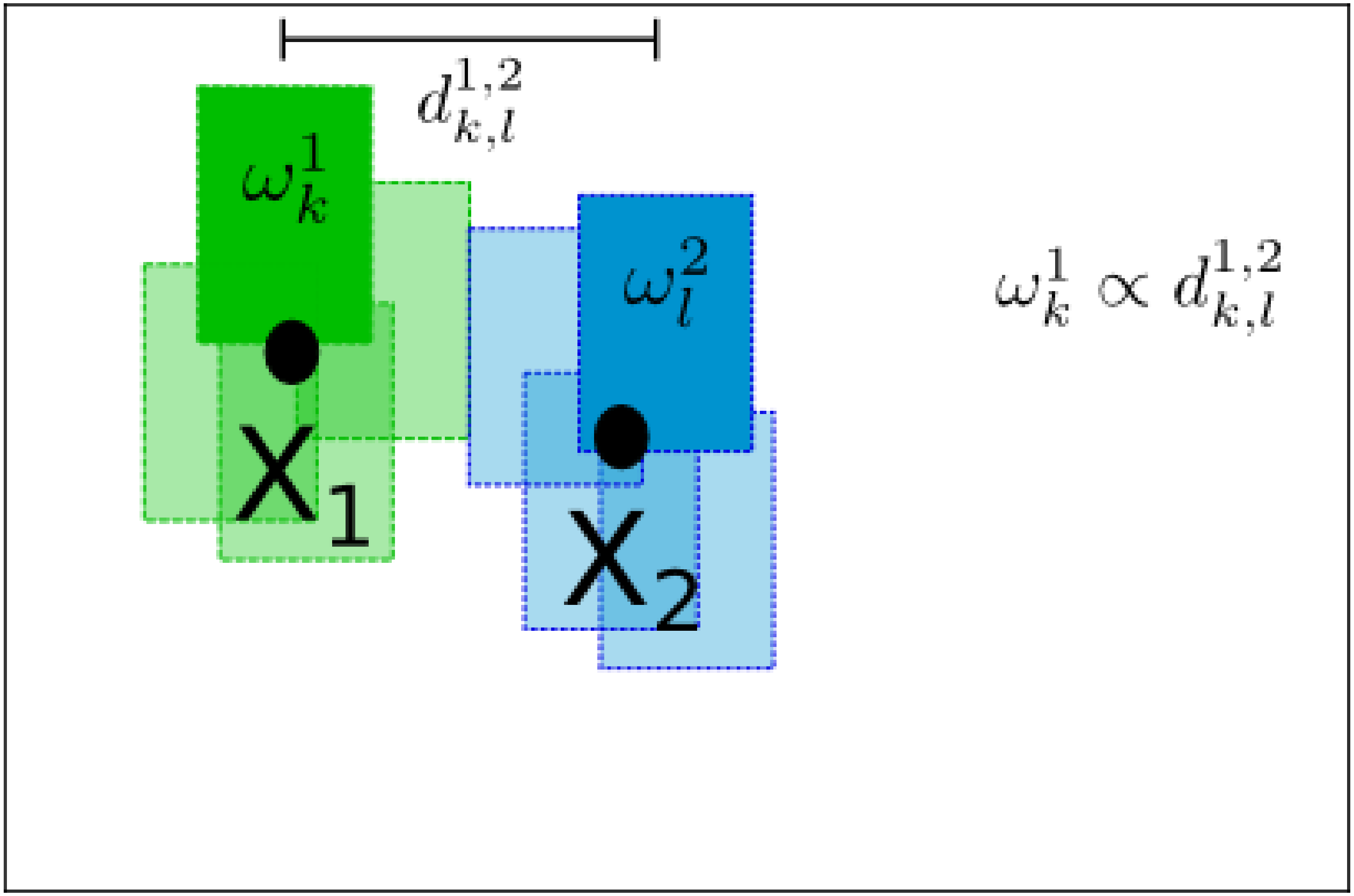
Zhu et al.

Experimental Results - Failures

- Double counting
- Occlusion
- Incorrect model selection

Zhu et al.

Interactive Likelihoods



- Particle weights proportional to distance to particles from other vectors
- Precludes samples from different vectors to overlap

The Multi-Bernoulli Filter

- A multi-Bernoulli RFS is given by the union of several independent Bernoulli RFS
- A Bernoulli RFS is empty with probability r or a singleton with probability 1-r
- A non-empty Bernoulli RFS has a spatial distribution p(.) over the state space
- The cardinality distribution of a multi-Bernoulli RFS depends on the existence probability of each individual Bernoulli RFS
- A labeled multi-Bernoulli RFS is a multi-Bernoulli RFS in which the state space is augmented with unique labels