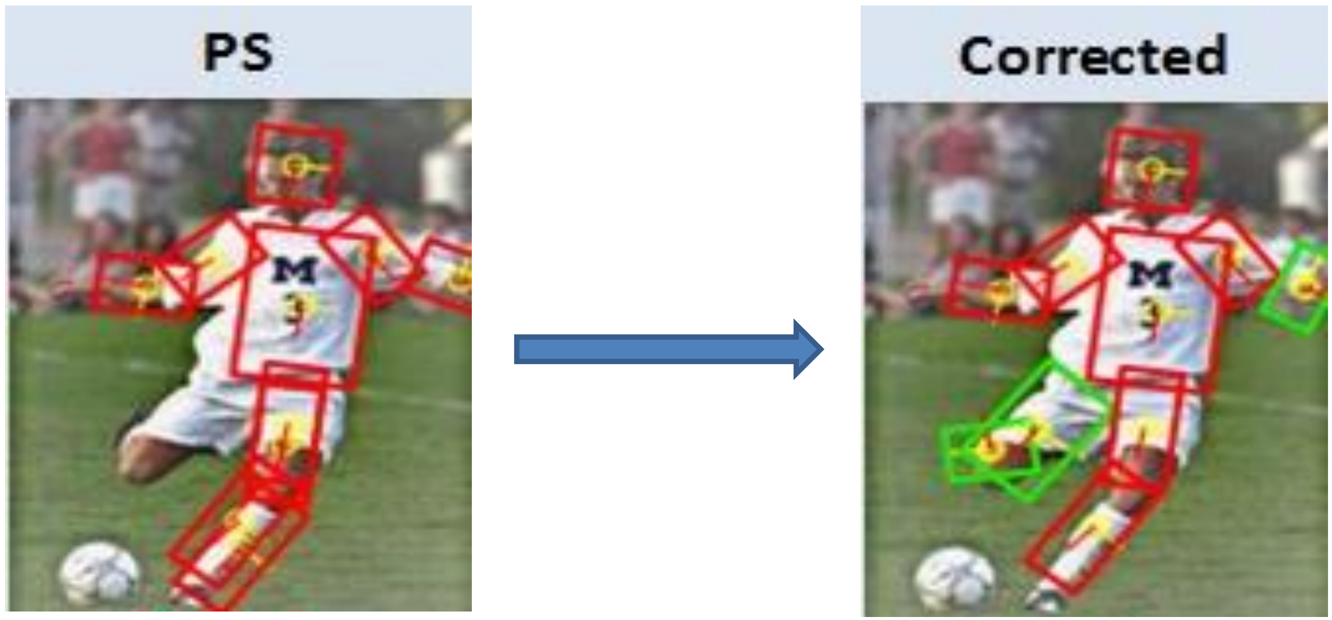
Regression based Pose Estimation with Automatic Occlusion Detection and Rectification

Ibrahim Radwan*1, Abhinav Dhall*2, Jyoti Joshi1 and Roland Goecke1,2

ibrahim.radwan@canberra.edu.au, abhinav.dhall@anu.edu.au, jyoti.joshi@canberra.edu.au, roland.goecke@ieee.org

*FIRST Equal Authors [1]University of Canberra [2]Australian National University

Goal



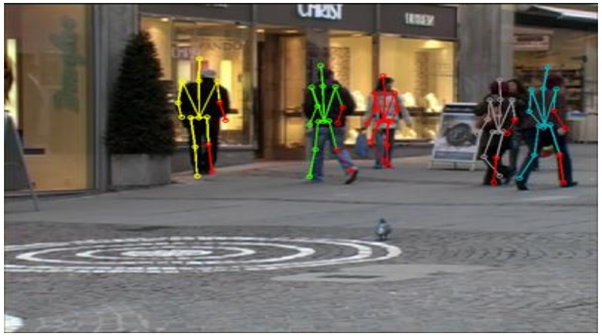
Articulated Pose Estimation

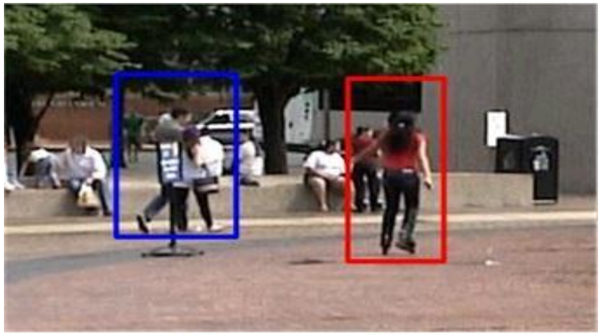
Detect the occluded parts and rectify their position and orientation

Articulated Pose Estimation

- Pose Estimation:
  - Active topic of computer vision and multimedia retrieval
  - Task of identifying the object (e.g. human) pose by determining the position and orientation relative to some coordinate system
- Challenges:
  - Large degree of freedom for articulated objects
  - Cluttered backgrounds

Articulated Pose Estimation (2)

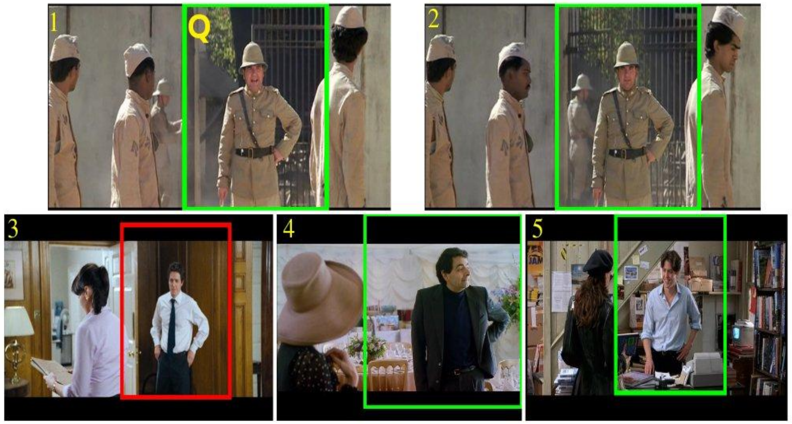




walk skate

- **Applications:**
  - Search by pose
  - Surveillance
  - Pedestrian detection
  - Video indexing
  - Sport analysis
  - Human action recognition

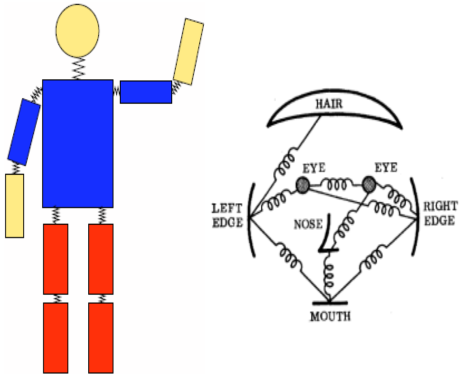
query



results

Articulated Pose Estimation (3)

- Approaches:
  1. Inferring Poses from skeletal Points
  2. Model-based methods ➔ Pictorial Structure



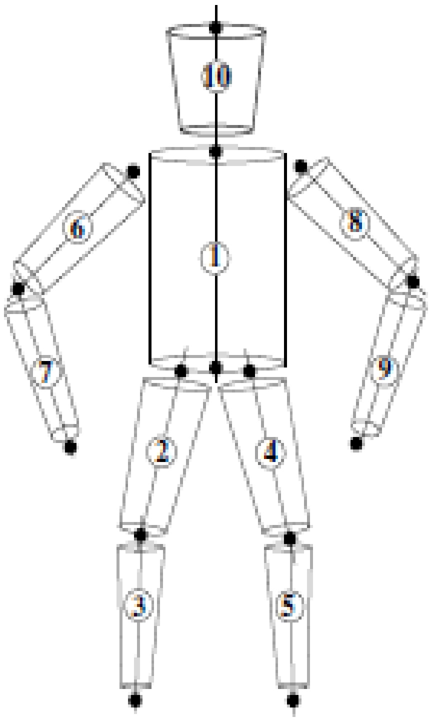
Fischler & Elschlager 1973

Felzenszwalb & Huttenlocher 2005

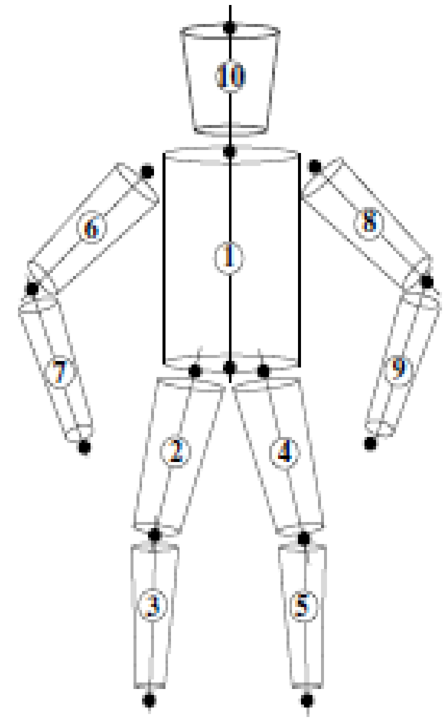
- Pictorial Structure:
  - Collection of parts with connections between certain pairs of parts.
  - **Some previous Work**
    - *Felzenszwalb et al. (IJCV 2005),*
    - *Ramanan (NIPS 2006)*
    - *Felzenszwalb et al. (CVPR 2008),*
    - *Ferrari et al. (CVPR 2008), Ferrari et al. (CVPR 2009)*
    - *Andriluka et al. (CVPR 2009)*
    - *Yang et al. (CVPR 2011)*

Pictorial Structure Model

- Combination of two models:
  - Appearance model
    - Each part, $d_i = (x, y, \theta)$
    - Measure the likelihood of each part i using:
      - $p(I|l_i, d_i) = \alpha_i \phi(I, l_i)$
      - $p(I|L, D) = \prod_{i \in V} p(I|l_i, d_i)$
        - » α_i: learnt template for part i,
        - » $\phi(I, l_i)$: features of the image patch with configuration l_i



Pictorial Structure Model (2)



- Combination of two models:
  - Configuration model
    - Learn spatial relationship between two adjacent parts $i\ and\ j$
    - Measure the probability of a certain configuration :
      - $p(l_i|l_j) = \beta_{ij}\varphi(l_i, l_j)$
      - $p(L) = \prod_{ij \in E} p(l_i|l_j)$
        - » β_{ij}: learnt spatial relationship between i, j,
        - » $\varphi(l_i, l_j)$: spatial features between parts i, j

- Posterior:

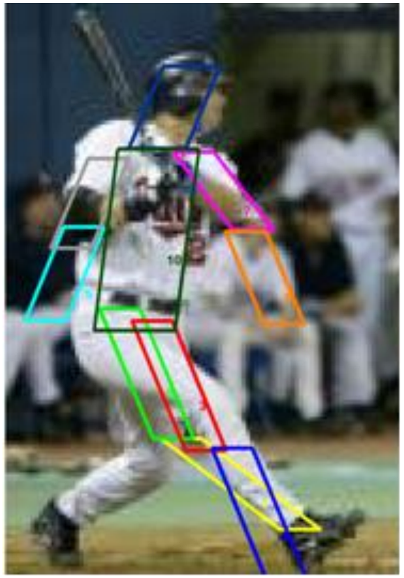
$$p(L|I, D) \propto p(I|L, D)p(L)$$

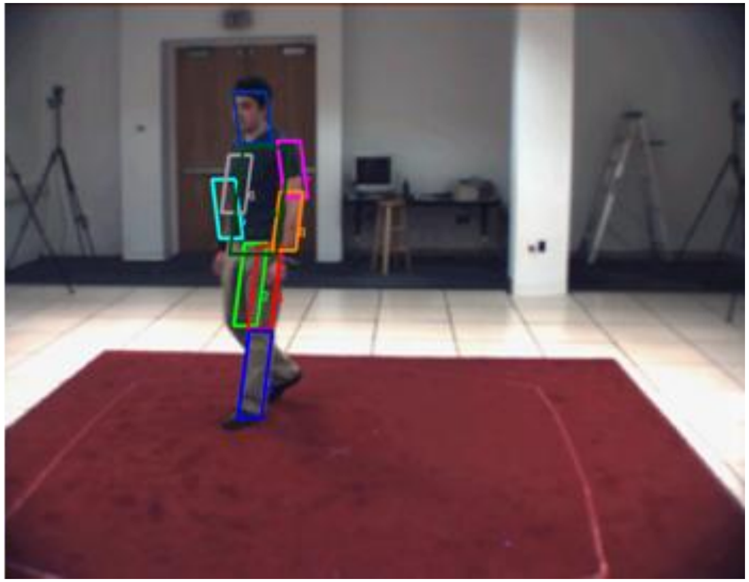
Pictorial Structure Model (3)

- One of the main Problems in PS is:
  - Self Occlusion

    Due to large degree of freedom

PS models are not able to correctly estimate the human pose in highly self-occluded scenes



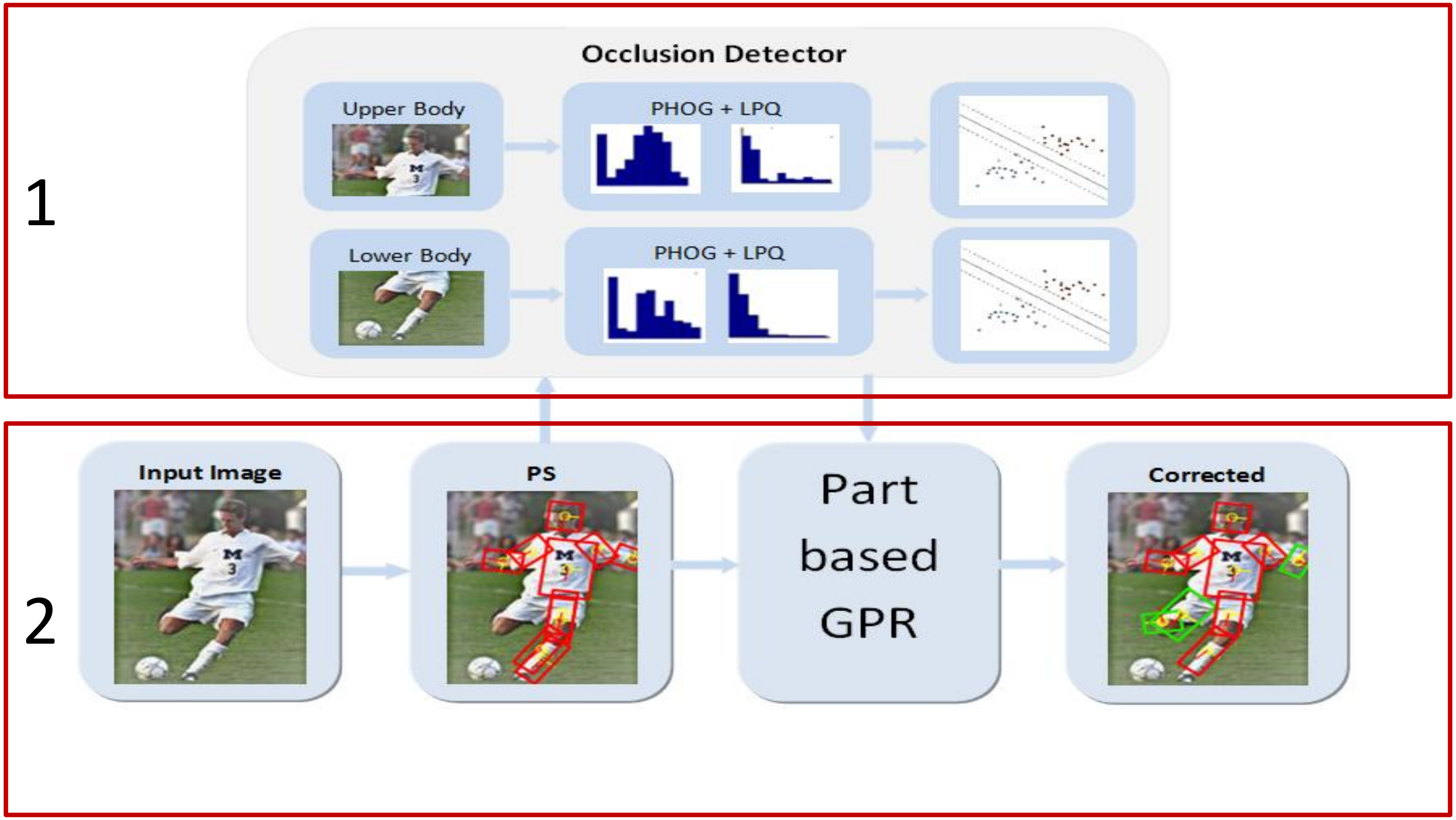


Motivations

1) How can we detect occlusion in a given image?
2) If there is occlusion, how can we identify the body parts responsible for it?
3) How can we rectify the occluded part's position?

Our Method

1- Occlusion detection 2- Part-by-part rectification

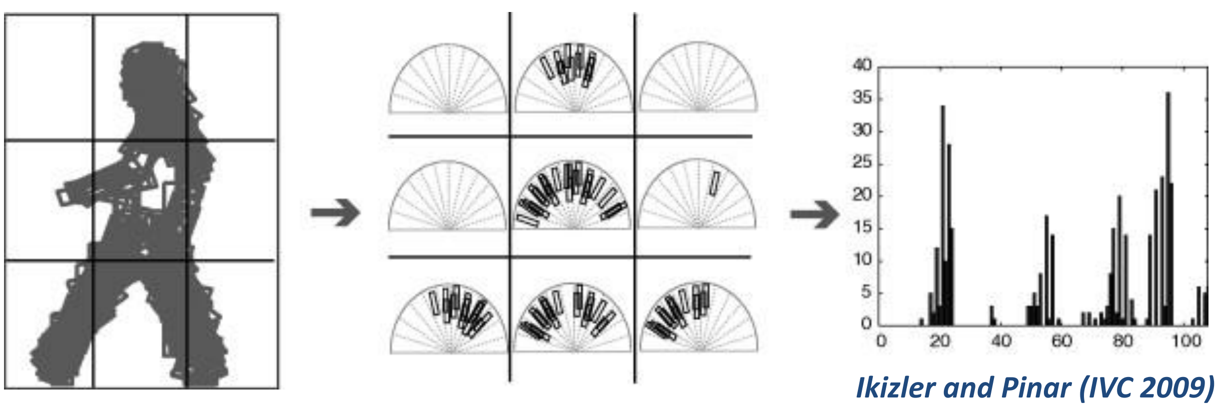


Occlusion Detection

A novel self-occlusion detection approach

- Does not rely on background subtraction for input images
- Two binary models corresponding to the upper and lower body, respectively
    1. Pyramid of Histogram of Oriented Gradients (PHOG)
    2. Local Phase Quantisation (LPQ)
- The output from the two descriptors is combined
- A non-linear Support Vector Machine (SVM) is learnt, Radial Basis Function

PHOG descriptor example



Ikizler and Pinar (IVC 2009)

Rectification

Train phase

Input:

$$D = [d_1, \dots, d_m], PS\ Features$$
$$D' = [d'_1, \dots, d'_m], GT\ Features$$

For each part *i* training set is

$$\tau_i = \{(d_j, d'_{i,j})\}_{j=1}^{m},$$

Learning

Learn a non-linear mapping function

$$w_i\colon D^{nxm} \rightarrow D'^{\,ixm}, i = 1, \dots, q$$

We use GPR to estimate the w_i $mapping$

Results

$$q\ models\ for\ q\ parts$$

Prediction phase

Input:

$D = [d_1, \dots, d_m], PS\ Features$

$P \leftarrow parts\ in\ ROI\ contain\ occlusion$

Rectify

$i = 0$

$p(\hat{L_i}) = p(L_i), initially$

$for\ i = 1\ to\ size\{P\}$

1. $i \leftarrow i + 1$
2. Select part with minimum appearance score
3. Load GPR model for that part
4. Predict d' for that part
5. Estimate new $p(\hat{L_i})$
6. Stop if no improvement in $p(\hat{L_i})$

$Endfor$

Experiments

Datasets:

1. BUFFY dataset
2. People dataset
3. HumanEva dataset

Evaluating on:

- Full body pose {10 parts: torso, head, left and right upper/lower arms, and left and right upper/lower legs}
- Upper body pose {6 parts : torso, head, left and right upper/lower arms}

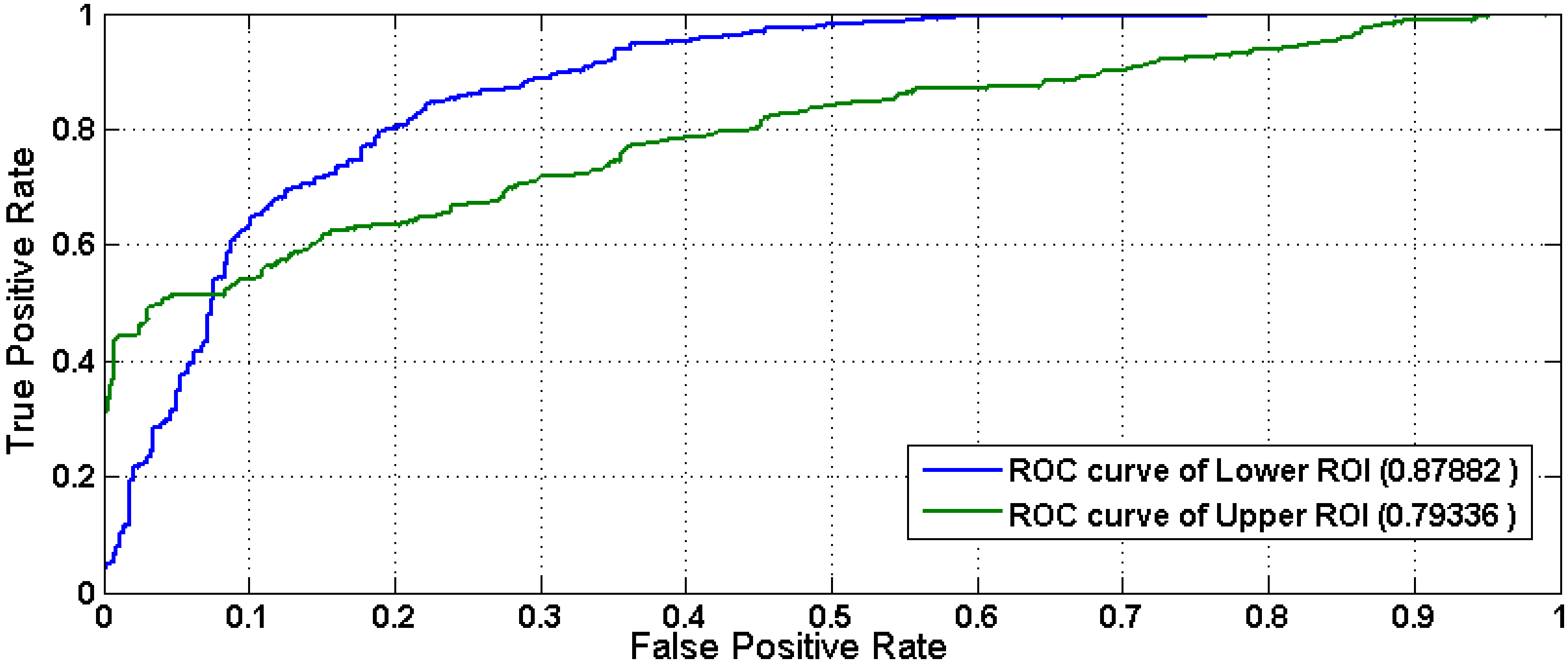
Experiments

Evaluation Criteria:

- **Detection rate**
    - Indicates the number of detected stick figures (PASCAL VOC criterion).
- **Percentage of Correctly estimated Body Parts (PCP)**
    - Measures the intersection ratio between the detected body part window with the ground truth one
- **Accuracy**
    - Detection rate x PCP

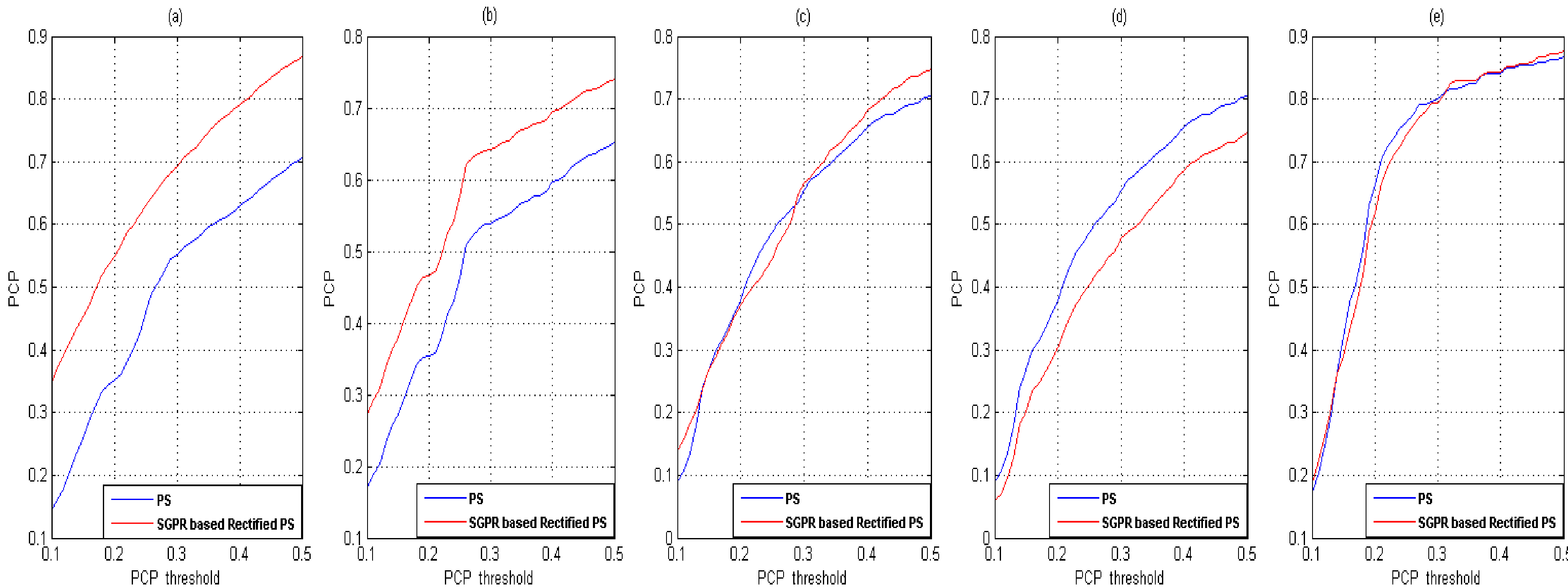
Using these evaluation criteria, we measure the performance of our part-by-part rectification method against the a state-of-the-art PS method

Experiments



Occlusion Detection Results

Experiments



PCP performance of our 2 frameworks against Andriluka et al. (PS)

(a) HumanEva dataset - full body, (b) HumanEva dataset - upper body,
(c) People dataset - full body, (d) People dataset – upper body,
(e) Buffy datatset - upper body

Experiments

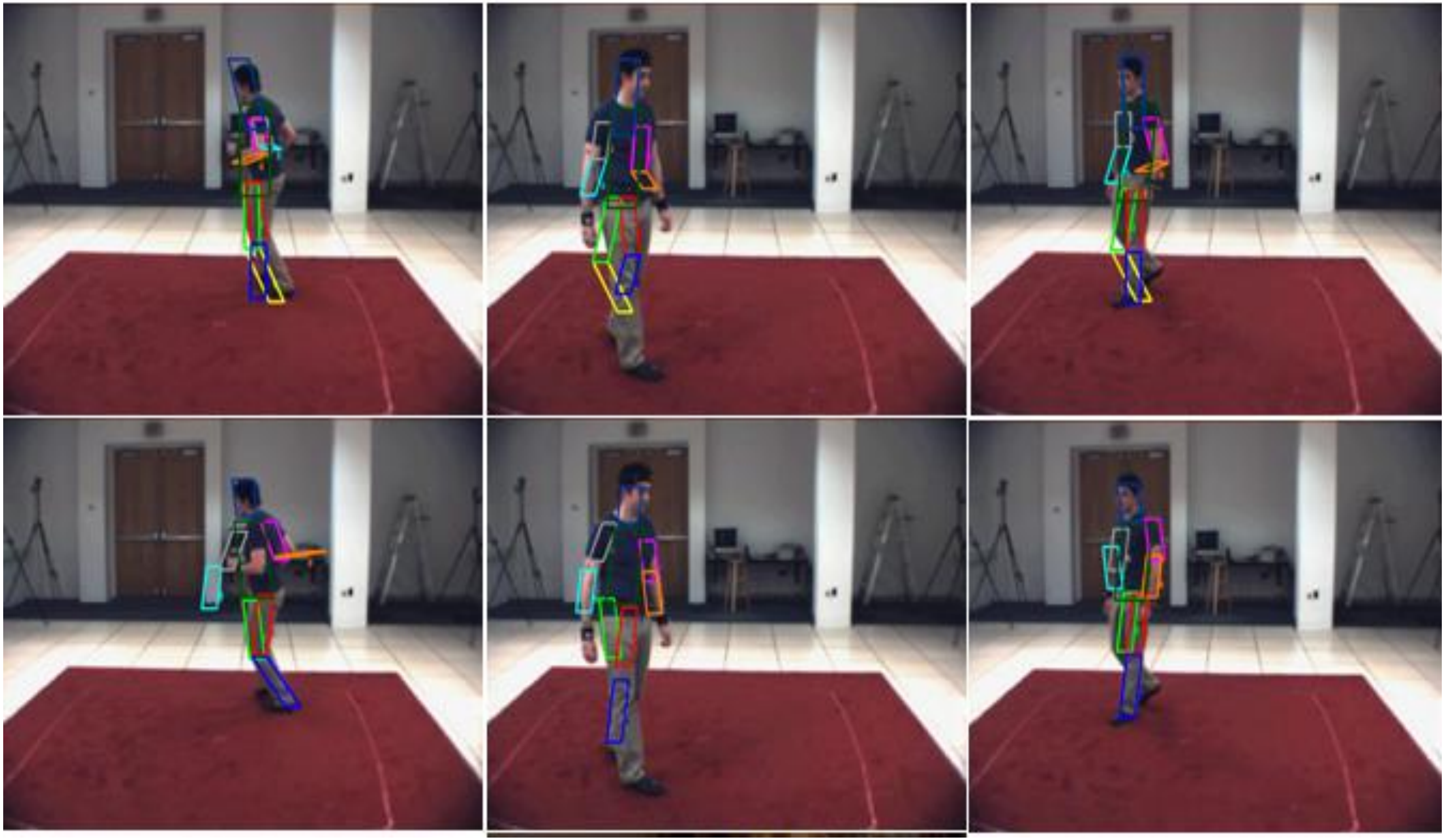
Database	Type	(PS) PCP	(Our) PCP	(PS) DetRate	(Our) DetRate	(PS) Accuracy	(Our) Accuracy
HumanEva	full body upper body	70.54% 65.19%	86.70% 74.10%	85.20% 91.50%	92.09% 94.00%	60.10% 59.64%	79.48% 69.65%
People	full body upper body	70.56% 70.56%	74.62% 64.55%	84.71% 84.71%	94.12% 90.59%	59.77% 59.72%	70.23% 58.47%
Buffy	full upper body	– 86.67%	– 87.72%	– 85.11%	– 88.64%	– 73.76%	– 77.75%

Comparison of PCP, detection rate and total accuracy between the proposed approach and Andriluka et al. (PS)

Experiments

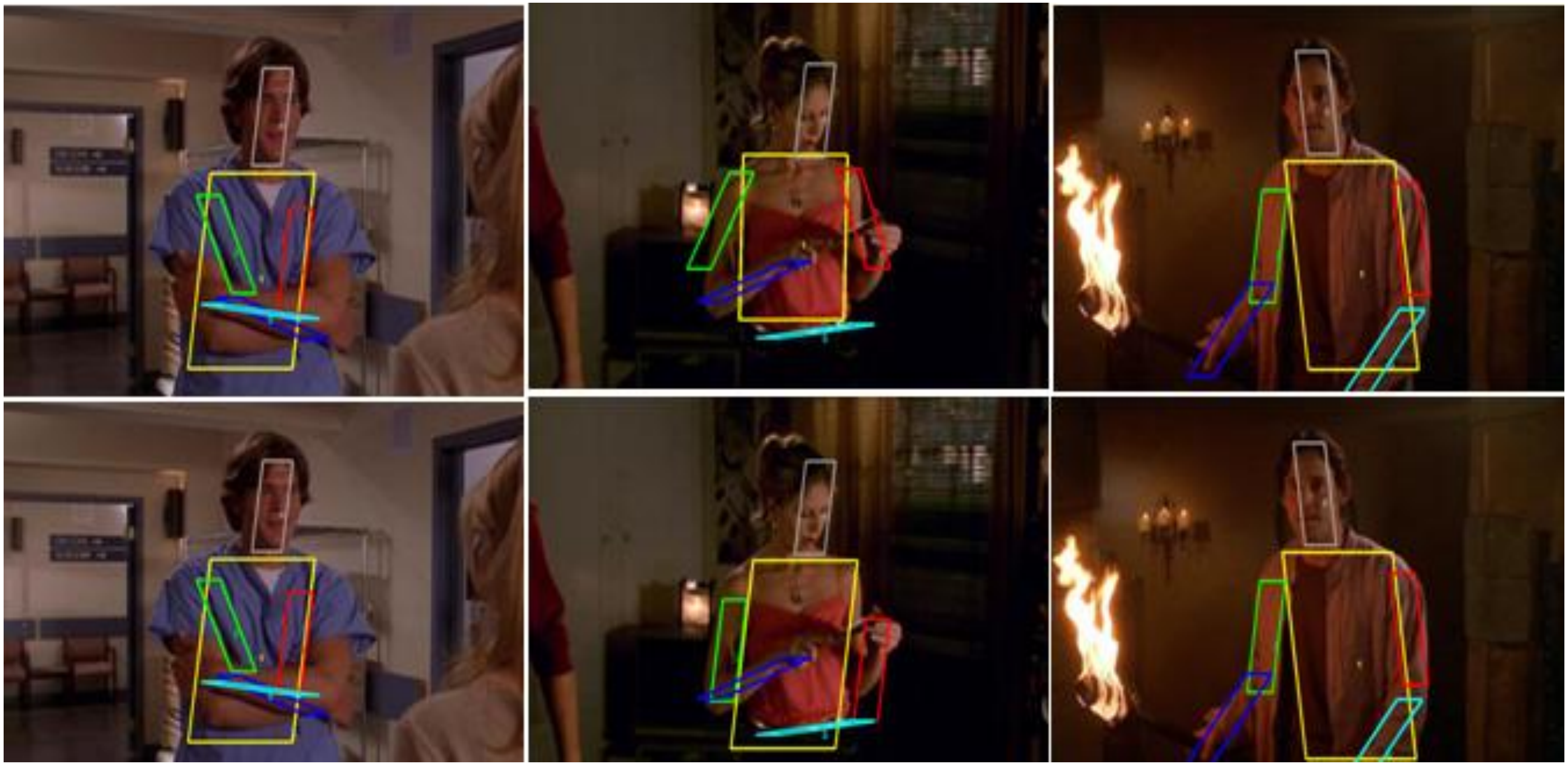
People dataset

Experiments



HumanEva dataset

Experiments



Buffy dataset

Conclusion

- The aim of that work is to produce a robust framework for correcting Pictorial Structure based pose estimation.
- Two steps
  - Occlusion detection
  - Part-by-part rectification
- Our method overcome one of the state-of-the-art articulated pose estimation method.

Future work

1. Different regression methods
2. Extending to Inter-Occlusion
3. Adding temporal information

References

[1] M. Andriluka, S. Roth, and B. Schiele. “Pictorial structures revisited: People detection and articulated pose estimation”. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2009.

[2] Y. Yang, D. Ramanan. "Articulated Pose Estimation using Flexible Mixtures of Parts". IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2011.

[3] L. Sigal and M. J. Black “Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation”. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2006.

[4] P. F. Felzenszwalb and D. P. Huttenlocher. “Pictorial structures for object recognition”. Int. J. Computer Vision, 61(1):55–79, 2005.

[5] N. Dalal and B. Triggs. “Histograms of oriented gradients for human detection”. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2005.

[6] P. F. Felzenszwalb, D. McAllester, and D. Ramanan. “A discriminatively trained, multiscale, deformable part model”. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2008.

Thank you

http://ise.canberra.edu.au/ibrahim/

Ibrahim.radwan@canberra.edu.au