

Holistic Human Pose Estimation with Regression Forests

Vasileios Belagiannis¹, Christian Amann¹, Nassir Navab¹, Slobodan Ilic^{1,2}

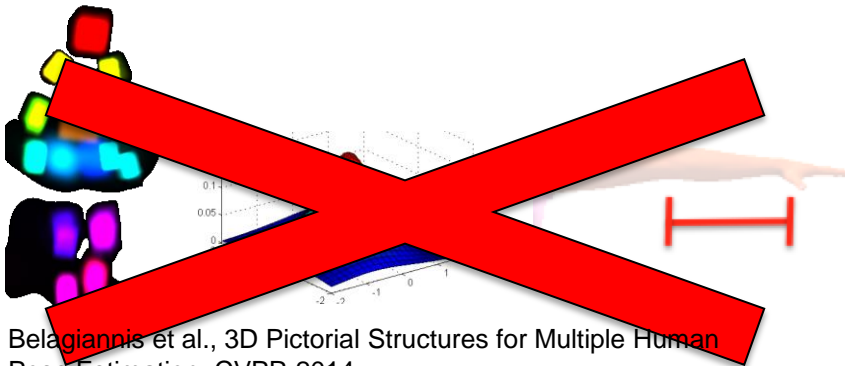
¹Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

²Siemens AG, CT RTC SET INT-DE, Germany



Motivation

- One-Shot 2D human pose estimation
- Less hand-crafted features

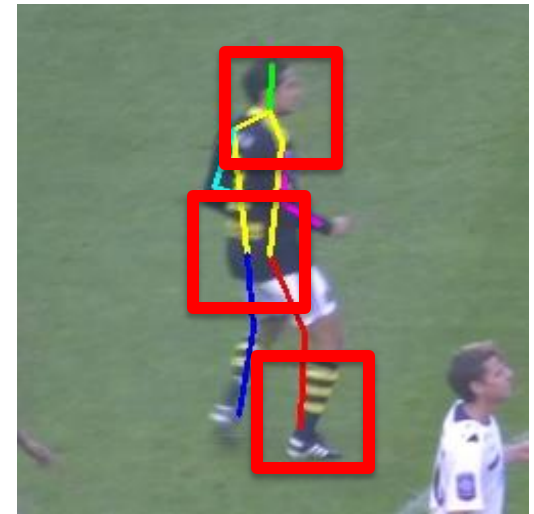


Belagiannis et al., 3D Pictorial Structures for Multiple Human Pose Estimation, CVPR 2014.



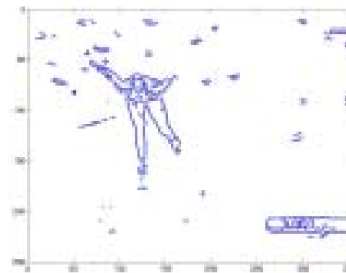
Main Idea

- Associate the body pose with image features
- Regress the human body joint offsets
- Problems
 - Huge appearance variation
 - Ambiguity between appearance & geometric pose
 - Computational cost of mode seeking

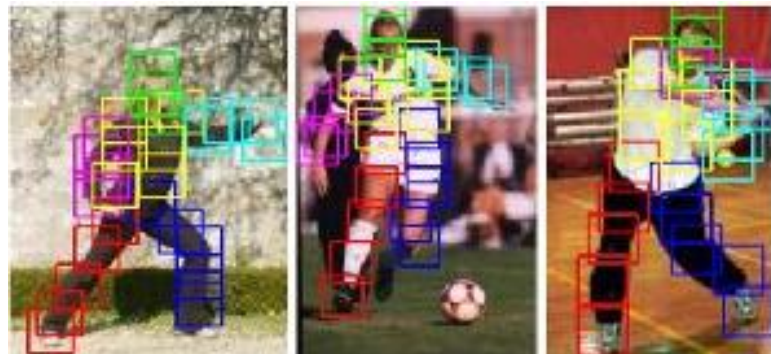


Related Work

- Holistic approaches
 - + Skeleton inference in one step
 - Require complete data
- Part-Based approaches
 - + Rich appearance features
 - Rely on complex models



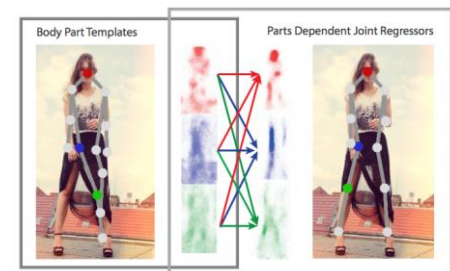
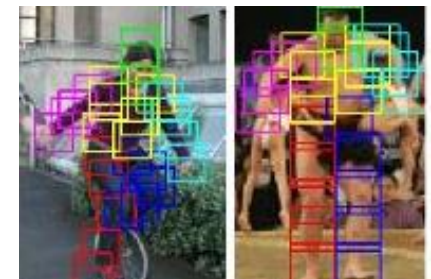
G. Mori and J. Malik. Estimating human body configurations using shape context matching. In ECCV, 2002.



Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In CVPR, 2011.

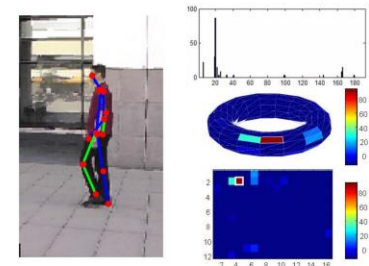
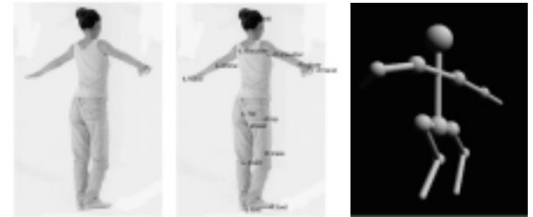
Related Work (Part-based)

- Andriluka, M., Roth, S., Schiele, B., Pictorial structures revisited: People detection and articulated pose estimation, In CVPR 2009.
- Yang, Y., Ramanan, D., Articulated pose estimation with flexible mixtures-of- parts, In CVPR 2011.
- Dantone, M., Gall, J., Leistner, C., Van Gool, L., Human pose estimation using body parts dependent joint regressors, In CVPR 2013.



Related Work (Holistic)

- Mori, G., Malik, J., Estimating human body configurations using shape context matching, In ECCV 2002.
- Rogez, G., Rihan, J., Ramalingam, S., Orrite, C., Torr, P.H., Randomized trees for human pose detection. In CVPR 2008.
- Girshick, R., Shotton, J., Kohli, P., Criminisi, A., Fitzgibbon, A., Efficient regression of general-activity human poses from depth images, In ICCV 2011.



Method (Regression forest)

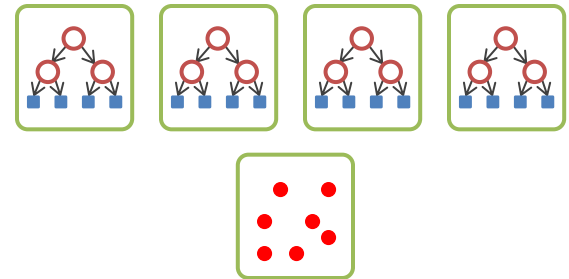
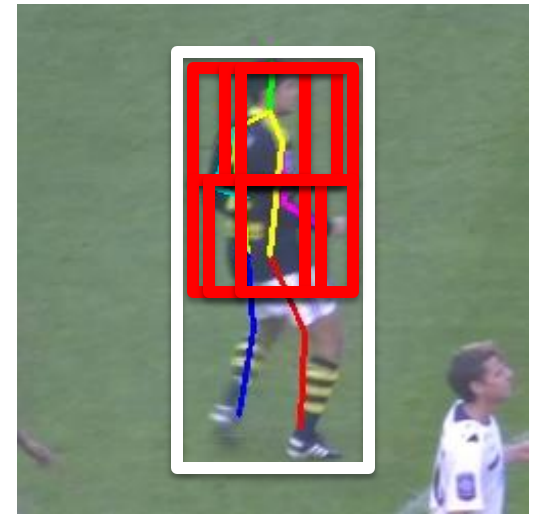
- Ensemble of trees
- Continuous output
- **Contribution:** Mapping between image patches (**HOG features**) & the parameter space (**N joints in the 2D space**)



Phil Cutler – Source:
<http://www.stat.berkeley.edu/~breiman/RandomForests/>

Method (Prediction)

- Bounding-box localization
 - Rescaled
- Random and dense sampling
 - HOG feature extraction
- Vote aggregation
- Mode estimation
 - **Contribution:** *dense-window* algorithm



Motivation

Related Work

Method

Training

Evaluation

Conclusion

Method (Forest elements)

- **Input:** pool of randomly extracted image patches P with associated skeleton joint offsets
- **Goal:** node creation for each tree

Entropy (joint- and mean-offsets)

$$H(P) = \sum_{p \in P} \sum_j \|\mathbf{v}_{p,j} - \boldsymbol{\mu}_j\|_2^2$$

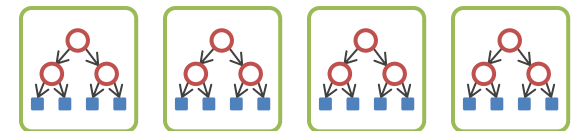
Split function (pool of random tests)

$$\theta^* = \arg \max_{\theta} g(\theta)$$

Info-gain (best test)

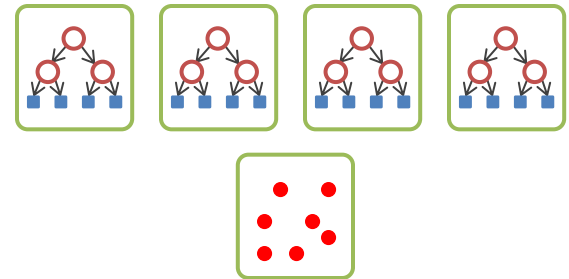
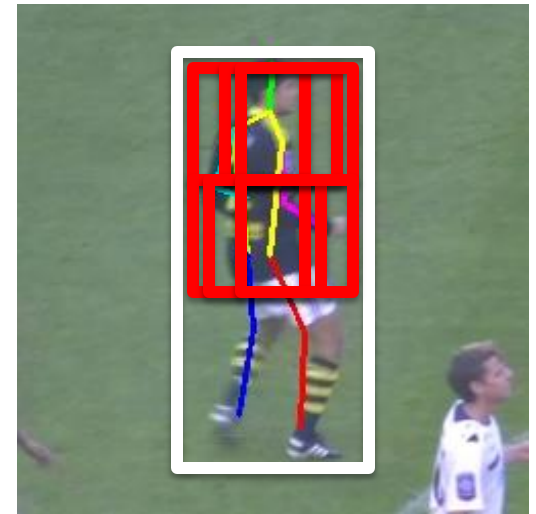
$$g(\theta) = H(P) - \sum_{i \in \{l,r\}} \frac{|P_i(\theta)|}{|P|} H(P_i(\theta))$$

Forest formation



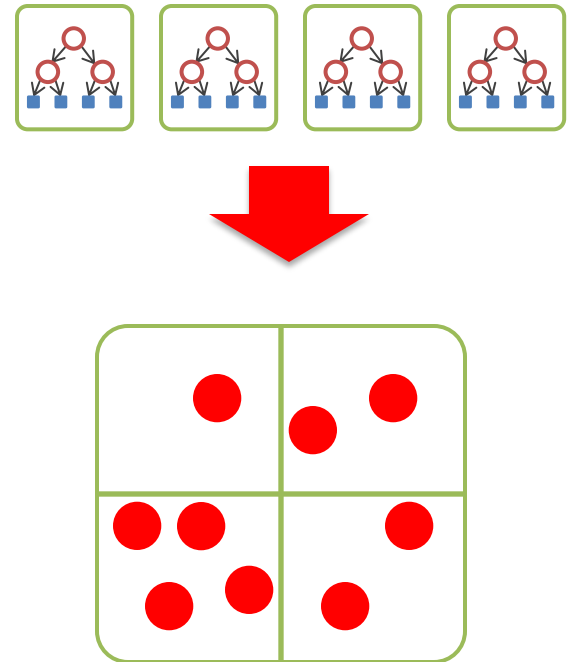
Method (Prediction)

- Bounding-box localization
 - Rescaled
- Random and dense sampling
 - HOG feature extraction
- Vote aggregation
- Mode estimation
 - **Contribution:** *dense-window* algorithm



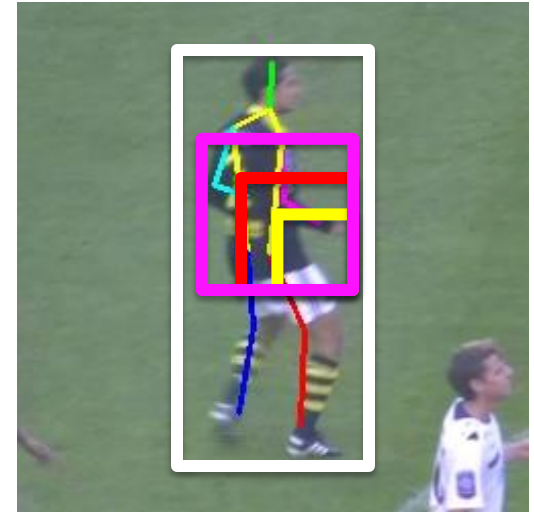
Method (Prediction – Mode Estimation)

- Forest leaves: joint-offsets
- *Dense-window* algorithm
 - Mode estimation of a density function
 - Integral matrices
 - Deterministic convergence
 - Dependence: a sliding window
 - Scalability: number of predictions



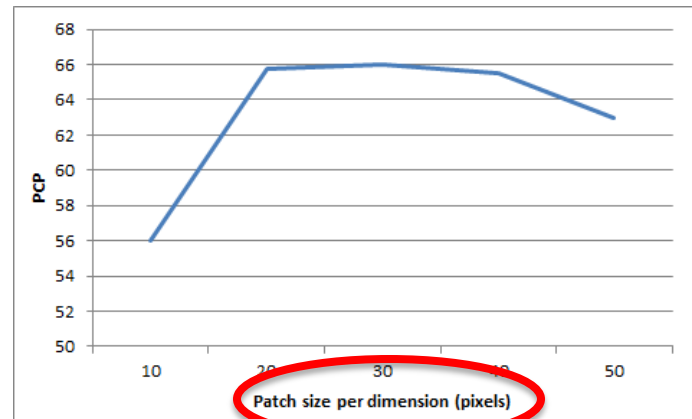
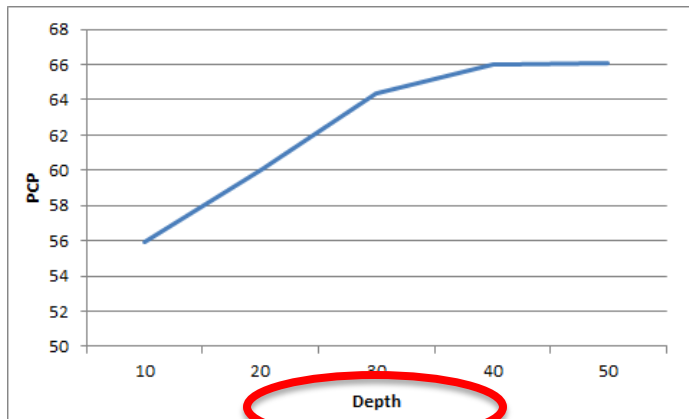
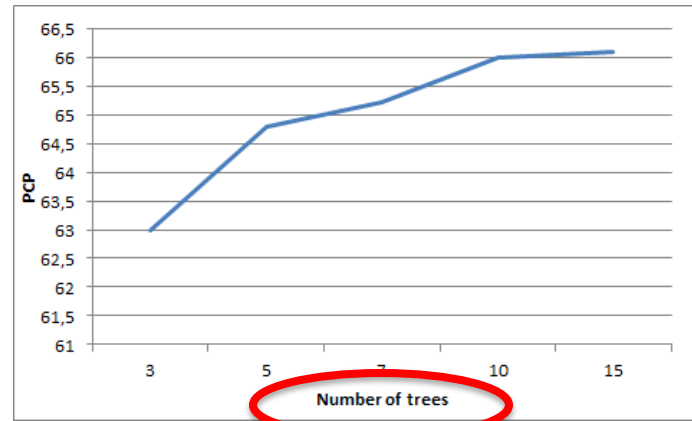
Method (Parameters)

- Scale Invariance
 - Bounding-box normalization
- Image Patches
 - Fixed Size
- Threshold ρ
 - Local joint votes



Forest Parameters

- Number of trees
- Depth of a tree
- Patch size



Motivation

Human Model

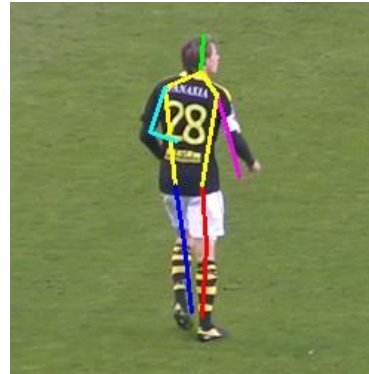
Training

Evaluation

Conclusion

Evaluation (Datasets)

- Football
- Image Parse
- Volleyball



Evaluation: Football Dataset



PCP Scores	Head	Torso	Upp. Arm	Low. Arm	Upp. Leg	Low. Leg	Avg.
Our method	0.86	0.98	0.88	0.57	0.92	0.80	0.84
Yang & Ramanan [3]	0.84	0.98	0.86	0.55	0.89	0.73	0.80
Kazemi et al. [19]	0.94	0.96	0.90	0.69	0.94	0.84	0.87
Kazemi et al. [19] + Prior	0.96	0.98	0.93	0.71	0.97	0.88	0.89

Motivation

Human Model

Training

Evaluation

Conclusion

Evaluation: Image Parse Dataset



PCP Scores	Torso	Upp. Leg	Low. Leg	Upp. Arm	Low. Arm	Head	Avg.
Our method	88.8	80.9	72.8	58.2	27.5	74.1	67.1
Andriluka et al.[4]	86.3	66.3	60.0	54.6	35.6	72.7	59.2
Yang & Ramanan [3]	82.9	69.0	63.9	55.1	35.4	77.6	60.7
Pischulin et al. [2]	92.2	74.6	63.7	54.9	39.8	70.7	62.9
Pischulin et al. [33] + [2]	90.7	80.0	70.0	59.3	37.1	77.6	66.1
Johnson & Ever.[8]	87.6	74.7	67.1	67.3	45.8	76.8	67.4

Motivation

Human Model

Training

Evaluation

Conclusion

Evaluation: Volleyball Dataset

- Proposed dataset
- Train on a game
- Test on a different



PCP Scores	Head	Torso	Upp. Arm	Low. Arm	Upp. Leg	Low. Leg	Avg.
Our method	97.5	81.4	54.4	19.3	65.1	81.2	63.8
Yang & Ramanan [3]	76.1	80.5	40.7	33.7	52.4	70.5	59.0

Motivation

Human Model

Training

Evaluation

Conclusion

Conclusion

- One-shot 2D human pose estimation
- Appearance mapping to body poses using image patches
- Efficient prediction with the *dense-window* algorithm
- State-of-the-art results only with HOG features – Random Forest

Future Work

- Learn jointly the appearance features and classifier parameters
- Learn the body structure

Thank you!

- Volleyball dataset available at: <http://campar.in.tum.de/Chair/SingleHumanPose>