Articulated Part-based Model for Joint Object Detection and Pose Estimation

Min Sun Silvio Savarese

Dept. of Electrical and Computer Engineering, University of Michigan at Ann Arbor, USA

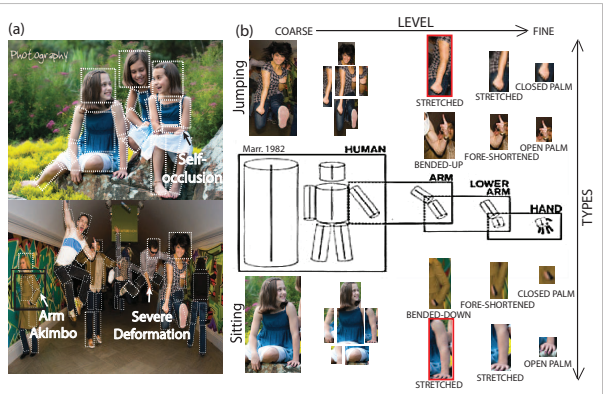
VISION LAB

1. Overview

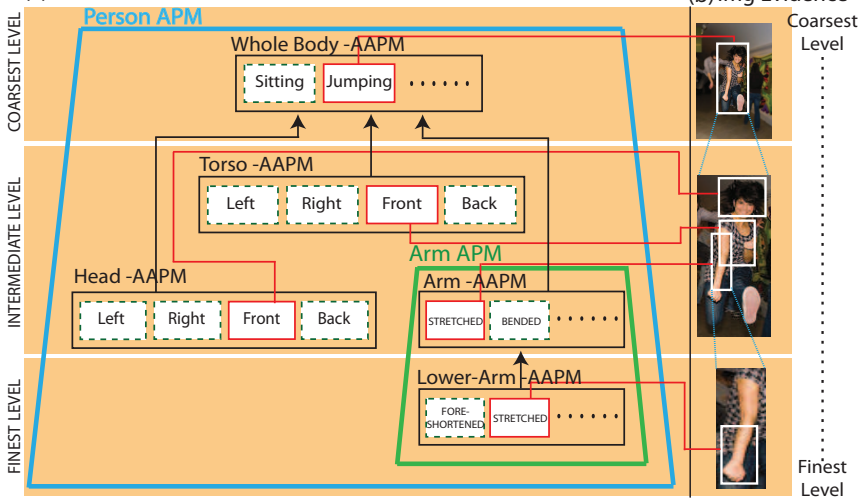
-Goal: Joint Object Detection and Pose Estimation

-Articulated Part-based Model:

- recursive coarse-to-fine representation
- multiple part-types
- parents-child relationship



2. Model



-Matching Score

- Appearance Score: $f^A(h;I) = A^T \psi_a(h,I)$
- Deformation Score: $f^D(h,\hat{h}) = -d^T \psi_d(h,\hat{h})$

-Model Properties

- Sublinearity
- Efficient Exact Inference

-Score Aggregation

- Child location selection: $f_{c,s_c}(\hat{h}_c, I) = \max_{h_c} f_{c,s_c}(h_c, I) + f^D(h_c, \hat{h}_c)$
- Childd alignment: $f_{c,s_c}(T(h_i, t_{i,c}^{s_i,s_c}), I)$; $T(h,t) = (x - t_x, y - t_y, L - t_L, \theta - t_\theta)$
- Child type selection: $f_c(h_i, I) = \max_{s_c} f_{c,s_c}(T(h_i, t_{i,c}^{s_i,s_c}), I) + b_{i,c}^{s_i,s_c}$
- Aggregation: $f_{i,s_i}(h_i, I) = f_{i,s_i}^A(h_i, I) + \sum_c f_c(h_i; I)$

[1] B. Sapp, A. Toshev, and B. Taskar. Cascaded models for articulated pose estimation. ECCV, 2010.
[2] M. Eichner and V. Ferrari. Better appearance models for pictorial structures. BMVC, 2009.
[3] L. Bourdev, S. Maji, T. Brox, and J. Malik. Detecting people using mutually consistent poselet activations. In ECCV, 2010.
[4] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.
[5] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The PASCAL VOC2007 Results.

Correspondence
sunmin@eecs.umich.edu

3. Learning

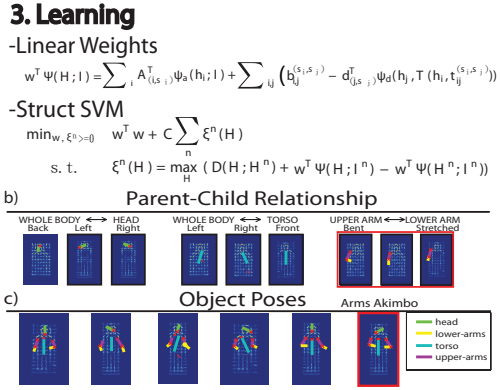
-Linear Weights

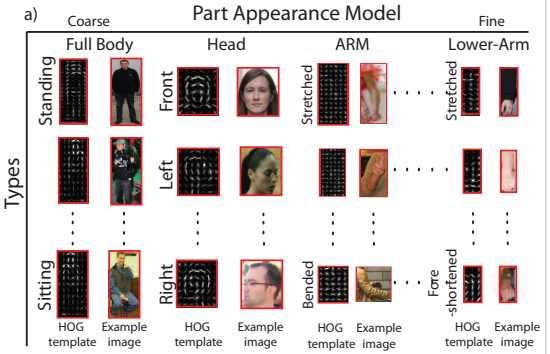
$$w^T \Psi(H;I) = \sum_i A_{(i,s_i)}^T \psi_a(h_i;I) + \sum_{ij} \left(b_{ij}^{(s_i,s_j)} - d_{(i,s_i)(j,s_j)}^T \psi_d(h_j, T(h_i, t_{ij}^{(s_i,s_j)})) \right)$$

-Struct SVM

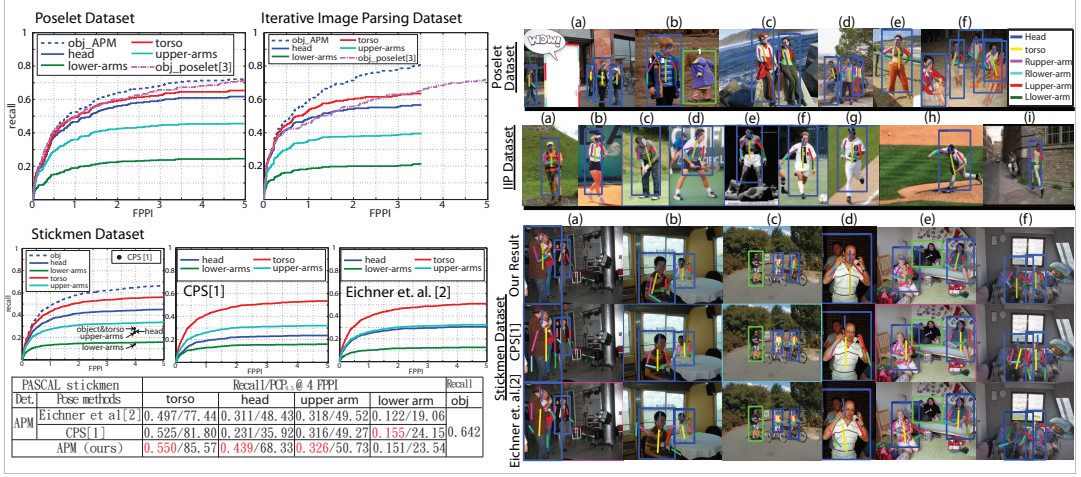
$$\min_{w,\xi^n \geq 0} \; w^T w + C \sum_n \xi^n(H)$$

$$\text{s. t.} \quad \xi^n(H) = \max_H \left(D(H;H^n) + w^T \Psi(H;I^n) - w^T \Psi(H^n;I^n) \right)$$



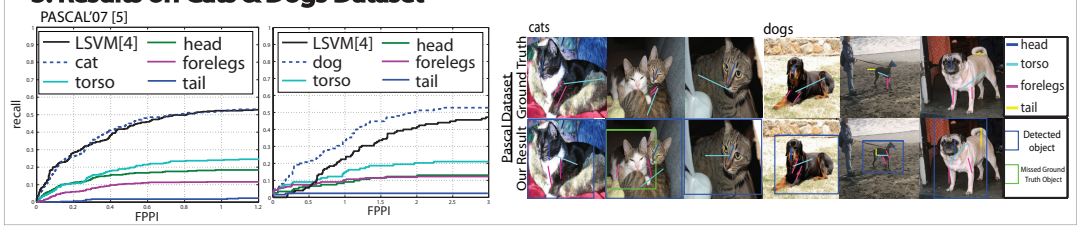


4. Results on Human Dataset



PASCAL stickmen		Recall/PCP$_{1/2}$ @ 4 FPPI				
Det.	Pose methods	torso	head	upper arm	lower arm	obj
APM	Eichner et al[2]	0.497/77.44	0.311/48.43	0.318/49.52	0.122/19.06	
	CPS[1]	0.525/81.80	0.231/35.92	0.316/49.27	0.155/24.15	0.642
	APM (ours)	0.550/85.57	0.439/68.33	0.326/50.73	0.151/23.54	

5. Results on Cats & Dogs Dataset



6. Conclusion

-Improvement in both object detection and pose estimation: recursive coarse-to-fine and multiple part-type representation

- Novel performance measure: the part recall vs. FPPI curve

Acknowledgments

ONR grant
N000141110389