Antonio Hernández-Vela
Advisors: Sergio Escalera, Stan Sclaroff

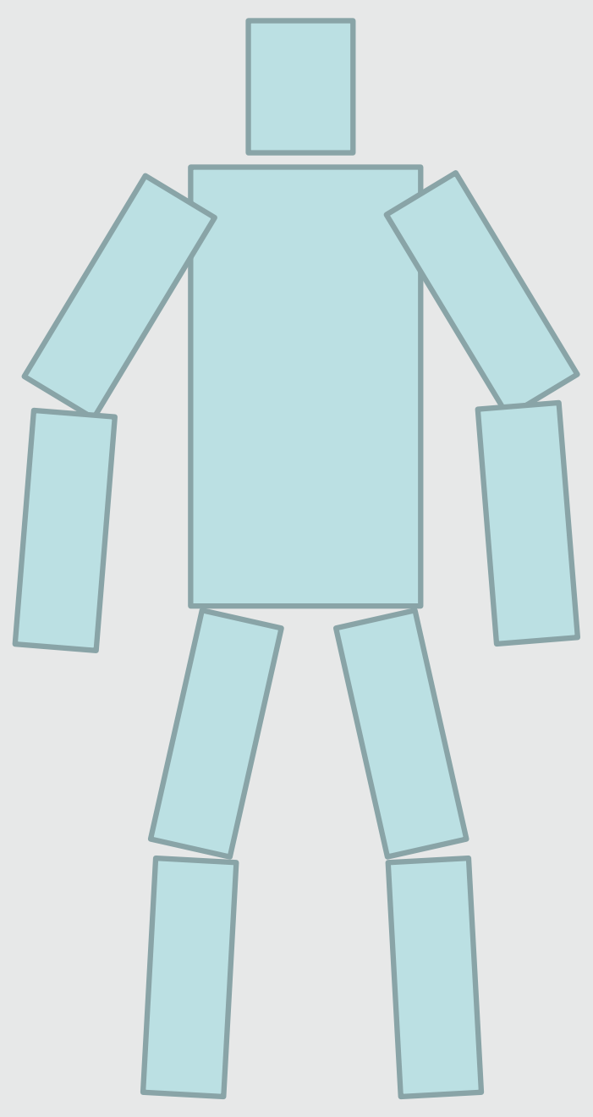
Universitat de Barcelona
BOSTON UNIVERSITY
Centre de Visió per Computador
Universitat Autònoma de Barcelona

Contextual Part Rescoring for Human Pose Estimation

Abstract

A contextual rescoring method is proposed for improving the detection of body joints of a pictorial structure model for human pose estimation. A set of mid-level parts is incorporated in the model, and their detections are used to extract spatial and score-related features relative to other body joint hypotheses. A technique is proposed for the automatic discovery of a compact subset of poselets that covers a set of validation images while maximizing precision. A rescoring mechanism is defined as a set-based boosting classifier that computes a new score for body joint detections, given its relationship to detections of other body joints and mid-level parts in the image. This new score complements the unary potential of a discriminatively trained pictorial structure model

Pictorial Structures



Energy Function

$$E(L;D,\beta) = \sum_{m=1}^{M} E^u(l_m;D,\beta^u) + \sum_{n\sim m} E^p(l_n,l_m;\beta^p).$$

$$E^u(l_m;D,\beta^u) = \log \phi^u(l_m;D), \quad \forall m = 1,\ldots,M,$$

$$E^p(l_n,l_m;\beta^p) = \left\langle \beta^p_{n,m}, \phi^p_{n,m}(l_n,l_m) \right\rangle, \quad \forall n \sim m.$$

Poselet [1] selection

Weighted set cover in validation set

$$\text{minimize} \sum_{\hat{j}} (1-\text{Prec}(\hat{j}))\mathbf{x}_{\hat{j}}$$

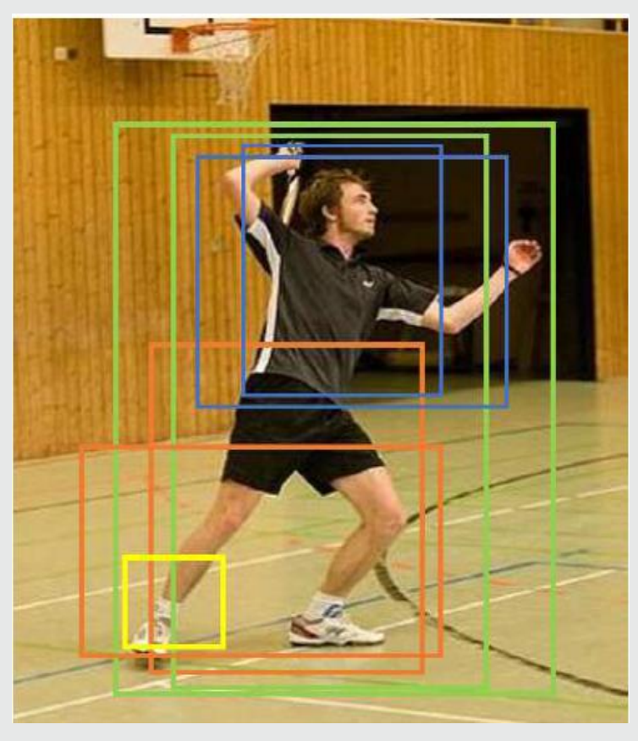
$$\text{subject to} \sum_{\hat{j}:A_{n\hat{j}}=1} \mathbf{x}_{\hat{j}} \geq 1\ \forall n,\ \mathbf{x}_{\hat{j}} \in \{0,1\},$$

poselet $\hat{j}$
n-th validation image

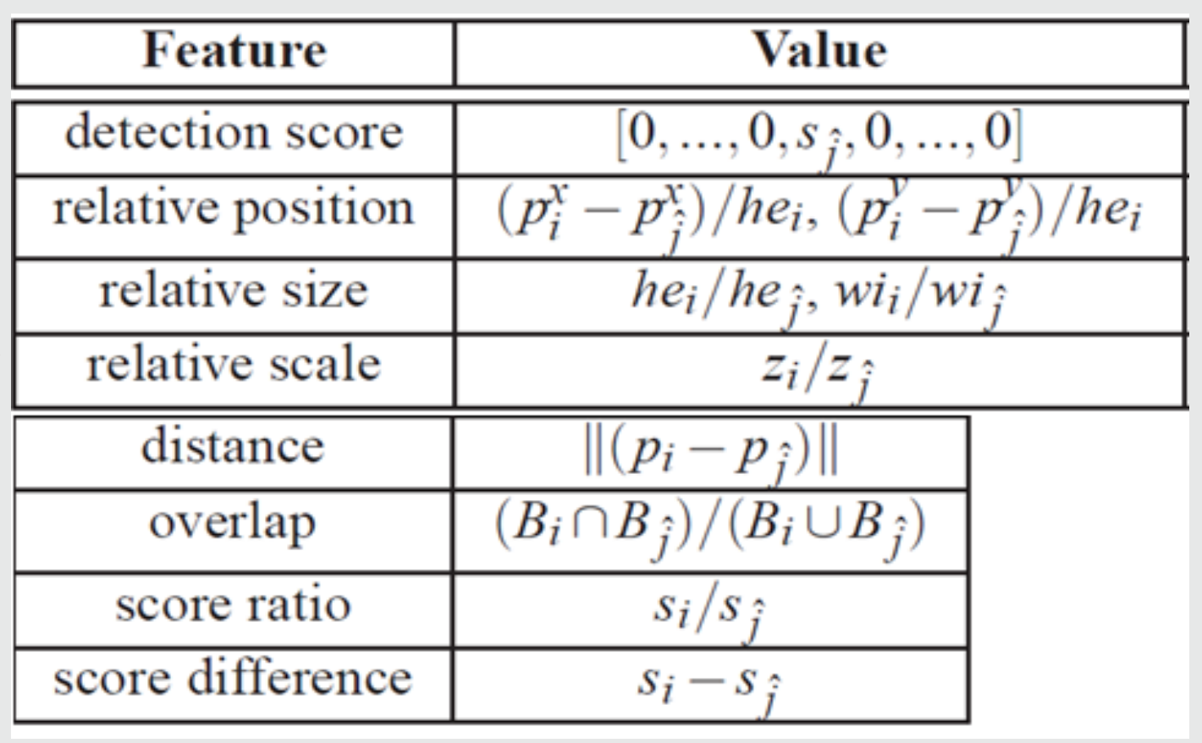
Automatic Poselet selection

Contextual Rescoring

Mid-level context



Contextual features



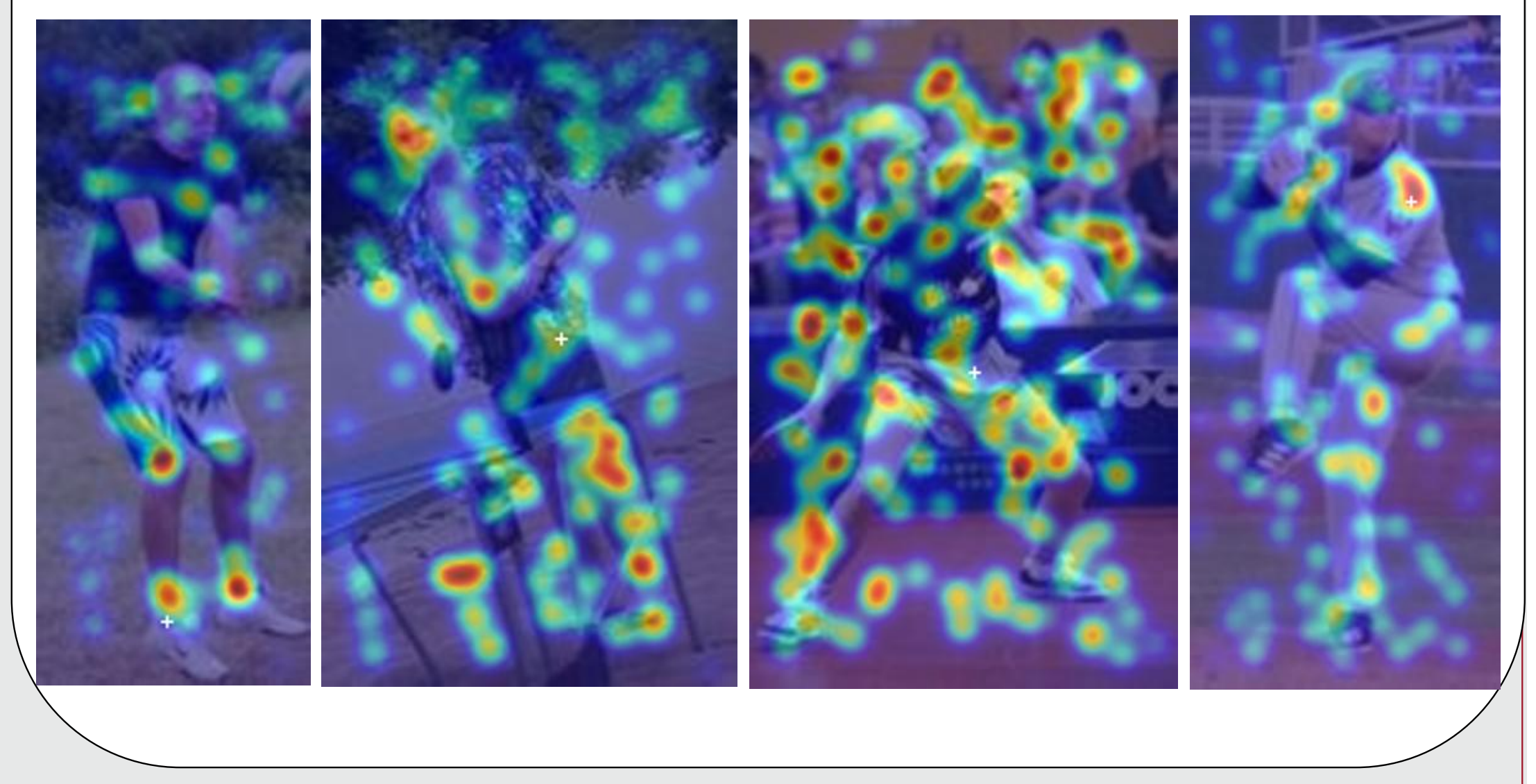
Feature	Value
detection score	$[0,\ldots,0,s_{\hat{j}},0,\ldots,0]$
relative position	$(p_i^x - p_{\hat{j}}^x)/he_i,\ (p_i^y - p_{\hat{j}}^y)/he_i$
relative size	$he_i/he_{\hat{j}},\ wi_i/wi_{\hat{j}}$
relative scale	$z_i/z_{\hat{j}}$
distance	$\|(p_i - p_{\hat{j}})\|$
overlap	$(B_i \cap B_{\hat{j}})/(B_i \cup B_{\hat{j}})$
score ratio	$s_i/s_{\hat{j}}$
score difference	$s_i - s_{\hat{j}}$

SetBoost [2] rescoring function

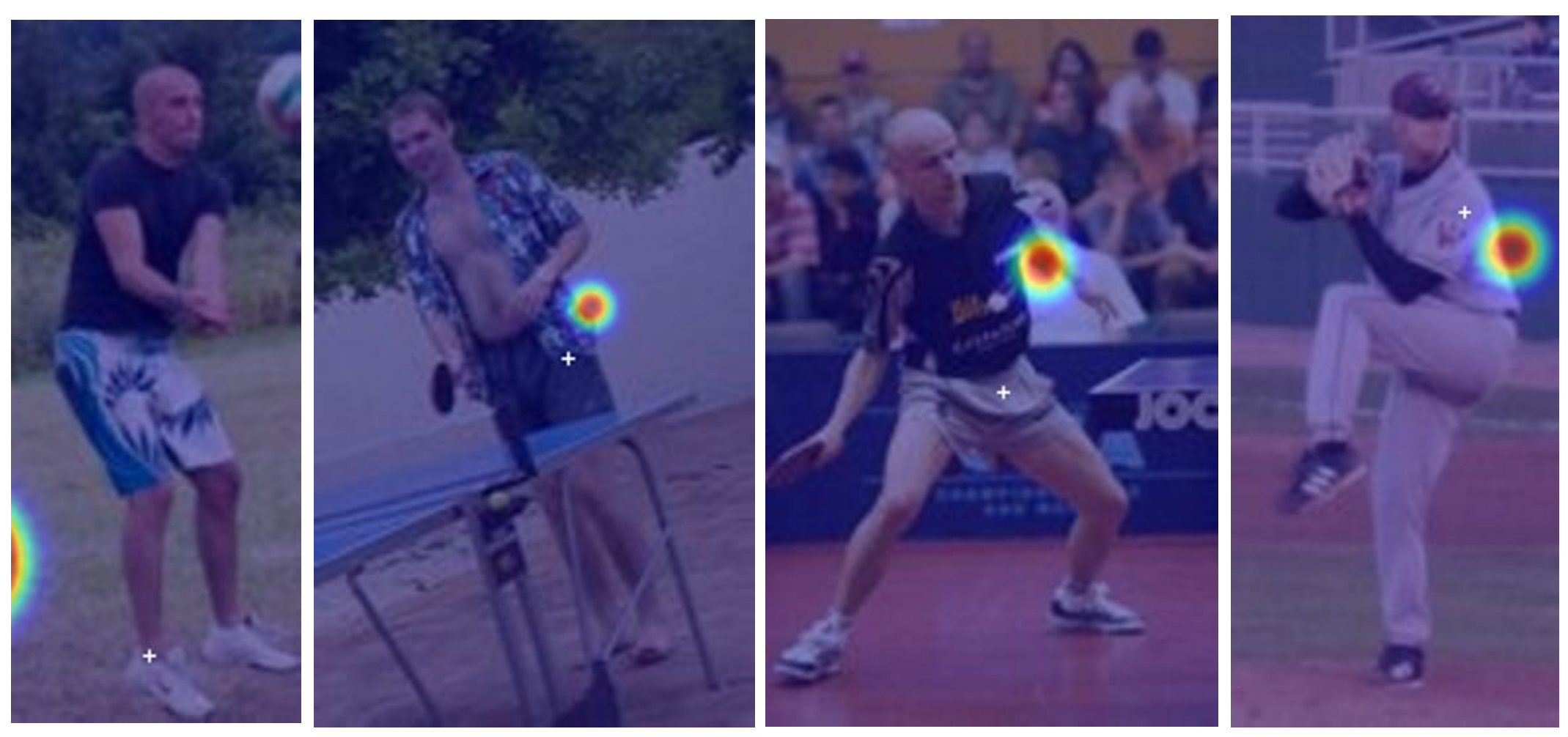
$$R(C) = \sum_{\theta=1}^{\Theta} Q_\theta(C)$$

$$Q_\theta(C) = \alpha_\theta \sum_{c\in C} k_c \cdot q_\theta(c)$$

Unary potential: local appearance model

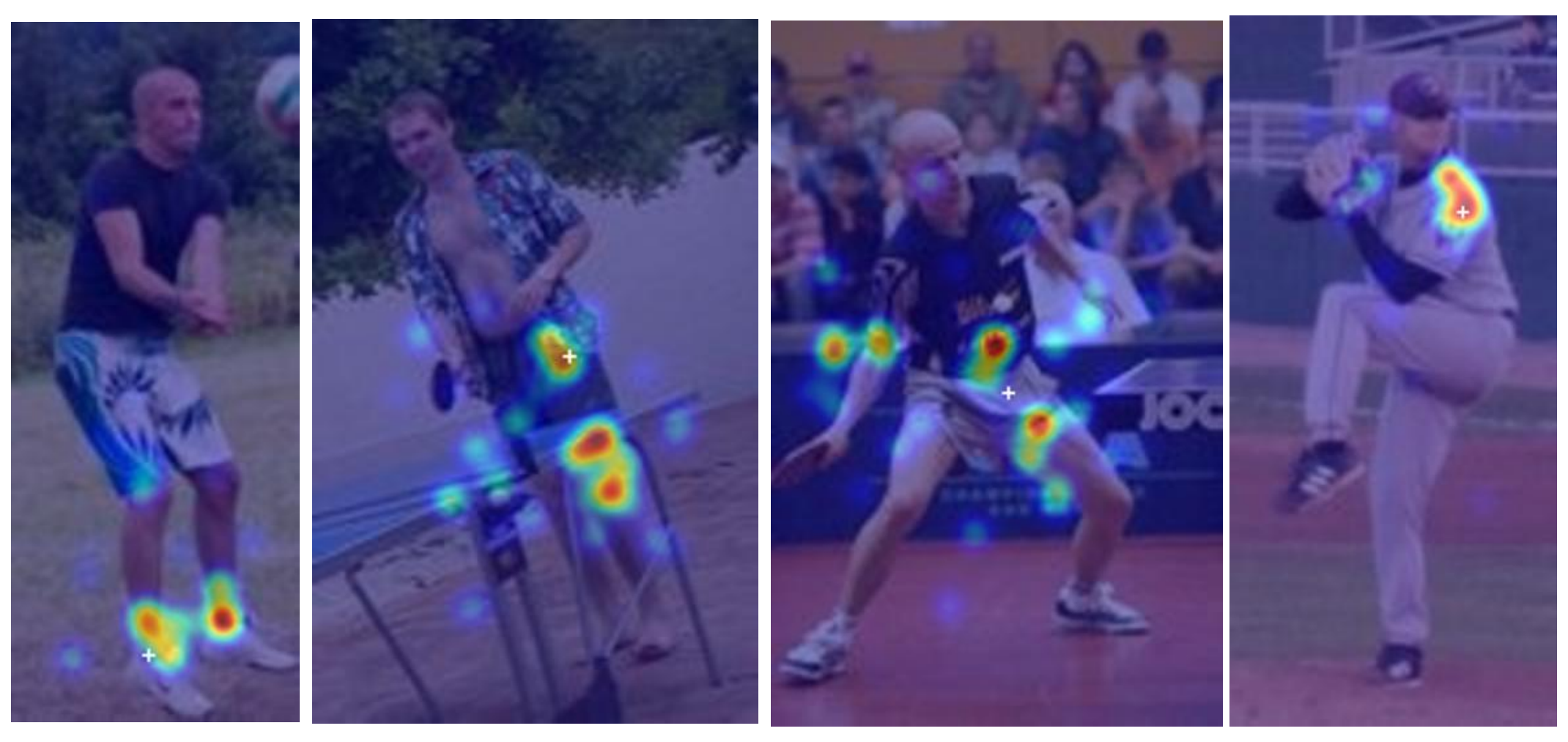


Gaussian-shaped extra unary potential [3]



~1,000 poselets

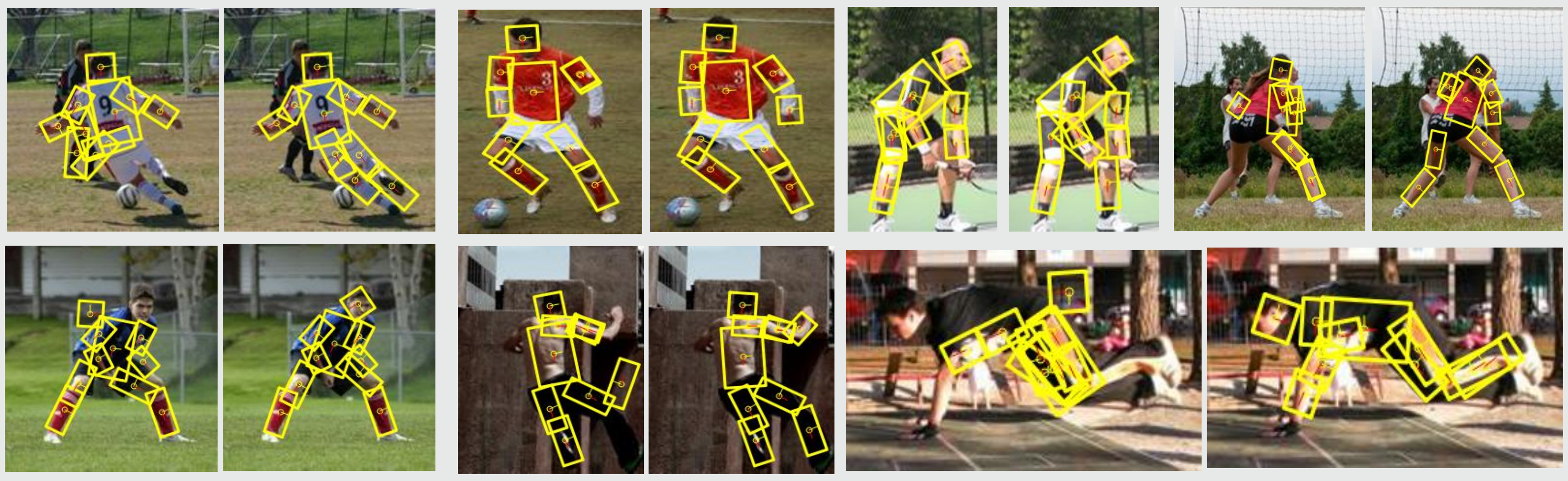
SetBoost rescoring



~50 poselets

Results: PCP on LSP Dataset [4]

Method	Torso	Upper Leg		Lower Leg		Upper Arm		Forearm		Head	Mean
[3] (+1000 poselets)	83.00	75.20	73.80	70.40	67.40	57.00	37.30	40.60	37.30	66.30	62.42
[3] (54 poselets)	83.30	71.50	72.30	66.80	66.40	52.90	36.50	37.80	36.50	67.70	60.58
[3] + setboost (54 poselets)	**85.60**	**75.60**	**74.90**	70.20	**69.20**	**57.60**	37.40	**41.10**	37.40	74.00	63.87
[3] + setboost (47 poselets covering)	85.50	75.40	74.80	**70.30**	68.90	**57.60**	**37.70**	41.00	**37.70**	**74.50**	**63.95**



Qualitative results. Left: [3] (+1,000 poselets), Right: [3] + setboost (47 poselets covering)

Biography

Antonio Hernández-Vela received the B.S. degree in computer science and the M.S. degree in computer vision and artificial intelligence, both from the Universitat Autònoma de Barcelona (UAB), in 2009 and 2010, respectively. He is currently working toward the P.h.D. degree in mathematics on human pose recovery and behavior analysis at the Universitat de Barcelona. His main research interests include human pose recovery, gesture recognition and behavior analysis.

[1] Lubomir Bourdev, Subhransu Maji, Thomas Brox, and Jitendra Malik. Detecting people using mutually consistent poselet activations. In ECCV 2010.
[2] Ramazan Gokberk Cinbis and Stan Sclaroff. Contextual object detection using set-based classification. In ECCV 2012.
[3] Leonid Pishchulin et al. Strong appearance and expressive spatial models for human pose estimation. In ICCV 2013.
[4] Wang, Y., Tran, D., Liao, Z.: Learning hierarchical poselets for human parsing. In: CVPR 2011.