HuPBA

Universitat de Barcelona

CVC Centre de Visió per Computador

ECCV '14
Zürich, September 6-12th, 2014

European Conference on Computer Vision

Learning To Segment Humans By Stacking Their Body Parts

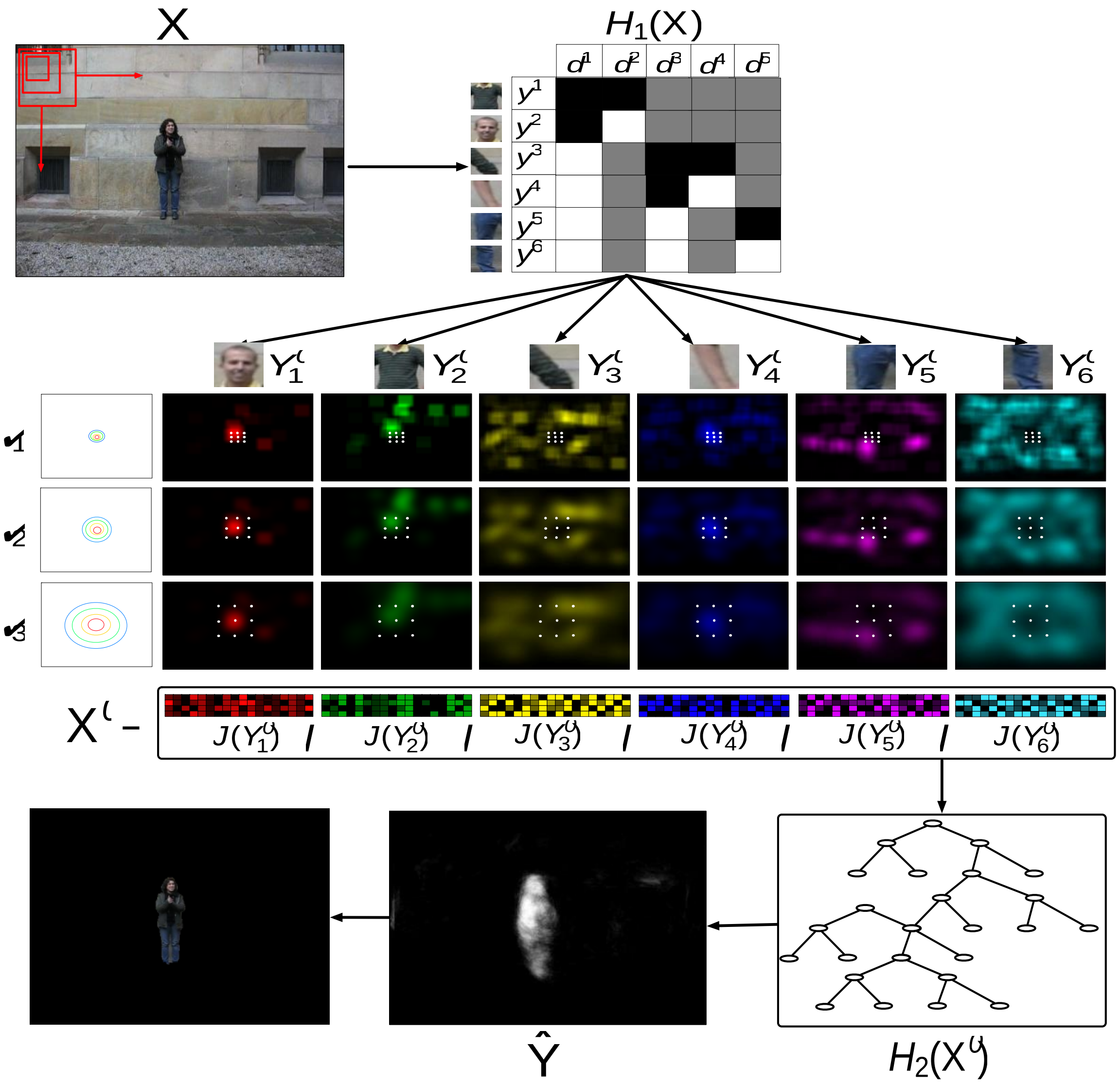
Eloi Puertas, Miguel Ángel Bautista, Daniel Sanchez, Sergio Escalera and Oriol Pujol

University of Barcelona and Computer Vision Center

Outline

Outline

1. Motivation.
2. Methodology.
3. Results.
4. Conclusions.



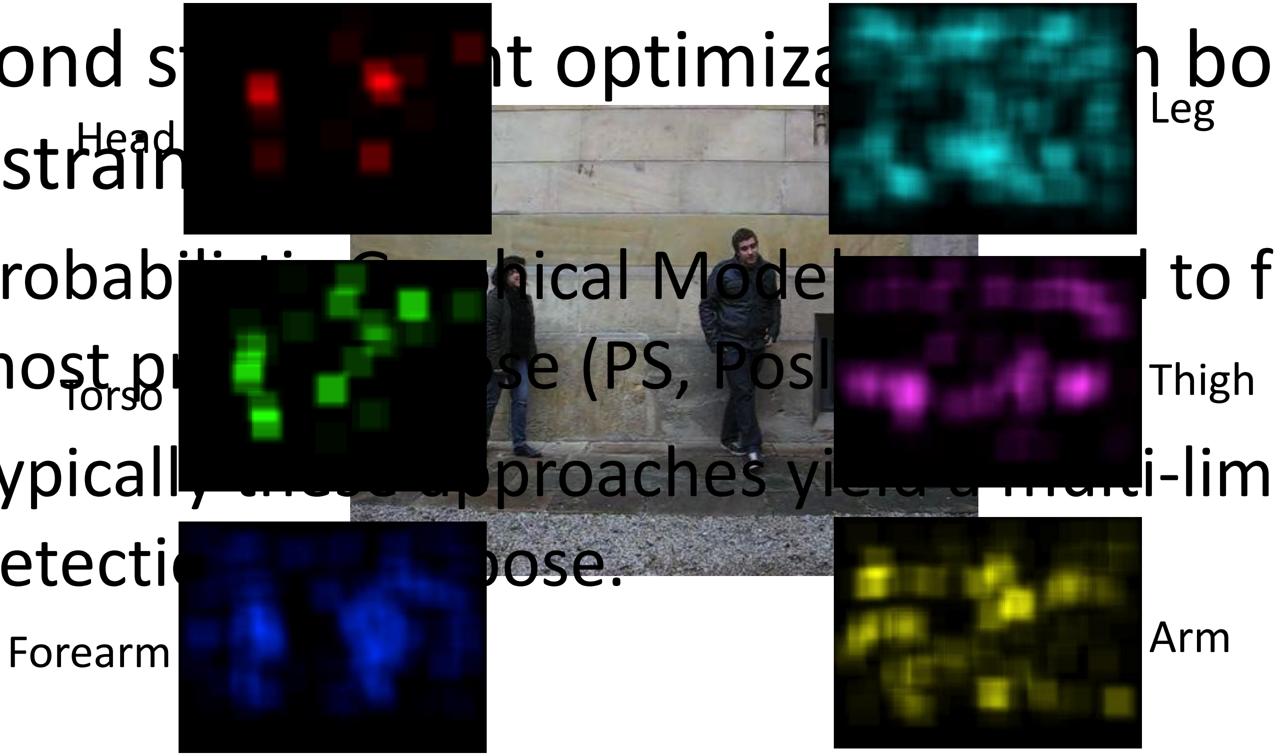
Segmenting humans in still images

- Problem:
  - **Segmenting the human** body (not the body-parts) in still **RGB images.**
  - **Several people** can appear portraying a **wide range of poses.**
- Approaches:
  - **One stage:**
    - Dalal & Triggs (HoG+SVM).
  - **Two stage**:
    - Andriluka, Roth & Schiele (Pictorial Structure)[1].
    - Bourdev, Maji, Brox & Malik (Poselets) [2].
    - Hernandez, Zlateva, Marinov, Reyes, Radeva, Dimov & Escalera (Graph Cuts) [3].

1. Andriluka, M., Roth, S., Schiele, B.: Pictorial structures revisited: People detection and articulated pose estimation. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. pp. 1014–1021. IEEE (2009)
2. Bourdev, L., Maji, S., Brox, T., Malik, J.: Detecting people using mutually consis- tent poselet activations. In: Computer Vision–ECCV 2010, pp. 168–181. Springer (2010)
3. 3Hernandez-Vela, A., Zlateva, N., Marinov, A., Reyes, M., Radeva, P., Dimov, D., Escalera, S.: Graph cuts optimization for multi-limb human segmentation in depth maps. In: CVPR. pp. 726–732 (2012)

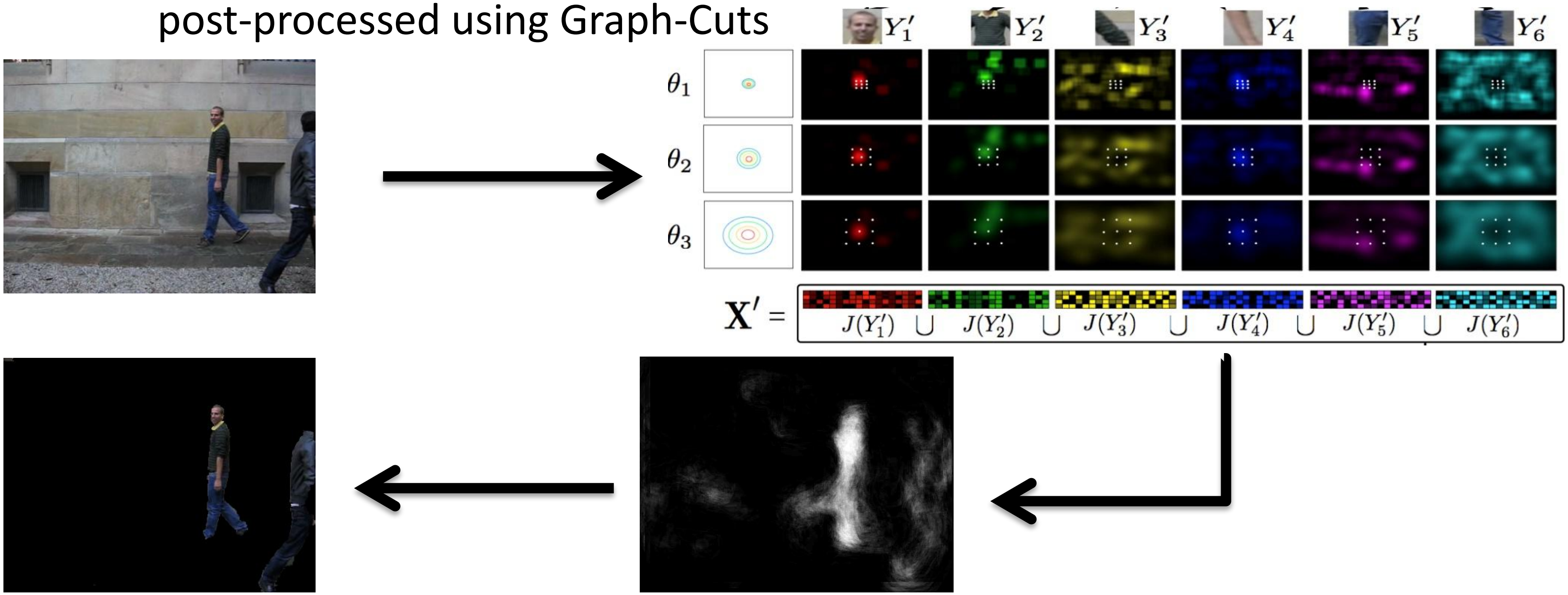
Two-Stage approach for human (multi-limb) segmentation

- First stage (body-part detection)
  - Use “unexpensive” classifiers to learn body parts: SVM, Adaboost, Cascading Classifiers, etc.
  - A large/noisy set of candidate parts is obtained.
- Second s t optimiza n body constrain
  - Probabi hical Mode l to find the most p se (PS, Posl
  - Typicall proaches yi i-limb detecti ose.



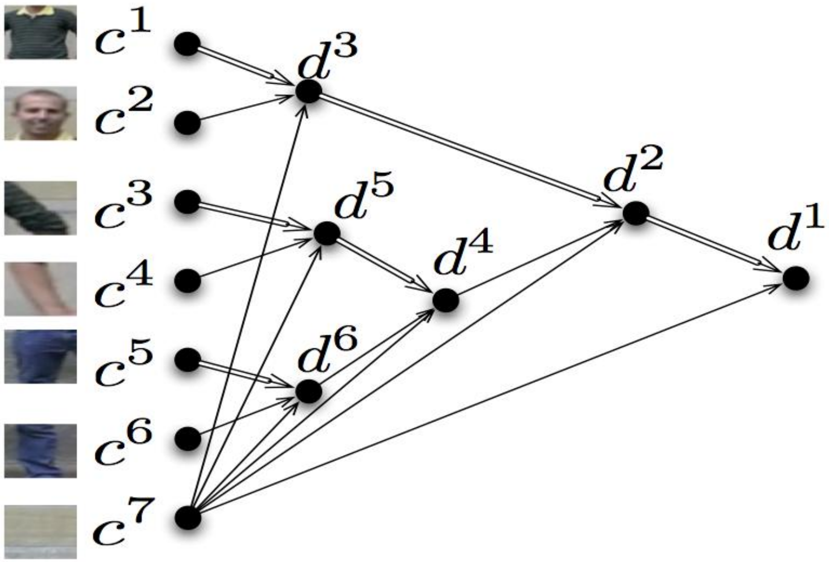
Segmenting humans by stacking body-parts

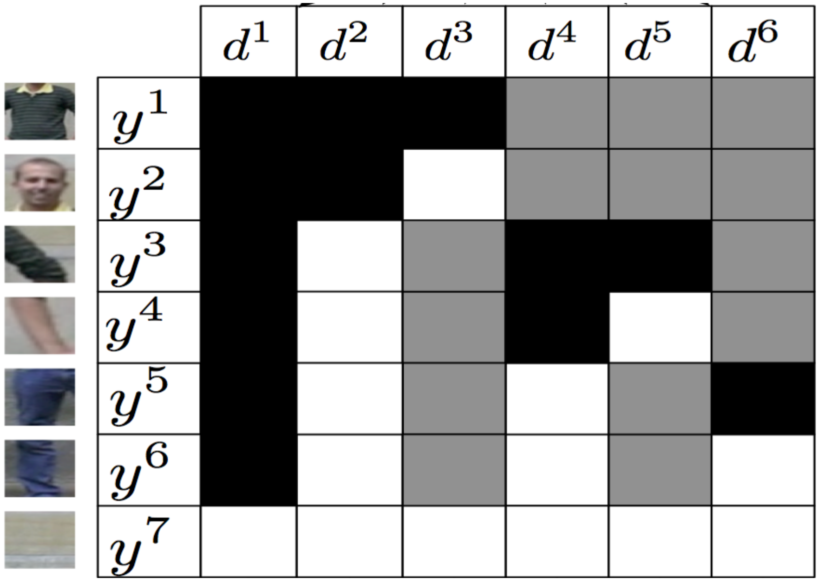
- **Our goal**
  - **Improve the binary segmentation** of the human body in RGB images by **learning context-aware features**.
- **Our proposal**
  - Define a **two stage** scheme where an extended feature set is learned.
  - Use the **Multi-Scale Stacked Sequential Learning** framework (MSSL) to build the extended feature set.
  - Obtain a prior **pixel-wise binary classification** of the image which is post-processed using Graph-Cuts


Y'_1
Y'_2
Y'_3
Y'_4
Y'_5
Y'_6
θ_1
θ_2
θ_3
$\mathbf{X}' = J(Y'_1) \cup J(Y'_2) \cup J(Y'_3) \cup J(Y'_4) \cup J(Y'_5) \cup J(Y'_6)$

Stage One (H_1): Soft Body Parts Detectors and Error-Correcting Output Codes (TR)

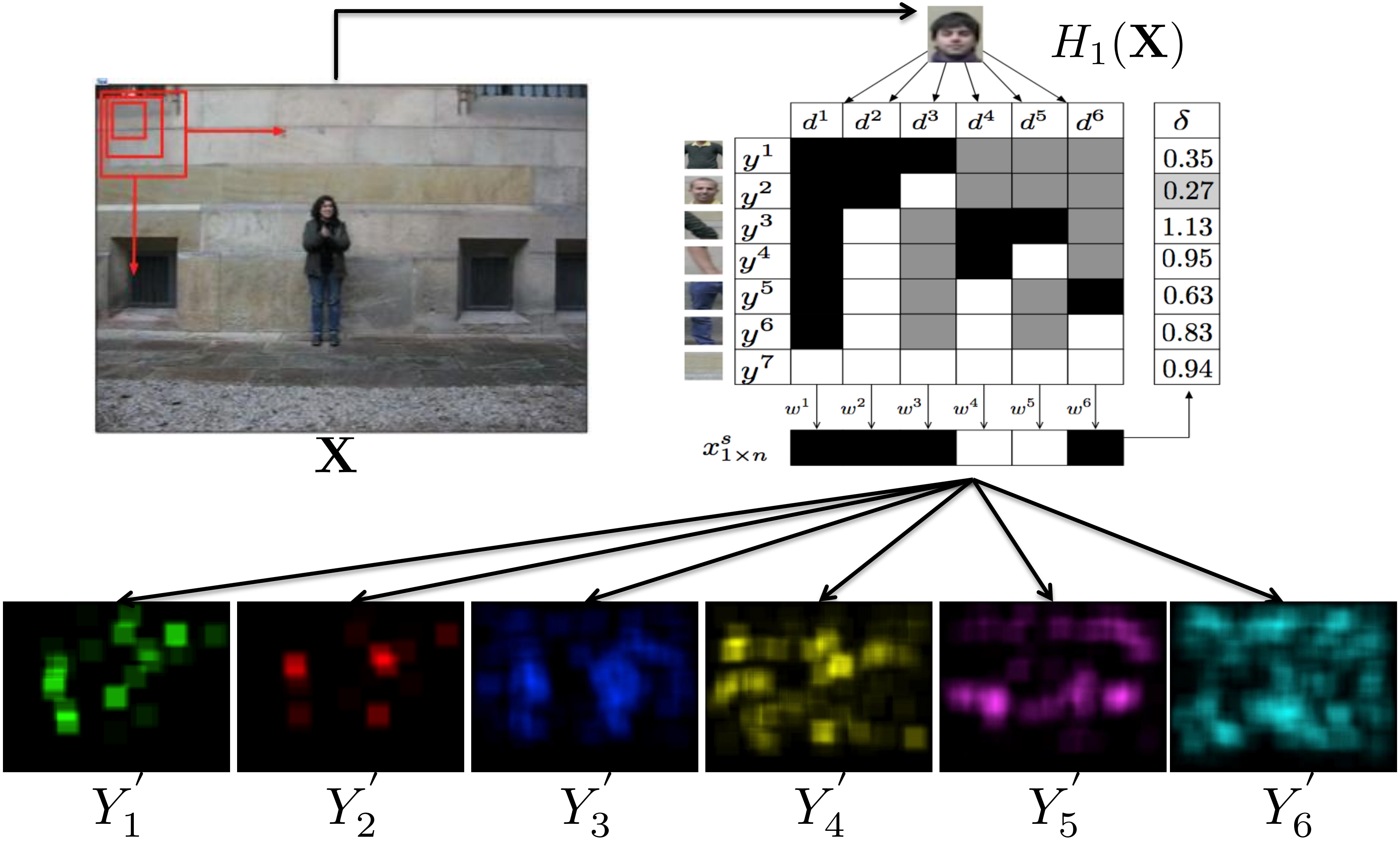
- Multi-class body-part detection based on **Error-Correcting Output Codes** (ECOC) and Soft Body Part Detectors (Cascading classifiers+Adaboost).
- **Problem-dependent coding** for body-part learning, where difficult dichotomies have few classes.
- Each d^i denotes a **dichotomy** (binary body part classification problem), that is coded within the ECOC coding matrix.





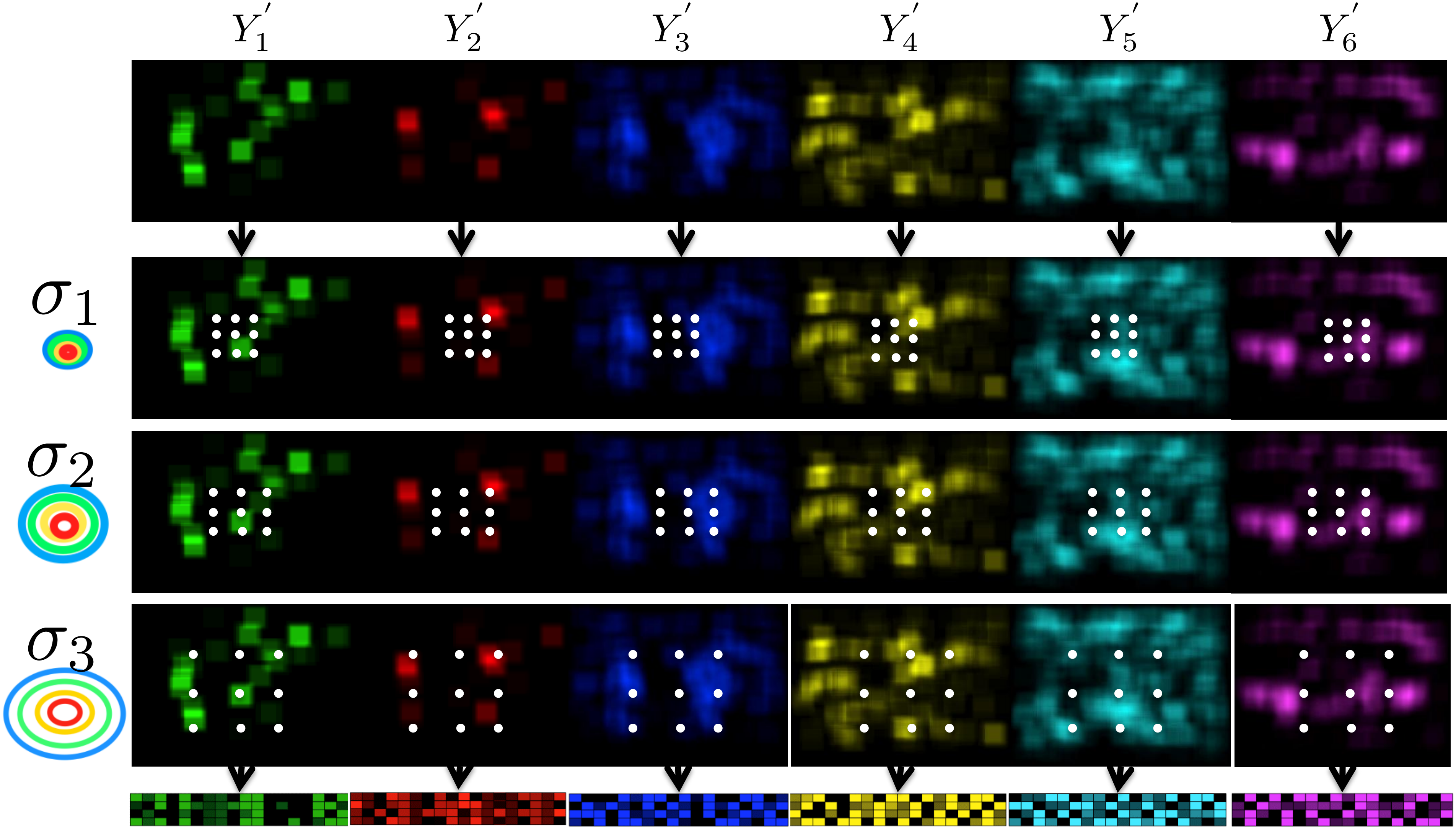
	d^1	d^2	d^3	d^4	d^5	d^6
y^1	■	■	■	▨	▨	▨
y^2	■	■	□	▨	▨	▨
y^3	■	□	▨	■	■	▨
y^4	■	□	▨	■	□	▨
y^5	■	□	▨	□	▨	■
y^6	■	□	▨	□	▨	□
y^7	□	□	□	□	□	□

Stage One (H_1): Soft Body Parts Detectors and Error-Correcting Output Codes (TR)


$H_1(\mathbf{X})$
$\mathbf{X}$
w^1 w^2 w^3 w^4 w^5 w^6
$x^s_{1\times n}$
Y'_1 Y'_2 Y'_3 Y'_4 Y'_5 Y'_6

	d^1	d^2	d^3	d^4	d^5	d^6	δ
y^1	■	■	■	▨	▨	▨	0.35
y^2	■	■	□	▨	▨	▨	0.27
y^3	■	□	▨	■	■	▨	1.13
y^4	■	□	▨	■	□	▨	0.95
y^5	■	□	▨	□	▨	■	0.63
y^6	■	□	▨	□	▨	□	0.83
y^7	□	□	□	□	□	□	0.94
$x^s_{1\times n}$	■	■	■	□	□	■	

Stage Two: Fusing Limb Likelihood Maps Using MSSL (TR)

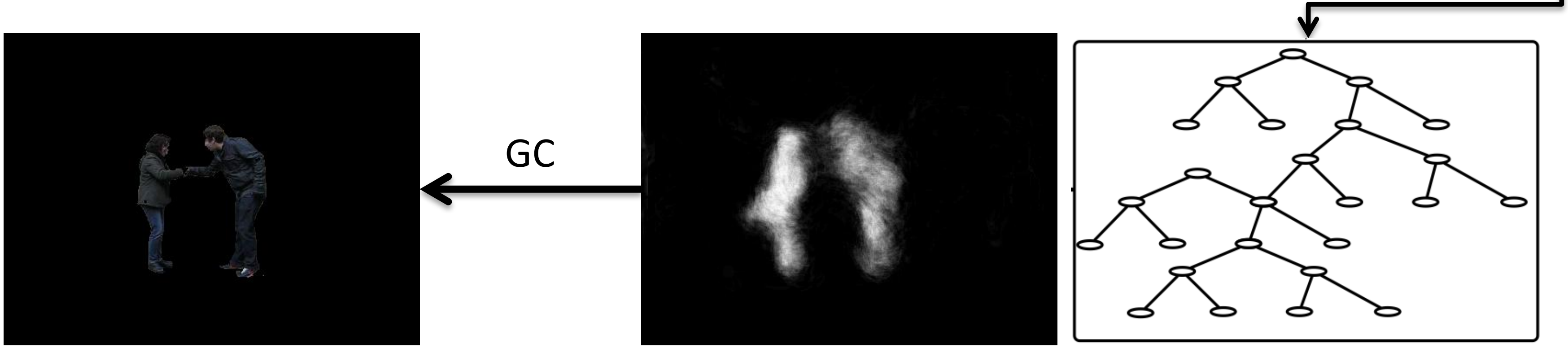


$$\mathbf{X}' = J(Y_1') \cup J(Y_2') \cup J(Y_3') \cup J(Y_4') \cup J(Y_5') \cup J(Y_6') \in \mathbb{R}^{\#N \times \#S \times \#Y}$$

Stage Two: Fusing Limb Likelihood Maps Using MSSL (TR & TE)

- The **extended feature set** $\mathbf{X}'$ encodes for each sampled pixel a concatenation of the **probability of neighbouring pixels** to belong to a certain **body part**.
- Then we use a **Random Forest** classifier $H_2(\mathbf{X}')$ to learn the **pixel-wise classification problem** (person vs. background), which output is then optimized by means of Graph Cuts.

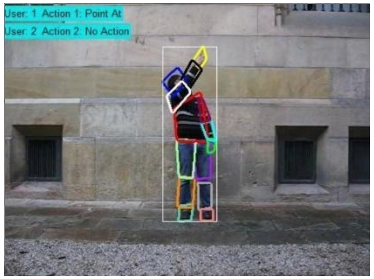
$$\mathbf{X}' = \boxed{J(Y_1') \;\cup\; J(Y_2') \;\cup\; J(Y_3') \;\cup\; J(Y_4') \;\cup\; J(Y_5') \;\cup\; J(Y_6')}$$

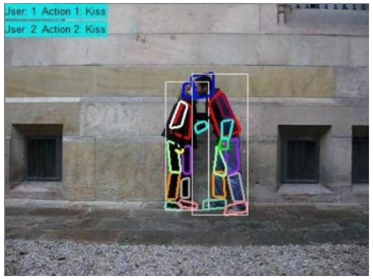


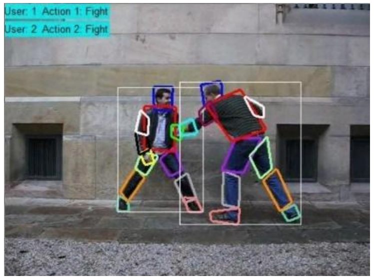
$H_2(\mathbf{X}')$

Experimental Settings I

- **Dataset:**
  - We used **HuPBA 8k+ dataset** which contains more than **8000 labeled** images at pixel precision, including more than **120000 manually labeled samples** of 14 different limbs.
  - We reduced the number of limbs from the 14 available in the dataset to 6: **head, torso, forearms, arms, thighs and legs**.







- **Methods:**
  - SBP-ECOC (H_1) + MSSL-RF (H_2) + Graph cut.
  - SBP-ECOC (H_1) + MSSL-RF (H_2) + GMM-Graph cut (Grabcut).
  - SBP-ECOC (H_1) + Graph Cut.
  - SBP-ECOC (H_1) + GMM-Graph Cut (Grabcut).

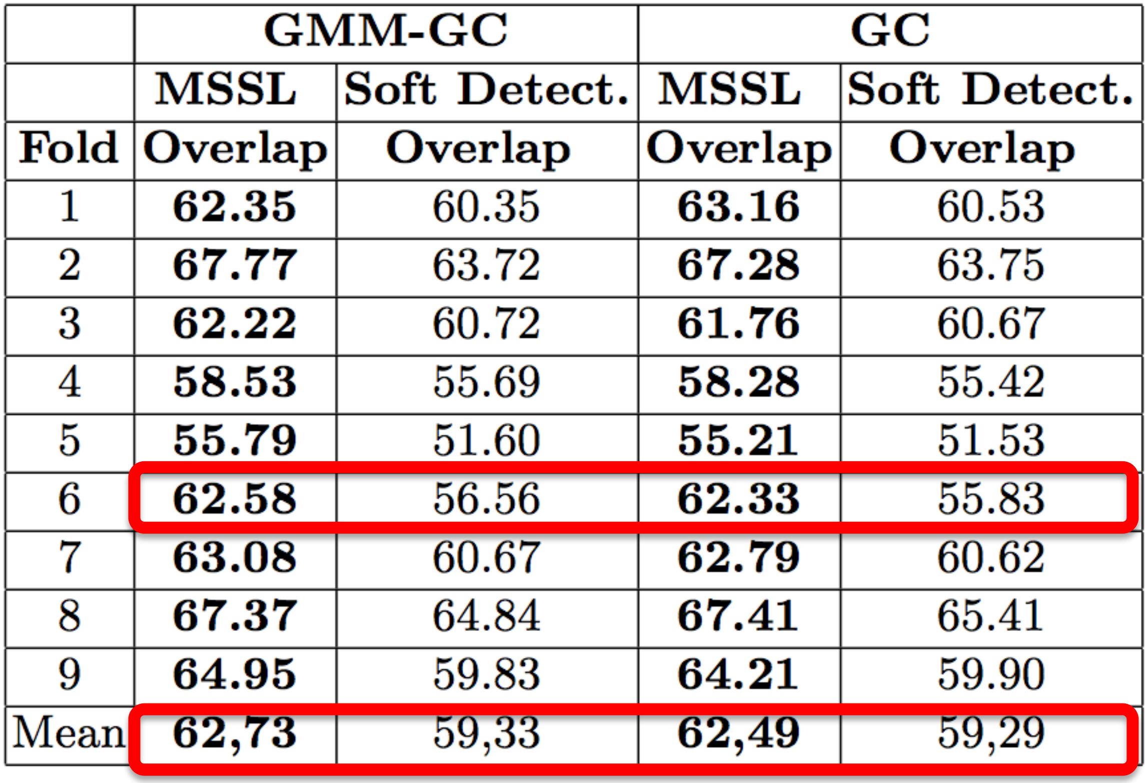
Experimental Settings II

- **Settings:**
  - We used the standard **Cascade of Classifiers** based on **AdaBoost and Haar-like features** as our body part multi-class classifier H_1, forcing a 0.99 false positive rate during 8 stages.
  - In the second stage, we performed **3-scale Gaussian decomposition** with $\sigma \in [8, 16, 32]$ for each body part.
  - We used a **Random Forest with 50 decision trees**, as H_2 classifier.
  - In a post-processing stage**, binary Graph Cuts with a GMM color modeling** (we experimentally set 3 components) were applied.
- **Validation Protocol:**
  - We used **9-fold cross-validation (leave one sequence out)**.
  - We used the **Jaccard Index of overlapping** as our results measurement.

$$J = \frac{A \cap B}{A \cup B}$$

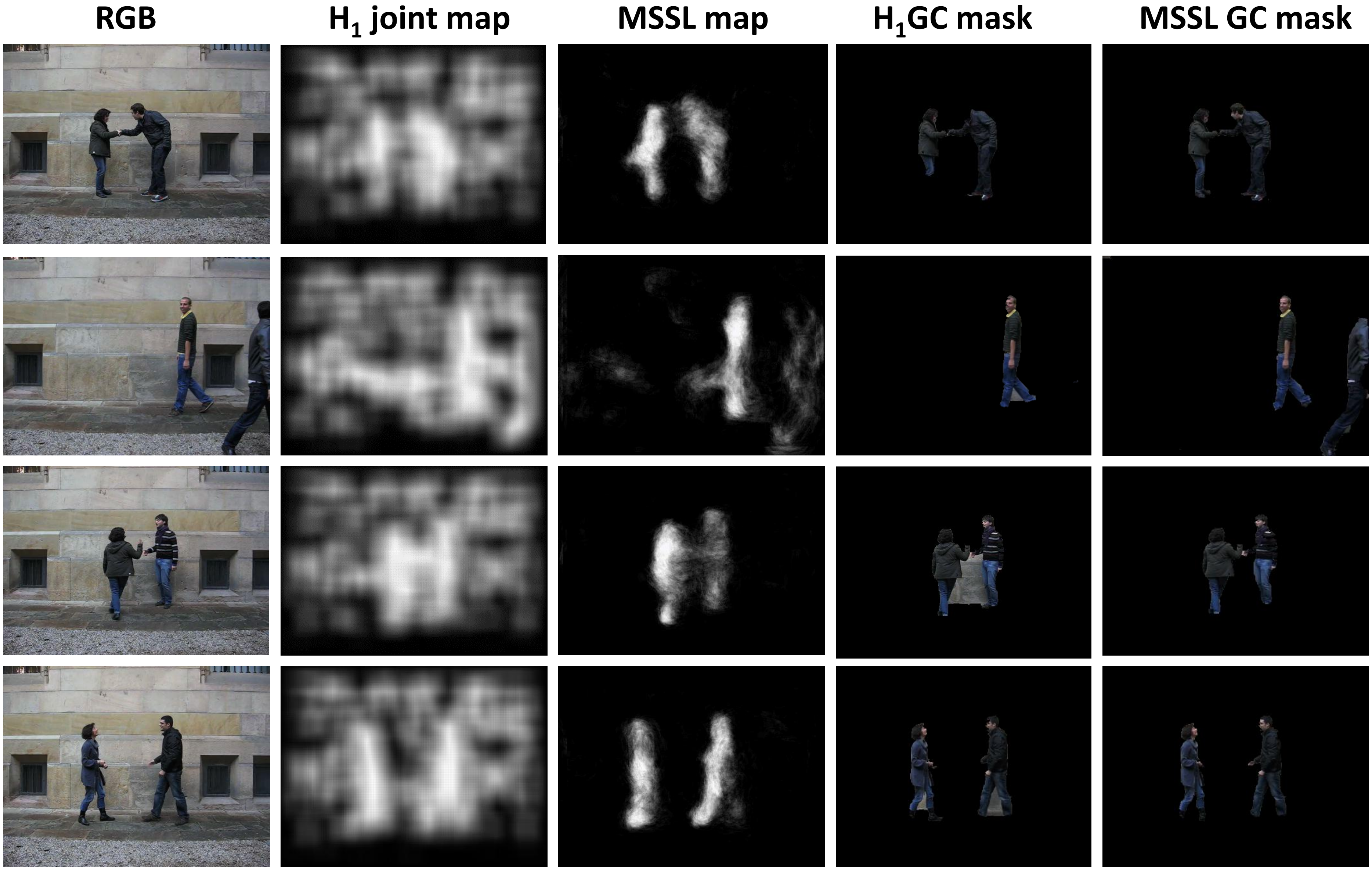
Quantitive Results

- When applying MSSL we find a consistent **improve in overlap of at least 3% in mean**.
- For certain folds the **improvements reach 5%**.



	GMM-GC		GC	
	MSSL	Soft Detect.	MSSL	Soft Detect.
Fold	Overlap	Overlap	Overlap	Overlap
1	**62.35**	60.35	**63.16**	60.53
2	**67.77**	63.72	**67.28**	63.75
3	**62.22**	60.72	**61.76**	60.67
4	**58.53**	55.69	**58.28**	55.42
5	**55.79**	51.60	**55.21**	51.53
6	**62.58**	56.56	**62.33**	55.83
7	**63.08**	60.67	**62.79**	60.62
8	**67.37**	64.84	**67.41**	65.41
9	**64.95**	59.83	**64.21**	59.90
Mean	**62,73**	59,33	**62,49**	59,29

Qualitative Results



Conclusions & Future Work

- We presented a **two-stage scheme based on the MSSL** framework for the **segmentation of the human body** in still images.
- **MSSL encodes extended feature** set using **contextual information** of human limbs.
- Our proposal was tested on a large dataset obtaining **significant segmentation improvement** over baseline methodologies.
- We are currently **extending the MSSL framework to the multi-limb case**, in which two multi-class classifiers will be concatenated to obtain a **body-aware segmentation**.

THANK YOU!

QUESTIONS?

Learning To Segment Humans By Stacking Their Body Parts

Universitat de Barcelona

HuPBA

CVC
Centre de Visió per Computador

ECCV '14
Zürich, September 6-12th, 2014

European Conference on Computer Vision