ECOC and Graph Cuts Segmentation of Human Limbs

Author: Daniel Sánchez

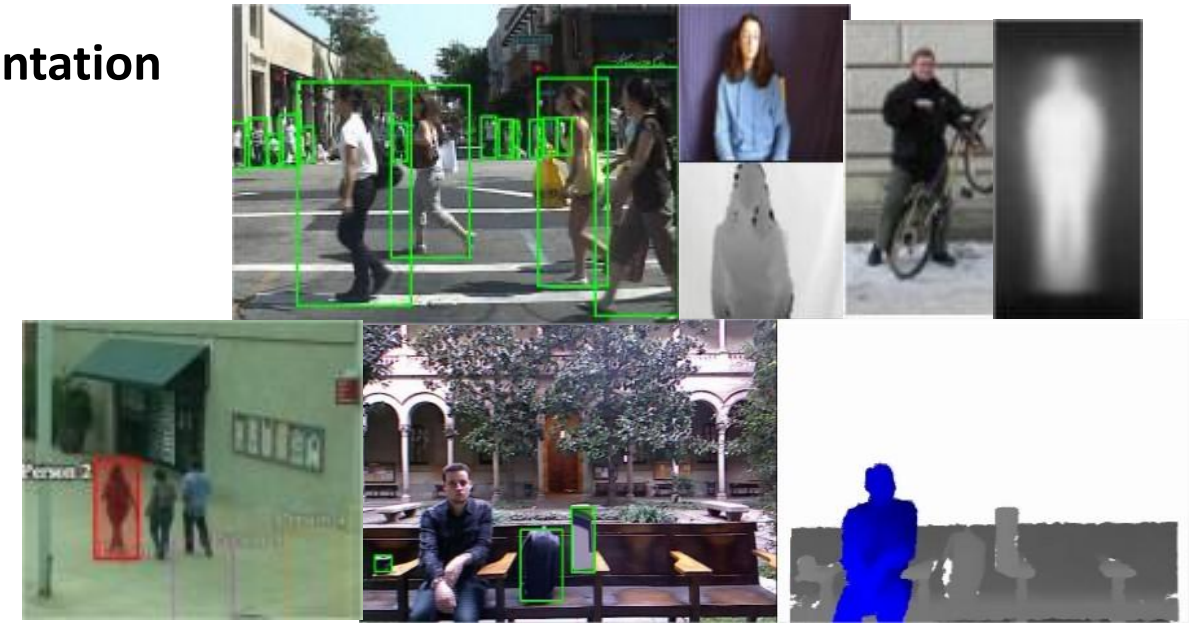
Advisors: Sergio Escalera and Miguel Ángel Bautista

Outline

1. Motivation
2. HuPBA 8k+ dataset
3. Proposal
4. Results
5. Conclusions

Motivation

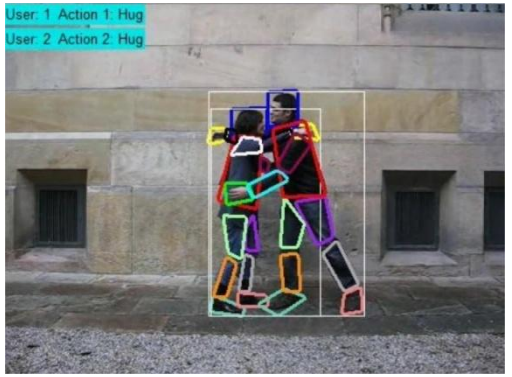
- **User Detection/Segmentation**

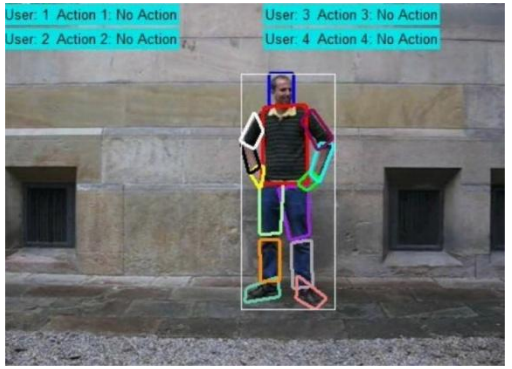


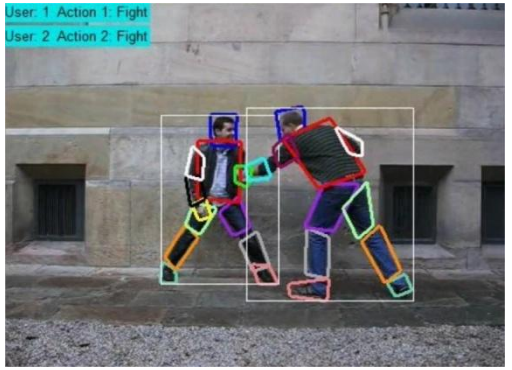
- **Applications:** medicine, photography, sign language…

HuPBA 8k+ dataset

- **Tasks:** Multi-limb human pose detection, segmentation, action /gesture recognition.
- 9 actors, 14 limbs categories, 11 gesture categories (isolated and collaborative actions).
- More than 8.000 frames, 120.000 manual labeled limbs at pixel precision.
- Limbs categories: Head, Torso, L/R forearms, L/R arms, L/R hands, L/R thighs, L/R legs and L/R foot.
- Action categories: Wave, Point, Clap, Crouch, Jump, Walk, Run, Shake Hands, Hug, Kiss and Fight.



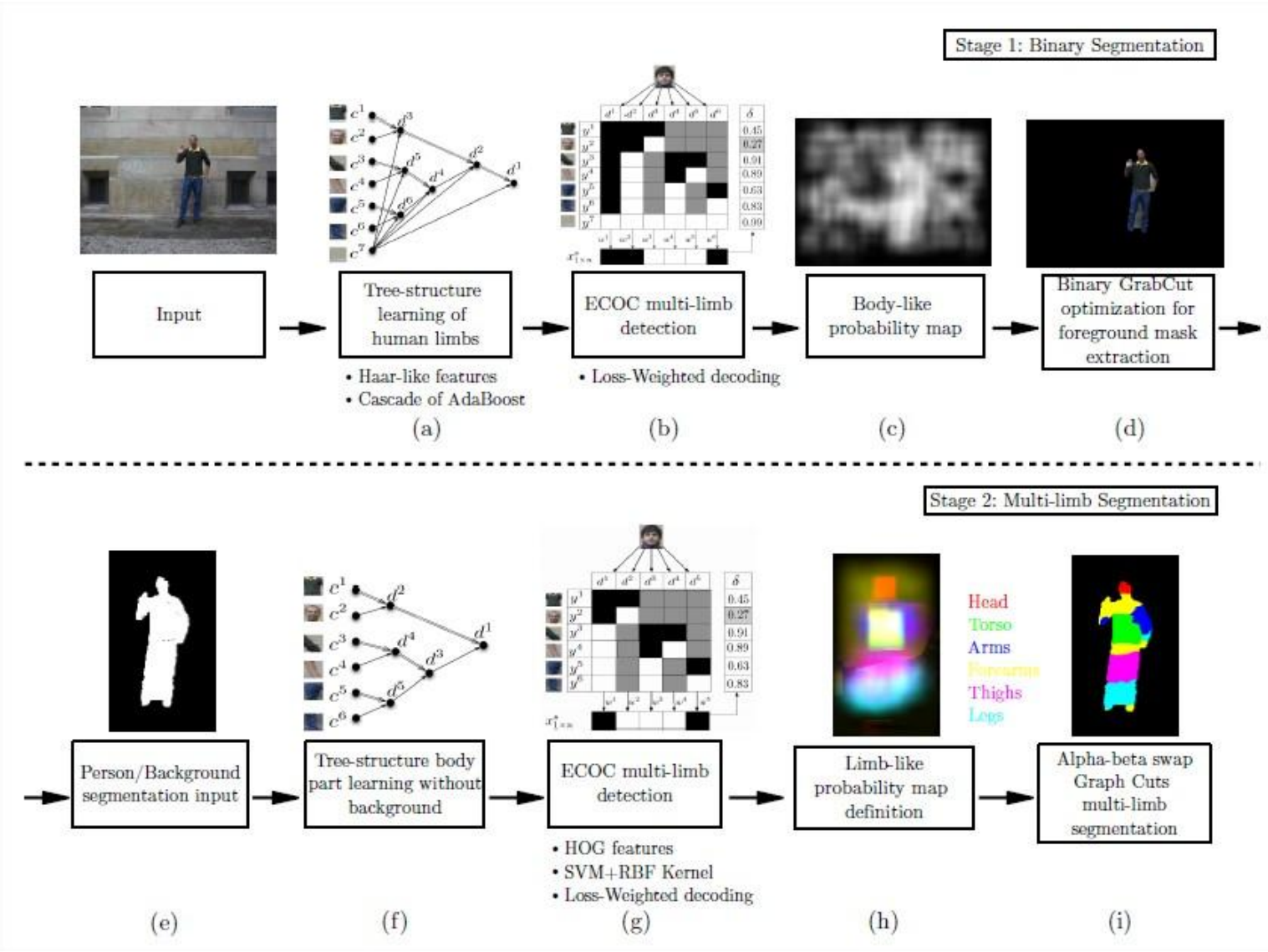


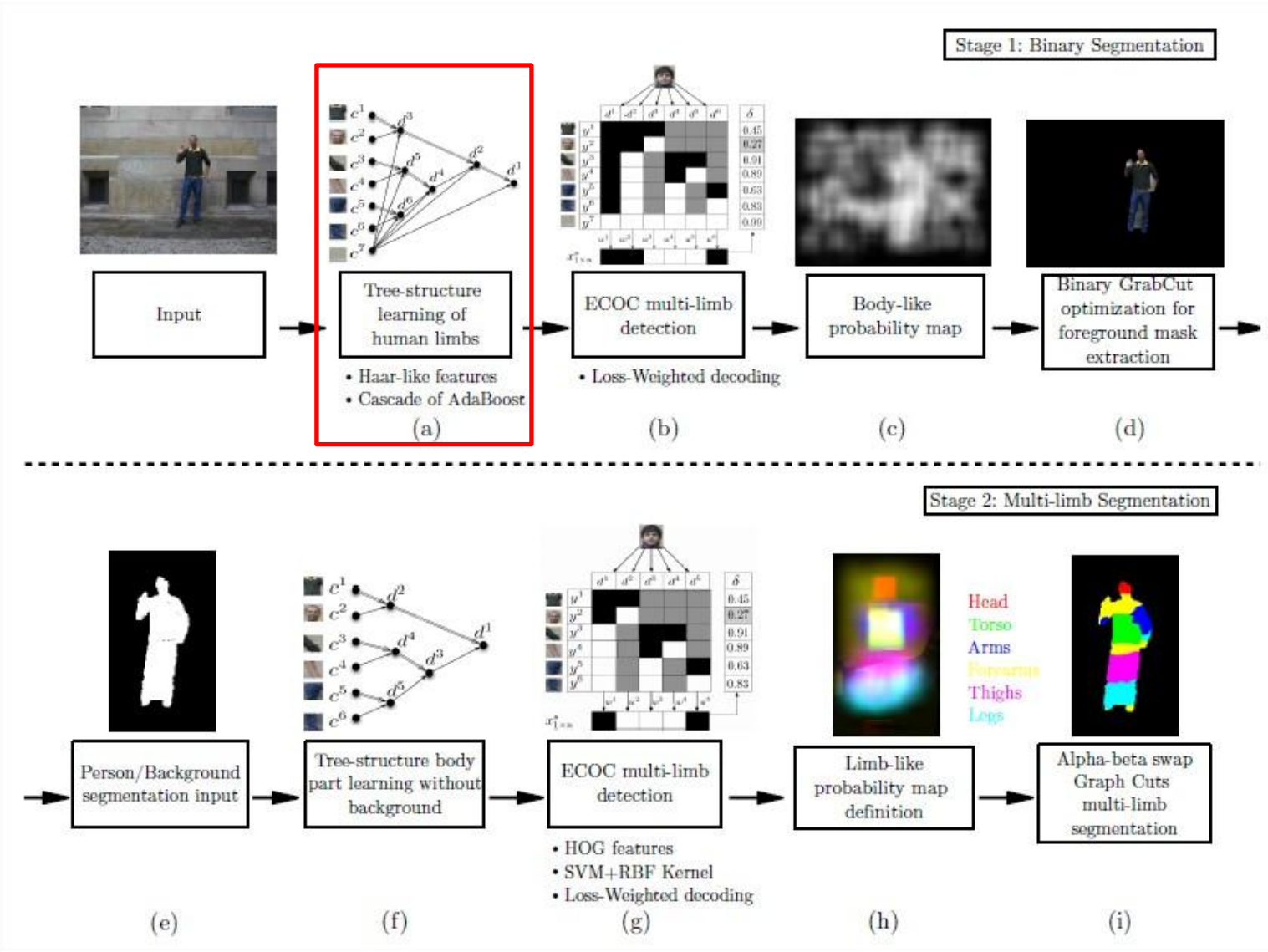


HuPBA 8k+ dataset

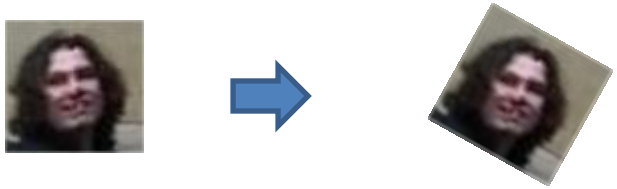
- Compare with some publicly available datasets. These public datasets are chosen taking into account the variability of limbs and gestures/actions.

	HuPBA	PARSE	BUFFY	UIUC people	LEEDS SPORTS	HW	MMGR13	H.Actions	Pascal VOC
Labeling at pixel precision	Yes	No	No	No	No	-	No	No	Yes
Number of limbs	14	10	6	14	14	-	16	-	5
Number of labeled limbs	124 761	3 050	4 488	18 186	28 000	-	27 532 800	-	8 500
Number of frames	8 234	305	748	1 299	2 000	-	1 720 800	-	1 218
Full body	Yes	Yes	No	Yes	Yes	-	Yes	Yes	Yes
Limb annotation	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes
Gesture annotation	Yes	No	No	No	No	Yes	Yes	Yes	No
Number of gestures	11	-	-	-	-	8	20	6	-
Number of gesture samples	235	-	-	-	-	430	13 858	600	-

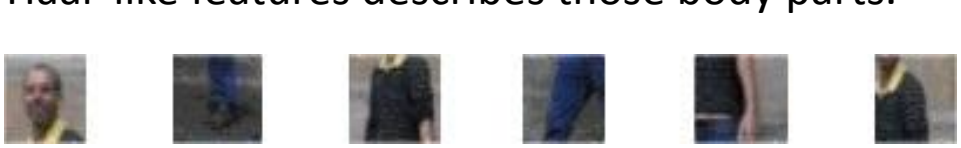

Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Alpha-beta swap Graph Cuts multi-limb segmentation
Head
Torso
Arms
Forearms
Thighs
Legs
(e)
(f)
(g)
(h)
(i)

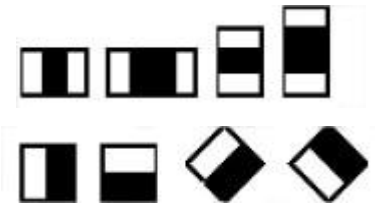

Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
(a)
ECOC multi-limb detection
• Loss-Weighted decoding
(b)
Body-like probability map
(c)
Binary GrabCut optimization for foreground mask extraction
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
(e)
Tree-structure body part learning without background
(f)
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
(g)
Limb-like probability map definition
Head
Torso
Arms
Forearms
Thighs
Legs
(h)
Alpha-beta swap Graph Cuts multi-limb segmentation
(i)

- Body parts rotational invariant by computing dominant orientation.

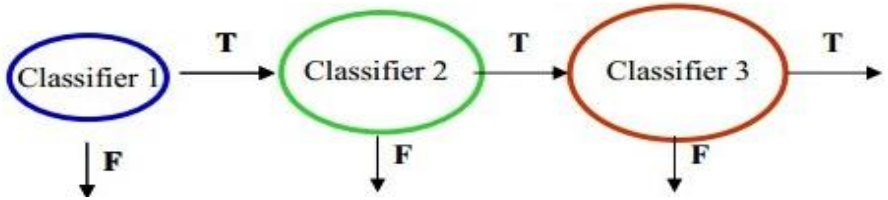


- Haar-like features describes those body parts.



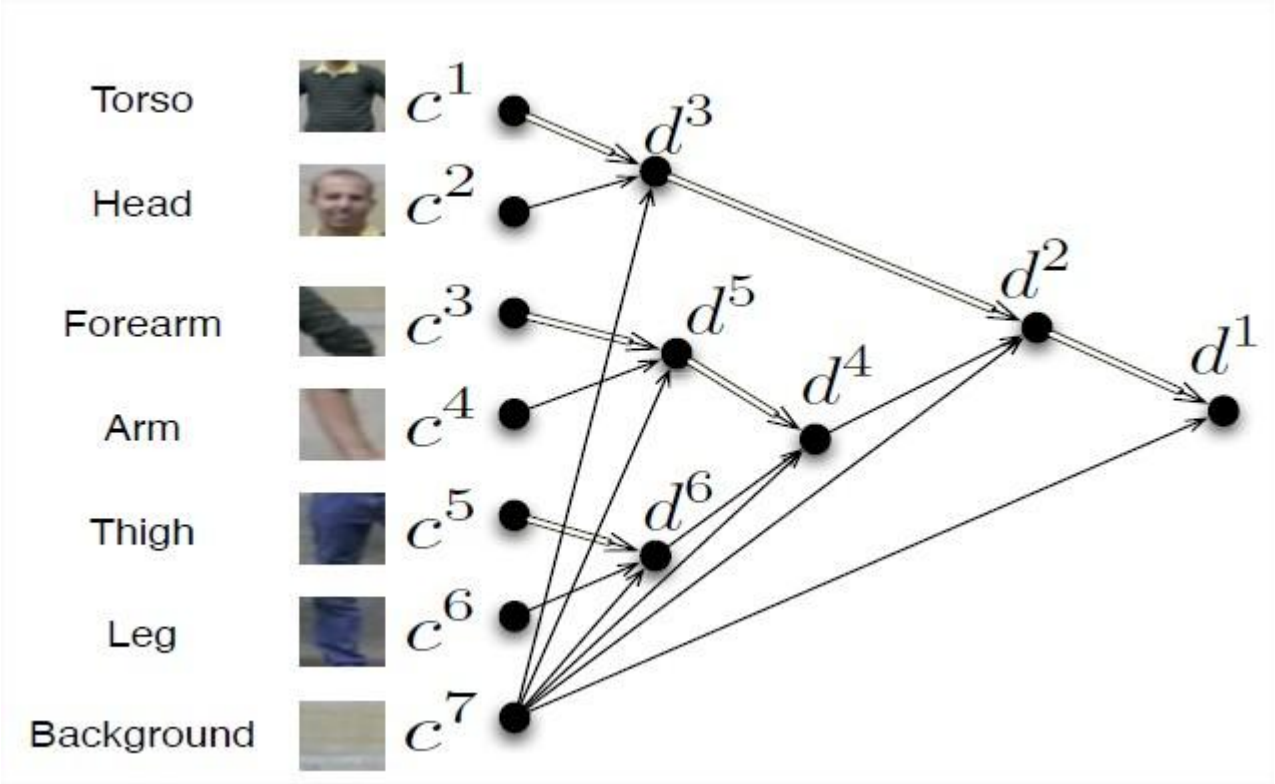


- Adaboost as the base classifier in the cascade architecture.

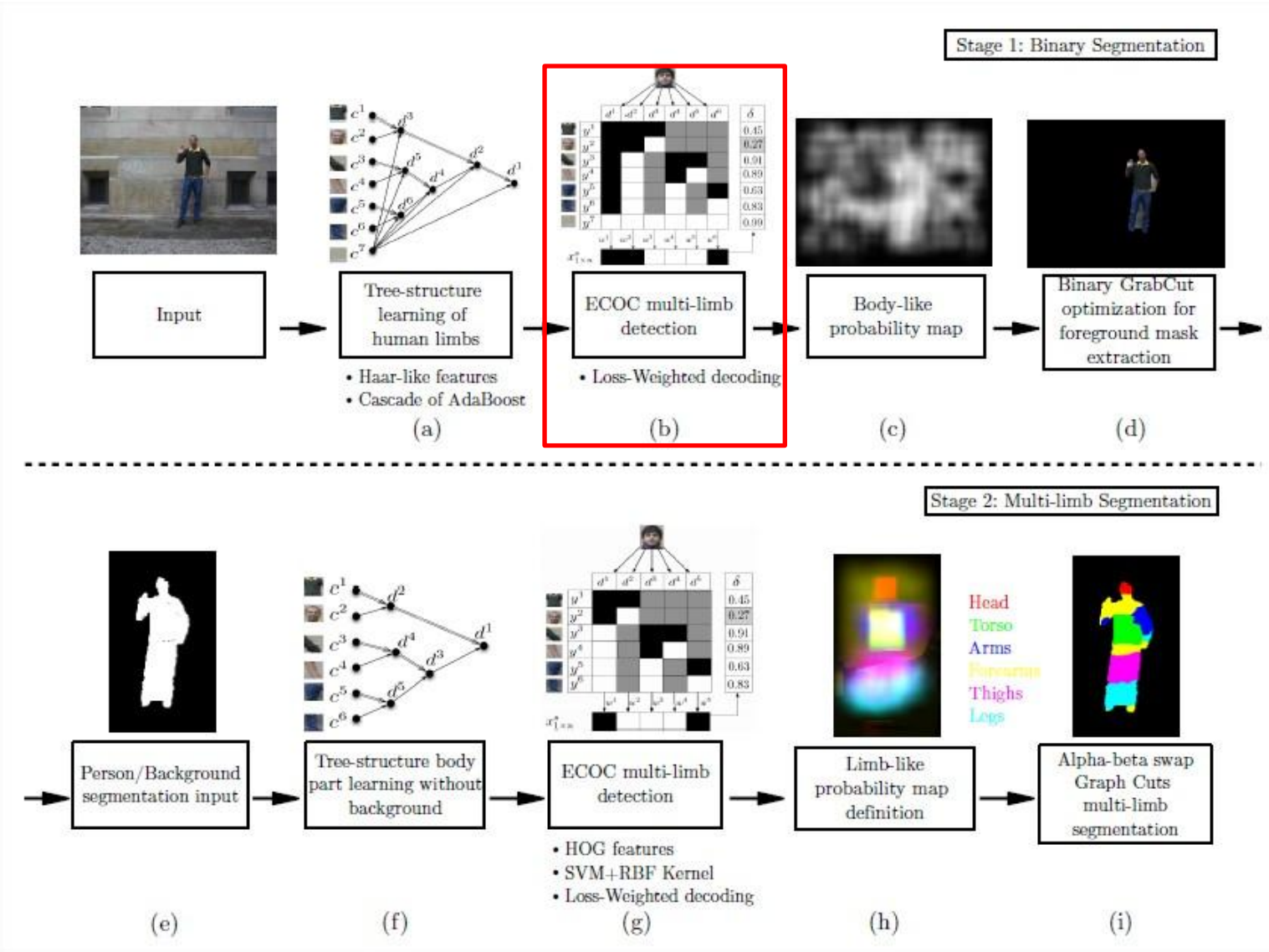


P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR, Vol. 1, 2001.
Y. Freund, R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, in: EuroCOLT, 1995, pp. 23-37.

- Define the groups of limbs to be learnt by each individual cascade.

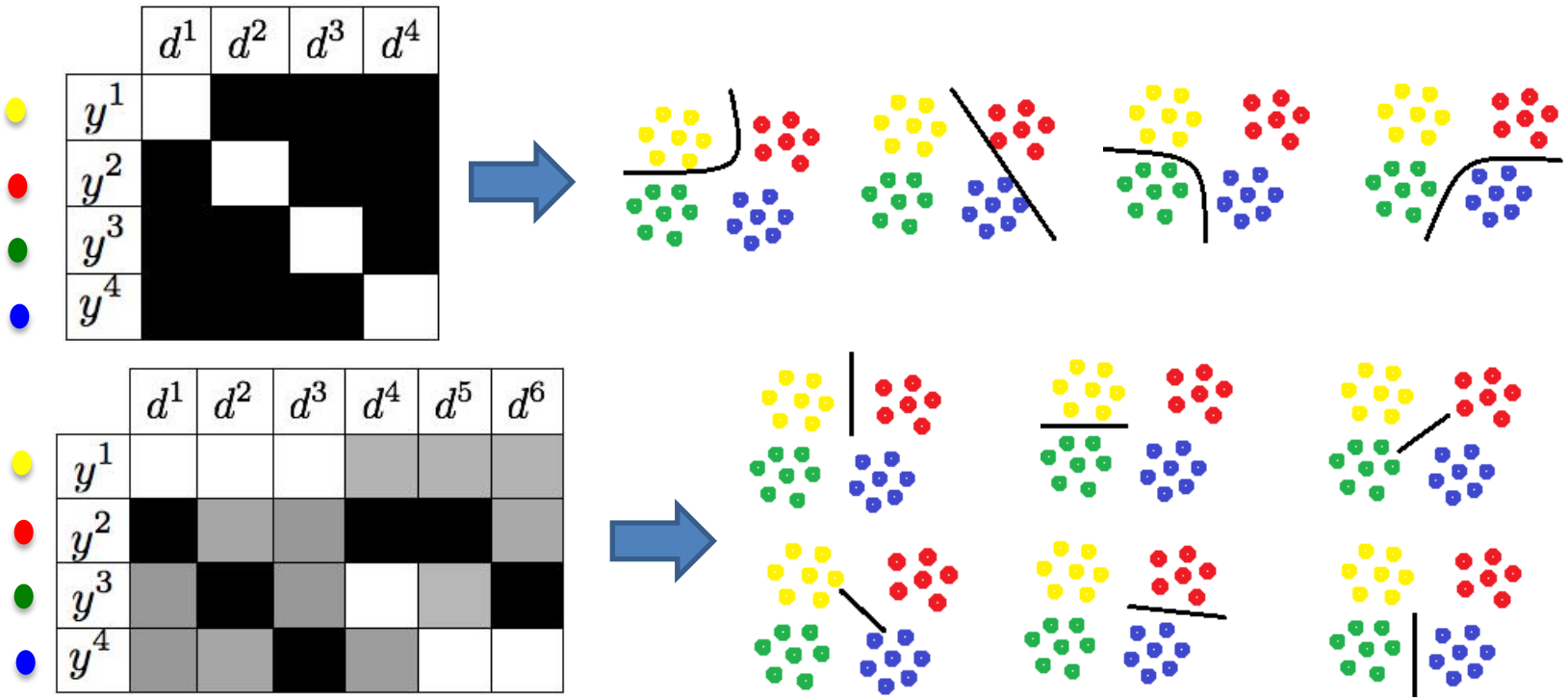


S. Escalera, O. Pujol, P. Radeva, On the decoding process in ternary error-correcting output codes, PAMI 32 (2010) 120-134.


Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
(a)
ECOC multi-limb detection
• Loss-Weighted decoding
(b)
Body-like probability map
(c)
Binary GrabCut optimization for foreground mask extraction
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
(e)
Tree-structure body part learning without background
(f)
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
(g)
Limb-like probability map definition
Head
Torso
Arms
Forearms
Thighs
Legs
(h)
Alpha-beta swap Graph Cuts multi-limb segmentation
(i)

- In classification tasks, the goal is to classify an object among a certain number of possible categories.

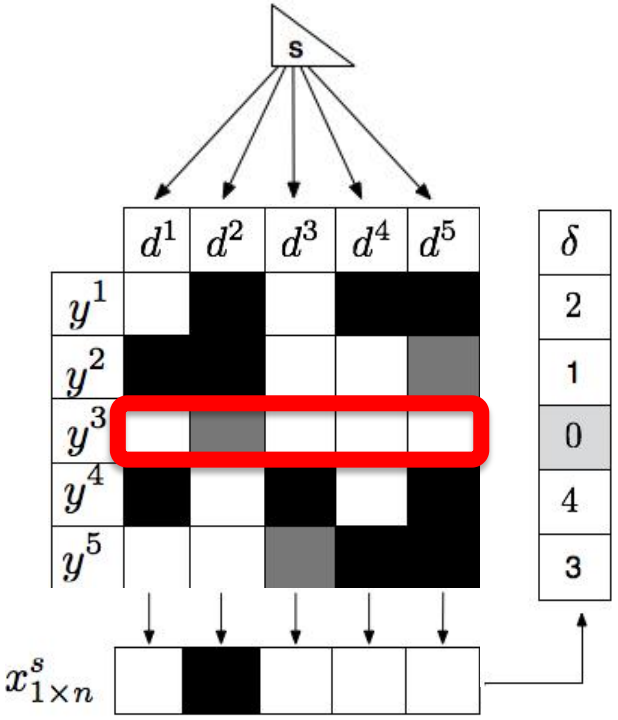
- This framework is composed of two different steps :
  - **Coding** : Decompose a given *N*-class problem into a set of *n* binary problems.
  - **Decoding** : Given a test sample *s*, determine its category.



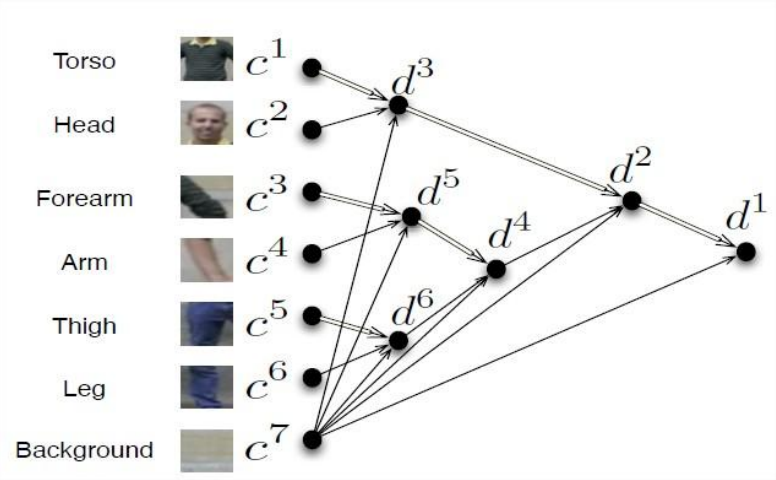
- At the decoding step a new sample *s* is classified by comparing the binary responses to the rows of *M* by means of a decoding measure δ.

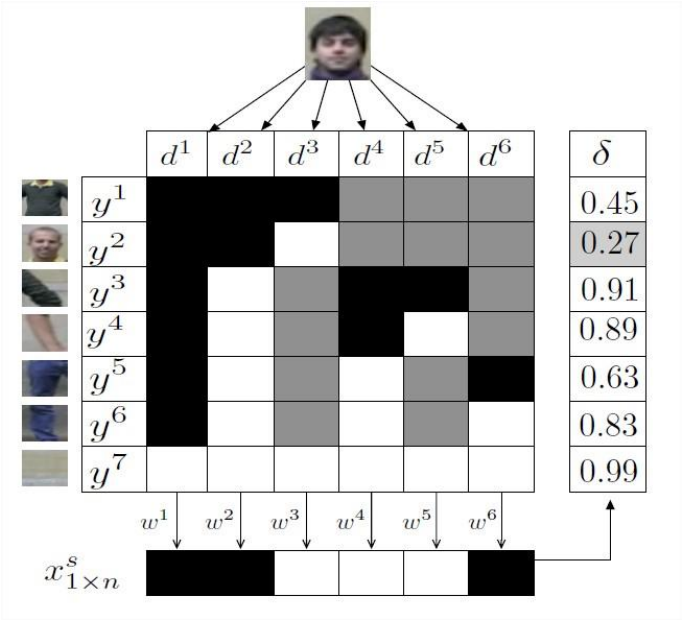
- Different types of decoding based on the distance used (i.e. Hamming, Euclidean, etc.)

$$\arg\min_i \delta(x^s, y^i)$$



- We propose to use a predefined coding matrix in which each dichotomy is obtained from the body part tree-structure.



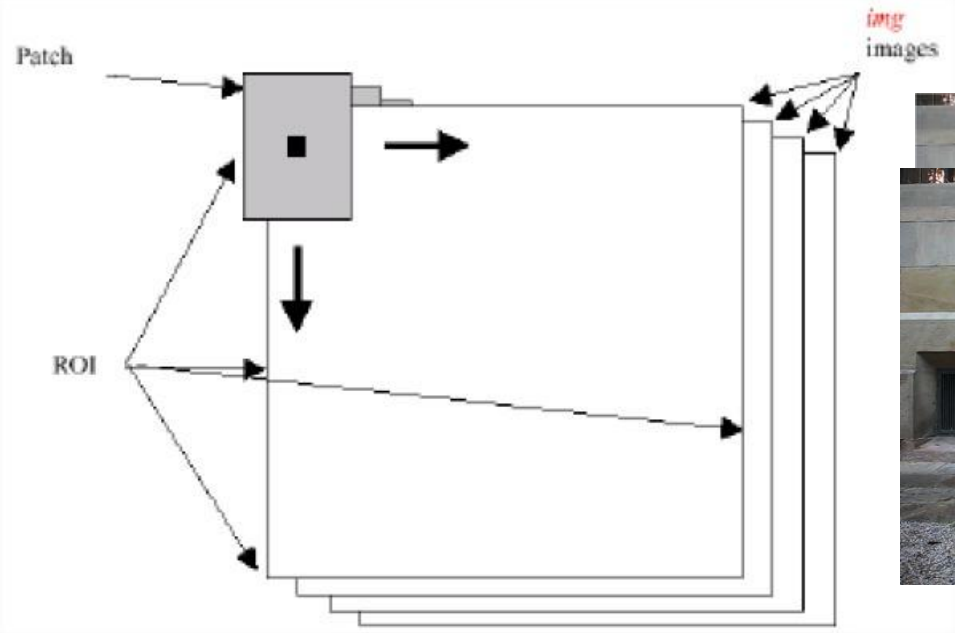


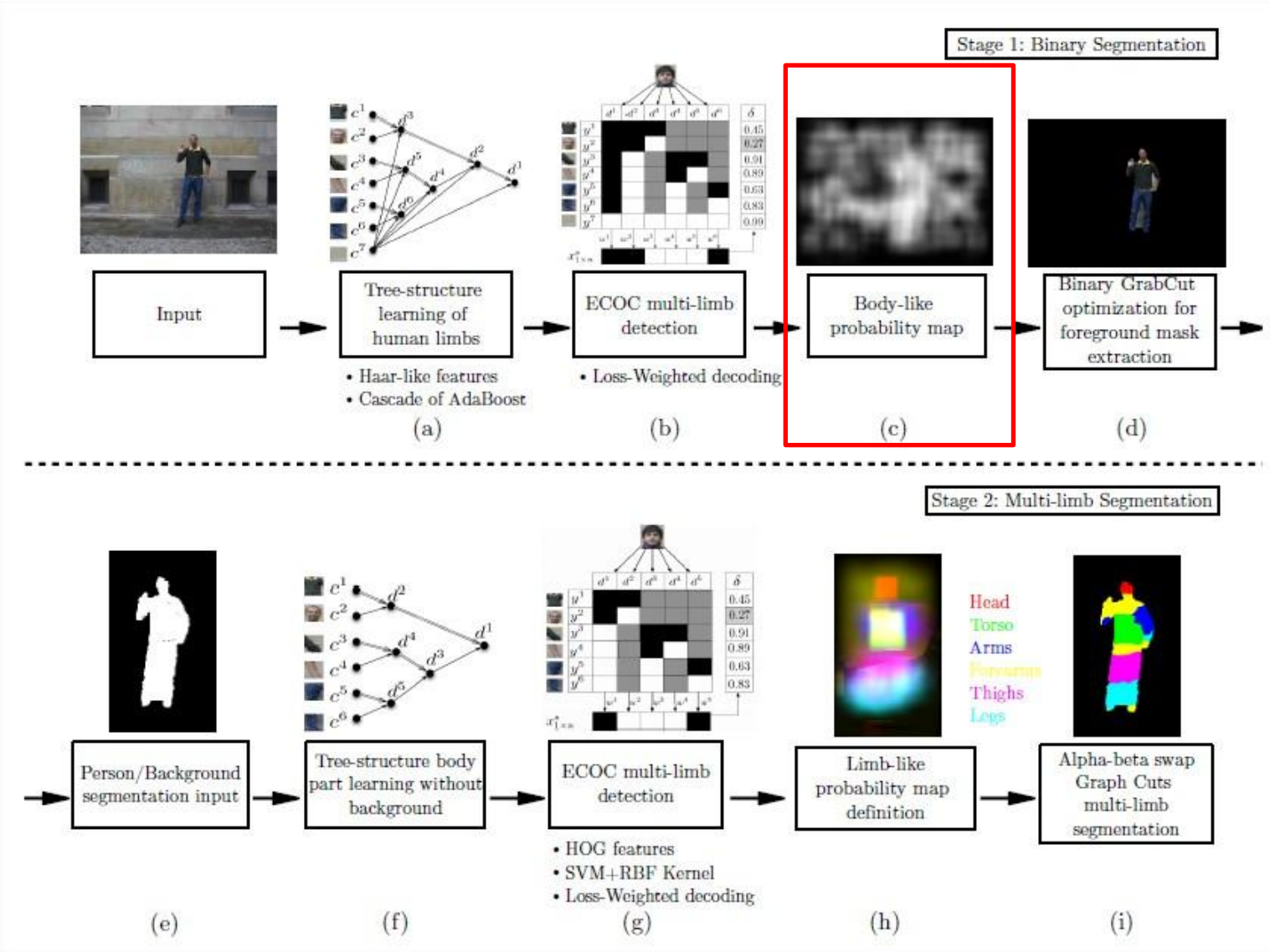
- Then, each cascade will give us its prediction and decoding ECOC step will be applied.
  - Loss-weighted decoding using cascade of classifier weights (takes into account classifiers performance)

S. Escalera, D. Tax, O. Pujol, P. Radeva, R. Duin, Subclass problem-dependent design of error-correcting output codes, PAMI 30 (6) (2008) 1-14.

M. A. Bautista, S. Escalera, X. Baro, P. Radeva, J. Vitria, O. Pujol, Minimal design of error-correcting output codes, Pattern Recogn. Lett. 33 (6) (2012) 693-702.

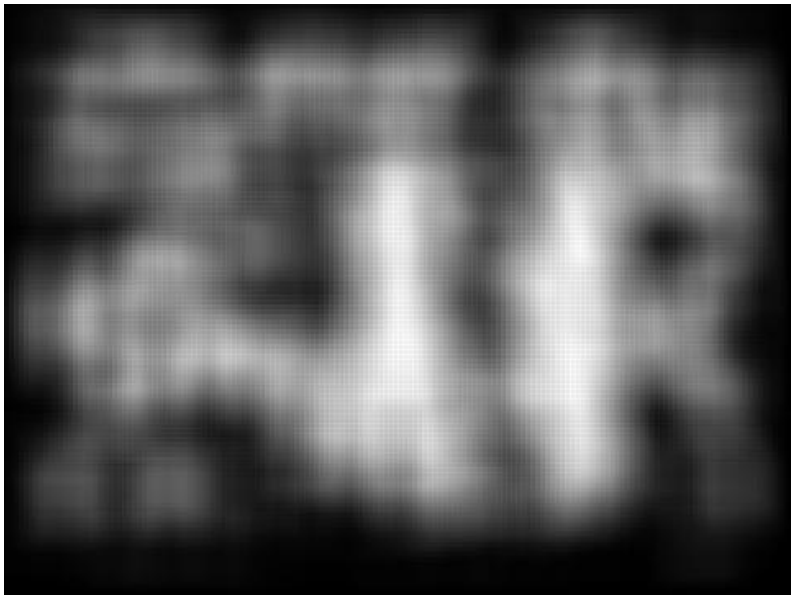
- In order to classify a new sample we apply a sliding window over the image:

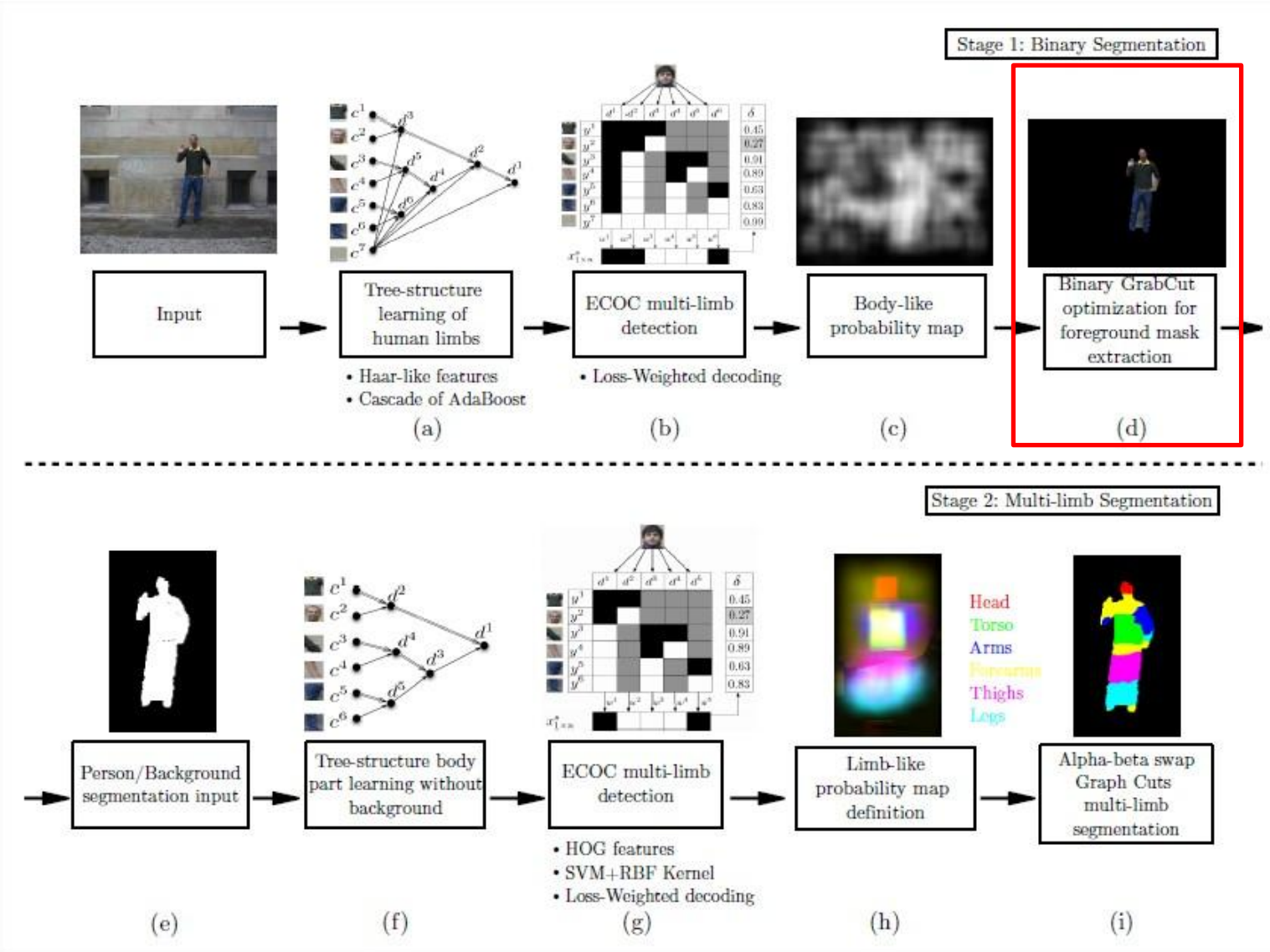



Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Alpha-beta swap Graph Cuts multi-limb segmentation
Head
Torso
Arms
Forearms
Thighs
Legs
(e)
(f)
(g)
(h)
(i)

Body-like probability map

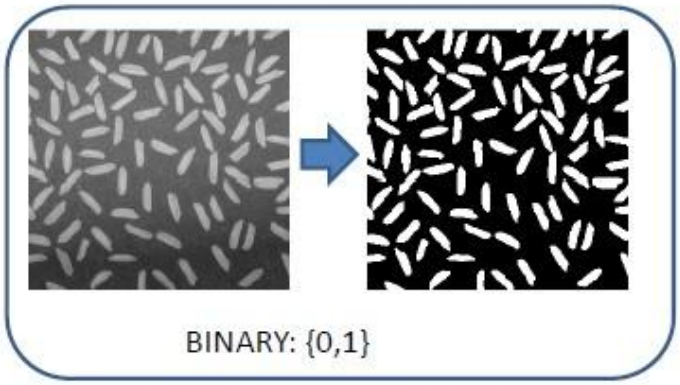
- A body-like probability map $P^{bl} \in [0,1]^{l \times w}$ is build




Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Head
Torso
Arms
Forearms
Thighs
Legs
Alpha-beta swap Graph Cuts multi-limb segmentation
(e)
(f)
(g)
(h)
(i)

GrabCut optimization for foreground mask extraction

- Image Segmentation == **Image labeling!**

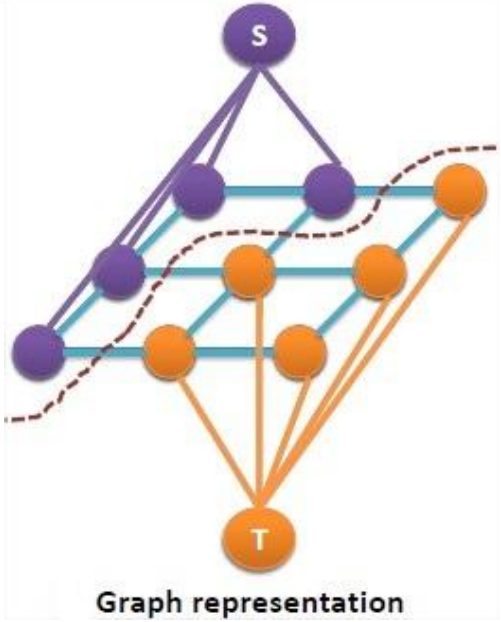


- Graph Cuts (Energy minimization)

$$\mathbf{E}(\boldsymbol{\alpha}, \mathbf{u}, \boldsymbol{\theta}, \mathbf{z}) = \mathbf{U}(\boldsymbol{\alpha}, \mathbf{u}, \boldsymbol{\theta}, \mathbf{z}) + \lambda \mathbf{V}(\boldsymbol{\alpha}, \mathbf{z})$$

Unary Potential

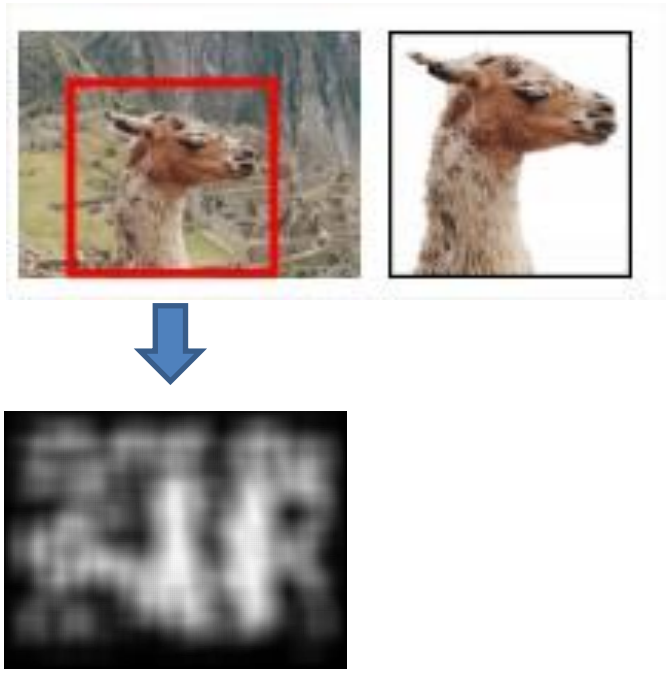
Pair-wise Potential



Yuri Y. Boykov and Marie-Pierre Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images", International Conference on Computer Vision, 2001

GrabCut optimization for foreground mask extraction

- User interaction by superimposed user input, background brush and so on.

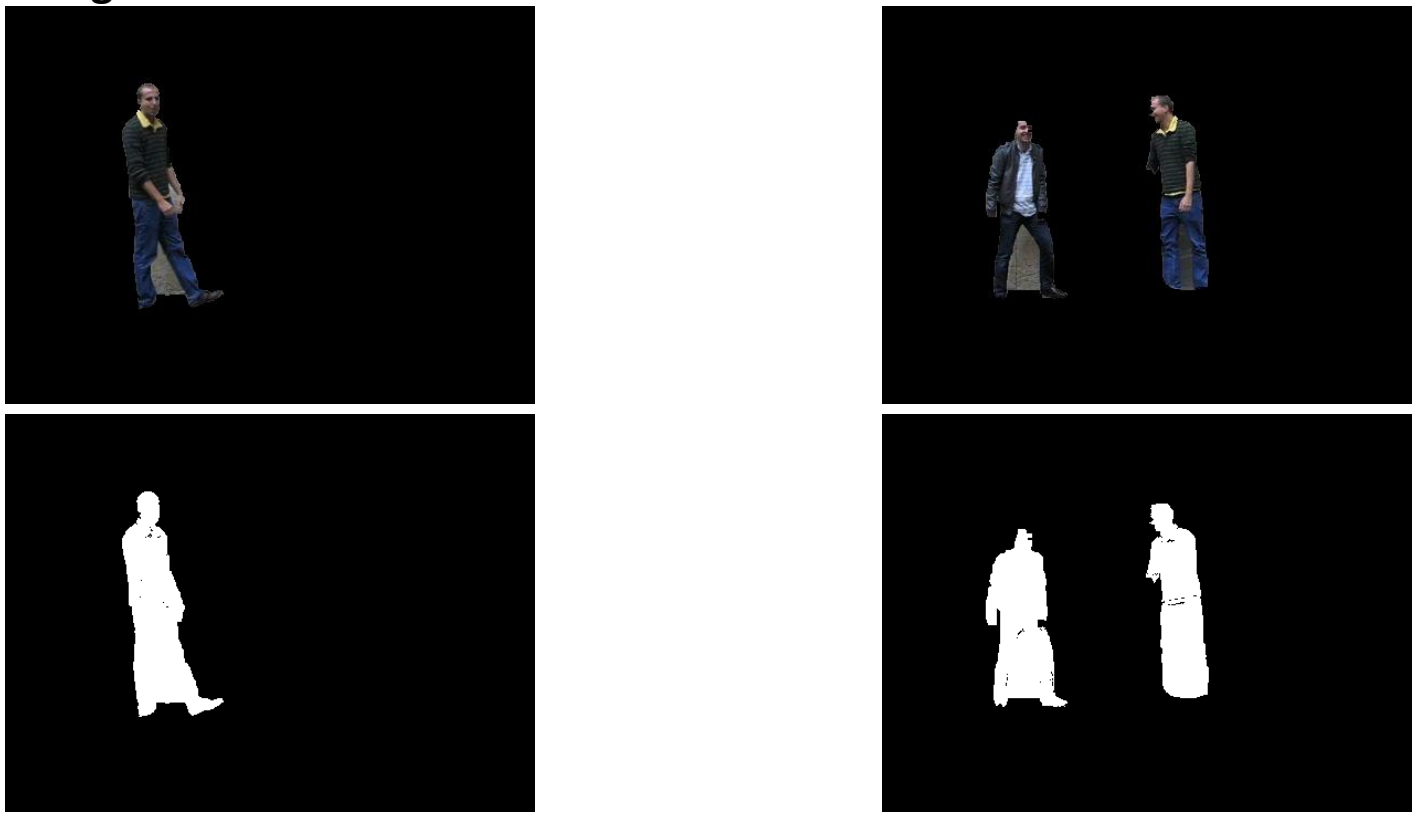


- We omit the classical interaction…

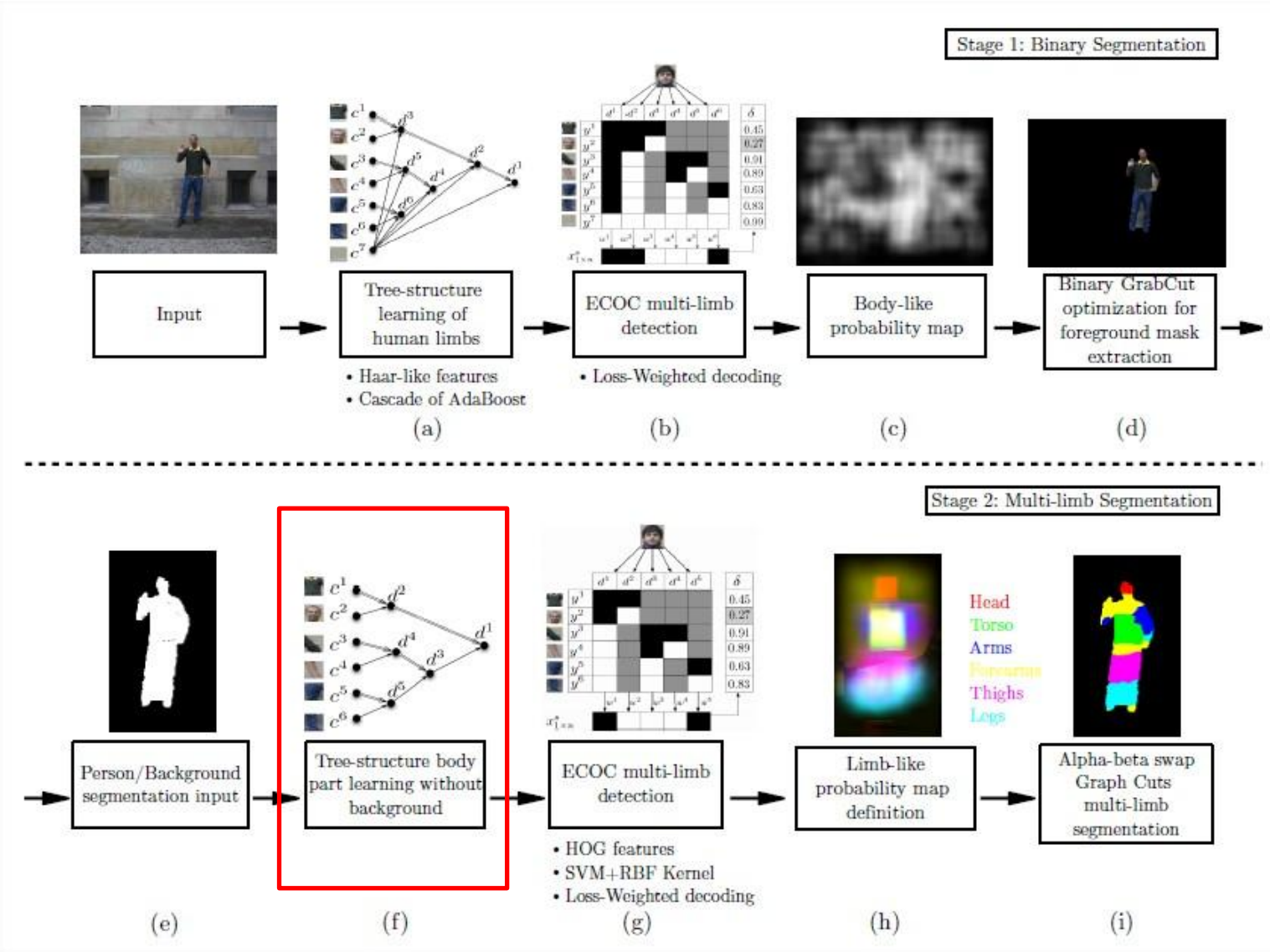
Yuri Y. Boykov and Marie-Pierre Jolly, “Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images”, International Conference on Computer Vision, 2001

GrabCut optimization for foreground mask extraction

- Binary segmentation by means of background and foreground segmentation.
  - **Background**
  - **Foreground**

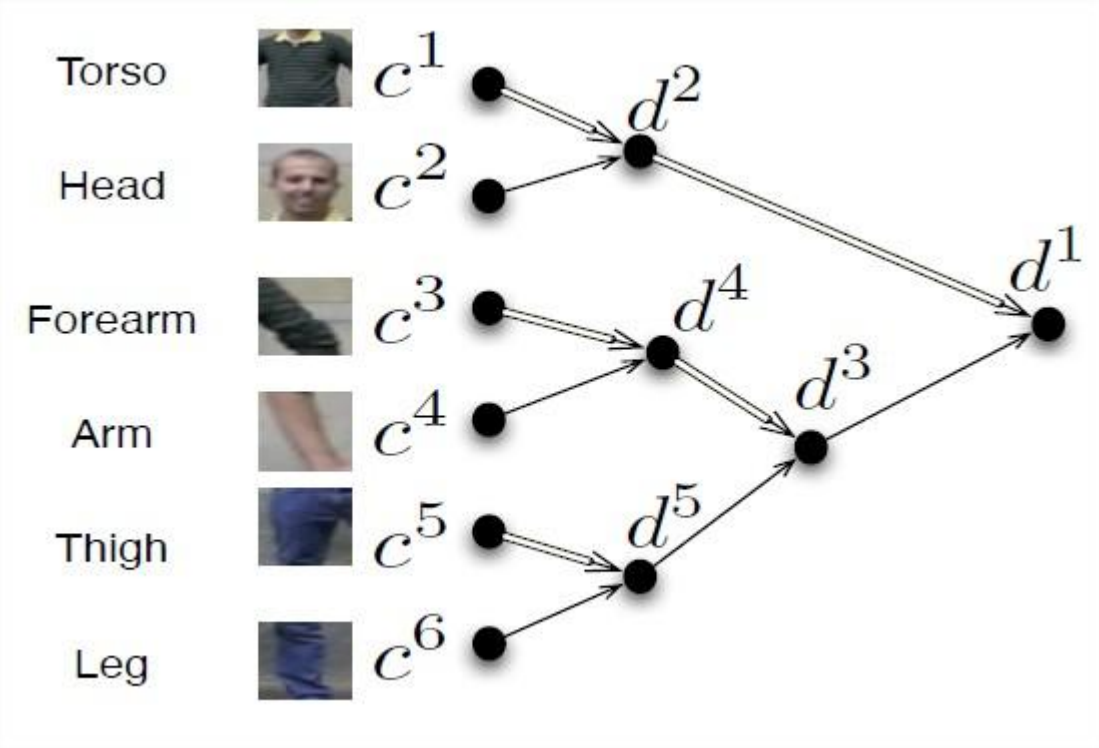


A. Hernandez-Vela, N. Zlateva, A. Marinov, M. Reyes, P. Radeva, D. Dimov, S. Escalera, Graph cuts optimization for multi-limb human segmentation in depth maps, in: CVPR, 2012, pp. 726-732.

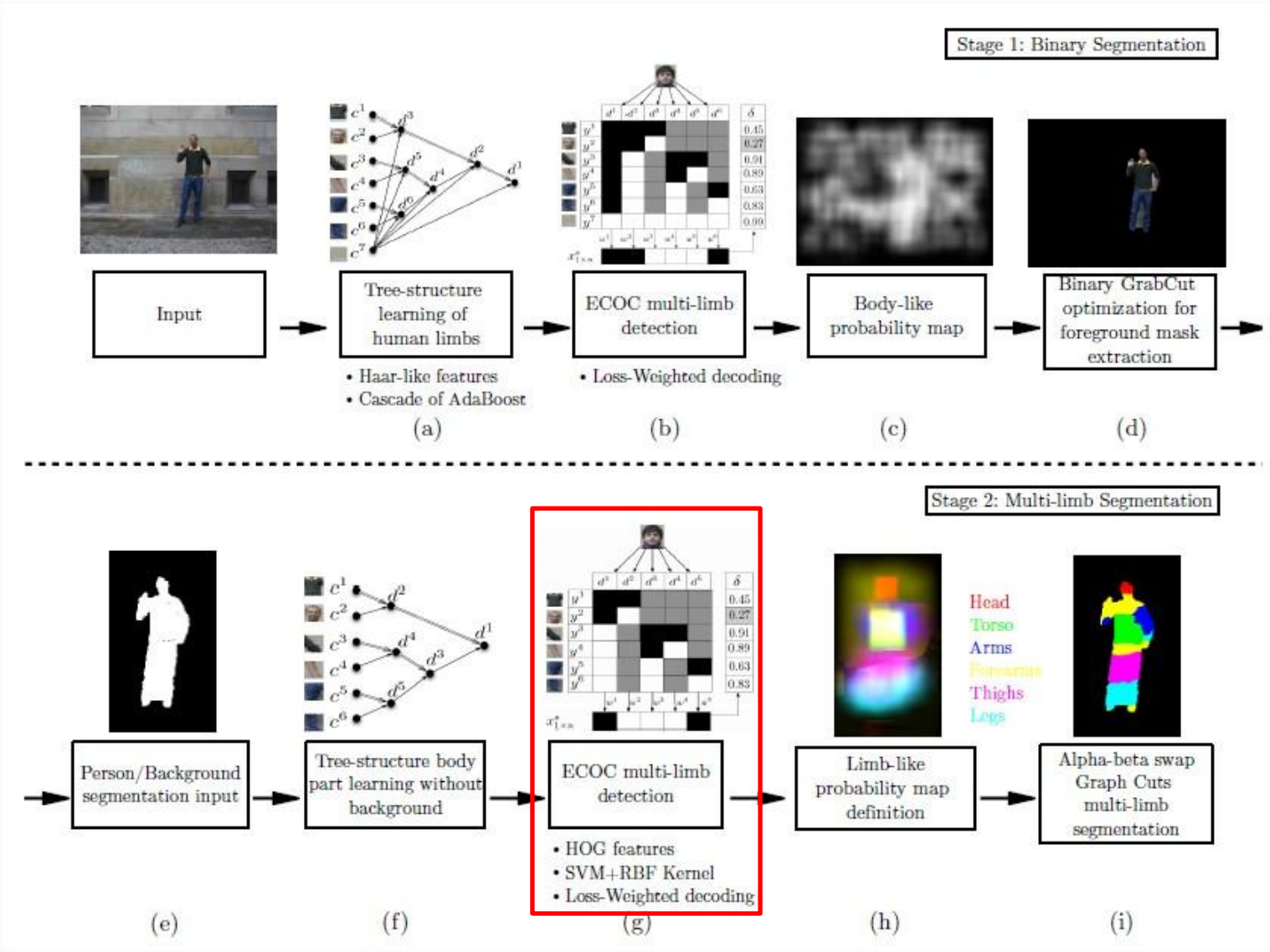

Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Head
Torso
Arms
Forearms
Thighs
Legs
Alpha-beta swap Graph Cuts multi-limb segmentation
(e)
(f)
(g)
(h)
(i)

Tree-structure body part learning without background

- Define the groups of limbs to be learnt by each individual cascade without background.

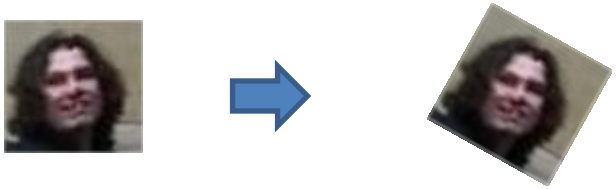


S. Escalera, O. Pujol, P. Radeva, On the decoding process in ternary error-correcting output codes, PAMI 32 (2010) 120-134.


Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Head
Torso
Arms
Forearms
Thighs
Legs
Alpha-beta swap Graph Cuts multi-limb segmentation
(e)
(f)
(g)
(h)
(i)

Body part learning using SVM classifiers

- Body parts rotational invariant by computing dominant orientation.

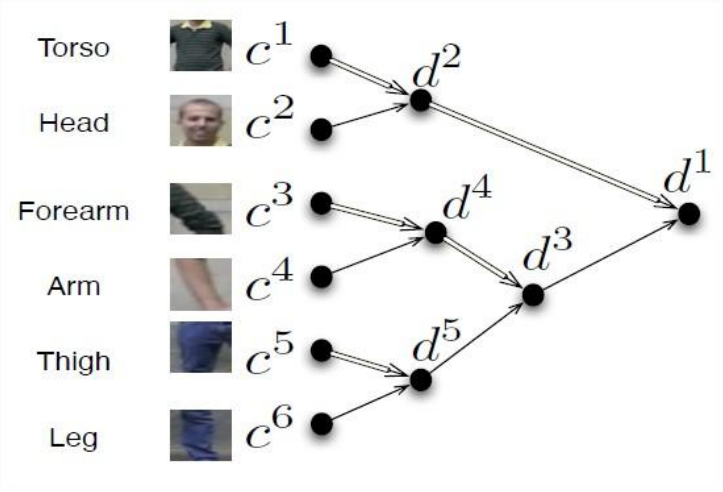


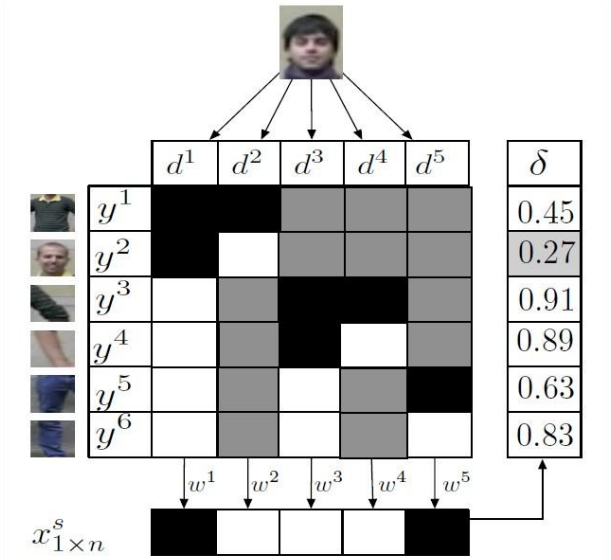
- HOG features describes those body parts.

- SVM classifiers with Generalized Gaussian RBF Kernel based on Chi-squared distance.

N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, volume 1, pages 886 {893 vol. 1, 2005.
Jianchao Yang, Kai Yu, Yihong Gong, and Thomas Huang. Linear spatial pyramid matching using sparse coding for image classication. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 1794{1801. IEEE, 2009.

- We propose to use a predefined coding matrix in which each dichotomy is obtained from the body part tree-structure without background.

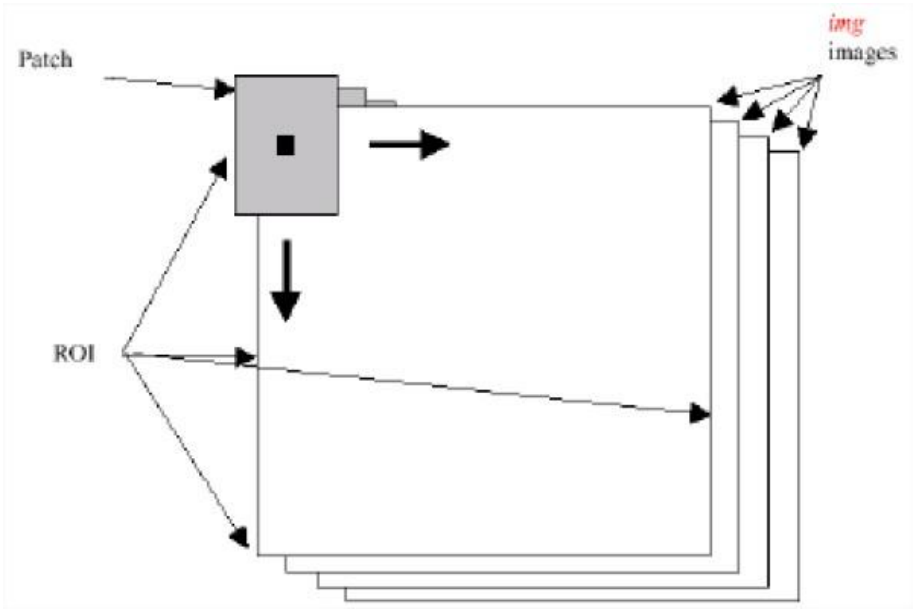


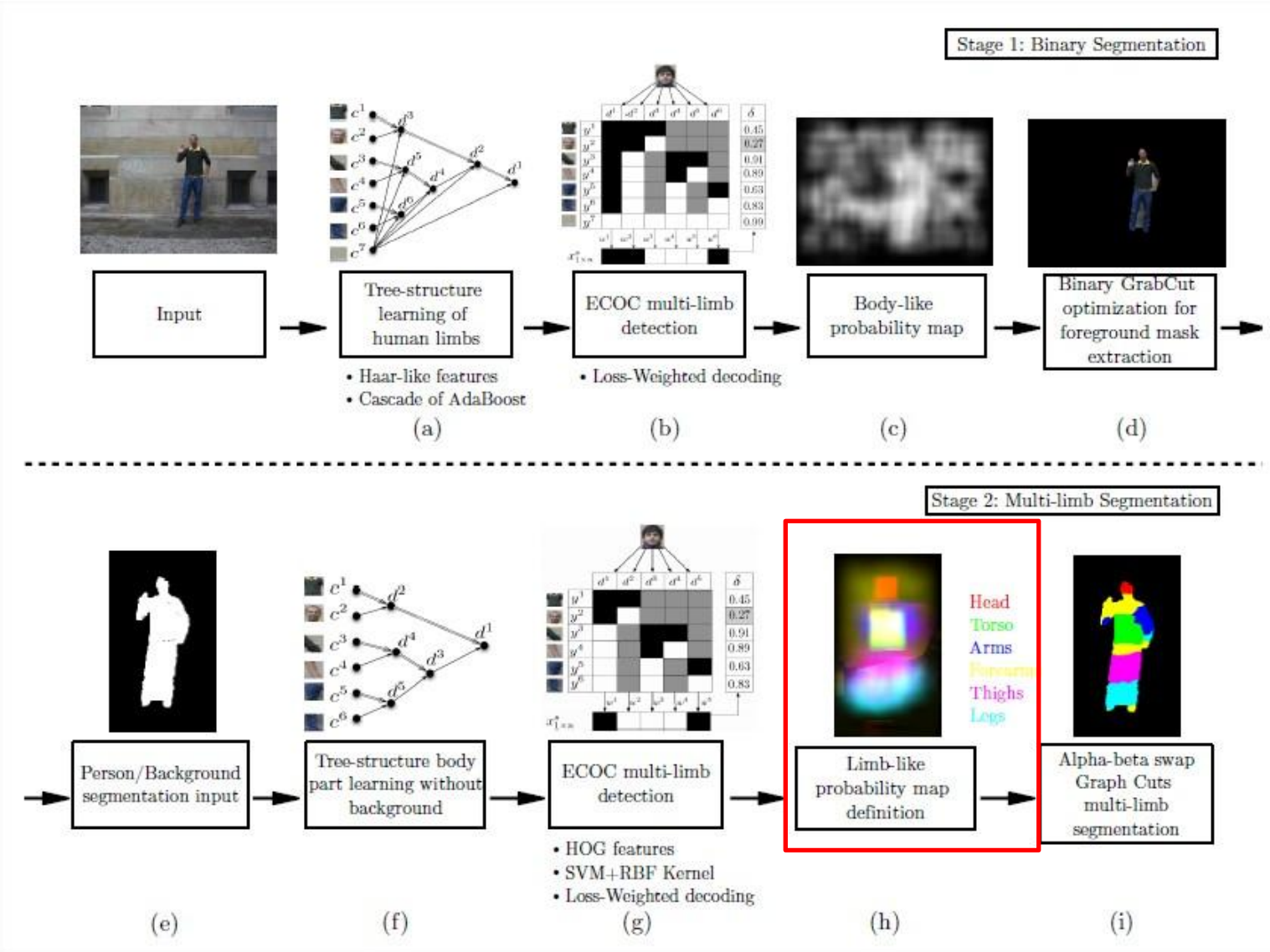


- Then, each cascade will give us its prediction and decoding ECOC step will be applied.
  - Loss-weighted decoding using cascade of classifier weights (takes into account classifiers performance)

S. Escalera, D. Tax, O. Pujol, P. Radeva, R. Duin, Subclass problem-dependent design of error-correcting output codes, PAMI 30 (6) (2008) 1-14.
M. A. Bautista, S. Escalera, X. Baro, P. Radeva, J. Vitria, O. Pujol, Minimal design of error-correcting output codes, Pattern Recogn. Lett. 33 (6) (2012) 693-702.

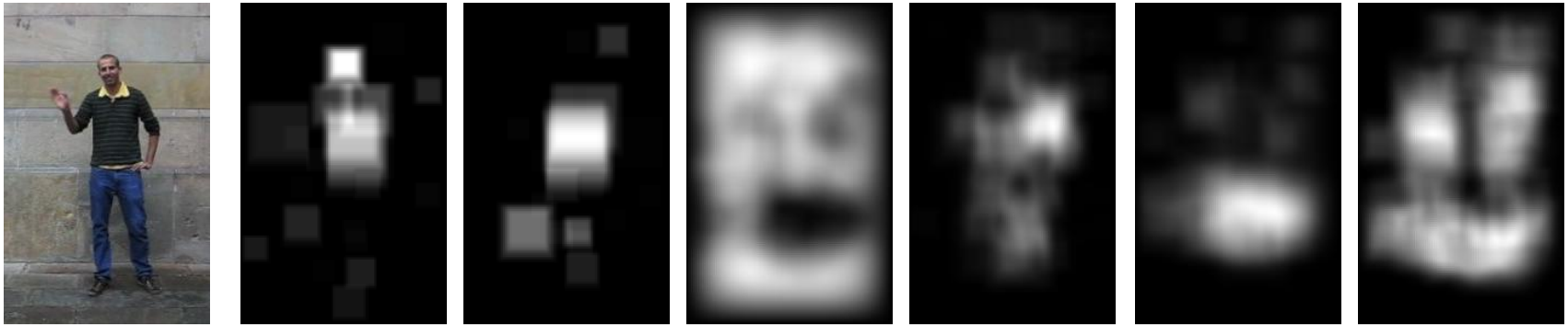
- In order to classify a new sample we apply a sliding window over the image considering the binary mask:




Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Head
Torso
Arms
Thighs
Alpha-beta swap Graph Cuts multi-limb segmentation
(e)
(f)
(g)
(h)
(i)

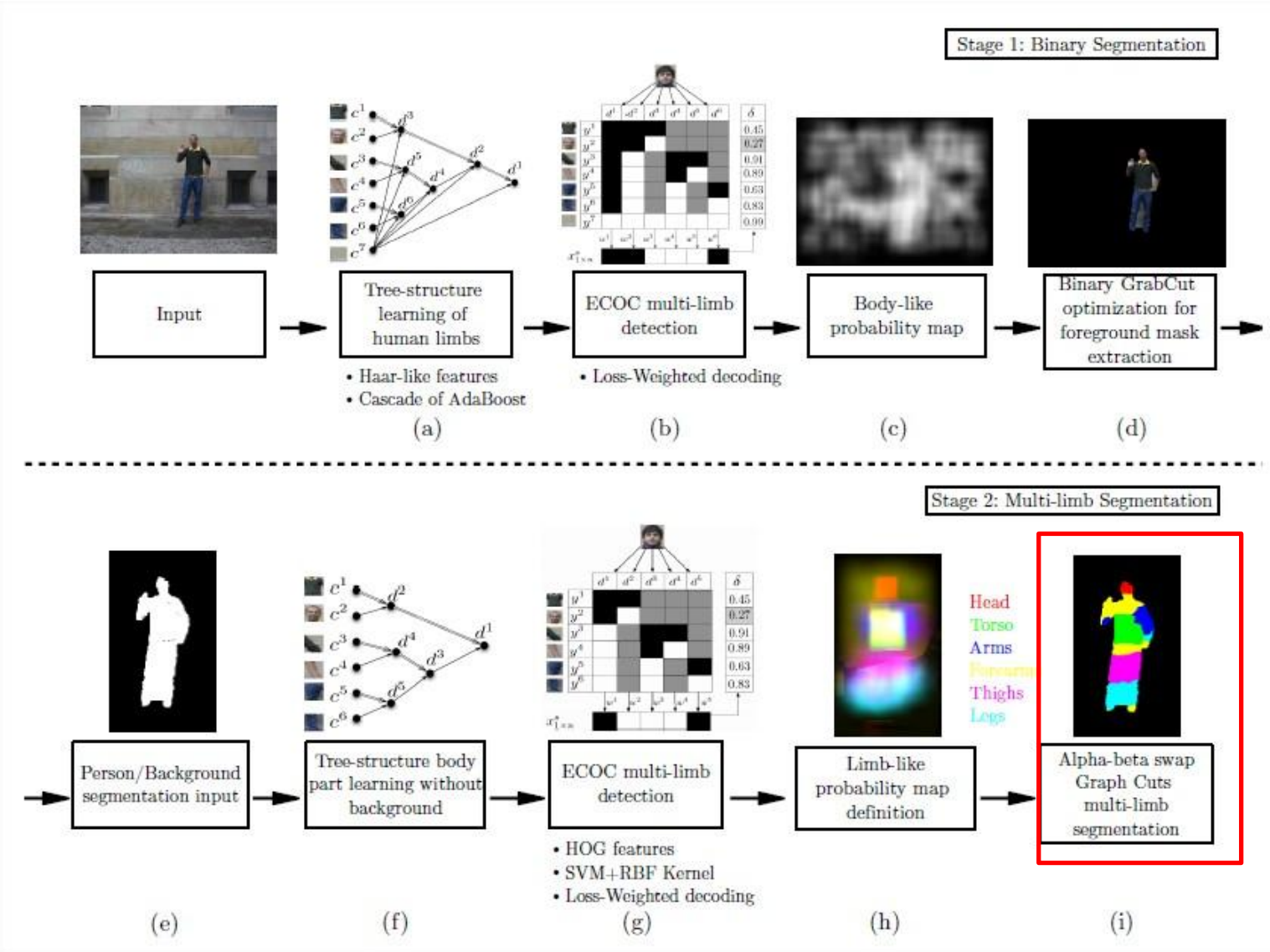
Limb-like probability map definition

- Limb-like probability maps $P_c^{pb} \epsilon [0,1]^{l \times w}$ are build



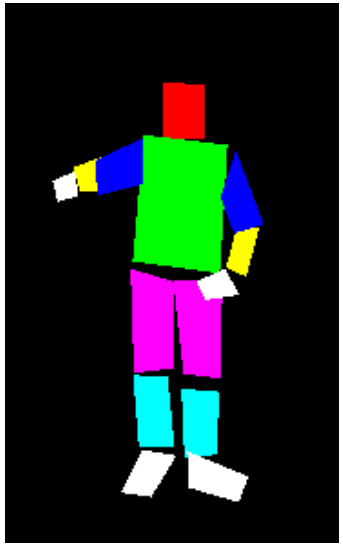
RGB Head Torso Arms Forearms Thighs Legs

- Haar-like based on AdaBoost gave us an accurate and efficient initialization of human regions for binary user segmentation.
- HOG-SVM is applied in a reduced region of the image, providing better estimates of human limb locations.


Stage 1: Binary Segmentation
Input
Tree-structure learning of human limbs
• Haar-like features
• Cascade of AdaBoost
ECOC multi-limb detection
• Loss-Weighted decoding
Body-like probability map
Binary GrabCut optimization for foreground mask extraction
(a)
(b)
(c)
(d)
Stage 2: Multi-limb Segmentation
Person/Background segmentation input
Tree-structure body part learning without background
ECOC multi-limb detection
• HOG features
• SVM+RBF Kernel
• Loss-Weighted decoding
Limb-like probability map definition
Head
Torso
Arms
Thighs
Alpha-beta swap Graph Cuts multi-limb segmentation
(e)
(f)
(g)
(h)
(i)

Alpha-beta swap Graph Cuts multi-limb segmentation

- Multi-label segmentation problem → Graph Cuts alpha-beta swap
- Segmentation by combining all pair labels $(\alpha_q, \alpha_m), \{m, q\} \in \{1, 2, ..., 6\}$
- A predefined pair-wise cost function $\Omega(c_q, c_m)$ penalizes relations between labels taking into account the natural constraints of the human limbs.

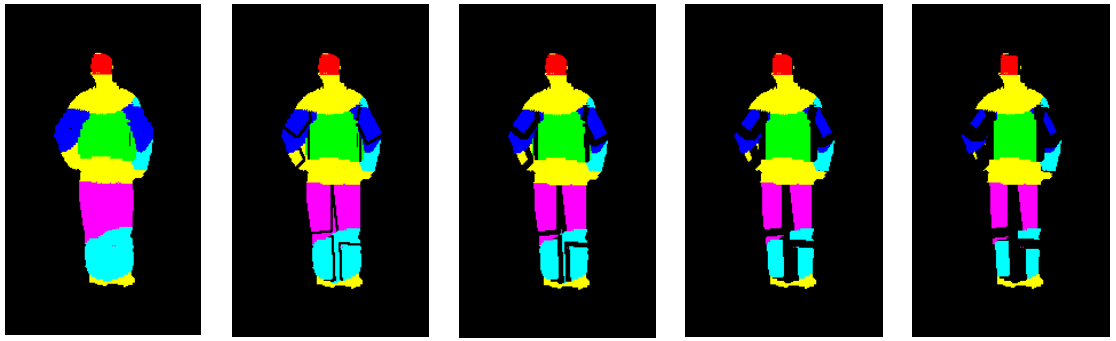


Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23(11):1222{1239, 2001.

- **DATA**
  - HuPBA 8k+: Combining by symmetry we obtain 6 limbs categories.
  - Limb regions are scale and rotational invariant for training.
  - Resize all limb sample to a 32x32 pixels region for computational purposes.
  - Limbs categories and samples: head (9.000), torso(9.000), left/right arms(14000), left/right forearms(15200), left/right thighs(16900), left/right legs(17200).
  - Action categories and samples: Wave(10), Point(13), Clap(15), Crouch(10), Jump(20), Walk(72), Run(17), Shake Hands(), Hug(18), Kiss(18), Fight(18).

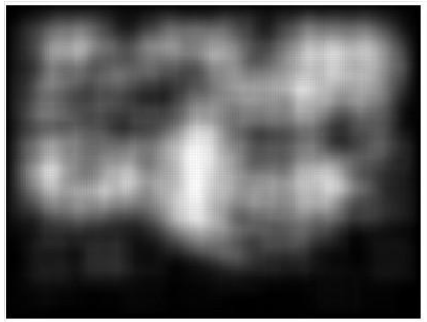
- **FIRST STAGE (Binary Segmentation)**
  - Haar-like fatures + AdaBoost: forced a 0.9 false positive rate and maximum of 0.4 false alarm rate during 8 stages.
  - Test: Sliding window approach with an initial patch size of 32x32 pixels up to 60x60 pixels.
  - Use of Graph Cut for binary segmentation tuned via cross-validation.
- **SECOND STAGE (Multi-limb Segmentation)**
  - HOG descriptor: 32x32 window size, 16x16 block size, 8x8 block stride, 8x8 cell size and 8 for number of bins.
  - SVMs with a Generalized Gaussian RBF kernel based on Chi-squared. The parameters of the kernel were tuned via cross-validation.
  - Model selection was done via a leave-one-sequence-out cross-svalidation.
  - Multi-limb segmentation: Alpha-beta GraphCut procedure, we set a 8x8 neighboring grid and tuned the λ parameter using cross-validation.

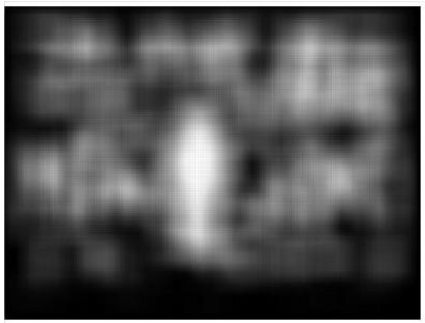
- **GESTURE RECOGNITION**
  - Feature vector of a frame: concatenation of the 6 limb-like probability maps, resizing each one of them to a 40x20 pixels region and vectorizing that region. Obtaining a final vector of d = 40x20x6 = 4800 dimensions, which is then reduced to d = 150 dimensions using a Random Projection.
  - Cost-threshold and the action/gesture model for both DTW experiments was obtained by cross-validation on training data, using a leave-one-sequence-out procedure.
  - Each HMM and its corresponding probability-threshold was obtained by cross-validation on training data, using a leave-one-sequence-out procedure.
- **EVALUATION:**
  - Jaccard Index overlapping, $J = \frac{A \bigcap B}{A \bigcup B}$
  - Do not care value

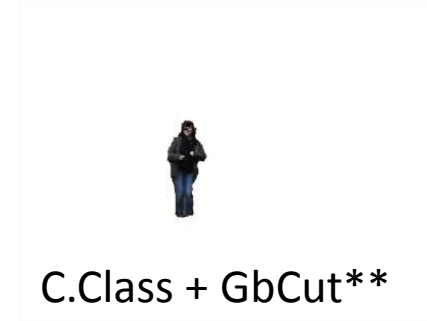


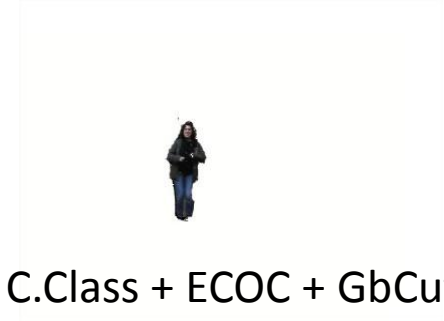
Results: Binary Segmentation

- We compare three methods:









P.Detector + GbCut* | C.Class + GbCut** | C.Class + ECOC + GbCut (our proposal)

P.Detector+GbCut	C.Class+GbCut	ECOC+GbCut
49.60 ± 20.45	58.26 ± 17.31	$\mathbf{61.79 \pm 14.02}$

* N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: CVPR, Vol. 1, 2005, pp. 886-893 vol. 1.

** P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: CVPR, Vol. 1, 2001.

Results: Multi-limb Segmentation

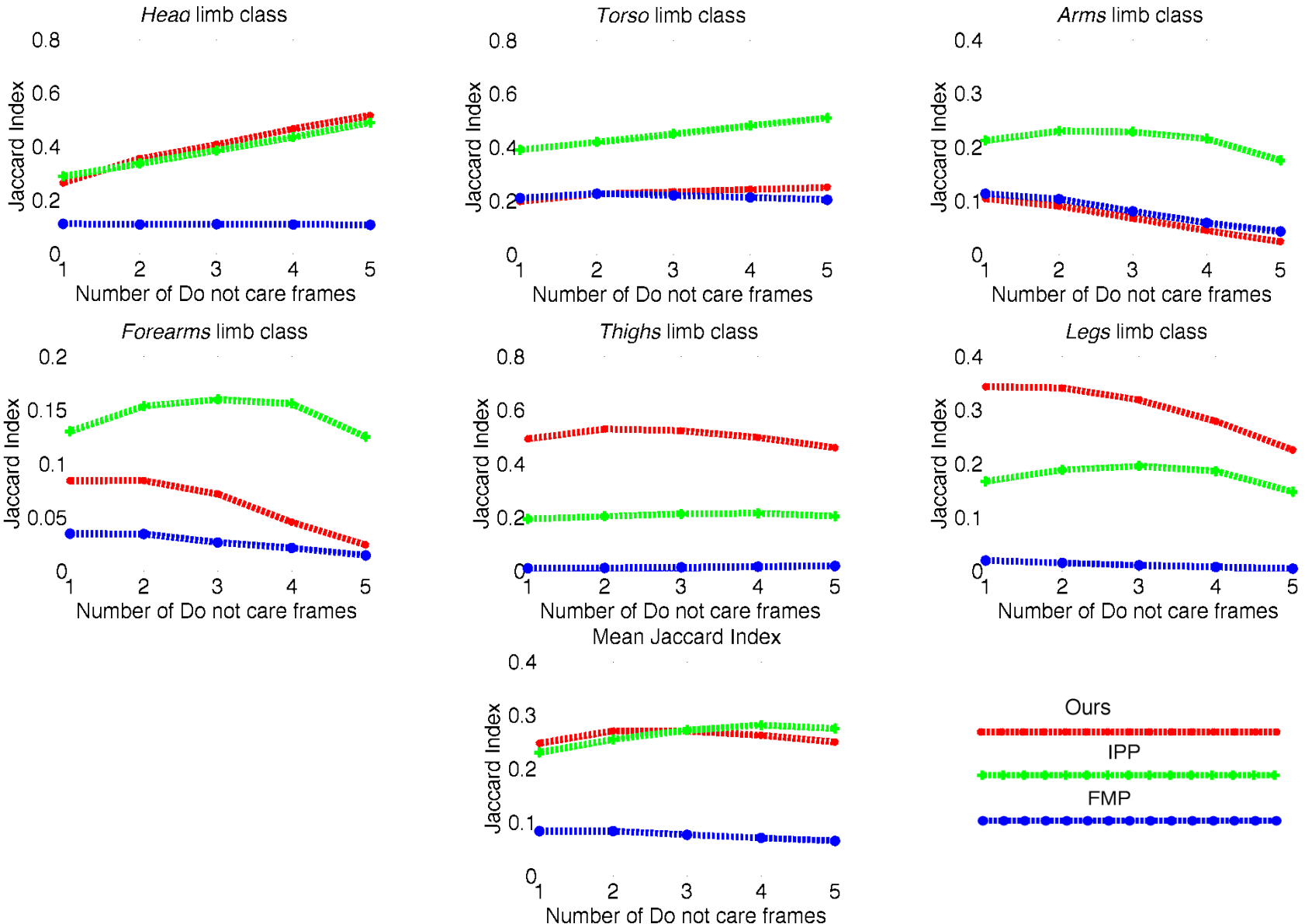
- We compare three methods:
- ECOC + GraphCut (our proposal)
- Flexible-mixture-parts(FMP)*
- Iterative Parsing Process(IPP)**

RGB GT ECOC + GraphCut FMP IPP

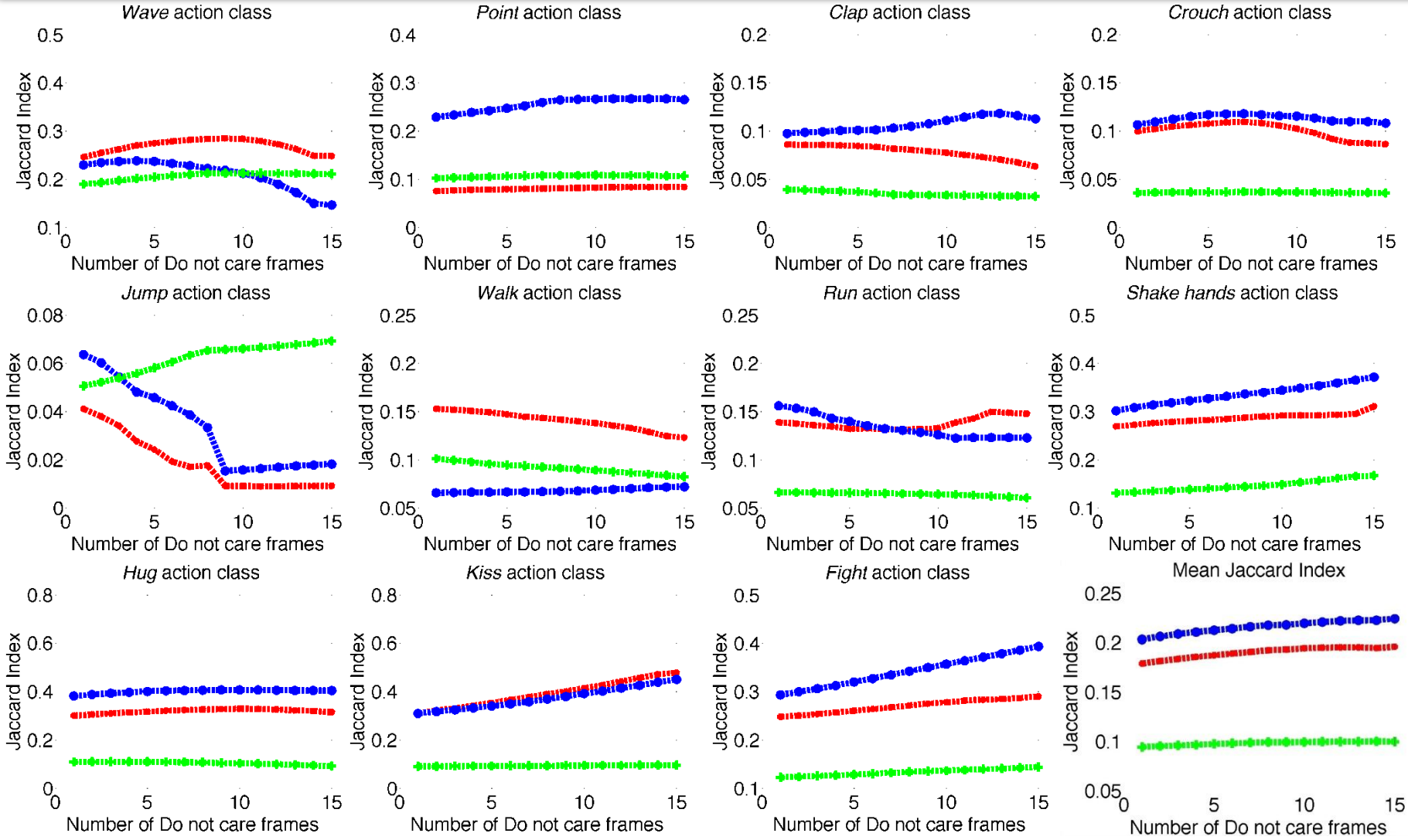
* Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1385{1392. IEEE, 2011.

** Deva Ramanan. Learning to parse images of articulated bodies. In Advances in neural information processing systems, pages 1129{1136, 2006.

Results: Multi-limb Segmentation



Results: Action Recognition



Conclusions

- We introduce the **HuPBA 8k+ dataset, the largest RGB labeled dataset of human limbs**, with more than 120000 manually annotated limbs. The data set also includes frame-level annotation for 11 action/gesture categories.

- We propose a **two stage approach based on ECOC and Graph Cuts** for the segmentation of human limbs in RGB images.

- The proposed method is compared with state-of-the-art methods for human pose estimation obtaining very satisfying results.

- We provide with a baseline for Action Recognition in the novel dataset.

- Work published in IBPRIA 2013. Sánchez Daniel, Ortega Juan Carlos, Bautista Miguel Angel, Escalera Sergio. Human Body Segmentation with Multi-limb Error-Correcting Output Codes Detection and Graph Cuts Optimization. In *Proceedings of IbPRIA,* page 50--58, year 2013.

- Selected from IBPRIA and extended version submitted to Neurocomputing journal.

- The novel data set is currently used for the ChaLearn posture-gesture recognition challenge and workshop at European Conference on Computer Vision 2014 by the HuPBA group of the University of Barcelona.

ECOC and Graph Cuts Segmentation of Human Limbs

Thank you!

Author: Daniel Sánchez

Advisors: Sergio Escalera and Miguel Ángel Bautista