Detector of Facial Landmarks Learned by the Structured Output SVM

Michal Uřičář, Vojtěch Franc and Václav Hlaváč

Department of Cybernetics, Faculty Electrical Engineering
Czech Technical University in Prague
Center for Machine Perception

The first two authors were supported by EC project FP7-ICT-247525 HUMAVIPS. The second author was also supported by EC project PERG04-GA-2008-239455 SEMISOL. The last author was supported by the Czech Ministry of Education project 1M0567.

{uricamic, xfrancv, hlavac}@cmp.felk.cvut.cz

February 22, 2012

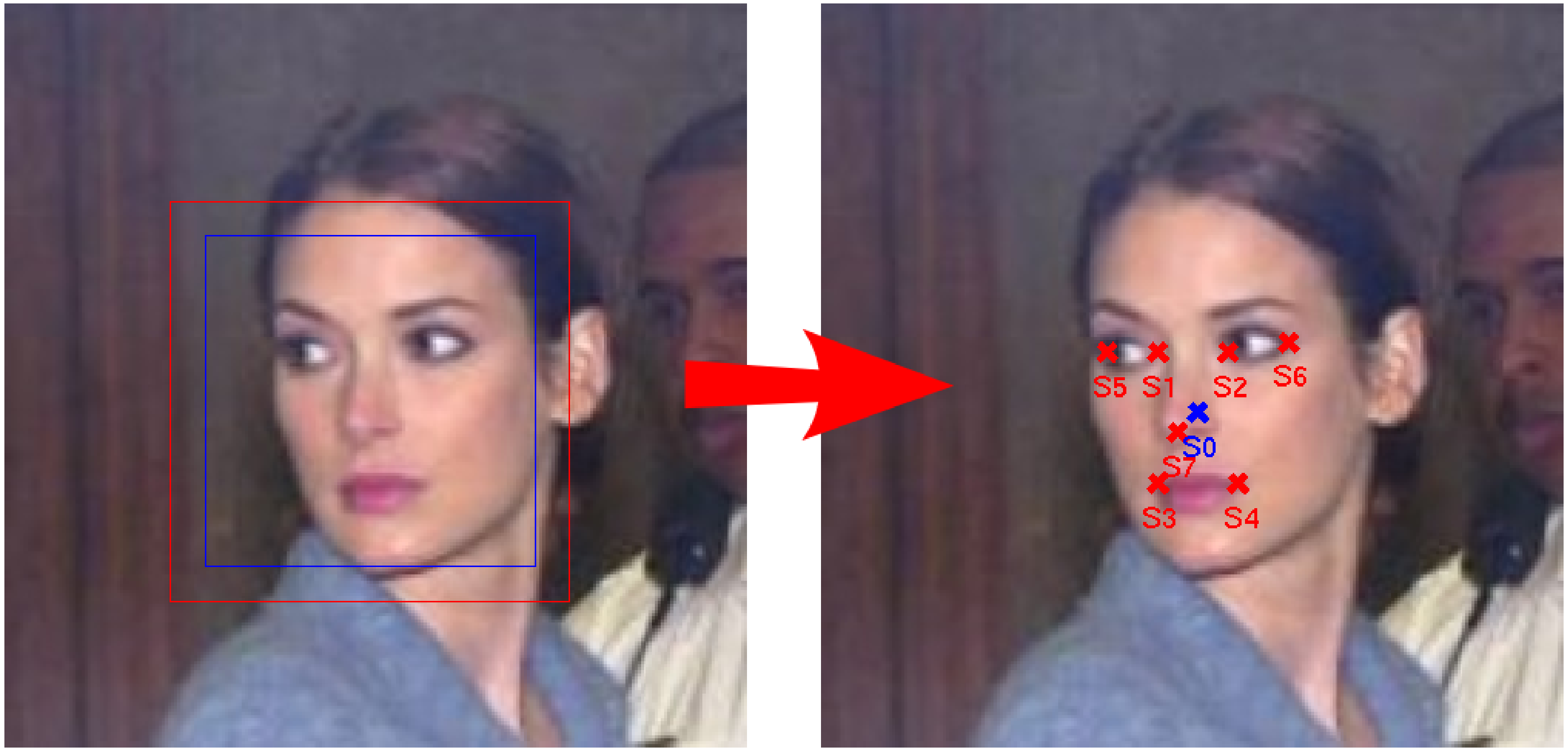
Outline

- Motivation
- Structured output classifier for facial landmark detection
- Large margin approach to structured output learning
- Competing methods
- Evaluation procedure
- Results
- Demo
- Conclusions
- References

Motivation

Face registration

- Essential part of face recognition (identity, gender, . . .)
- Quality of recognition and further processing depends on the quality of registration

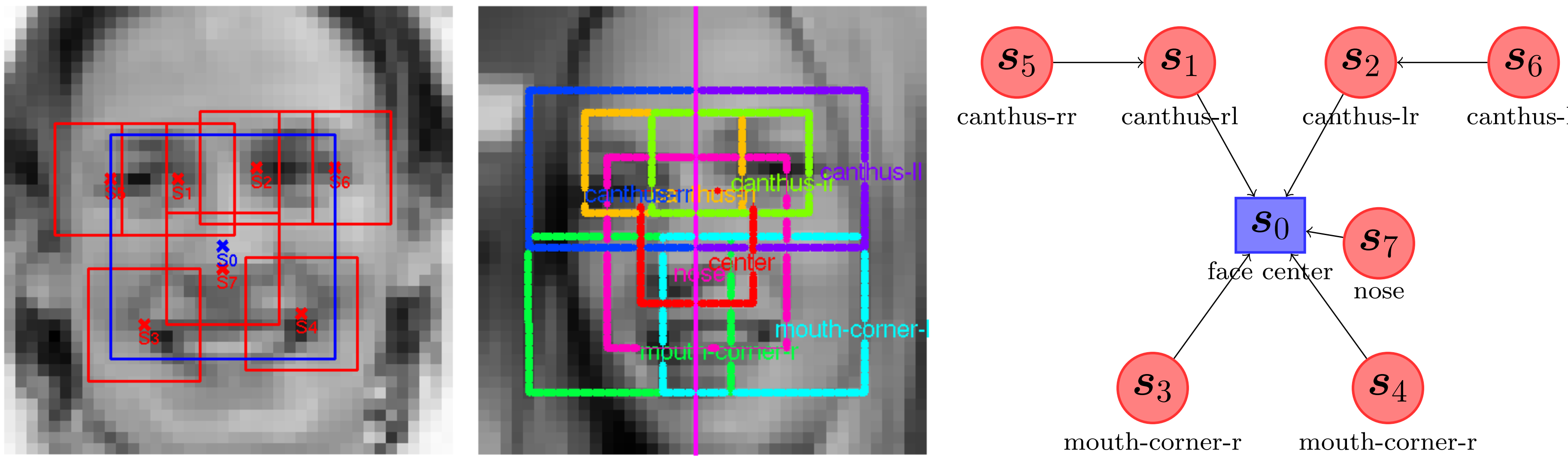


(a) Example of detected landmarks

Structured output classifier for facial landmark detection

Input: image I (normalized image frame), M components $\boldsymbol{s} = (\boldsymbol{s}_0, \ldots, \boldsymbol{s}_{M-1})$ and search space $\mathcal{S} = (\boldsymbol{S}_0 \times \cdots \times \boldsymbol{S}_{M-1})$.

Output: estimation of landmarks positions $\hat{\boldsymbol{s}}$.



(b) Landmarks & Components (c) Search spaces (d) Graph constraints

Configuration of M landmarks is described by graph $G = (V, E)$.

The landmark positions are estimated from image I by maximizing the cost function:

$$f(I, \boldsymbol{s}) = \sum_{i=0}^{M-1} q_i(I, \boldsymbol{s}_i) + \sum_{i,j \in E} g_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \left\langle \boldsymbol{w}, \boldsymbol{\Psi}(I, \boldsymbol{s}) \right\rangle$$

Structured output classifier for facial landmark detection

The cost function consist of two terms:

- appearance model $q_i(I, \mathbf{s}_i) = \left\langle \mathbf{w}_i^q, \mathbf{\Psi}_i^q(I, \mathbf{s}_i) \right\rangle$, $\mathbf{\Psi}_i^q(I, \mathbf{s}_i)$ . . . LBP features.
- deformation cost $g_i(\mathbf{s}_0, \mathbf{s}_i) = \left\langle \mathbf{w}_i^g, \mathbf{\Psi}_i^g(\mathbf{s}_0, \mathbf{s}_i) \right\rangle$, $\mathbf{\Psi}_i^g(\mathbf{s}_0, \mathbf{s}_i)$ . . . displacement vector

Displacement vector [Felzenszwalb et al., 2009]:

$$\begin{aligned} \mathbf{\Psi}_i^g(\boldsymbol{s}_0, \boldsymbol{s}_i) &= (dx, dy, dx^2, dy^2) \\ (dx, dy) &= (x_i, y_i) - (x_0, y_0) \end{aligned}$$

The registration is estimated from image I by maximizing the cost function:

$$\hat{\mathbf{s}} \in \arg\max_{\mathbf{s} \in \mathcal{S}} f(I, \mathbf{s})$$

$$\hat{\boldsymbol{s}} = \boldsymbol{h}(I, \boldsymbol{w}) = \arg\max_{\boldsymbol{s} \in \mathcal{S}} \left[\sum_{i=0}^{M-1} q_i(I, \boldsymbol{s}_i; \boldsymbol{w}) + \sum_{i,j \in E} g_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{w}) \right]$$

Large margin approach to structured output learning [Tsochantaridis et al., 2005]

Joint parameter vector $\boldsymbol{w}$ is given by solving the convex minimization problem:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}\in\mathbb{R}^n} \left[\frac{\lambda}{2}\|\boldsymbol{w}\|^2 + R(\boldsymbol{w})\right],$$

where

$$R(\boldsymbol{w}) = \frac{1}{m}\sum_{i=1}^{m} \max_{\boldsymbol{s}\in\mathcal{S}} \left(L(\boldsymbol{s}^i, \boldsymbol{s}) + \langle \boldsymbol{w}, \boldsymbol{\Psi}(I^i, \boldsymbol{s})\rangle\right) - \frac{1}{m}\sum_{i=1}^{m} \langle \boldsymbol{w}, \boldsymbol{\Psi}(I^i, \boldsymbol{s}^i)\rangle \ .$$

- $R(\boldsymbol{w})$ . . . convex upper bound on the empirical risk.
- λ . . . regularization term — prevents overfitting (set experimentally during validation).

Loss function (arbitrary):

$$L(\boldsymbol{s}, \boldsymbol{s}^*) = \frac{1}{M}\sum_{j=0}^{M-1} \|\boldsymbol{s}_j - \boldsymbol{s}_j^*\|$$

- For solving optimization task, we use the BMRM solver [Teo et al., 2010].
- This method converges in $\mathcal{O}(1/\epsilon)$ steps to ϵ precision.

Competing methods

We compare our proposed detector with detectors based on the following approaches:

- **Active Appearance Models (AAM)** [Cootes et al., 2001]

  Detector based on slightly modified publicly available MATLAB code [Kroon, 2010].

- **Independently trained binary SVMs**

  This detector is formed by binary (i.e. standard two-class) SVM classifiers trained independently for each facial landmark.

  For training we use the SVM solver implemented in LIBOCAS [Franc and Sonnenburg, 2010].

- **Detector of Everingham**

  Publicly available DPM based detector [Everingham et al., 2008].

  Strong competitor.

LFW face database [Huang et al., 2007]

- 13233 annotated face images
- Dimension of images 250×250 px
- 10 landmarks (centers of eyes, canthi, mouth, mouth corners and nose)
- Contains people of various ethnicity

Example images from LFW face database

Active Appearance Models

- Competing method — detector build on AAM [Cootes et al., 2001].
- Publicly available code [Kroon, 2010].
- Need for a different training database.

Training database for AAM

- IMM Face database [Nordstrøm et al., 2004], 240 faces (6 images per person).
- Dimension of images 640×480 px.
- 58 manually annotated points on each face, poor variance in ethnicity.
- Expensive data

Examples of faces from IMM database with annotation

Evaluation procedure

Partitioning of the LFW database:

Data set	Training	Validation	Testing
Percentage	60%	20%	20%
# of examples	6,919	2,307	2,316

Evaluation procedure:

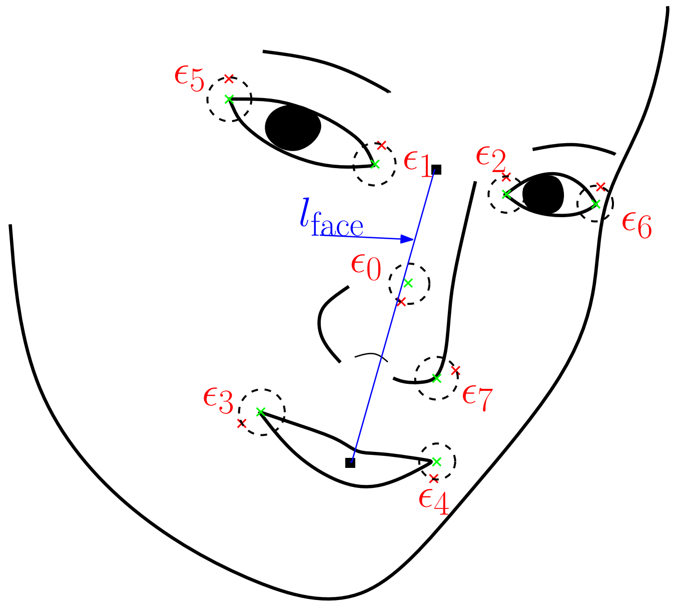
- Find $\boldsymbol{w}$ for $\lambda \in \Lambda = \{10, 1, 10^{-1}, 10^{-2}, 10^{-3}\}$ on TRN set.
- Select optimal $\lambda^* = \arg\min_{\lambda\in\Lambda} R_{\text{VAL}}\big(\boldsymbol{w}(\lambda)\big)$ according to validation risk:

$$R_{\text{VAL}}\big(\boldsymbol{w}(\lambda)\big) = \frac{1}{p}\sum_{i=1}^{p} L(\boldsymbol{s}^i, \hat{\boldsymbol{s}}^i) \quad \text{where}$$

$$\hat{\boldsymbol{s}}^i = \arg\max_{\boldsymbol{s}\in\mathcal{S}} \big\langle \boldsymbol{w}(\lambda), \boldsymbol{\Psi}(I^i, \boldsymbol{s})\big\rangle$$

- Compute the test risk on the TST set

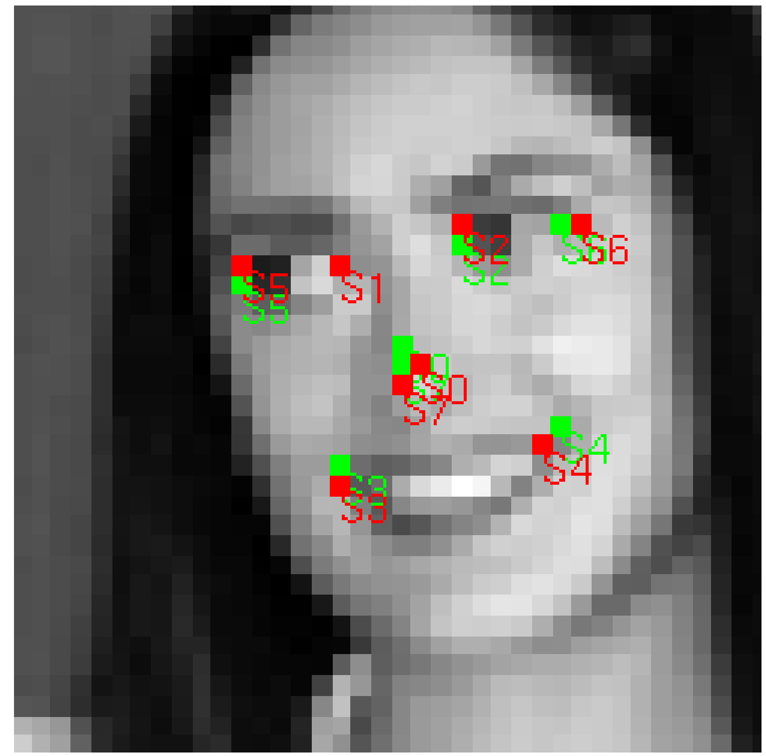
Performance measurements



$$L(\mathbf{s},\hat{\mathbf{s}}) = \frac{1}{l_{\text{face}}}\frac{\epsilon_0+\cdots+\epsilon_8}{8}$$

$$L^{\max}(\mathbf{s},\hat{\mathbf{s}}) = \frac{1}{l_{\text{face}}}\max\{\epsilon_0,\ldots,\epsilon_8\}$$

(e) Illustration of accuracy statistics.

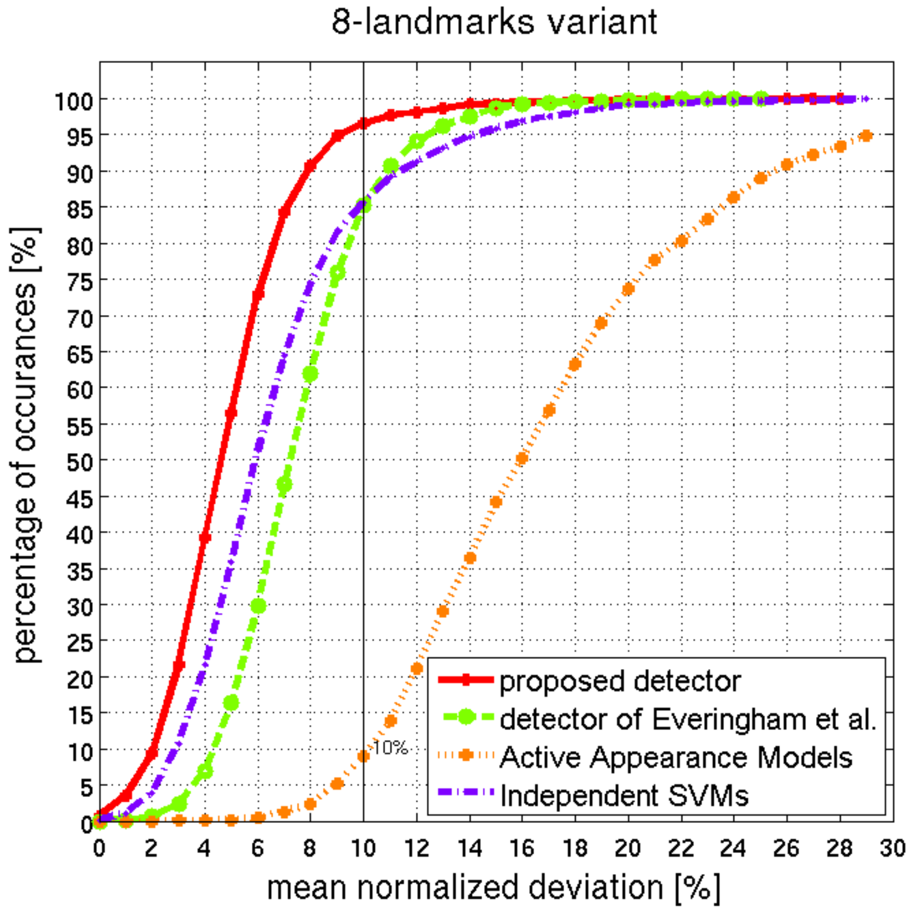


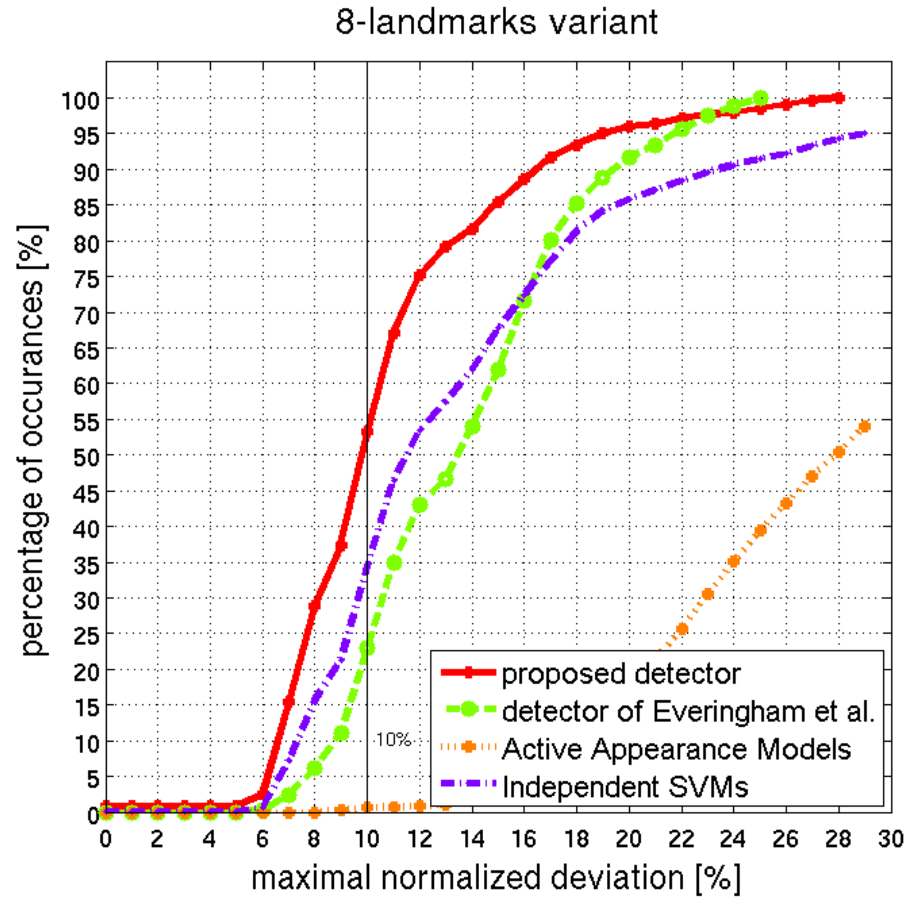
(f) Example detection with mean normalized deviation equal to 10%

Results:

	R_{TST}	$R_{\text{TST}}^{\max}$
AAM	17.60	31.27
Independent SVMs	7.20	18.36
Everingham et. al.	8.00	15.95
proposed detector	**5.46**	**12.42**

Results





Detail around 10%:

AAM:	$8.98\ \%$	AAM:	$0.62\ \%$
Independently trained SVMs:	$85.66\ \%$	Independently trained SVMs:	$34.50\ \%$
Everingham:	$22.93\ \%$	Everingham:	$22.93\ \%$
proposed detector:	$\mathbf{96.59}\ \%$	**proposed detector**:	$\mathbf{53.23}\ \%$

Demo 1

CNN Anchorwoman (resolution 640×360 px).

Demo 2

Movie “In Bruges” (resolution 720×304 px).

Demo 3

Video captured by the head camera of humanoid robot NAO (resolution 320×240 px).

Demo 4

Live Demo

Conclusions

Main contribution

- DPM based facial landmark detector
- Structured Output classification with arbitrary loss function
- One-stage learning of the appearance model & deformation cost
- Performance evaluation and comparison with competing methods

flandmark

- Open-source library implementing proposed detector
- Written in C with interface to MATLAB
- Learning written solely in MATLAB (also part of flandmark)
- Real-time detection on a standard PC
- Demo applications with OpenCV
- Fully annotated LFW database
- homepage: `http://cmp.felk.cvut.cz/~uricamic/flandmark/`

References

[Cootes et al., 2001] Cootes, T., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. IEEE Trans. Pattern Analysis and Machine Intelligence, 23(6):681–685.

[Everingham et al., 2008] Everingham, M., Sivic, J., and Zisserman, A. (2008). Willow project, automatic naming of characters in tv video. MATLAB implementation, www: http://www.robots.ox.ac.uk/~vgg/research/nface/index.html.

[Felzenszwalb et al., 2009] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2009). Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 99(1).

[Franc and Sonnenburg, 2010] Franc, V. and Sonnenburg, S. (2010). Libocas — library implementing ocas solver for training linear svm classifiers from large-scale data. www: http://cmp.felk.cvut.cz/~xfrancv/ocas/html/index.html.

[Huang et al., 2007] Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst.

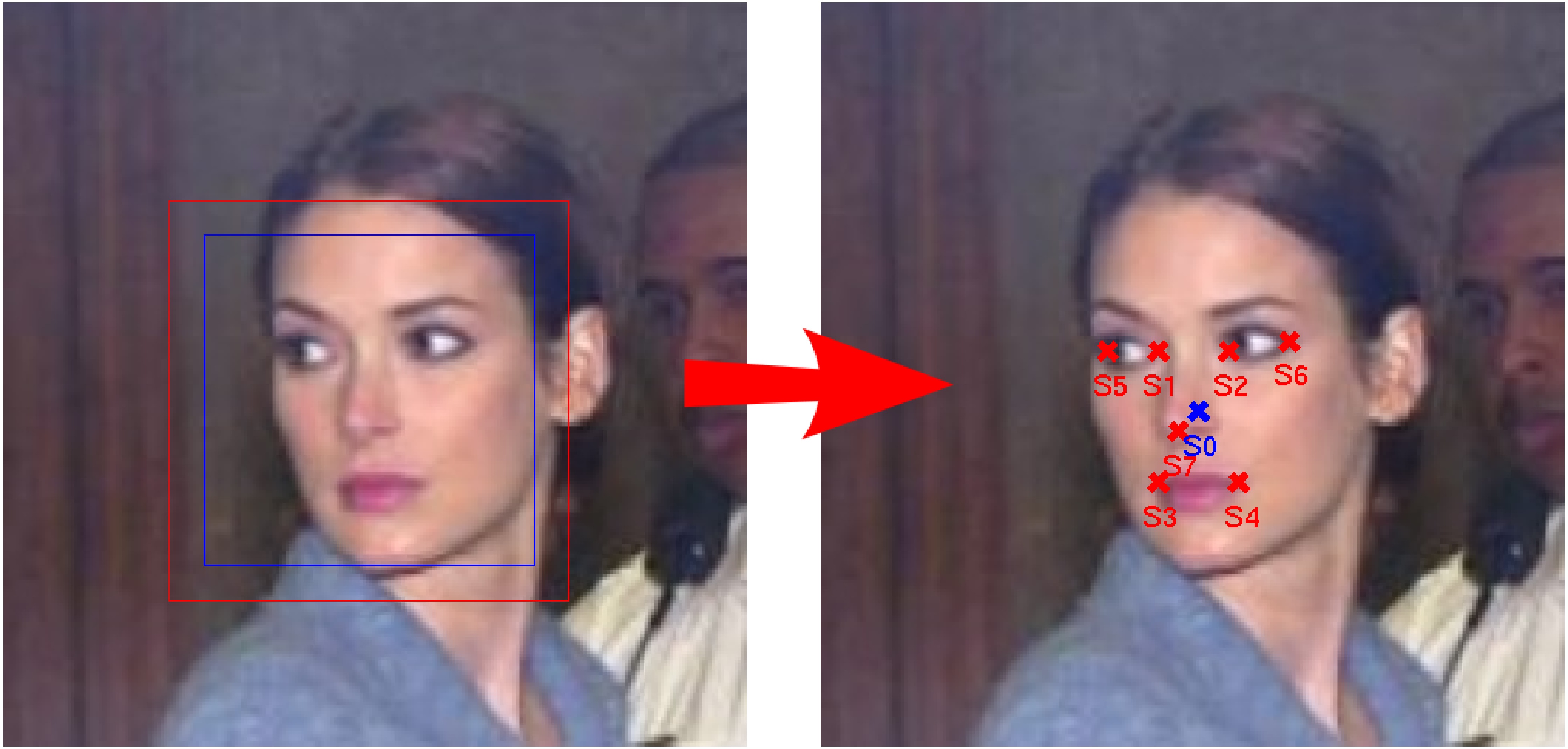
[Kroon, 2010] Kroon, D.-J. (2010). Active shape model (ASM) and active appearance model (AAM). MATLAB Central, www: http://www.mathworks.com/matlabcentral/fileexchange/26706-active-shape-model-asm-and-active-appearance--model-aam.

[Nordstrøm et al., 2004] Nordstrøm, M. M., Larsen, M., Sierakowski, J., and Stegmann, M. B. (2004). The IMM face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

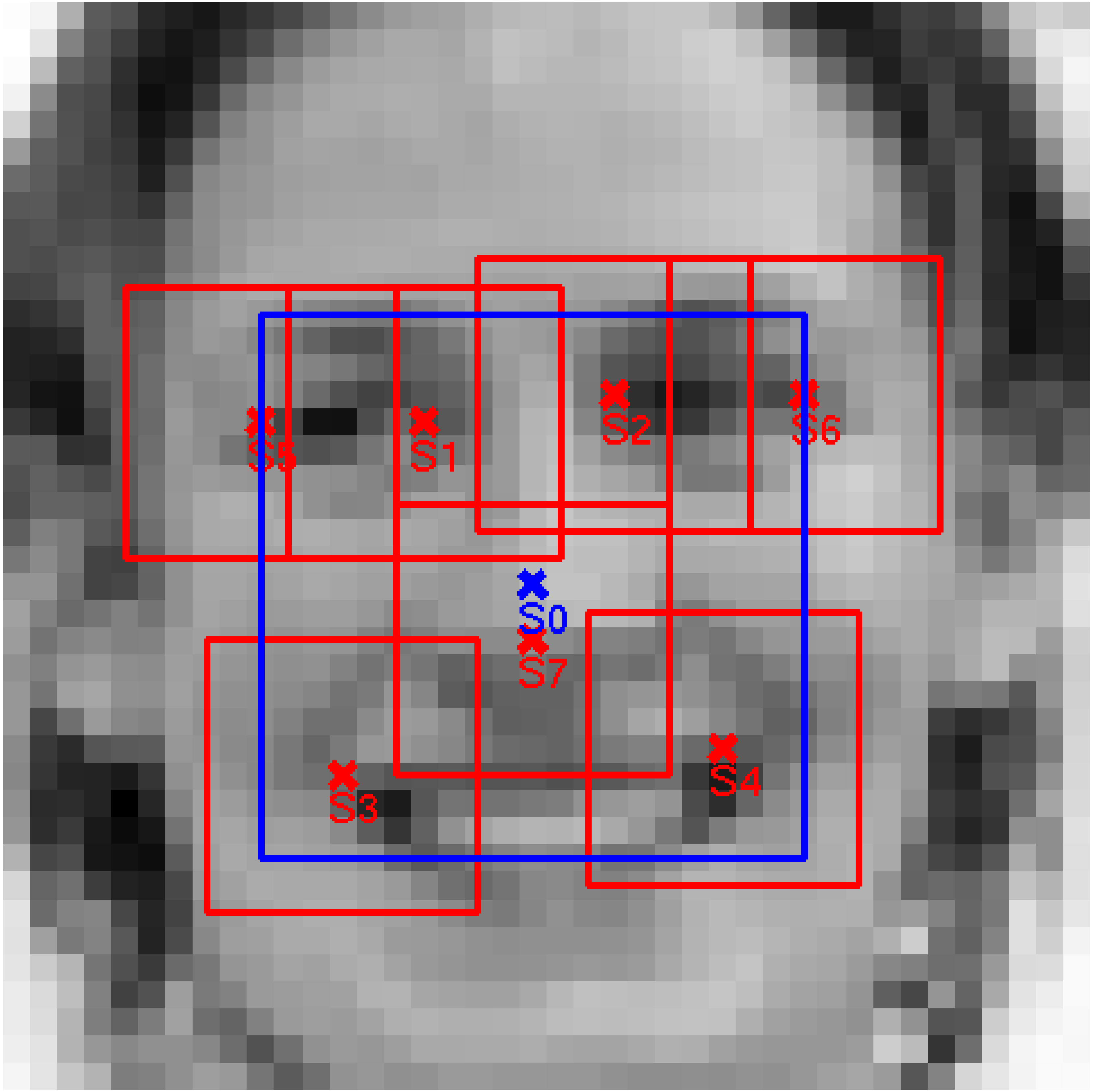
[Teo et al., 2010] Teo, C. H., Vishwanthan, S., Smola, A. J., and Le, Q. V. (2010). Bundle methods for regularized risk minimization. J. Mach. Learn. Res., 11:311–365.

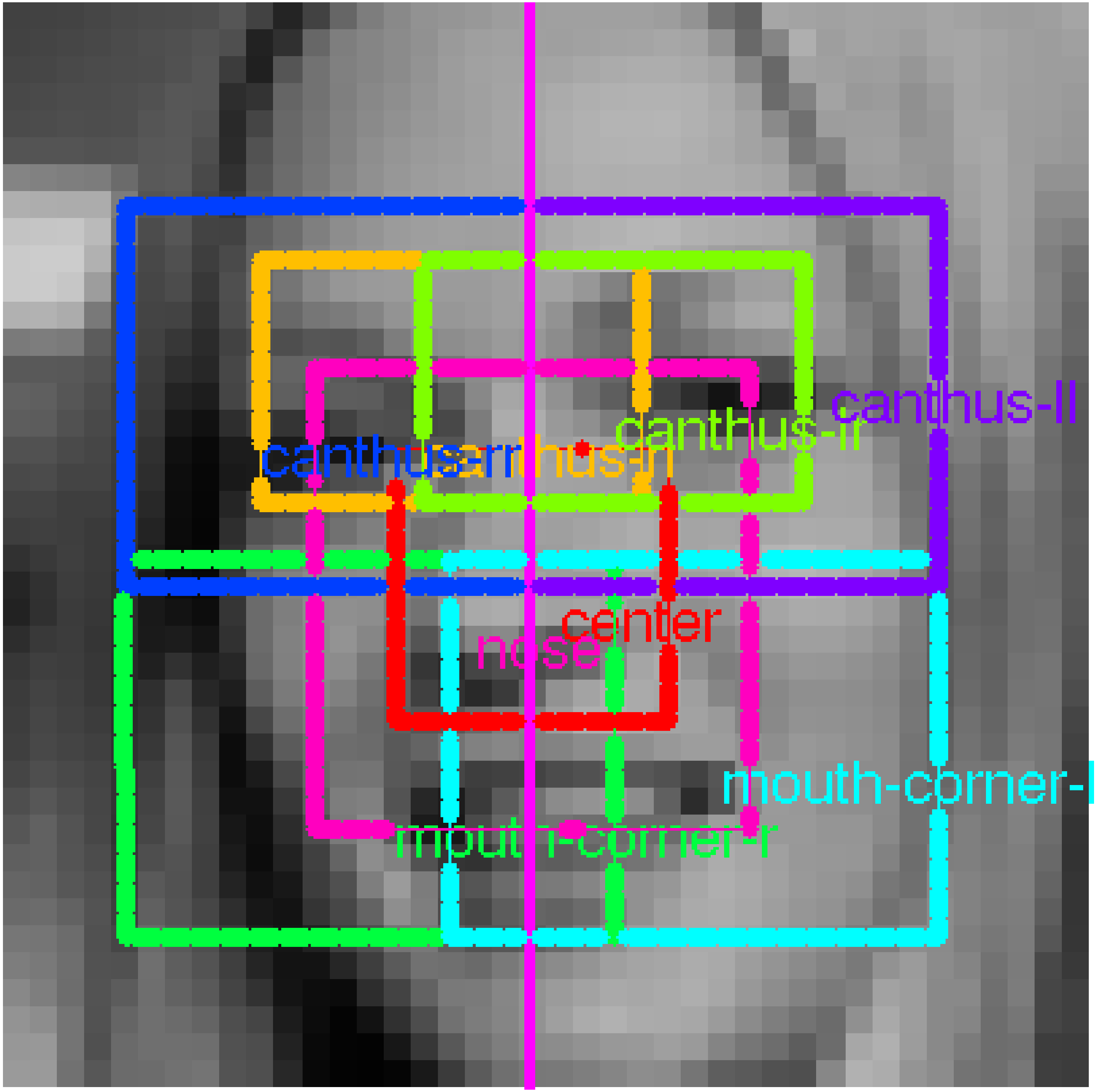
[Tsochantaridis et al., 2005] Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y., and Singer, Y. (2005). Large margin methods for structured and interdependent output variables. Journal of Machine Learning Research, 6:1453–1484.

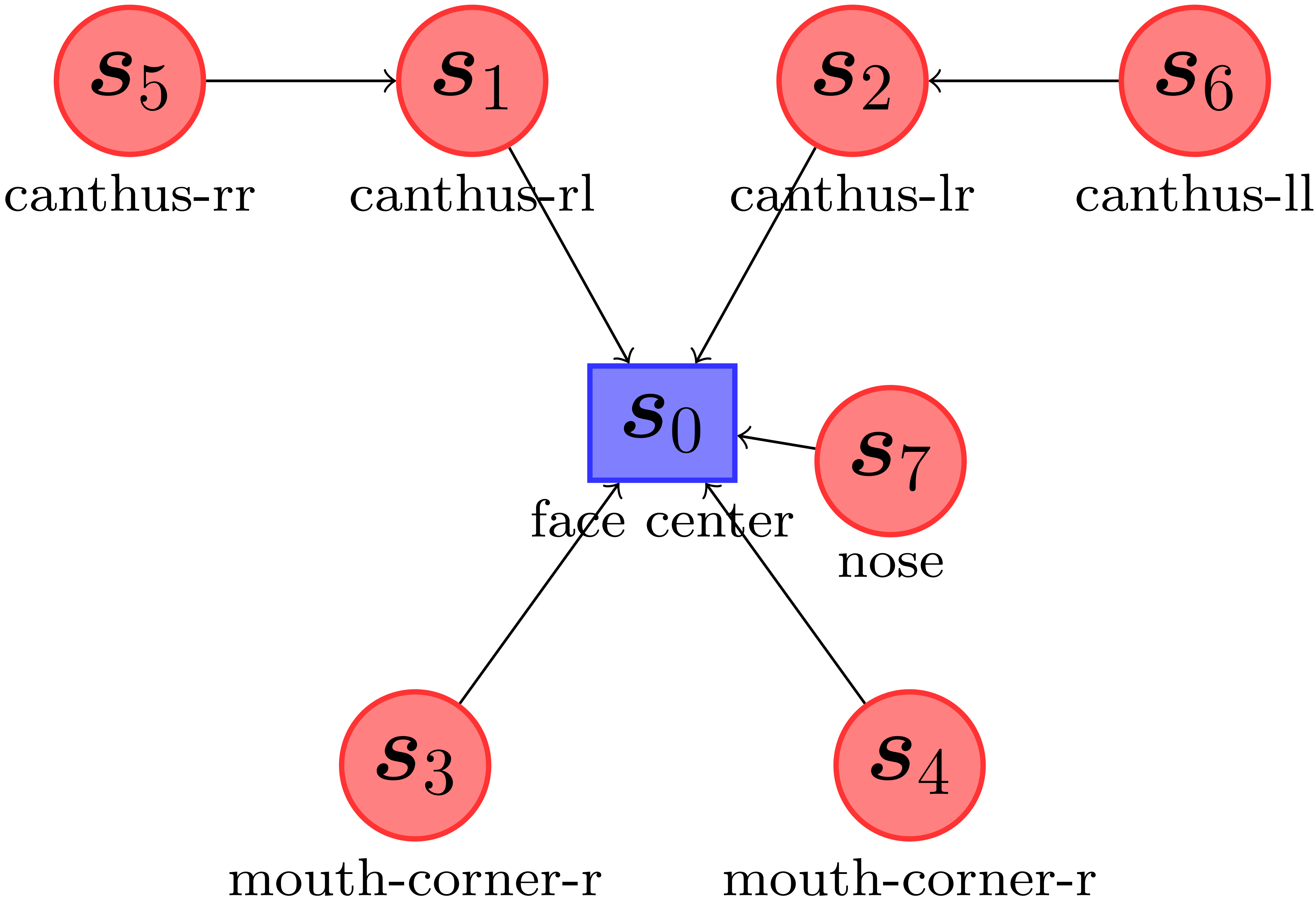
Thank You for your Attention



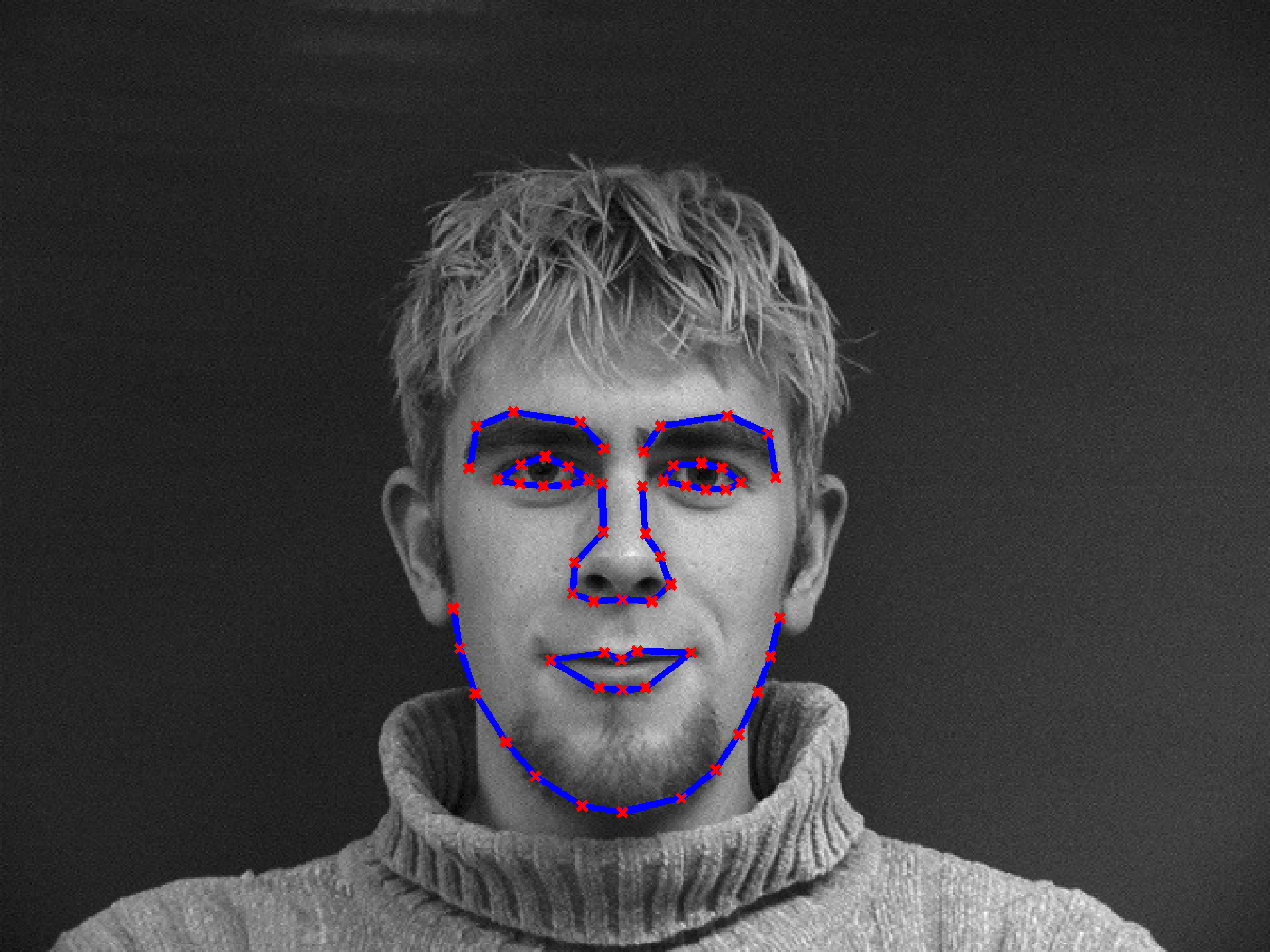
S5
S1
S2
S6
S0
S7
S3
S4

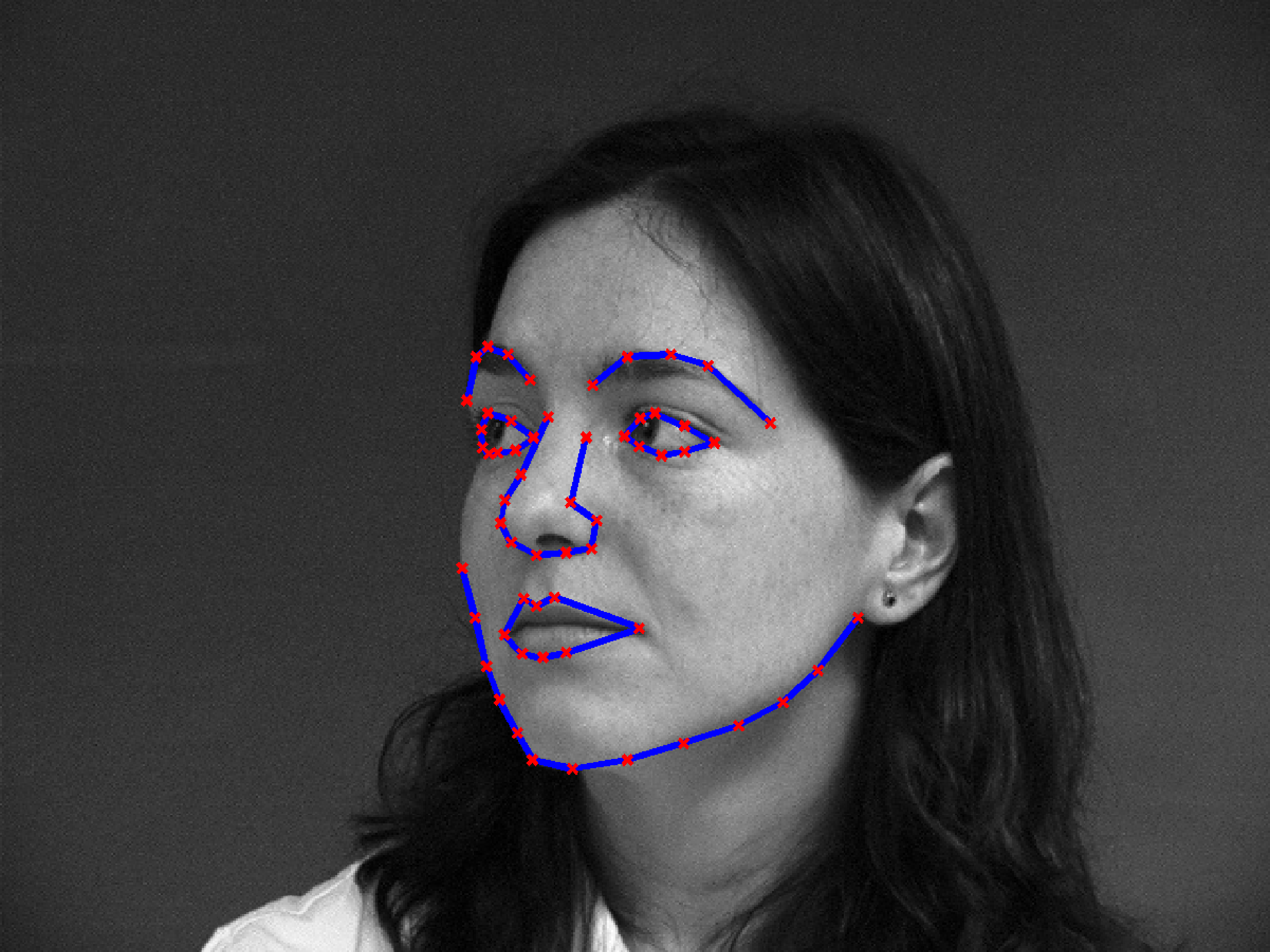

S5
S1
S2
S6
S0
S7
S3
S4

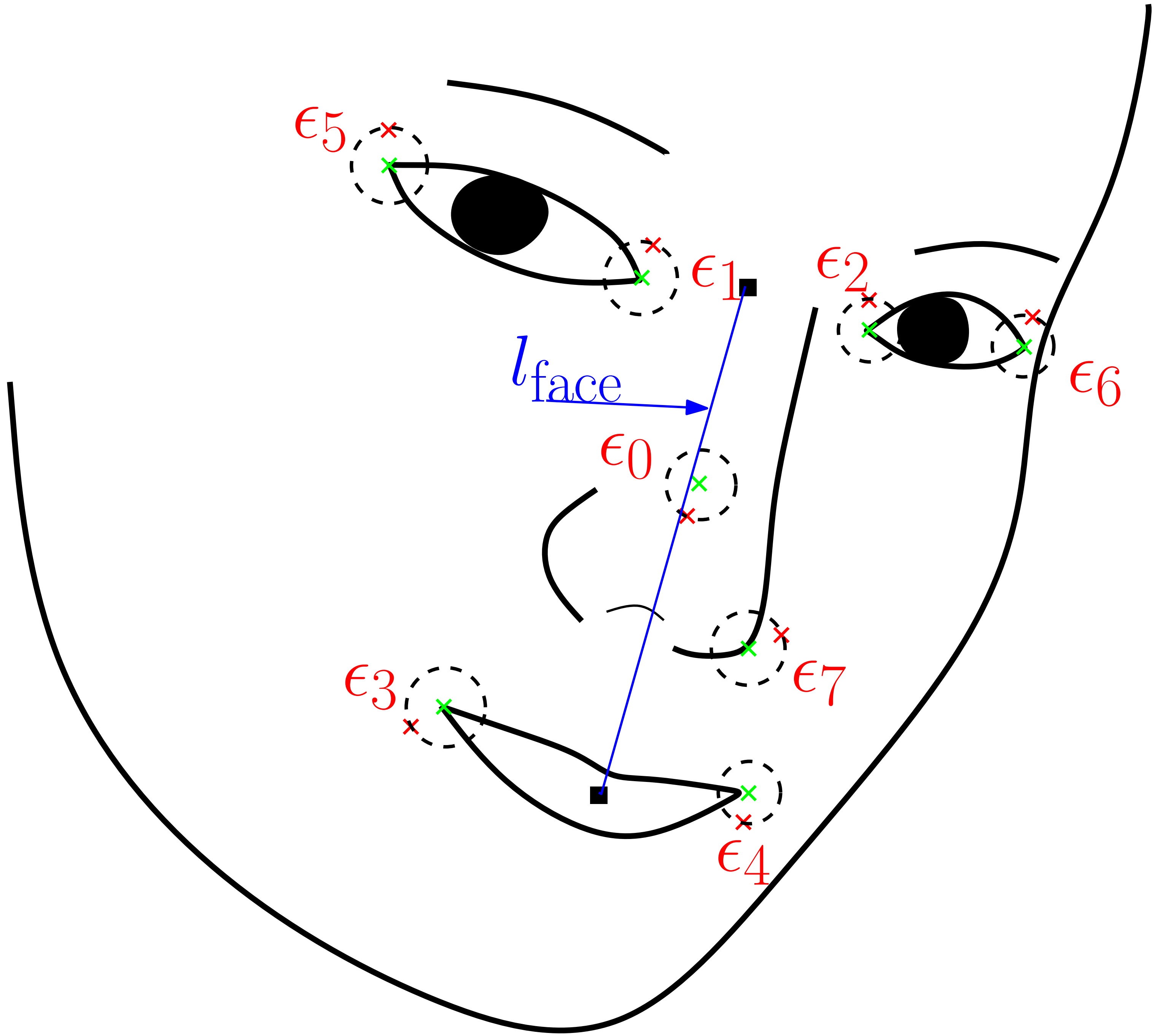

canthus-ll
canthus-lr
canthus-rr
canthus-rl
center
nose
mouth-corner-l
mouth-corner-r



$\boldsymbol{s}_5$
$\boldsymbol{s}_1$
$\boldsymbol{s}_2$
$\boldsymbol{s}_6$
canthus-rr
canthus-rl
canthus-lr
canthus-ll
$\boldsymbol{s}_0$
face center
$\boldsymbol{s}_7$
nose
$\boldsymbol{s}_3$
$\boldsymbol{s}_4$
mouth-corner-r
mouth-corner-r



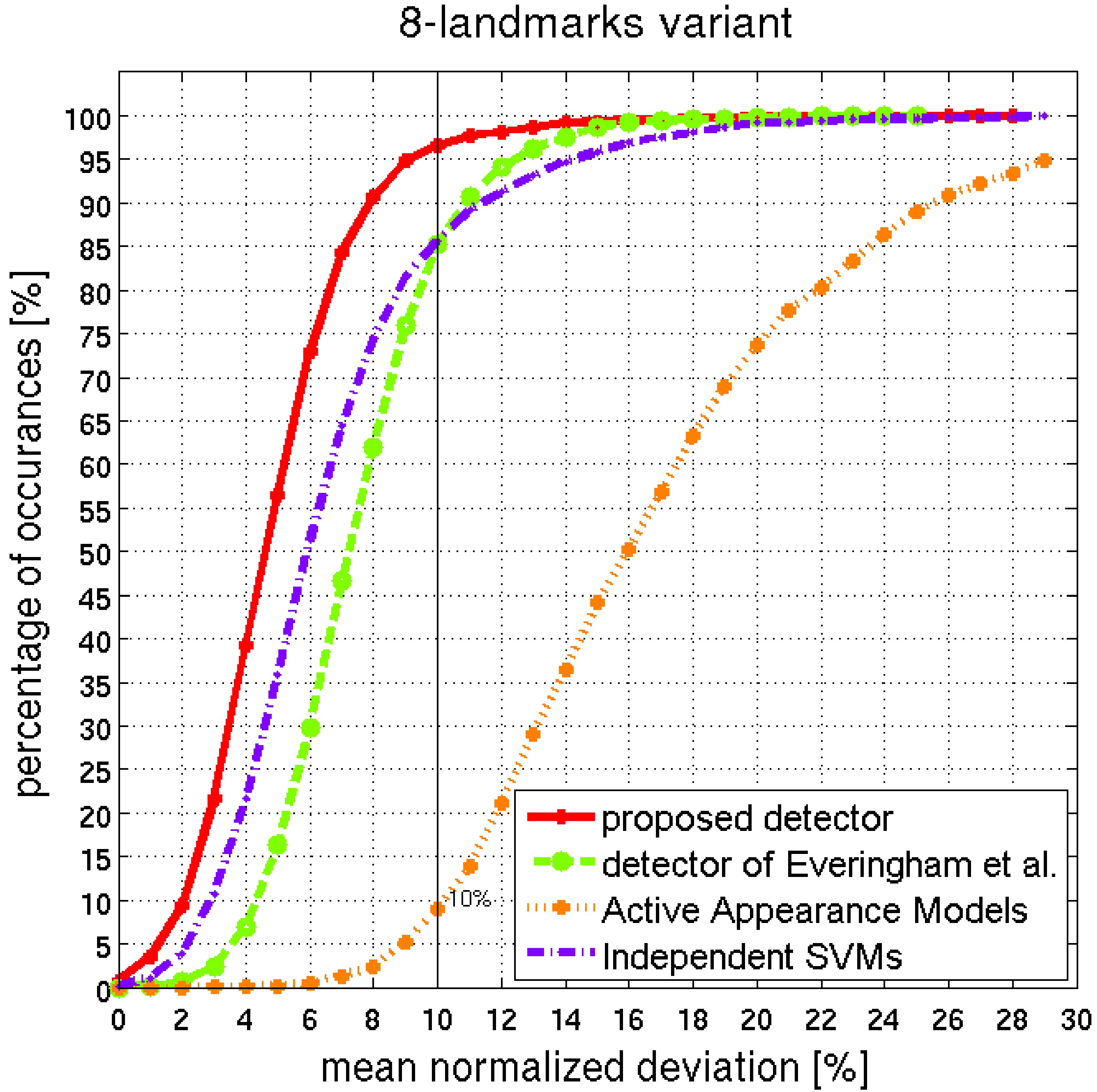




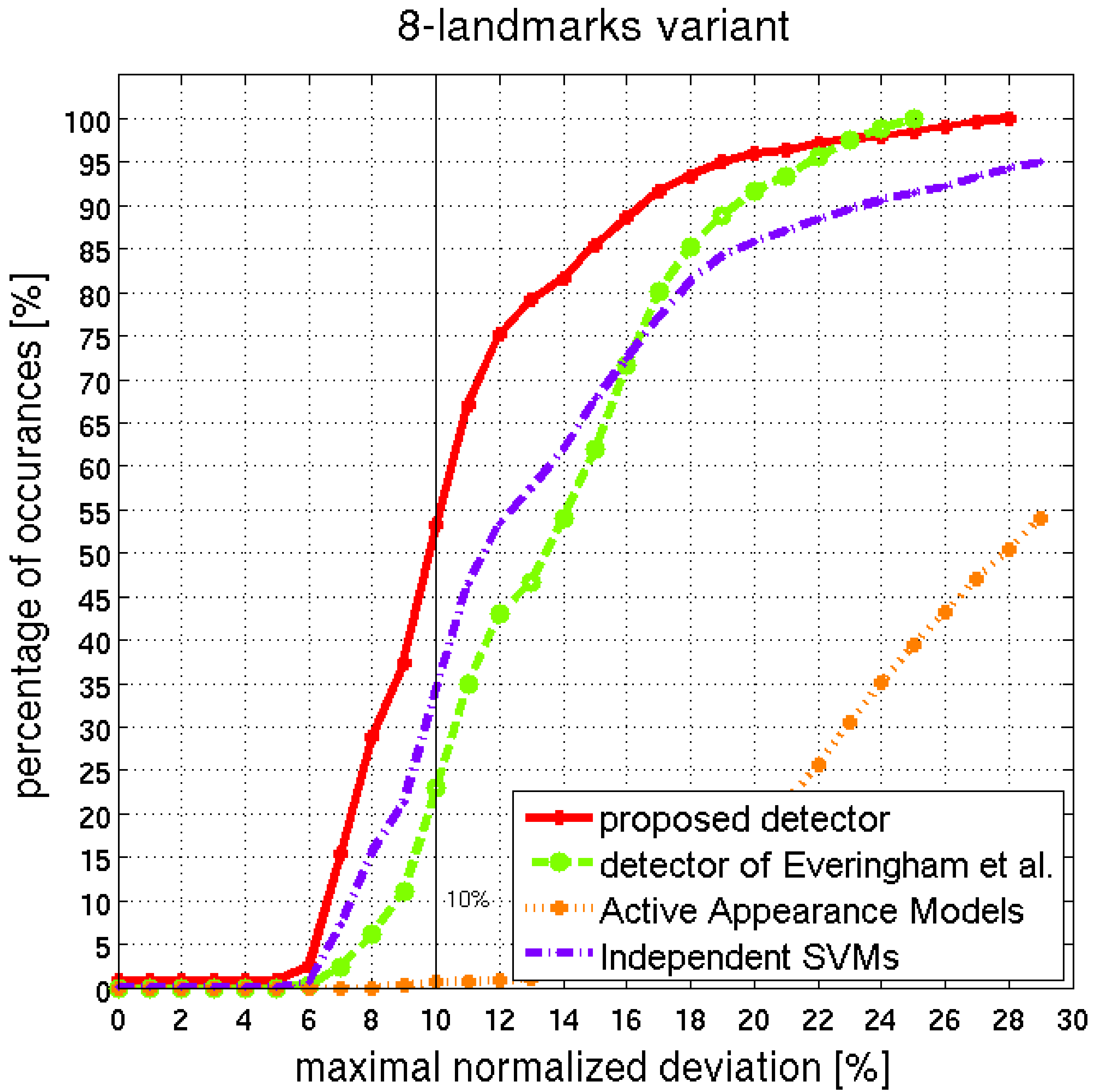
$$L(\mathbf{s}, \hat{\mathbf{s}}) = \frac{1}{l_{\text{face}}} \frac{\epsilon_0 + \cdots + \epsilon_8}{8}$$

$$L^{\max}(\mathbf{s}, \hat{\mathbf{s}}) = \frac{1}{l_{\text{face}}} \max\{\epsilon_0, \ldots, \epsilon_8\}$$


S5
S1
S2
S6
S4



8-landmarks variant
percentage of occurances [%]
mean normalized deviation [%]
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
10%
proposed detector
detector of Everingham et al.
Active Appearance Models
Independent SVMs



8-landmarks variant
percentage of occurances [%]
maximal normalized deviation [%]
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
5
15
25
35
40
45
50
55
60
65
70
75
80
85
90
95
100
10%
proposed detector
detector of Everingham et al.
Active Appearance Models
Independent SVMs