Cascaded Shape Space Pruning for Robust Facial Landmark Detection

Xiaowei Zhao, Shiguang Shan, Xiujuan Chai, Xilin Chen
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
{xiaowei.zhao, shiguang.shan, xiujuan.chai, xilin.chen}@vipl.ict.ac.cn; http://vipl.ict.ac.cn

1. Motivation & Contribution

◆Motivation

Limitations of existing methods proposed for obtaining the ***globally*** optimal landmarks configuration from all possible shapes:

- ➢For Viola *et al*.'s cascaded AdaBoost framework, each landmark is ***individually*** detected, which causes ***ambiguous*** detections.
- ➢The ASMs and AAMs are often prone to ***locally*** optimal solution. A good initialization is needed.
- ➢For the promising regression-based method, it is still challenging to ***directly*** predict an accurate shape from the ***complex*** image appearances [Saragih *et al*. CVPR2011].

◆Contribution

- ➢A ***discriminative structure classifier*** is presented to ***jointly assess*** the configurations of landmarks.
- ➢A novel ***coarse-to-fine shape space pruning*** algorithm is proposed to ***progressively*** filter out the incorrect candidate shapes.

2. Method Overview

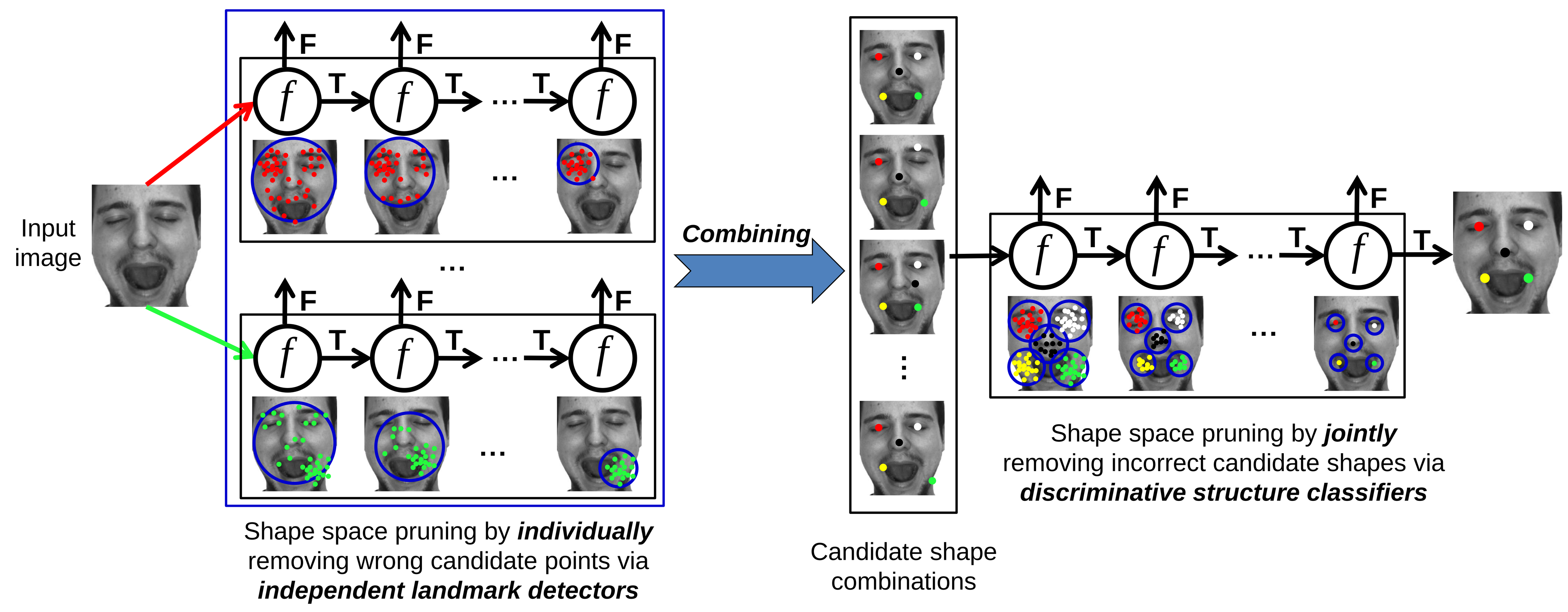


Fig.1 Overview of the proposed cascaded shape space pruning algorithm

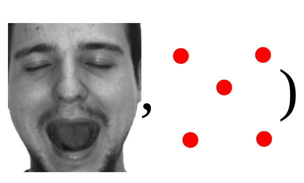
3. Cascaded Shape Space Pruning

◆Step (a): Shape space pruning by individually removing landmark candidates

- ➢The candidate shape space is firstly pruned by individually removing impossible positions of each landmark with separate landmark detector. [Viola *et al.* CVPR2001].

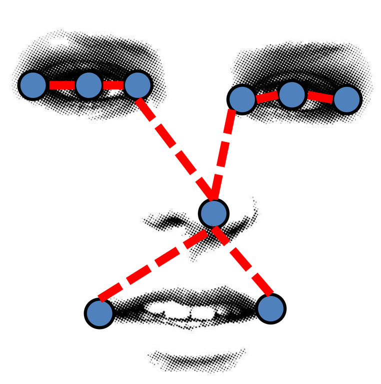
◆Step (b): Shape space pruning by jointly removing candidate landmark positions

➢***Face shape assessment via discriminative structure classifier***



$$F(I, s) = \langle w_{tex}, \Psi_{tex}(I,s)\rangle + \langle w_{shape}, \Psi_{shape}(I,s)\rangle$$

Local descriptor (HOG,LBP,SIFT, *etc.*)



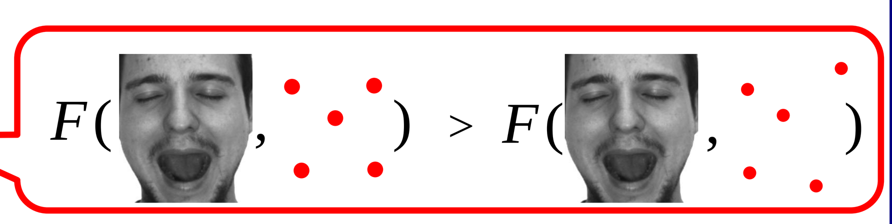
$T=(V,E)$

$$\langle w_{tex}, \Psi_{tex}(I,s)\rangle = \sum_{i\in V}\langle w^i_{tex}, \Psi^i_{tex}(I,s_i)\rangle$$
$$\langle w_{shape}, \Psi_{shape}(I,s)\rangle = \sum_{jk\in E}\langle w^{jk}_{shape}, \Psi^{jk}_{shape}(I,s_j,s_k)\rangle$$
$$\Psi^{jk}_{shape}(I,s_j,s_k) = (dx, dy, dx^2, dy^2)$$
$$(dx, dy) = (s_{jx}, s_{jy}) - (s_{kx}, s_{ky})$$

➢***Learning of discriminative structure classifier***

- The Structured Output SVM is exploited to learn the model parameters w, where the learned model should satisfy that:

the score of a true shape should be greater than that of *any other* wrong shape



$$F(\cdot) > F(\cdot)$$

➢***Efficient cascaded shape space pruning***

- The landmarks configuration s^* with the highest score can be fast computed by dynamic programming.
- We quickly remove the most unconfident ones by the distance between them and s^*.
- We do not use the time-consuming one-by-one assessment scheme.

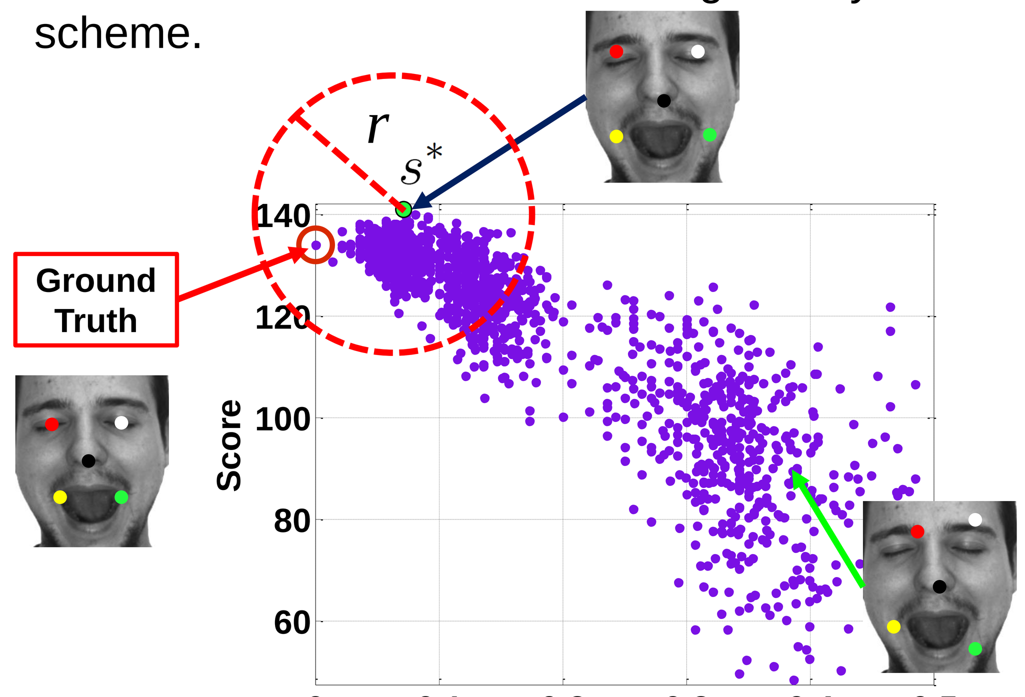


Fig.2 Score distribution of a real F(I, s) on a face image.

- More and more accurate shapes are predicted in the pruned shape space with finer and finer appearance features.

4. Experiments

◆Setup

➢Datasets

- The BioID and LFW face databases

➢Evaluation metric

- NRMSE (Normalized Root-Mean-Squared Error) is adopted as the error measure
- CDF (Cumulative Distribution Function) of NRMSE

$$\mathrm{CDF}(x) = \frac{\mathrm{num}(\mathrm{NRMSE} \le x)}{N}$$

x is the specified error, N is the number of test images.

◆Results

➢Analysis of cascaded shape space pruning

Landmark	Candidates Size			Maximum Deviation			Detection Rate (NRMSE≤0.10)		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
Left eye center	90	52	22	0.46	0.14	0.1	99.20%	98.20%	96.50%
Right eye center	85	51	22	0.41	0.14	0.1	99.10%	98.00%	96.10%
Nose tip	90	53	22	0.44	0.15	0.1	99.20%	98.10%	96.20%
Left mouth corner	87	52	22	0.44	0.15	0.1	99.20%	98.30%	96.40%
Right mouth corner	84	50	22	0.45	0.14	0.1	99.20%	98.20%	96.30%
Outer corner of left eye	88	50	22	0.47	0.14	0.1	99.20%	98.00%	96.10%
Inner corner of left eye	94	52	22	0.49	0.15	0.1	99.00%	97.60%	95.70%
Inner corner of right eye	97	52	22	0.5	0.15	0.1	99.00%	97.50%	95.50%
Outer corner of right eye	91	52	22	0.47	0.15	0.1	99.10%	98.00%	96.10%

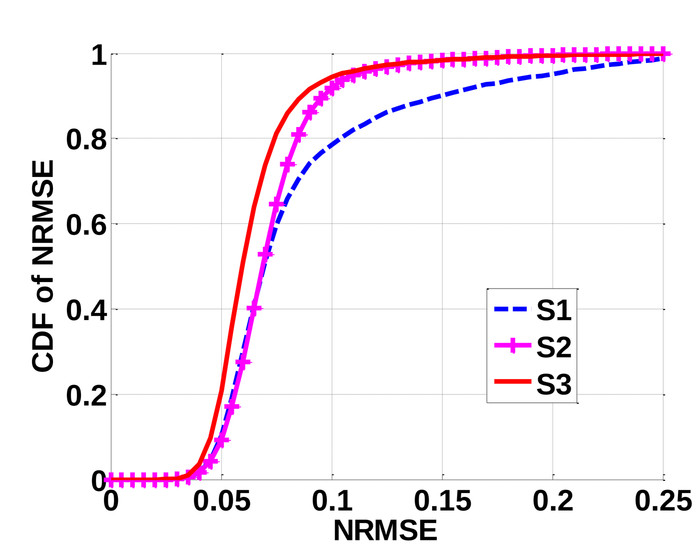


Fig.3 Localization performances in different stages on the LFW face database

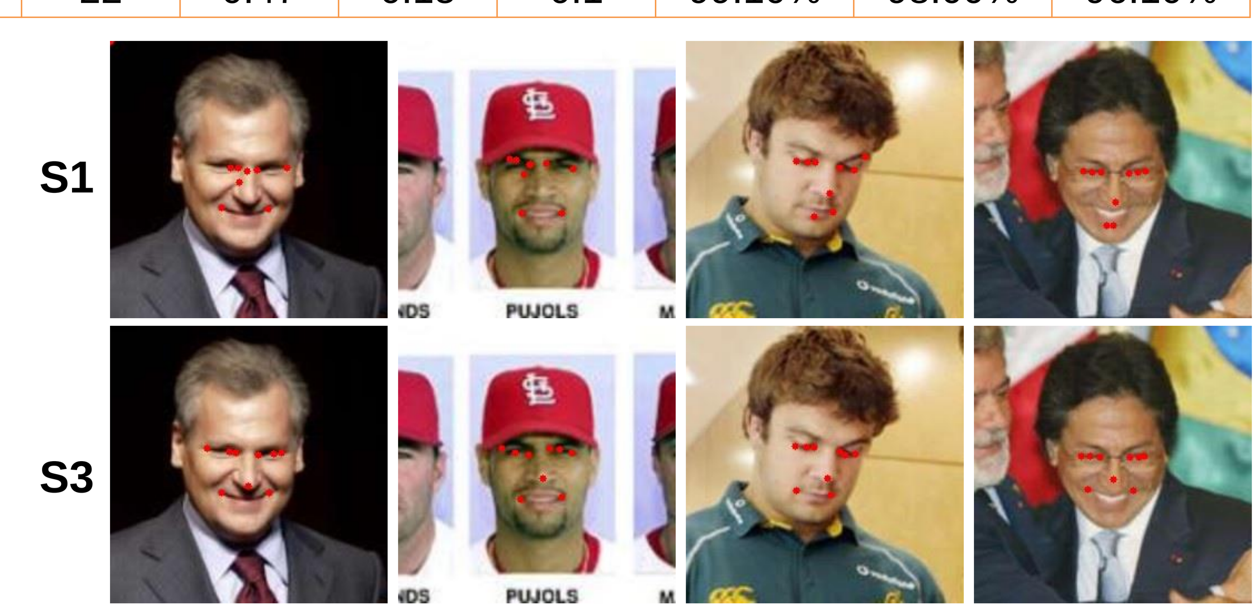


Fig.4 Localization results in different stages (S1&S3)

➢Comparison with state-of-the-art methods

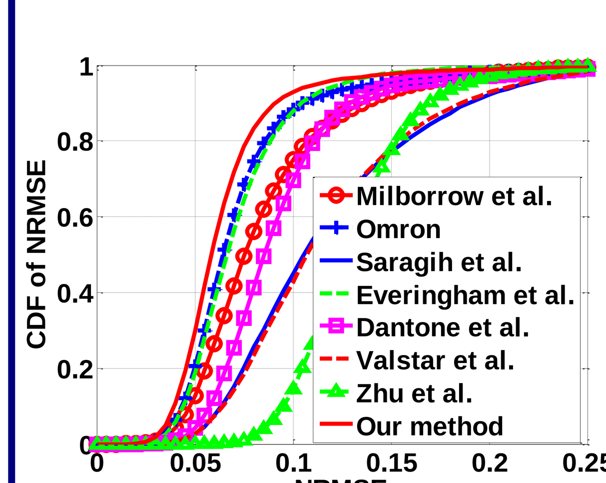


Fig.5 Results on the LFW database

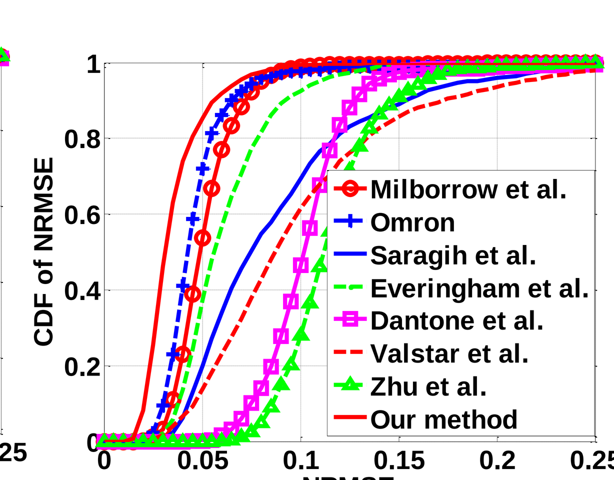


Fig.6 Results on the BioID database

Fig.7 Localization results on some challenging images

5. Summary

➢The positions of landmarks are not only individually evaluated by the local detectors but also jointly evaluated by the discriminative structure classifier.

➢The globally optimal configuration is progressively approximated by gradually filtering out the incorrect candidate configurations.

ICCV 2013 Sydney