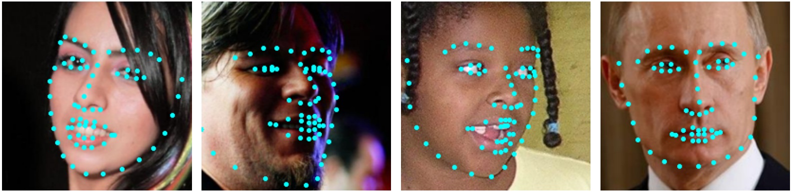
Continuous Supervised Descent Method for Facial Landmark Localisation

Marc Oliu, Ciprian A. Corneanu
Laszlo Jeni, Jeffrey Cohn
Takeo Kanade, Sergio Escalera

November 22, 2016
Asian Conference on Computer Vision, Taiwan

FACIAL LANDMARK LOCALISATION

PROBLEM DEFINITION



Facial landmark localisation (aka. face alignment) is a processing step common to many face analysis techniques. It locates a series of points of interest in a face image.

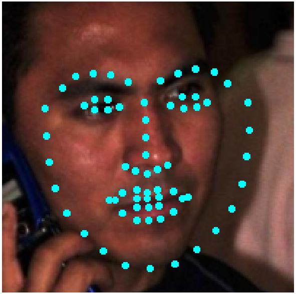
- Problem partially solved for near-frontal faces
- Some difficulties for extreme shadows and rotations
- The more robust approaches are expensive to train

FACIAL LANDMARK LOCALISATION

CASCADED REGRESSION

Usually solved by sequentially applying a series of regression functions f^i mapping the features Φ^i, extracted using the current shape estimate X^i, to the difference between the estimate and ground truth shapes $\Delta X^i = X^i - X^*$.

$$\begin{aligned} X^{i+1} &= X^i + \Delta X^i \\ &= X^i + f^i(\Phi^i) \end{aligned}$$

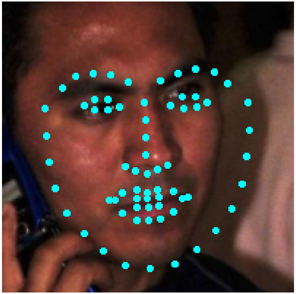


FACIAL LANDMARK LOCALISATION

CASCADED REGRESSION

Usually solved by sequentially applying a series of regression functions f^i mapping the features Φ^i, extracted using the current shape estimate X^i, to the difference between the estimate and ground truth shapes $\Delta X^i = X^i - X^*$.

$$\begin{aligned} X^{i+1} &= X^i + \Delta X^i \\ &= X^i + f^i(\Phi^i) \end{aligned}$$



FACIAL LANDMARK LOCALISATION

GLOBAL SUPERVISED DESCENT METHOD

Suppose an ideal function $\Delta X^i = f(\Phi)$ mapping the features Φ to targets ΔX^i. We can express it with as $\Delta X^i = \Phi^i W^i$, where $W^i = g(\Phi^i)$. Can we approximate the weights space?

GSDM solution: Partition the space into quadrants across a projected feature subspace $\widetilde{\Phi}^i = \Phi^i P$. Learn a linear regressor for each quadrant.

Xiong, X. & De la Torre, F. (2015). *Global supervised descent method*. In CVPR (2664-2673).

FACIAL LANDMARK LOCALISATION

GLOBAL SUPERVISED DESCENT METHOD

Advantages

- Adds robustness to the features main modes of variation
- Approximate $g(\Phi^i)$ non-linearly

Disadvantages

- Low granularity approximating $g(\Phi^i)$
- Number of weights grows exponentially wrt. $||\widetilde{\Phi}^i||$
- Logarithmic reduction on number of samples contributing to each weight

CONTINUOUS SUPERVISED DESCENT METHOD

SPACE OF LINEAR REGRESSORS

CSDM Solution: Define a linear regressor approximating $g(\Phi^i)$ given the feature subspace $\widetilde{\Phi}^i$.

This corresponds to a second order polynomial regression where the projection matrix P restricts the combination of variables in Φ^i.

$$\arg\min_{R_j^i} ||(\Delta\Phi^i \circ (\Delta\widetilde{\Phi}^i R_j^i))\mathbf{1}_{(k+1)} - \Delta X_j^i||_2^2$$

Continuous Supervised Descent Method

Space of linear regressors

CSDM Solution: Define a linear regressor approximating $g(\Phi^i)$ given the feature subspace $\widetilde{\Phi}^i$.

Which can be expressed as a linear regression problem by expanding the features using the Khatri-Rao product.

$$\arg\min_{R_j^i} ||(\Delta\widetilde{\Phi}^i \odot \Delta\Phi^i) vec(R_j^{i\intercal}) - \Delta X_j^i||_2^2$$

CONTINUOUS SUPERVISED DESCENT METHOD

ADVANTAGES AND DISADVANTAGES

Compared to the method most similar to ours, Global SDM, our approach has the following pros and cons.

Advantages

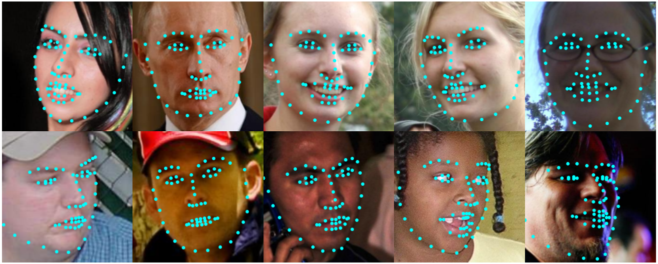
- Adds robustness to the features main modes of variation
- Continuous approximation of $g(\Phi^i)$
- Linear growth in number of parameters wrt. $||\widetilde{\Phi}^i||$
- All instances contribute to each parameter

Disadvantages

- Approximate $g(\Phi^i)$ linearly

DATASETS

300-W



- 3148 train and 689 test samples
- 68 facial landmarks
- No extreme face poses

Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). *300 faces in-the-wild challenge: The first facial landmark localization challenge.* ICCV Workshop (397-403).

DATASETS

PROPOSED: BU4DFE-SYNTHETIC

- 75k images, synthetically rotated from BU4DFE
- Rotations between ±90° in yaw and ±45° in pitch
- Backgrounds sampled from the Places-205 test set

Yin, L., Chen, X., Sun, Y., Worm, T., & Reale, M. (2008). *A high-resolution 3D dynamic facial expression database.* FG (1-6).

QUANTITATIVE RESULTS

COMPARISON TO THE STATE OF THE ART

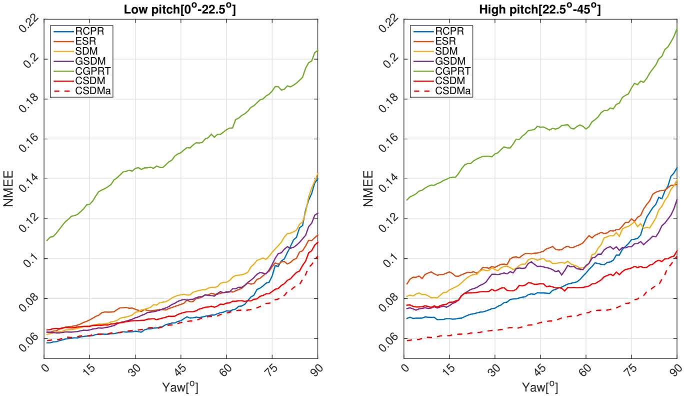
$$NMEE = \frac{1}{n}\frac{\sum_i ||x_i - x_i^*||_2}{||x_l^* - x_r^*||_2}$$

	ESR	RCPR	SDM	ERT	LBF	CGPRT	CFSS	GSDM	CSDM	CSDMa
300W	7.58	8.38	7.52	6.40	6.32	**5.71**	5.76	6.96	6.83	6.40
BU4DFE-S	9.45	8.61	9.57	-	-	15.81	-	9.01	8.28	**7.62**

Table: Comparison with state-of-the-art methods NMEE without (CSDM) and with multiple test initialisations (CSDMa).

QUANTITATIVE RESULTS

ROBUSTNESS TO POSE ON BU4DFE-S



QUANTITATIVE RESULTS

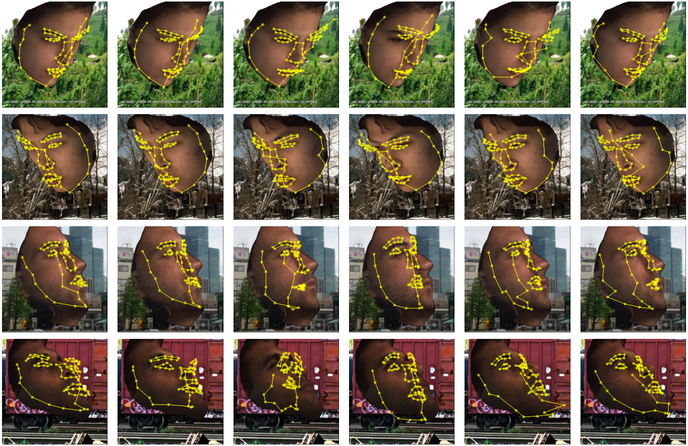
ERROR FOR EACH FACIAL REGION

	Close to frontal						
	ESR	RCPR	SDM	CGPRT	GSDM	CSDM	CSDMa
eyes	3.92	**3.38**	4.02	10.53	3.92	4.04	3.82
eyebrows	5.84	**5.17**	5.60	13.15	5.56	5.84	5.54
nose	6.03	5.59	5.60	10.30	5.51	5.58	**5.27**
mouth	5.46	4.28	4.47	10.91	4.27	4.40	**4.27**
contour	12.59	12.11	13.26	17.49	13.52	13.27	**12.04**
	Far from frontal						
	ESR	RCPR	SDM	CGPRT	GSDM	CSDM	CSDMa
eyes	6.94	6.11	6.76	14.29	6.25	5.55	**5.20**
eyebrows	9.01	8.02	8.50	17.73	8.12	7.20	**6.77**
nose	8.26	7.69	8.58	13.21	8.00	7.41	**6.99**
mouth	8.20	6.70	8.18	14.52	6.72	6.17	**5.84**
contour	17.30	17.19	18.54	22.43	18.53	17.20	**15.27**

Table: NMEE for different facial regions on BU4DFE-S. Results for both close to frontal (between $\pm 30°$ pitch, $\pm 15°$ yaw) and far from frontal head poses (between $\pm 90°$ pitch, $\pm 45°$ yaw).

Qualitative results

Test samples using different approaches



ESR RCPR SDM CGPRT GSDM CSDM

CONCLUSIONS

Contributions

- Natural generalisation of SDM
- Continuous, more adaptive approach to regressor selection

Strengths

- Highly robust to the head pose
- Smaller memory footprint
- Reduced need for training instances

Thank you

https://github.com/moliusimon/csdm