A CASCADED FRAMEWORK FOR MODEL-BASED 3D FACE RECONSTRUCTION

Pengrui Wang, Wujun Che, Bo Xu

Institute of Automation, Chinese Academy of Sciences, China

University of Chinese Academy of Sciences, China

1. Motivation & Contributions

1.1 Motivation

Parameter solving is the main process of model-based 3D face reconstruction which is widely used for performance-driven facial animation, pose or expression invariant face recognition and large-pose face alignment. Methods of parameter solving are generally two types: online optimization processes and regression. We aim to summarize a unified framework for some of the methods, which has the ability to utilize sophisticated face alignment methods for efficiency or accuracy and has intuitive theories to be understood.

1.2 Contributions

The contributions are **threefold**.

1. We proposed a cascade framework for the parameter solving. In our framework, face alignment methods that has landmark update estimation such as SDM, LBF and Deep Alignment Network can be incorporated. The presented framework is based on Gauss-Newton algorithm, which make it well-grounded.
2. Three kinds of methods derived from the framework, called Parameter Constrained Local Model (PCLM), Parameter Regression Method (PRM) and Parameter Augmented Regression Method (PARM).
3. PARM is a novel method we proposed. Different from some methods which add extra time-consuming features in regression approach, PARM supported by our framework, just uses straightforward parameters as additional feature.

2. Proposed Methods

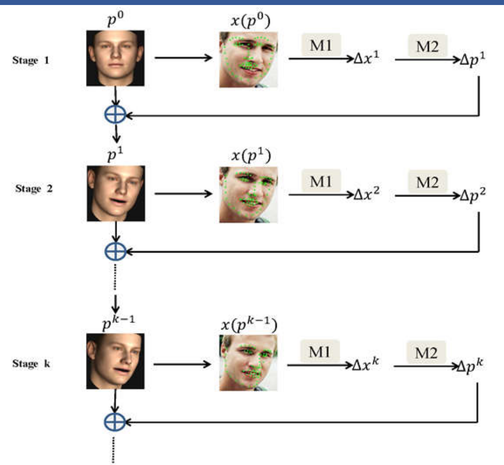

Fig. 1. Workflow of the proposed cascaded. "M1" and "M2" represent optional methods. "M1" can utilize part of mature face alignment methods. "M2" can be PCLM, PRM or PARM in this paper.

2.1 Proposed Framework

We choose 3DMM represents an individual's face, all the parameters we need to fit are : $\mathbf{p} = [s, \text{pitch}, \text{yaw}, \text{roll}, \mathbf{t}_{2d}, \alpha_{\text{id}}, \alpha_{\text{exp}}]$ They are scale, Euler angles, translation, identity and expression. Denote $\mathrm{x}(\mathbf{p})$ as the projected 3D face vertexes corresponding to sparse 2D landmarks in image plane controlled by $\mathbf{p}$. For the task of aligning $\mathrm{x}(\mathbf{p})$ to ground truth 2D face landmark y, essentially $\mathrm{x}(\mathbf{p}^*)$, the following function needs to be minimized :

$$\mathbf{p}^\star = \arg\min_{\mathbf{p}}\{\|\mathbf{y} - \mathbf{x}(\mathbf{p})\|_2^2 + r\|\mathbf{p}\|_W^2\}$$

where $\|p\|_W$ is short for $\sqrt{\mathbf{p}^T W \mathbf{p}}$.

According to Gauss-Newton method, at each stage k,

$$\Delta\mathbf{p}^k = [(J^k)^T J^k + rW]^{-1}[(J^k)^T(\mathbf{x}(\mathbf{p}^\star) - \mathbf{x}(\mathbf{p}^k)) - rW\mathbf{p}^k]$$

Instinctively, we define $\mathbf{x}(\mathbf{p}^\star) = \mathbf{x}(\mathbf{p}^k) + \Delta\mathbf{x}^k + \Delta\mathbf{x}^k_{\text{bias}}$. Then, the final form of our framework is

$$\Delta\mathbf{p}^k = [(J^k)^T J^k + rW]^{-1}[(J^k)^T(\Delta\mathbf{x}^k + \Delta\mathbf{x}^k_{\text{bias}}) - rW\mathbf{p}^k]$$

2.2 Derivation Methods

PCLM: We omit the bias term, then it is similar to the 3D Constrained Local Model problem,

$$\Delta\mathbf{p}^k = [(J^k)^T J^k + rW]^{-1}[(J^k)^T \Delta\mathbf{x}^k - rW\mathbf{p}^k]$$

PRM: Just like the Supervised Descent Method [1] in face alignment, PRM tackles the problem by learning a sequence of regressors,

$$\Delta\mathbf{p}^k = R^k \Delta\mathbf{x}^k + \mathbf{b}^k$$

PARM: To have a guide and automatically adjustment for regressors in PRM, we propose parameter augmented Regression method,

$$\Delta\mathbf{p}^k = R^k \begin{pmatrix} \Delta\mathbf{x}^k \\ \mathbf{p}^k \\ 1 \end{pmatrix}$$

It's reasonable, because our framework can be write as,

$$\Delta\mathbf{p}^k = A\Delta\mathbf{x}^k + B\mathbf{p}^k + C\Delta\mathbf{x}^k_{\text{bias}}$$

The augmented parameter can be seen as the same function of PNCC in [2], but it doesn't need extra calculations.

3. Experiments

3.1 Experimental Setups

We use dataset 300W-3D as the training set and dataset AFLW2000-3D as the testing set. The 3D face model we used is the same with [2]. Δx is estimated use one stage of LBF.

Evaluation Criteria: We choose 68 landmarks alignment error as our e-valuation criteria. The alignment accuracy is the average of all (visible and invisible) landmarks error normalized by the square root of the face bounding box size, called Normalized Mean Error with All (NMEA)

$$\text{NMEA} = \frac{1}{N_t}\sum_i^{N_t}\left(\frac{1}{L_i N}\right)\|\mathbf{x}(\mathbf{p}_{gt}) - \mathbf{x}(\mathbf{p}_i)\|$$

3.2 Experimental Results

◆ Experiments on PCLM, PRM, PARM

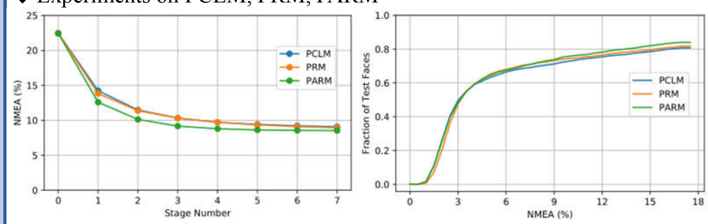

Fig. 2. (left) Testing error on AFLW2000-3D with the increase of iteration stage. (right) Cumulative distribution curves of facial landmark detection results when stage k=7.

Fig. 2 shows that the three methods are feasible, and PARM is better than others. We owe this to the augmented parameter which provides effective information for parameter fitting, while PCLM and PRM are only depend on the previous predicted landmarks increments.

Table 1. Mean Norm Error on Different Types of Parameters

	Pose	Identity	Expression
PCLM	0.411	**3.562**	0.130
PRM	0.404	4.079	0.129
PARM	**0.398**	4.443	**0.128**

We calculate the mean error norm for different types of face parameters. Tab. 1 shows the results, where the dim of pose, identity and expression are 3, 199 and 29 respectively. It seems predicting more accurate pose (Euler angles), which has a larger impact on landmarks than other parameters, that makes PARM better performance.

◆ Comparison with face alignment method LBF

Table 2. The NMEA(%) results on AFLW2000-3D

	PCLM	PRM	PARM	LBF
stage=5	9.42	9.34	**8.63**	**8.14**
stage=7	9.07	8.95	**8.53**	**8.12**

Tab. 2 shows that our 3D parameter fitting methods are inferior to LBF on the performance of landmark alignment error. It should be noticed that our proposed methods pay more attention to parameter fitting, which is a more difficult problem, while LBF search the best landmark positions directly.

We test the speed of 5 stages PARM, it runs over 90 fps on a quad-core Intel Core i7-2600K (3.4GHz) CPU, while PCLM is 35 fps.

4. References

[1] Xuehan Xiong and Fernando De La Torre, "Supervised descent method and its applications to face alignment," in IEEE CVPR, 2013.

[2] Xiangyu Zhu, Zhen Lei, Junjie Yan, Dong Yi, and Stan Z Li, "High-fidelity pose and expression normalization for face recognition in the wild," in IEEE CVPR, 2016.

Interactive Digital Technology Research Center, Chinese Academy of Sciences, Beijing, China

Mail: wangpengrui2015@ia.ac.cn