

A CASCADED FRAMEWORK FOR MODEL-BASED 3D FACE RECONSTRUCTION

Pengrui Wang^{1,2}, Wujun Che¹, Bo Xu¹

¹Institute of Automation, Chinese Academy of Sciences, China

²University of Chinese Academy of Sciences, China

{wangpengrui2015, wujun.che, xubo}@ia.ac.cn

ABSTRACT

This paper presents a general framework for model-based 3D face reconstruction from a single image, which can incorporate mature face alignment methods and utilize their properties. In the proposed framework, the final model parameters, i.e., mostly including pose, identity and expression, are achieved by estimating updating the face landmarks and 3D face model parameter alternately. In addition, we propose the parameter augmented regression method (PARM) as an novel derivation of the framework. Compared with existing methods, PARM is able to utilize mature face alignment methods and use fairly simple features in addition to image appearances for the reconstruction task. Experiments on three derivation methods of the framework show that the proposed framework is feasible and PARM is quite an effective and fast method. With face alignment method LBF, PARM can run over 90 fps on a desktop.

Index Terms— Model-based 3D face reconstruction, face alignment, cascaded regression, supervised descent method

1. INTRODUCTION

3D face models have been widely employed in various research fields such as performance-driven facial animation [1, 2], pose or expression invariant face recognition [3, 4] and large-pose face alignment [5, 6]. In these works, 3D faces are reconstructed from a given 3D face model and a RGB face image. The widely used face models, such as 3D Morphable Model (3DMM) [7, 8] and blend-shapes model [9], are controlled by the parameters, which contains two parts: rigid parameter (pose) and nonrigid parameters (identity and expression). Thus the main process of the model-based 3D face reconstruction is to obtain the model parameter which can drive the model to simulate real faces, i.e., parameter fitting.

Typically, parameter fitting methods could be divided into two types, one is solving parameters by fitting 3D point distribution model to 2D points, the other is obtaining parameters by regression approaches.

Solving methods usually require online optimization processes. Generally, they solve parameters after a 2D face alignment method [1, 2] or iteratively execute face shape regression and parameters solving [10, 11]. However, these methods have some disadvantages: (1) because the number of visible landmarks and the spatial distribution of the landmarks are pose dependent, producing credible landmarks directly is infeasible for most existing 2D face alignment approaches; (2) the optimization approaches for parameter fitting are

usually time-consuming and are hard to utilize the information from the training set; (3) 2D landmarks provided by a separate method is inconsistent with the 3D model vertexes, which may conflict the later solving process.

Recently, researchers [5, 12, 6] tackle large pose face alignment problem with cascaded parameter regression in an intermediate manner. These methods can be directly used for parameter fitting. Without a separate 2D face alignment, these method focus on 3D fitting and avoid drawbacks brought from 2D face alignment. Besides, obtaining parameter regressively can take the whole training set information into account. Although possessing these advantages, they are hard to utilize mature face alignment methods and lack of intuitive theories to be understood.

In this paper, we present a unified framework for the parameter fitting problem, which aims to overcome the mentioned shortcomings. Our framework is based on the equation of parameter update which is an key step in fitting a 3D point distribution model to 2D points [13] using Gauss-Newton algorithm. In addition, our framework is a cascade model, where any face alignment approach that has landmark update estimation such as SDM [14], LBF [15], shape augment regression [16] and Deep Alignment Network [17] can be incorporated. Under the novel framework, two basic methods and PARM are proposed. PARM explicitly combines previous estimated parameters with information come from appearance to better the regression performance. Compared with regression methods, such as [6] and [12], which add extra time-consuming features for parameter fitting, PARM, supported by our framework, just uses straightforward parameters as additional feature. Thus makes it a fast method.

The rest of the paper is organized as follows. Section 2 briefly reviews related works. Section 3 gives a detailed introduction to our framework and proposed methods. In Section 4, we present the experiments and the evaluation results. And finally, we conclude the paper with some discussions of future work in section 5.

2. RELATED WORK

General Face Alignment: 2D/3D face alignment [18, 19, 20] aims to locate a sparse set of fiducial facial landmarks. Quite a few of systems have been reported and could be classified into three major categories: holistic methods [21], constrained local model (CLM) [22] and the regression based methods [14, 15]. The holistic methods jointly model the whole face region appearance and global face shape, while CLM learns a set of local experts or regressors and constrains them using various shape models. The regression based methods directly learn the mapping from facial images to landmark positions. The mapping cascades a list of weak regressors to reduce alignment error progressively. All these methods that estimate landmark updates can be incorporated into our framework.

This work was supported by National Natural Science Foundation of China under No.61471359, by National Key Technology R&D Program of China under No. 2015BAH53F01 and by Beijing Digital Content Engineering Technology Research Center No. Z171100002217015.

Face Alignment use 3D Models: Recently, researches on large-pose, pose-invariant and joint face reconstruction face alignment methods [5, 6, 12, 23] begin to use 3D face models. These works are closely related to ours.

[23] proposes a joint face alignment and 3D face reconstruction method. It applies two sets of cascaded regressors iteratively and alternately, one for updating 2D landmarks and the other for updating 3D face shape, they are correlated via a 3D-to-2D mapping matrix. This work directly reconstructs 3D face shape while our framework is focus on parameter fitting. Besides, our framework applies parameter updating rather than shape updating and doesn't need 3D-to-2D mapping matrix.

[6] and [12] design cascaded Coupled-Regressor to fit projection parameter and shape parameter iteratively for face alignment. [5] and [6] adopts CNN to fit parameter directly with specifically designed input features, such as PNCC in [5], PAWF and D3PF in [6], for large pose face alignment. All these methods fit parameters without obvious 2D/3D landmark alignment process. Meanwhile, it seems that the novel appearance features they designed lack of intuitive understanding for parameter fitting. However, our PRAM with clear theoretical explanation shows that the additional features can be as simple as the pure parameter.

3. PROPOSED METHOD

In order to better understand our framework, we first introduce 3D Morphable Model. Then, we give detailed introduction to our framework. Finally, we introduce three kinds of methods derived from the proposed framework. They are Parameter Constrained Local Model (PCLM), Parameter Regression Method (PRM) and Parameter Augmented Regression Method (PARM). Among them, PARM is a very novel method for parameter fitting task. The overview of our proposed framework is shown in Fig. 1.

3.1. 3D Morphable Model

We choose 3DMM [3] represents an individual's face:

$$S = S_0 + A_{id}\alpha_{id} + A_{exp}\alpha_{exp}, \quad (1)$$

where S is the 3D shape, S_0 is the mean shape, A_{id} is the identity basis, A_{exp} is the expression basis, α_{id} is the identity parameter and α_{exp} is the expression parameter. In our work, A_{id} and A_{exp} come from BFM [24] and FaceWarehouse [9] respectively. The dim of α_{id} and α_{exp} are 199 and 29 respectively. The 3D face is then projected onto the image plane with Weak Perspective Projection, i.e.,

$$V = sR_{2D}S + t_{2d}, \quad (2)$$

where V is the 2D positions of model vertexes in image, s is the scale parameter, R_{2D} is the first two rows of a 3×3 rotation matrix controlled by Euler angle: [pitch, yaw, roll], and t_{2d} is the translation vector. We denote $\mathbf{p} = [s, \text{pitch}, \text{yaw}, \text{roll}, t_{2d}, \alpha_{id}, \alpha_{exp}]^T$ as the collection of all the model parameter. Let \mathbf{d} indicates the indexes of 3D face vertexes that corresponding to sparse 2D landmarks, then the 2D landmarks under a given \mathbf{p} are represented as $U(\mathbf{p}) = S(\mathbf{d})$. We reshape $U(\mathbf{p})$ as a vector $\mathbf{x}(\mathbf{p})$.

3.2. Proposed Framework

For the task of aligning $\mathbf{x}(\mathbf{p})$ to ground truth 2D face landmark \mathbf{y} , the following function needs to be minimized:

$$\mathbf{p}^* = \arg \min_{\mathbf{p}} \{\|\mathbf{y} - \mathbf{x}(\mathbf{p})\|_2^2 + r\|\mathbf{p}\|_W^2\}, \quad (3)$$

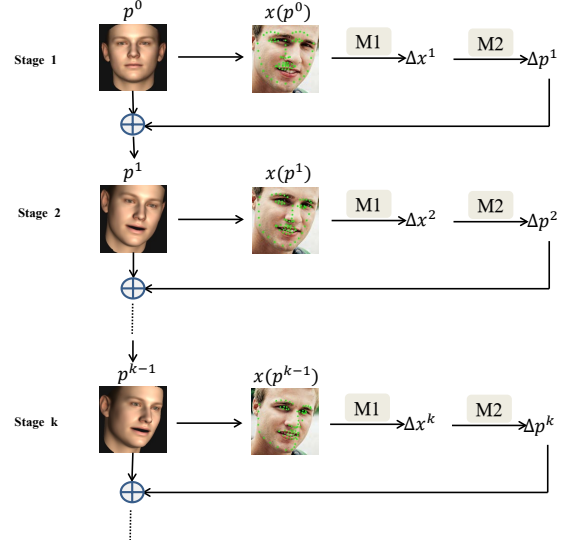


Fig. 1. Workflow of the proposed cascaded framework for model-based face reconstruction from an single image. It begins with an initial model parameter \mathbf{p}^0 , then the final parameter is \mathbf{p}^k adding a series of estimated parameter update $\Delta \mathbf{p}^k$. At stage k , $\Delta \mathbf{x}^k$ standing for landmark update is first estimated. "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.

where the notation $\|p\|_W$ is a shorthand for $\sqrt{\mathbf{p}^T W \mathbf{p}}$. Above, $\|\mathbf{p}\|_W^2$ is the regularization term, which helps avoid overfitting and W here is usually diagonal matrix to punish \mathbf{p} with prior information. The r controls the tradeoff between the landmark placement error and penalising unlikely faces [13].

According to the Gauss-Newton method, given an appropriate initial \mathbf{p} , it is possible to find $\Delta \mathbf{p}$ in the direction of the optimal solution. This leads to next estimate of \mathbf{p} , which can be used for next iteration. By use of Taylor expansion, the $\Delta \mathbf{p}$ which brings us closer to the solution is:

$$\Delta \mathbf{p} = \arg \min_{\Delta \mathbf{p}} \{\|\mathbf{y} - (\mathbf{x}(\mathbf{p}) + J\Delta \mathbf{p})\|_2^2 + r\|\mathbf{p} + \Delta \mathbf{p}\|_W^2\}, \quad (4)$$

where $J = \frac{\partial \mathbf{x}(\mathbf{p})}{\partial \mathbf{p}^T}$ is the Jacobian of $\mathbf{x}(\mathbf{p})$ evaluated at \mathbf{p} . The solution of $\Delta \mathbf{p}$ is:

$$\Delta \mathbf{p} = (J^T J + rW)^{-1} [J^T (\mathbf{y} - \mathbf{x}(\mathbf{p})) - rW \mathbf{p}]. \quad (5)$$

In our framework, \mathbf{y} is $\mathbf{x}(\mathbf{p}^*)$ essentially. We then rewrite the equation in a cascaded form. Thus, at each stage k ,

$$\Delta \mathbf{p}^k = [(J^k)^T J^k + rW]^{-1} [(J^k)^T (\mathbf{x}(\mathbf{p}^*) - \mathbf{x}(\mathbf{p}^k)) - rW \mathbf{p}^k], \quad (6)$$

Therefor, \mathbf{p}^* can be estimated by an initial \mathbf{p}^0 adding a series of $\Delta \mathbf{p}$.

The remaining issue is that $\mathbf{x}(\mathbf{p}^*)$ is unknown. However, $\mathbf{x}(\mathbf{p}^*)$ can be seen as the ground truth in face alignment problem, we can leverage the cascade face alignment methods, which solved face alignment as:

$$\mathbf{x}^* = \mathbf{x}_{init} + \Delta \mathbf{x}^1 + \Delta \mathbf{x}^2 + \Delta \mathbf{x}^3 + \dots \quad (7)$$

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

$$\Delta \mathbf{p}^k = [(J^k)^T J^k + rW]^{-1} [(J^k)^T (\Delta \mathbf{x}^k + \Delta \mathbf{x}_{bias}^k) - rW \mathbf{p}^k] \quad (8)$$

Equ. 8 shows that the final \mathbf{p} can be estimated iteratively. In each iteration, $\Delta\mathbf{x}^k$ can be obtained firstly utilizing face alignment method which has landmark update estimation. Fig. 1 visualizes the workflow of our framework.

3.3. Derivation Methods

Here we introduce the derivation methods, PCLM, PRM and PARM. **PCLM:** When we omit the bias term $\Delta\mathbf{x}_{\text{bias}}^k$, the problem becomes,

$$\Delta\mathbf{p}^k = [(J^k)^T J^k + rW]^{-1} [(J^k)^T \Delta\mathbf{x}^k - rW\mathbf{p}^k]. \quad (9)$$

This is similar to the 3D Constrained Local Model problem and can be solved using the Gauss-Newton method as in [13, 10, 11]. However, rather than using local experts, $\Delta\mathbf{x}^k$ can be estimated by any cascade face alignment method with landmark update prediction.

PRM: Just like the Supervised Descent Method [14] in face alignment, we propose PRM which tackles the problem by learning a sequence of regressors. Thus the form of our framework is a regression approach,

$$\Delta\mathbf{p}^k = R^k \Delta\mathbf{x}^k + \mathbf{b}^k, \quad (10)$$

where R and \mathbf{b} make up the regressors. The regression approach merges prior information, such as W and r , into regressors, they are not needed to be manually setting anymore. Meanwhile, PRM doesn't need to compute J , which is possibly time-consuming and even doesn't exist. The regressors can be very complex, a more general form of PRM is,

$$\Delta\mathbf{p}^k = R_{\text{general}}^k(\Delta\mathbf{x}^k), \quad (11)$$

where R_{general}^k can be a weight matrix or a deep neural network. It should be mentioned that if $\Delta\mathbf{x}^k$ is merged in the R_{general}^k and doesn't estimate obviously, PRM becomes the method similar to [5].

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

Equ. 8 shows that \mathbf{p}^k is an essential term for estimating $\Delta\mathbf{p}^k$. It's straightforward to improve the PRM by adding an extra term \mathbf{p}^k . This leads to our PARM,

$$\Delta\mathbf{p}^k = R^k \begin{pmatrix} \Delta\mathbf{x}^k \\ \mathbf{p}^k \\ 1 \end{pmatrix}, \quad (12)$$

or a general form, $\Delta\mathbf{p}^k = R_{\text{general}}^k(\Delta\mathbf{x}^k, \mathbf{p}^k)$. Next, we show that the augmented parameter is supported by our framework.

The framework Equ. 8 can be write as the form:

$$\Delta\mathbf{p}^k = A\Delta\mathbf{x}^k + B\mathbf{p}^k + C\Delta\mathbf{x}_{\text{bias}}^k, \quad (13)$$

where A , B and C represent matrix. We still adopt the idea in SDM and handle the problem in a regression manner. Then $C\Delta\mathbf{x}_{\text{bias}}^k$ can be learned as a bias vector, such leads to Equ. 12. What should be emphasized is that it is the item $\Delta\mathbf{x}_{\text{bias}}^k$ that makes the PARM a possible methods, this is unrevealed in previous literature.

With the new term added, regressors could change directly according to different parameters, which would make better estimation of PRM. The augmented parameter can be seen as the same function of PNCC in [5], but it doesn't need extra calculations.

4. EXPERIMENTS

In this section, we first discuss experiment protocols, followed by implementation details, and then we evaluate the performance of the three methods introduced in Sec. 3.3.

4.1. Protocols

Datasets: There are quite a few datasets can be directly used for our framework. The datasets should at least cover face images and their corresponding parameters for a chosen 3D face model. However, [5] constructed some datasets (300W-3D, 300W-LP and AFLW2000-3D) which are suitable for us. In our experiments, 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 [5], which is described in Subsection 3.1.

300W-3D is the 300W dataset [25] with its 3D parameters. 300W is original made from multiple alignments datasets with 68 landmarks. Here, the used datasets are LFPW, AFW, Helen and IBUG, totally 3837 images.

Original AFLW [26] contains in-the-wild faces with large-pose variations (yaw from -90° to 90°). AFLW2000-3D is constructed from the first 2000 AFLW samples.

Experiment setup: In our experiments, $\Delta\mathbf{x}$ is estimated use one stage of LBF [15]. For PRM and PARM, we use a set of liner regressors to estimate $\Delta\mathbf{p}$. All the liner regressors for landmark update and parameter update predictions are trained by the LIBLINEAR [27] package. For PCLM, we use the Gauss-Newton method the same as [13] to calculate $\Delta\mathbf{p}$ after landmark update estimation. We train 20 trees with depth of 5 for every landmark in each stage to generate local binary feats, details are in [15].

Evaluation Criteria: We choose landmark alignment error as our evaluation criteria, where the ground truth landmarks are obtained according to the ground truth parameter \mathbf{p}_{gt} . We use the 68 landmarks in AFLW2000-3D for evaluation. 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), i.e.,

$$\text{NMEA} = \frac{1}{N_t} \sum_i \left(\frac{1}{L_i N} \right) \| \mathbf{x}(\mathbf{p}_{gt}) - \mathbf{x}(\mathbf{p}_i) \|, \quad (14)$$

where N_t is the number of testing images, N is the number of landmark (here is 68), and L_i is the square root of face bounding box size of image i .

4.2. Implementation Details

Parameter Normalization: Faces in images have considerable difference in size and location, which possibly damages performance of the parameter fitting task. To avoid the bad effects, we use face bounding boxes to normalize the parameters. Specially, assuming the face bounding box in image is $bbox = [x_c, y_c, w, h]$, where (x_c, y_c) stands for the center position, w and h stand for the weight and height respectively, the normalized parameter \mathbf{p}_{norm} is similar with \mathbf{p} expect replacing s with s/\sqrt{wh} and replacing $\mathbf{t}_{2d} - (x_c, y_c)/\sqrt{wh}$.

Parameter Initialization: We denote \bar{s} and $\bar{\mathbf{t}}_{2d}$ to represent the mean scale and mean translation of the normalized parameters, which can be calculated from training set. Then the initial parameter \mathbf{p}_{init} we used is $[\bar{s}, 0, 0, 0, \bar{\mathbf{t}}_{2d}, \mathbf{0}^T, \mathbf{0}^T]$.

Estimating Invisible Landmarks: Face has a variety of poses. Part of landmarks are invisible under a pose, which could generate unsuitable features for $\Delta\mathbf{x}$ estimation. In this work, an invisible landmark is detected according to the average surface normal around the landmark in the deformed 3D shape S .

4.3. Comparison Experiments

Experiments on PCLM, PRM, PARM: In this experiment, we aim to compare the performance of our three implementations (PCLM, PRM, PARM) based on our proposed framework. The experimental results are shown in Fig. 2 and Fig. 3.

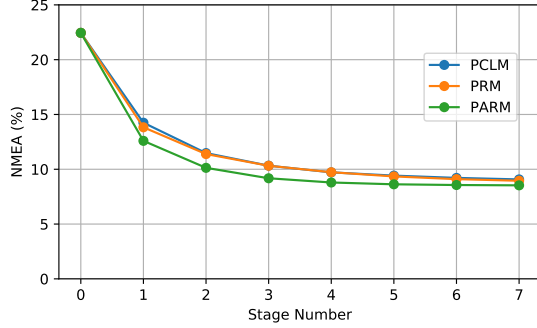


Fig. 2. Testing error with PCLM, PRM and PARM on AFLW2000-3D with the increase of iteration stage.

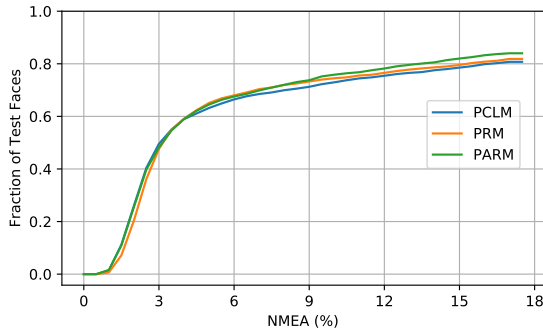


Fig. 3. Cumulative distribution curves of facial landmark detection results about PCLM, PRM and PARM respectively when stage $k=7$.

The results indicate that our proposed framework are feasible, because all of the implementations improve the performance with iteration increasing. There is also an obvious observation that the performance on PARM is better than others. We owe this to the augmented parameter. The augmented parameter has a guide or provides an 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

To have a further understanding of the effects brought from the augmented parameter, 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.

Some reconstruction results generated by PARM are visualized in Fig. 4. It shows that PARM has relatively accurate estimation on pose and expression, but the identity estimation is a little bit unsatisfied. This may caused by the imprecise face alignment method and the unconsidered factors that different element in \mathbf{p} has different importance on the reconstructed result.

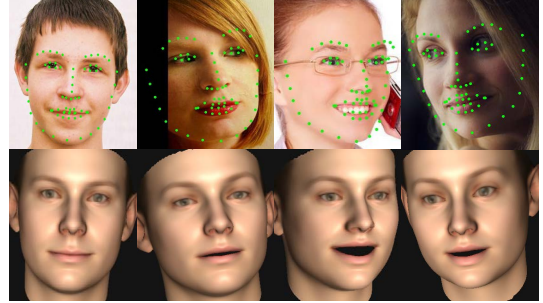


Fig. 4. Part of reconstruction 3D face model with PARM.

Comparison with face alignment method LBF: To have a comparison with methods which focus on 2D face alignment, it is fair to compare our methods with LBF, for the sake of that they have the same methods on landmark increments prediction.

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 also test reconstruction speeds of five stages PCLM and PARM with LBF incorporated in. Results show that PARM running over 90 fps, which is nearly 3 times faster than PCLM running at 35 fps. The methods test on a quad-core Intel Core i7-2600K (3.4GHz) CPU with not optimized c++ code. That shows our PARM is also a fast method for the face reconstruction task.

5. CONCLUSION

In this paper, we present a novel framework for model-based 3D face reconstruction with theoretical support. Under our framework, part of mature face alignment methods can be integrated in. For instance, LBF can be use to accelerate the methods based on our framework. Meanwhile, we introduce three feasible methods: PCLM, PRM and PARM. Experiments tell that PARM is an efficient method.

In the future, we would employ new technical routes based on our framework to improve the performance. We should use more accurate method to predict $\Delta\mathbf{x}$ and more powerful regressors, like deep neural networks, to predict $\Delta\mathbf{p}$. We also plan to take different items in parameters into account. Specifically, we would use the weighted parameter distance cost function [5] to optimize the regressors. Furthermore, it is worth exploring the strategy on facial reconstruction if multiple images with one single person were given.

6. REFERENCES

- [1] Chen Cao, Yanlin Weng, Stephen Lin, and Kun Zhou, “3d shape regression for real-time facial animation,” *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 41, 2013.
- [2] Chen Cao, Qiming Hou, and Kun Zhou, “Displaced dynamic expression regression for real-time facial tracking and animation,” *ACM Transactions on Graphics*, vol. 33, no. 4, pp. 43, 2014.
- [3] Volker Blanz and Thomas Vetter, “Face recognition based on fitting a 3d morphable model,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.
- [4] 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 Conference on Computer Vision and Pattern Recognition*, 2015, pp. 787–796.
- [5] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li, “Face alignment across large poses: A 3d solution,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 146–155.
- [6] Amin Jourabloo and Xiaoming Liu, “Large-pose face alignment via cnn-based dense 3d model fitting,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4188–4196.
- [7] Volker Blanz and Thomas Vetter, “A morphable model for the synthesis of 3d faces,” in *Annual Conference on Computer Graphics and Interactive Techniques*, 1999, pp. 187–194.
- [8] James Booth, Epameinondas Antonakos, Stylianos Ploumpis, George Trigeorgis, Yannis Panagakis, and Stefanos Zafeiriou, “3d face morphable models” in-the-wild,” *arXiv preprint arXiv:1701.05360*, 2017.
- [9] Chen Cao, Yanlin Weng, Shun Zhou, Yiyong Tong, and Kun Zhou, “Facewarehouse: A 3d facial expression database for visual computing,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2014.
- [10] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency, “3d constrained local model for rigid and non-rigid facial tracking,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2610–2617.
- [11] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency, “Constrained local neural fields for robust facial landmark detection in the wild,” in *IEEE International Conference on Computer Vision Workshops*, 2013, pp. 354–361.
- [12] Amin Jourabloo and Xiaoming Liu, “Pose-invariant 3d face alignment,” in *IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [13] Tadas Baltrušaitis, “Automatic facial expression analysis,” *University of Cambridge*, 2014.
- [14] Xuehan Xiong and Fernando De La Torre, “Supervised descent method and its applications to face alignment,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 532–539.
- [15] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun, “Face alignment at 3000 fps via regressing local binary features,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1685–1692.
- [16] Yue Wu and Qiang Ji, “Shape augmented regression method for face alignment,” in *IEEE International Conference on Computer Vision Workshop*, 2015, pp. 979–985.
- [17] Marek Kowalski, Jacek Naruniec, and Tomasz Trzcinski, “Deep alignment network: A convolutional neural network for robust face alignment,” in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 2034–2043.
- [18] Adrian Bulat and Georgios Tzimiropoulos, “How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks),” *arXiv preprint arXiv:1703.07332*, 2017.
- [19] Adrian Bulat and Georgios Tzimiropoulos, “Two-stage convolutional part heatmap regression for the 1st 3d face alignment in the wild (3dfaw) challenge,” 2016.
- [20] Ruiqi Zhao, Yan Wang, C. Fabian Benitez-Quiroz, Yaojie Liu, and Aleix M. Martinez, “Fast and precise face alignment and 3d shape reconstruction from a single 2d image,” in *The Workshop on 3d Face Alignment in the Wild*, 2016, pp. 590–603.
- [21] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor, “Active appearance models,” in *European Conference on Computer Vision*, 1998, pp. 484–498.
- [22] David Cristinacce and Timothy F. Cootes, “Feature detection and tracking with constrained local models,” in *British Machine Vision Conference*, 2006, pp. 929–938.
- [23] Feng Liu, Dan Zeng, Qijun Zhao, and Xiaoming Liu, “Joint face alignment and 3d face reconstruction,” in *European Conference on Computer Vision*, 2016, pp. 545–560.
- [24] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter, “A 3d face model for pose and illumination invariant face recognition,” in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2009, pp. 296–301.
- [25] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, “300 faces in-the-wild challenge: The first facial landmark localization challenge,” in *IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403.
- [26] Martin Koestinger, Paul Wohlhart, Peter M Roth, and Horst Bischof, “Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization,” in *IEEE International Conference on Computer Vision Workshops*. IEEE, 2011, pp. 2144–2151.
- [27] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, “Liblinear: A library for large linear classification,” *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.