Discriminative 3D Morphable Model Fitting

Xiangyu Zhu, Junjie Yan, Dong Yi, Zhen Lei and Stan Z. Li
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Haidian District, Beijing, China.

Abstract—This paper presents a novel discriminative method for estimating 3D shape from a single image with 3D Morphable Model (3DMM). Until now, most traditional 3DMM fitting methods depend on the analysis-by-synthesis framework which searches for the best parameters by minimizing the difference between the input image and the model appearance. They are highly sensitive to initialization and have to rely on the stochastic optimization to handle local minimum problem, which is usually a time-consuming process. To solve the problem, we find a different direction to estimate the shape parameters through learning a regressor instead of minimizing the appearance difference. Compared with the traditional analysis-by-synthesis framework, the new discriminative approach makes it possible to utilize large databases to train a robust fitting model which can reconstruct shape from image features accurately and efficiently. We compare our method with two popular 3DMM fitting algorithms on FRGC database. Experimental results show that our approach significantly outperforms the state-of-the-art in terms of efficiency, robustness and accuracy.

I. INTRODUCTION

Since the seminal work of Blanz and Vetter [9], the 3D Morphable Model (3DMM) has been widely used to estimate 3D shape from image data, with applications ranging from relighting [39], super-resolution [25] to pose robust face recognition [10]. Given a single face image under unknown pose and illumination, the 3D Morphable Model can solve its 3D shape, texture, pose and illumination parameters simultaneously following analysis-by-synthesis framework, where Gauss-Newton optimization is applied to minimize the difference between the synthetic image and the input image.

The original 3DMM has shown its robustness to complicated pose and illumination conditions and provides a promising way to face recognition in the wild due to the explicit pose and illumination recovery [10]. However, the fitting process of 3DMM is very time-consuming and suffers from local minimum problem just as other Gauss-Newton based methods. In the last decade, many researchers have made their efforts to improve the efficiency and accuracy of 3DMM fitting algorithm. On one hand, [30] extended the Inverse Compositional Image Alignment (ICIA) to 3DMM fitting, improving the efficiency by pre-computing the derivative of the cost function. [38], [2] adopted the spherical harmonic reflectance model, making the appearance estimation completely linear. [8], [26], [1] concentrated on estimating shape with a sparse set of 2D landmarks, providing an efficient linear method for shape fitting. On the other hand, [31] presented a Multi-Features Framework (MFF) to handle the local minimum problem, where a smoother objective function is constructed by considering contours, textured edges, specular highlights and pixel intensities jointly, leading to the state-of-the-art in both robustness and accuracy. Besides 3DMM, there have proposed some algorithms which can estimate 3D information from a single image. [23] used the CCA mapping to transform the image from the pixel space to the depth space, [22] adopted the shape-from-shading framework to recover the 3D shape, [21] used the SIFT-FLOW to estimate the depth image and [11] proposed a regression method to transfer the expression from video to 3D model.

However, existing methods still have some disadvantages. For landmark based algorithms [8], [26], [1], since the only input information is the landmark positions, they heavily depend on the face alignment algorithm. Unfortunately in many cases, even though the landmarks look fine on the image, they are not accurate enough for 3D shape fitting, especially those on the eyebrow, nose and contour. Thus directly estimating 3D shape from landmarks is usually unreliable. Traditional analysis-by-synthesis 3DMM fitting algorithms [10], [31] are mainly based on explicitly modelling the image formation process and estimating shape parameters by appearance fitting. It has been shown that these methods heavily rely on the quality of initialization and have to adopt the stochastic optimization to avoid local minimum [2]. As a result, most of them are computationally expensive and always need more than one minute to fit a single face image. Besides, most training sets of 3DMM are very small (100 to 500 samples) due to the difficulty in collecting complete face scans, thus the appearance model of 3DMM is always too weak to cover the large variations of face appearance, especially in the wild. Since the central problem of analysis-by-synthesis framework is fitting the model appearance to the input image, the weak expressive ability of appearance model will lead to non-accurate results.

In this paper, we discuss 3DMM in the context of face alignment and find a new direction to overcome the problems described above. Instead of the traditional analysis-by-synthesis framework, we propose a novel discriminative 3DMM fitting algorithm based on local features and cascade regression. In section 2, we introduce the 3D Morphable Model. In section 3 we highlight our motivation by briefly revisiting face alignment algorithms. Then we propose the discriminative 3DMM fitting in section 4 and discuss some implemental details in section 5. In the experiments, we show that our algorithm outperforms existing 3DMM fitting methods in both accuracy and efficiency.
II. 3D MoraPhable Model

3D Morphable Model [9] is constructed from a set of 3D face scans in dense correspondence. Each scan is represented by a shape-vector $S = (x_1, y_1, z_1, ..., x_n, y_n, z_n)$ and a texture-vector $T = (r_1, g_1, b_1, ..., r_n, g_n, b_n)$, which contain the coordinate and the color of each point respectively. The points are dense enough ($n > 10000$) to directly represent a human face. PCA is applied to decorrelate texture and shape vectors respectively and a 3D face can be described as:

$$S = \bar{s} + \sum_{i=1}^{m-1} \alpha_i \cdot s_i \quad T = \bar{t} + \sum_{i=1}^{m-1} \beta_i \cdot t_i$$  \hspace{1cm} (1)$$

where $m - 1$ is the number of eigenvectors, $\bar{s}$ and $\bar{t}$ are the means of shape and texture respectively, $s_i$ and $t_i$ are the $i$th eigenvectors, $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{m-1})$ and $\beta = (\beta_1, \beta_2, ..., \beta_{m-1})$ are shape and texture parameters determining $S$ and $T$. With the shape and texture model, a synthetic image can be generated by projecting the 3D face onto the image plane by weak perspective projection:

$$s_{2d} = fpr(\bar{s} + \sum_{i=1}^{m-1} \alpha_i \cdot s_i + t_{3d})$$  \hspace{1cm} (2)$$

where $s_{2d}$ is the image coordinates after projection, $f$ is the scale parameter, $p$ is the constant orthogonal projection matrix, $r$ is the rotation matrix and $t_{3d}$ is the translation vector in 3D space. We represent $\gamma = (f, r, t_{3d})$ as the pose parameters of 3DMM.

In the fitting process, the 3DMM solves the parameters by minimizing the Euclidean distance between the synthetic image and the input image:

$$E = \sum_{x,y} \|I_{input}(x,y) - I_{syn}(x,y)\|^2$$  \hspace{1cm} (3)$$

where $I_{input}(x,y)$ is the input image and $I_{syn}(x,y)$ is the 3DMM synthetic image. Usually, the stochastic Gauss-Newton method is adopted to minimize the cost function.

III. Motivation

In general, 3DMM can be seen as a branch of a facial alignment which concentrates on fitting a face model to the input image. Recently, the face alignment in 2D has greatly improved which concentrates on fitting a face model to the input image. Recently, the face alignment in 2D has greatly improved and the 3DMM synthetic image. Usually, the stochastic Gauss-Newton method is adopted to minimize the cost function.

The central problem of face alignment is minimizing the difference between the input image and the face model.

$$P = \arg \min_{P} \|Tr(I) - Tr(M(P))\|^2$$  \hspace{1cm} (4)$$

where $I$ is the input image, $M(P)$ is the face model with parameters $P$ which can give rise to an observed face image and $Tr$ is a transformation which is usually a feature extractor. Based on Equ. (4) explicitly or implicitly, the optimization can be summarized as an iterative updating process:

$$\Delta P_t = RF(I, P_{t-1}) \quad P_t = P_{t-1} + \Delta P_t$$  \hspace{1cm} (5)$$

where $P_t$ is the model parameters in the current iteration, $F(I, P_{t-1})$ is a feature extraction function depending on the input image $I$ and the parameters in the last iteration and $R$ is an updater that maps the features $F$ to the parameter update $\Delta P_t$. Three modules can be seen from Eqn. (5): the model parameter $P$ represents how to model a human face; the feature extractor $F(I, P_{t-1})$ represents what kind of information is used for alignment, and the updater $R$ represents how to use the information.

It is obvious that 3DMM also belongs to the face alignment framework, where $P$ is the PCA coefficients of shape and texture model, $F(I, P_{t-1})$ is either the pixel intensity [9], landmark [1] or Multi-Features [31] and $R$ is constructed from the Jacobian of a cost function like Eqn. (3). It is promising to discuss the achievements of 2D face alignment in recent years and introduce them into 3DMM fitting. In the last decade, a number of seminal works have been proposed to find accurate and robust fitting methods in 2D alignment. We will briefly review these works on the three topics of $P, F, R$ and extend their ideas to reinforce 3DMM fitting.

Active Appearance Model (AAM) AAM [13] which is characterized by its explicit shape and appearance model, has been widely used to match deformable objects to images in early years of face alignment. In AAM, the face shape is defined by a sparse set of landmarks and the appearance is based on the warped images on the reference frame. PCA is applied on landmark vectors and shape-free textures to construct face model just as 3DMM. In fitting process, AAM searches the best parameters that minimize the distance between the model instance and the input image by either the generative fitting [14], [24], [35] which obtains the updater from the Jacobian of the distance function or the discriminative fitting [13], [32], [15] which directly learns a regressor to map the image difference to the parameter update.

Constrained Local Model (CLM) CLM [17], [16], [33], [5], [20], [36], [7] represents an object using local image patches around landmarks. It inherits the PCA shape model from AAM, but discards the holistic appearance model and learns landmark detectors instead. During fitting process, the landmark detectors provide the response maps showing the distribution of probable landmark locations and then the shape parameters are estimated by maximizing the sum of responses of landmarks constrained by a priori. The fitting methods of CLM are also divided into generative methods [33], [20], [36] and discriminative methods [5], while the latter dramatically outstand in both accuracy and robustness.

Non Parameter Model (NPM) NPM [18], [12], [37] further removes any explicit PCA constrains on shape and directly uses landmark coordinates as shape model. Besides, robust features like HOG and SIFT are adopted to describe local spatially-coherent observations of landmarks which are proven to be more robust than image pixel and response map. Furthermore, the cascade regression [18], [6], where independent regressors are trained for each iteration, fully...
Thus we project both the input and the synthetic images into a new space with a transformation $T_r$, where the cost function is smoother.

$$P = \arg \min_p \| f(P) \|^2 \quad f(P) = T_r(I) - T_r(G(P))$$

(6)

where $P$ is the model parameters, $G(P)$ is the synthesis process that can generate an image from 3DMM and $T_r$ is an unknown transformation. For simplicity, we assume $T_r(G(\cdot))$ is differentiable and use Gauss-Newton method to optimize Eqn. (6).

From an initial estimate $P_0$, we apply Taylor expansion to $f(P)$ and minimize the cost function by equally optimizing the following function over $\Delta P$.

$$\arg \min_{\Delta P} f(P_0 + \Delta P)^T f(P_0 + \Delta P)$$

(7a)

$$f(P_0 + \Delta P) = f(P_0) + J_f \Delta P$$

(7b)

where $J_f$ is the Jacobian of $f$. Taking the derivation of Eqn. (7a) over $\Delta P$ and setting it to zero, we get an update to $P_0$.

$$\Delta P_0 = -(J_f^T J_f)^{-1} J_f^T f(P_0)$$

$$= -(J_f^T J_f)^{-1} J_f^T (T_r(I) - T_r(G(P_0)))$$

(8)

It is unlikely that the optimization can converge at a single iteration, thus Eqn. (8) is iterated by several times.

$$\Delta P_t = -(J_f^T J_f)^{-1} J_f^T (T_r(I) - T_r(G(P_{t-1})))$$

(9a)

$$P_t = P_{t-1} + \Delta P_t$$

(9b)

While in fact, we do not know the form of $T_r$ and cannot get the difference in the $T_r$ space. Note that $(T_r(I) - T_r(G(P_{t-1})))$ depends on the image $I$ and model parameters $P_{t-1}$, if we can extract features with $F(I, P_{t-1})$ that implicitly reflects the “goodness” of current fitting and learn a linear regressor $A$ to map the features to the difference in $T_r$ space, we can rewrite $(T_r(I) - T_r(G(P_{t-1})))$ as $AF(I, P_{t-1})$. Then Eqn. (9a) becomes:

$$\Delta P_t = -(J_f^T J_f)^{-1} J_f^T A|_{P=P_{t-1}} F(I, P_{t-1})$$

$$= R_t F(I, P_{t-1})$$

(10)

Note that we merge $A$ into the updater $R_t$ to directly map features to parameter update. According to the Supervised Descent Method (SDM) [37], we can get a list of $R = (R_1, \ldots, R_T)$ through learning instead of numerical approximation. During the testing process, the regressor list will give a sequence of descent directions so that the $P_0$ will converge to the ground truth.

$$P_t = P_{t-1} + R_t F(I, P_{t-1})$$

(11)

How to determine Eqn. (11) is the central problem of discriminative 3DMM fitting. This function depends on the model parameters $P$, regressor $R$, and features extractor $F(I, P)$.

Even though the appearance model of 3DMM is weak, the shape model can describe most of the real-world data due to the relative small variations of face shapes [22]. Besides, the utilizations the flexibility of NPM and keeps the robustness at the same time, making NPM dramatically outperforms any other models and shows the state-of-the-art in face alignment [37].

Table I summarizes AAM, CLM and NPM in the framework of face alignment. In the evolution of face alignment from AAM to NPM, the holistic PCA constraints are progressively removed and more robust features are adopted to handle complicated variations. With the removal of shape constraints, discriminative methods show better performance over generative methods, especially with cascade regression.

These achievements in 2D face alignment can provide promising clues for 3DMM. Current 3DMM fitting methods are based on modelling the physical process of forming a face image and minimizing the difference between the input image and the model appearance, which obviously belongs to the generative fitting framework [32]. It has been shown that this framework heavily relies on the simulation of the image forming process and the quality of initialization [24]. Besides, the minimization is very slow and easy to get stuck in local minimum [2]. Fortunately, these problems could be solved by discriminative fitting method which directly learns a regression based fitting model from a large training set just as NPM.

However, directly applying discriminative fitting framework has some difficulties. Firstly, unlike the sparse landmarks in 2D, the shape model of 3DMM is much denser with tens of thousands of points, which makes the regression matrix too large to learn. Secondly, there is no database containing visual image and complete face scan pairs, leading to the lack of training set for discriminative methods.

In the following sections, we will adopt discriminative methods to fit 3DMM and illustrate how to preprocess existing databases to construct a training set.

IV. DISCRIMINATIVE 3DMM FITTING

In this section, we show how to use the discriminative method instead of the analysis-by-synthesis framework to estimate shape parameters.

A. Derivation of Discriminative 3DMM Fitting

As in [37], we start from the traditional fitting process. Given an image $I$, we want to estimate its 3D shape by minimizing the difference between the synthetic and the input image. However in many cases, the optimization will converge to a local minimum far from the global one [31].
performance of CLM and NPM has shown that it is robust to directly estimate shape information without appearance fitting. Thus we discard the appearance model of 3DMM and only consider the shape PCA coefficients $\alpha$ and the weak perspective projection pose parameters $\gamma$ in Eq. (1)(2) and let $P = \{\alpha, \gamma\}$. In the next two subsections, we provide details of the feature extractor and the regression function.

B. Feature Extraction

This section illustrates how to extract features. The HOG features around landmark positions are used as the feature extractor. We mark a set of landmarks on the 3D model following the Multi-PIE [19] 68 points mark-up, as shown in Fig. 1. In each iteration $t$, with pose and shape parameters $P_{t-1} = \{\alpha_{t-1}, \gamma_{t-1}\}$, the 3D shape is constructed by Eq. (1) and projected to image plane by Eq. (2). Then the HOG on the landmark positions are extracted as the features in the current iteration:

$$F(I, P) = HOG(I, [fpr(\varnothing + \sum_{i=1}^{m-1} \alpha_i \cdot s_i + t_{3d})])$$  \hspace{1cm} (12)

where the $f$, $p$, $r$, $t_{3d}$, $s$, $\alpha$ have the same meanings as in Eq. (1)(2) and the subscript $l$ means only the landmark points are selected.

C. Learning for Regression

In this section, we describe how to learn the regressor list $R = (R_1, \ldots, R_T)$ in Eq. (11) from a training set. Note that Eq. (11) is in fact the process of cascade regression, thus we train independent $R_t$ for each iteration. Given a set of face images $\{I^1, \ldots, I^n\}$, their initial estimates $\{P^1_0, \ldots, P^n_0\}$ and ground truth parameters $\{P^1, \ldots, P^n\}$, we want to minimize the expected loss between the predicted update and the optimal update for all the training samples in each iteration:

$$\arg \min_{R_t} \sum_{i=1}^{n} \| (P^i - P^i_{t-1}) - R_t F(I^i, P^i_{t-1}) \|^2$$  \hspace{1cm} (13)

Writing Eq. (13) as matrix formation, we get:

$$\arg \min_{R_t} \left\| \begin{bmatrix} P^1 & \ldots & P^n \\ P^1_{t-1} & \ldots & P^n_{t-1} \end{bmatrix}^T - \begin{bmatrix} F(I^1, P^1_{t-1}) \\ \vdots \\ F(I^n, P^n_{t-1}) \end{bmatrix} \right\|^2$$  \hspace{1cm} (14a)

$$= \arg \min_{R_t} \| \Delta P_t - R_t F(I_{t-1}) \|^2$$  \hspace{1cm} (14b)

where $\Delta P_t$ is the optimal parameter update (ground truth minus current), and $F(I_{t-1})$ is the features extracted with current parameters for each training sample. Eq. (14b) can be solved directly by linear method:

$$R_t = \Delta P_t F^T((I_{t-1} F^T_{t-1} + \lambda E)^{-1}$$  \hspace{1cm} (15)

where $E$ is the identity matrix and $\lambda$ is the regularization term that avoids over fitting. Usually after each iteration, $P_t$ will be closer to $P^*$ than $P_{t-1}$, and with $P_t$ we have a new training set and can run another iteration with Eq. (15) until convergence. In our experiments, the algorithm converges in 4 to 5 steps.

V. DATA PRE-PROCESSING

To train a discriminative fitting model, we need a database with a large collection of visual images and corresponding 3D face shapes. However, unlike face alignment in 2D, the training set cannot be constructed by hand labelling because 3DMM shape model has tens of thousands of points. While using face scanners like Cyberware [34] or multiple ABW-3D [27] to collect complete face scans is so expensive and troublesome that the number of training samples is limited. The lack of training set is probably the main reason for the absence of discriminative fitting method in 3DMM.

A. Depth Image Registration

Compared with collecting complete face scans, only getting depth images is relatively easier and such work has been done in FRGC [28]. FRGC provides a large database with thousands of visual and depth image pairs in full correspondence, as shown in Fig. 2.
samples). To make the raw data usable, registering a template to depth images for filling holes, estimating missing patches and making every scans in full correspondence are necessary.

The optimal non-rigid ICP algorithm [4] can register a template to any target surface with the same semantic. It searches for the best deformation for each point by minimizing a cost function with 3 terms: the distance term which minimizes the distances between the template points and their closest target surface points, the stiffness term which penalises the differences of the transformations of neighbouring points, and the landmark term which guides the initialization and minimizes the distances between corresponding landmarks. Although the optimal non-rigid-ICP has shown its good performance in the construction of BFM face model [27], it has difficulty in handling large missing regions, because there are no closest points for the distance term in these regions and only stiffness constraint alone will give bad results as in Fig. 3(b).

To deal with large missing patches, we fit a 3DMM to constrain the template points falling onto the missing regions. Since the distance term provides a set of correspondences between template points and target points, a 3DMM can be fitted with the target point positions by common 3D-3DMM fitting methods [3]. Note that 3DMM is controlled by PCA coefficients, the missing regions are automatically estimated. In the registration process, for the template points having no closest points on the target surface, we find their closest points on the fitted 3DMM instead. Thus every point will have a distance term constraint. Fig. 3(c) shows the results of the new method, the filling of missing regions is smooth and looks reasonable.

The registered template can be seen as an approximation of the complete face scan. Since we have known the position of every point of 3DMM, we can get the best fitted pose and shape parameters through:

$$\text{arg min}_{(\alpha, \gamma)} \| \text{Rig} - f_r(s + \sum_{i=1}^{m-1} \alpha_i \cdot s_i + t_{3d}) \|^2$$  \hspace{1cm} (16)

where Rig is the point positions of registered template, $f_r, t_{3d}, s, \alpha, \gamma$ have the same meanings as in Equ. (1)(2). Fig. 4 shows the comparison of depth image, registered template and best fitted 3DMM. We can see that even though the best fitted 3DMM loses some details because of the limited expressive ability of PCA shape model, it is close to the depth image. In the training process, the best fitted parameters will be used as the target of regression.

B. Training Data Augmentation

Considering the success of 2D face alignment in recent years, we use the landmarks detected by SDM [37] to initialize pose and shape parameters. To achieve better generalisation ability, we augment the training set by randomly disturbing the bounding box and running SDM to get multiple groups of landmarks for each sample, as shown in Fig. 5.

It can be seen that the bounding box may affect landmark detection seriously and the eyebrow and border landmarks are always not accurate enough. The non-accuracy of initialization is very common in automated 3DMM reconstruction system and the augmentation is found to be helpful to achieve the robustness to rough initialization.

VI. EXPERIMENTS

We use the Basel Face Model (BFM) [27] as our 3D Morphable Model and conduct fitting experiments on the Spring2004range subset of Face Recognition Grand Challenge (FRGC) [28] database. The BFM provides a PCA shape model with 53490 vertices computed from 200 face scans. It can cover most of face shapes in the real world but cannot handle expressions. In our experiment, we eliminate the ear, neck and forehead regions because they are less important and easily occluded by hair and clothes. The processed model has 39226 vertices left. The Spring2004range has 2114 samples. Each sample consists of a frontal face image and a depth image with pixels in full correspondence. The faces with expression are discarded due to the limitation of BFM, with 1443 samples left.

For each sample in the database, the registration method in section 5.1 is used to get the target parameters and the face alignment algorithm is used to localize the landmarks for initialization. The landmarks are detected automatically using DPM face detector [40] and SDM face alignment [37], and most alignment results are accurate except for few samples (less than 10). As for error measure, we consider the depth image in FRGC as the ground truth shape of each sample. In the testing process, we first project the reconstructed 3D shape to a depth image, and the Root Mean Square Error
(RMSE) between the reconstructed depth image and the ground truth is measured as the fitting error. Invalid regions including holes and missing patches are ignored. Training and testing are conducted in 10-fold cross-validation without identity overlapping to measure all the 1443 samples in the database.

We compare our method with two popular 3DMM fitting algorithms. The first is the Multi-Features Framework (MFF) [31] which considers landmark, contour, textured edge and pixel intensity jointly to fit a 3DMM. It is the state-of-the-art of traditional analysis-by-synthesis based methods. The other is the landmark based fitting method [2] which only uses landmarks to estimate 3DMM shape. It heavily depends on the accuracy of landmark but is much more efficient than MFF. Besides, the method also claims to be the state-of-the-art on a synthetic database with provided landmarks.

Considering facial component region (including eye, nose and mouth regions of BFM in [27]) is more important in most face applications and the cheek area is always occluded by hair in FRGC, we conduct two experiments by computing error on full face and facial component area respectively. Fig. 6 shows the Cumulative Error Distribution (CED) curves of the three methods and Table II shows the mean error of all 1443 samples. Clearly our method outperforms both MFF and the landmark based method. Besides, Fig. 8 shows some fitting results of landmark based method, Multi-Features Framework, and discriminative 3DMM fitting. The resulting shapes are lightened by a frontal light to highlight the difference. Note that in the first row the MFF and landmark based method both fail because the non-accurate landmarks give a bad initialization. However, our method still converges to an accurate shape, showing the robustness to rough initialization. In Fig. 7, we show some fitting results of real world images. Although without ground truth shapes, we can see the fitting results are reasonable in visual sense.

In addition to the fitting accuracy, the running time performance of our method is also promising. It takes about 12 mins to train 6500 samples and 0.9s to fit a testing image on a 3.4GHz Intel Core i7 computer, which is close to the landmark based method (0.2s) and faster than MFF which needs about 69.58s to fit an image on 3.0GHz Intel Pentium IV [29].

Fig. 6. First: fitting accuracy on facial component region. Second: fitting accuracy on full face.

VII. CONCLUSIONS

We have proposed a novel discriminative 3DMM fitting algorithm, where the 3D shape is estimated by directly
regressing image features instead of minimizing a cost function with Gauss-Newton methods. The resulting method is highly accurate, robust and efficient. Experimental results on FRGC suggest that our approach significantly outperforms the popular landmark based method and the state-of-the-art Multi-Features Framework.

Despite the outstanding performance of our discriminative 3DMM fitting method, there is still large room for future improvements. For example, by training the fitting model on a database containing visual and depth image pairs in the wild, it may be able to handle more complicated scenarios.

<table>
<thead>
<tr>
<th>Region</th>
<th>Landmark</th>
<th>MultiFeatures</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facial Component</td>
<td>8.0837</td>
<td>7.8820</td>
<td>5.3505</td>
</tr>
<tr>
<td>Full Face</td>
<td>9.9956</td>
<td>9.5051</td>
<td>6.8849</td>
</tr>
</tbody>
</table>

TABLE II
MEAN ERROR OF LANDMARK BASED METHOD, MFF AND OUR METHOD

VIII. ACKNOWLEDGMENTS
This work was supported by the Chinese National Natural Science Foundation Projects #61203267, #61375037, #61473291, National Science and Technology Support Program Project #2013BAK02B01, Chinese Academy of Sciences Project No. KGZD-EW-102-2, and AuthenMetric R&D Funds.

REFERENCES


[37] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. CVPR, 2013.

