

3D Active Shape Model for Automatic Facial Landmark Location Trained with Automatically Generated Landmark Points

Dianle Zhou, Dijana Petrovska-Delacrétaz and Bernadette Dorizzi
Télécom SudParis TSP (ex GET-INT)

{Dianle.Zhou, Dijana.Petrovska, Bernadette.Dorizzi}@it-sudparis.eu

Abstract

In this paper, a 3D Active Shape Model (3DASM) algorithm is presented to automatically locate facial landmarks from different views. The 3DASM is trained by setting different shape and texture parameters of 3D Morphable Model (3DMM). Using 3DMM to synthesize training data offers us two advantages: first, few manual operations are needed, except labeling landmarks on the mean face of 3DMM. Second, since the learning data are directly from 3DMM, landmarks have one to one correspondence between the 2D points detected from the image and 3D points on 3DMM. This kind of correspondence will benefit 3D face reconstruction processing. During fitting, 3D rotation parameters are added comparing to 2D Active Shape Model (ASM). So we separate shape variations into intrinsic change (caused by the character of different person) and extrinsic change (caused by model projection). The experimental results show that our method is robust to pose variation.

1 Introduction

Finding the correct position of facial landmarks in 2D images is a crucial step for many face applications such as face recognition, modeling or tracking. We are interested in constructing a 3D face model from a single 2D image with pose variation. To achieve that goal, precise landmark location in 2D nonfrontal face images is required [10, 7].

A lot of algorithms are proposed for 2D facial landmark location, reviewed in [3, 9, 15], for images with frontal faces. Compared to frontal faces, automatic facial landmark location on nonfrontal faces is more difficult. Gu et al. [6], use a sparse 3D shape model (trained with synthesized faces) to align faces with dif-

ferent poses, but no evaluations are given. In [2], the authors use Support Vector Machines to detect seven facial landmarks. In [14], several view categories (frontal, half profile, full profile, etc.), with their own shape and texture models are used to build a view-based ASM.

Efraty et al. [4] create training landmarked samples from 3D face database. These landmarks are further employed to train a profile view ASM. In our previous work [15], we proposed the Combined ASM algorithm, which exploits Scale Invariant Feature Transform (SIFT) descriptors [8] as local texture model. We created two separate models related to the facial internal region and facial contour, in order to make the Active Shape Model more robust for nonfrontal faces.

All those systems need manually annotated landmarks for the training phase. Manually annotating landmarks is a tedious and time consuming task.

In this paper, we propose to use a 3D Morphable Model (3DMM) to automatically generate landmarks needed to train a 3D Active Shape Model (3DASM). The advantage of the 3DMM is that, once the set of landmark points is defined on the 3DMM, they can be automatically transposed to the newly generated 3D faces with this model. We use such automatically obtained landmarks to train a 3D Point Distribution Model and a 3D Local Texture Model, that compose our 3DASM. Compared to the 3DMM, the proposed 3DASM is lower resolution 3D deformable model with characteristic facial points. During landmark location, we use Gauss-Newton optimization to adapt our 3D model to 2D images. Unlike 2DASM which can only handle rotations which are present in the manually landmarked training 2D images, the proposed 3DASM can handle much bigger variations of out-of-plan rotations. The advantage of the proposed 3DASM method is that there is no need to manually annotate the landmarks in the training 2D images. It is only necessary to define the set of landmarks that are needed on the 3DMM, that

are going to be propagated automatically on any newly generated 3D faces related to the original 3DMM. Such automatically generated landmarks can serve as training examples.

The rest of paper is organized as follows: reminders of the classical Active Shape Model and the 3D Morphable Model, on which our algorithm is based, are given in Section 2. The proposed 3DASM construction and fitting are explained in Sections 3 and 4 respectively. The results are reported in Section 5. Finally, the conclusions and perspectives can be found in Section 6.

2 Background

Classical Active Shape Model (ASM) is a training based approach [3]. There are two statistical models that exploit the global shape variation and local texture prior knowledge of each training landmark in the fitting phase: Point Distribution Model (PDM) and Local Texture Model (LTM). The PDM is constructed using Principal Component Analysis (PCA) on the training set. A shape is represented as a vector of n landmarks in image coordinates: $S_{2D} = [x_1, y_1, \dots, x_n, y_n]$, so the model is represented by the mean shape \overline{S}_{2D} and deformation parameters p :

$$S_{2D} = (\overline{S}_{2D} + \Phi p), \quad (1)$$

where Φ is the eigenvector matrix of the shape space. PDM represents statistical variations from the training shapes, while LTM is used to update the position of each landmark depending on the Mahalanobis distance between the local texture of each landmark g_{2D} and the mean local texture \overline{g} . The Mahalanobis distance can be denoted as :

$$f(g_{2D}) = (g_{2D} - \overline{g})C_g^{-1}(g_{2D} - \overline{g})^T, \quad (2)$$

where C_g is the covariance matrix of the local texture for each landmark.

3D Morphable Model (3DMM) was introduced by Blanz and Vetter [1]. It is a parameterized model that can generate synthetic 3D faces constructed from a set of 3D facial scans. A vertex-to-vertex correspondence of all 3D training faces is a condition to build a properly working morphable model. Such models are based on the key observation that given two 3D faces, if they are previously registered, their linear interpolation (also known as ‘morph’) will still describe a human face, which make human faces lying in the 3D space intrinsically.

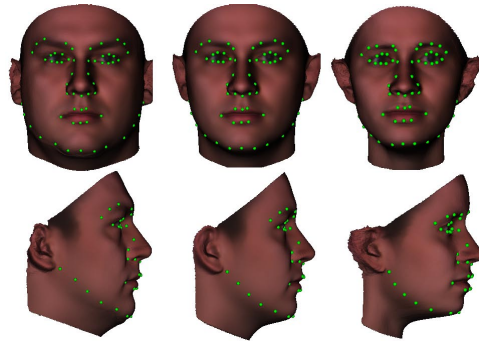


Figure 1. 3D Morphable Model from [1] and the 58 manually selected 3D landmarks. Middle: the average face model with 58 landmarks. Left and right: Change the first component of shape parameter and the correspondent 3D landmarks.

3 3D Active Shape Model (3DASM) Construction Using Automatically Generated Landmarks with a 3D Morphable Model (3DMM)

The 3D Active Shape Model construction using 3D Morphable Model is done as follows:

1. On the average shape of 3DMM, select the vertex points corresponding to the desired landmarks;
2. Choose different sets of shape and texture parameters to generate 3D face data. To simplify the processing, shape and texture parameters of the 100 training faces of the 3DMM are used in this paper. Randomly choosing parameters that resemble to human faces is also possible, but how choose to such parameters is out of scope of this paper;
3. From those synthesized 3D faces, use the positions of selected vertices to construct the 3DPDM (see Section 3.1);
4. Render images from all 3D faces in 7 views by setting roll angle to $(-90, -60, -30, 0, 30, 60, 90)$ degrees;
5. For each view of step 4, extract the SIFT descriptors from previously synthesized images;
6. Build a 2DLTM for each landmark, which composed our 3D view-based Local Texture Model, see Section 3.2.

For comparison purposes, we manually selected on the 3D average face of the 3DMM, the same 58 vertices as the ones defined in the IMM database [12] (see

Figure 1). Thanks to the morphing characteristics of 3DMM, by setting different shape and texture parameters we can obtain different 3D faces with the 3D position of the 58 vertices previously defined. Those vertices could be considered as landmarks in a 3D space, as shown in Figure 1. Like 2DASM, our 3DASM is composed of a 3D Point Distribution Model (3DPDM) and a 3D Local Texture Model (3DLTM), in order to handle statistical information of the 3D shape geometry and texture variations for each landmark.

3.1 3D Point Distribution Model

The 3D Point Distribution Model (3DPDM) is very similar to the 2DPDM, already mentioned in Section 2, except the addition of the z coordinate. A 3D shape can be described by a vector of 3D coordinates $S_{3D} = [x_1, y_1, z_1, \dots, x_n, y_n, z_n]$, and here we use $n = 58$ landmarks. The 3DPDM is obtained from the PCA spaces of the 3D faces with the automatically generated landmarks:

$$S_{3D} = (\overline{S_{3D}} + \Phi_{3D}p_{3D}), \quad (3)$$

where $\overline{S_{3D}}$ is the mean shape in the 3D space, and Φ_{3D} is the eigenvector matrix.

A 2DPDM deals both with intrinsic changes (caused by the change of expression and different persons) and extrinsic changes (caused by camera projection) with a single model. While 3DPDM reflects only the intrinsic changes. The extrinsic change is handled by the camera model and 3D geometric transformation parameters.

The detected 2D shape located in images $S_{2D}^* = [x_1^*, y_1^* \dots x_n^*, y_n^*]$ could be considered as the observation of the 3DPDM projection on the 2D image plan:

$$S_{2D}^{pro} = P(sR(\overline{S_{3D}} + \Phi_{3D}p_{3D}) + t), \quad (4)$$

where P is a projection matrix, R is a 3×3 rotation matrix, t is a translation vector, and s is the scale parameter. In this paper we assume an orthogonal projection

$$\text{where } P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

3.2 3D view-based Local Texture Model

Using SIFT descriptor as Local Texture Model in ASM can make the algorithm more robust to out-of-plan rotations [15]. In this work we use the same SIFT descriptor as [15] but extended it by using view-based statistical models for each landmark. This enables us to increase the ability to deal with large rotations and avoid

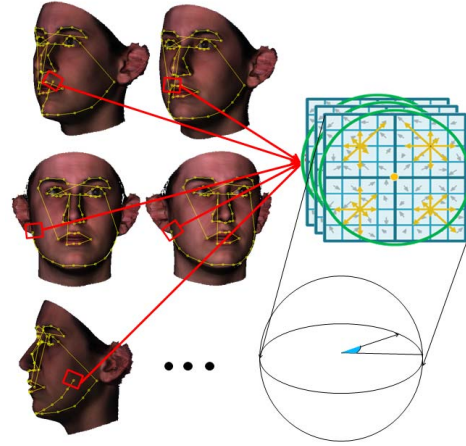


Figure 2. Illustration of 3D Local Texture Model. For each landmark, one 2DLTM is built separately for each view-point. 7 view-based 2DLTM compose our 3DLTM.

the self-occlusion problem during fitting. The concept of 3DLTM is described in Figure 2. In order to describe the landmarks from different view points, we render all 3D faces in 7 different views by setting roll angle to be: $(-90, -60, -30, 0, 30, 60, 90)$. For each landmark, a view-based 2DLTM $(\overline{g^v}, C_g^v)$ is trained for each view v separately. During the fitting (of the 3DASM to 2D images), the LTM are chosen dynamically according to the current pose (see Section 4). As shown in Figure 2, for occlusion points, we extract the SIFT descriptor from training images at the positions where those 3D points are projected. This kind of “virtual” descriptor can make all the landmarks have a uniform presentation, and solves the self-occlusion problem during the fitting procedure.

The advantages of using 3DMM to automatically generate training landmarks are the following:

1. We can generate as much landmarks as the resolution of the 3DMM is capable of with few manual operations.
2. We can also generate faces with landmarks for different subjects with pose, illumination and expression variabilities (if they are present in the 3DMM).
3. Since the 3DPDM and the 3DLTM are directly generated from the 3DMM, the 2D landmarks can be considered as a projection of 3D vertices. This is a strong advantage for 3D face reconstruction.

4 2D Landmark Location: Fitting the 3D Active Shape Model to 2D images

Once the 3DASM is trained, in this section, we explain how to exploit it for landmark location in 2D images (by a fitting procedure).

We construct a two-layered Gaussian pyramid, and apply the alignment algorithm sequentially from the coarsest to the finest layer. Then the algorithm goes iteratively as follow:

1. **Local Search:** For the i^{th} landmark point, compute Mahalanobis distance using the local texture model of the current view around the current location, then select the best candidate (x_i^*, y_i^*) which has the smallest distance as new location.
2. **Parameters Estimate:** The estimation of 3D shape (p_{3D}) and pose parameters ($\mathbf{R}, s, \mathbf{t}$) from 2D shape $S_{2D}^* = [x_1^*, y_1^*, \dots, x_i^*, y_i^*]$ is an ill-posed problem. We consider it as an over-constrained non-linear optimization problem that can be solved by generalized Gauss-Newton iterations as described in 4.1.
3. **Texture Model Update:** The view-based local texture models (\bar{g}^v, C_g^v) is chosen according to pose parameters obtained from step 2.

4.1 Shape and Pose Parameter Optimization

Given the observation shape S_{2D}^* and the 3DPDM S_{3D} , the objective function for the optimization is:

$$E = E_d + \lambda E_p; \quad (5)$$

$$E_d = \sum w_i \|S_{2D_i}^* - S_{2D_i}^{pro}\|; \quad (6)$$

$$E_p = \sum_{j=1}^m \frac{p_{3D}^2}{\delta_j^2}; \quad (7)$$

where E_d is the error between the observation shape and the projected shape. The contribution of the i^{th} difference is weighted with a point specific weight w_i , that is inversely proportional to the Mahalanobis distance of each landmark on the observation shape. The purpose of this weight is to define the quality of the location of each landmark. The weights w_i are normalized between (0.1, 1) and updated dynamically during the fitting. The E_p specifies the a priori term, which constraints the shape deformation to reasonable values. δ_j is the variance (i.e. eigenvalue) associated with the j^{th} eigenshape of the 3DPDM.

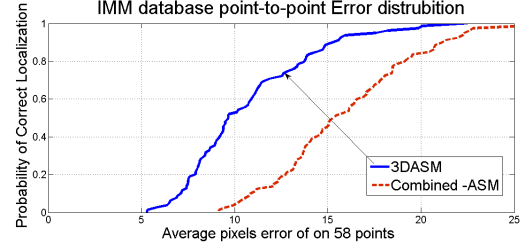


Figure 3. Comparison of 3DASM and Combined ASM on the subset of IMM Database.

At the beginning of optimization, we set λ equal to infinity, so only the pose parameters are optimized. After that λ is set such as E_d is proportional to E_p .

5 Databases and Experimental Results

In this section we first give some details about the database used to automatically generate new landmarks. The evaluation databases are summarized after, followed by the experimental results.

Training database: In this paper, we use the 3D Morphable Model (3DMM) provided with the USF Human-ID 3D Database [1], which is built from 3D scans of 100 individuals. Our method requires to annotate only once the 58 landmarks on the 3DMM. We then automatically "propagate" these 58 landmarks on the shape and texture parameters provided by the USF Human-ID 3D Database to synthesize the 100 face models with the 58 3D-landmarks as our training data.

Evaluation Databases: In order to validate the performance of using automatically generated landmarks for the training phase of the proposed 3DASM, we compared it with our previously reported similar experiments for 2D landmark detection, with the Combined ASM (CASM) method [15]. The CASM was trained with manually annotated images. The test databases include, IMM [12] and PIE [11] database. The part of the IMM database was taken in order to make the comparison with our previous CASM results. As the proposed 3DASM method should be more robust to pose variations, we evaluated it also on the PIE database, which contains more pose variability and for which ground truth information about some landmarks is also available.

Evaluation on the IMM database: From IMM Database we have chosen 80 images which have pose

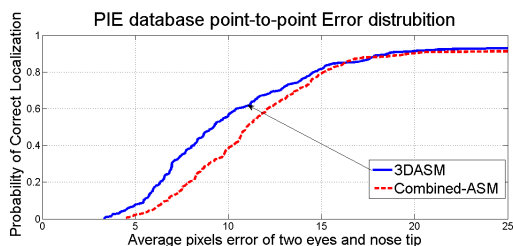


Figure 4. Comparison of 3DASM and Combined ASM on the subset of PIE Database.

variation, and the landmark location result is evaluated on all the 58 points for each image. We take the mean error of the 58 points. From Figure 3, we can see the proposed 3DASM gives better performance than the Combined ASM, therefore validating our proposal of using automatically generated landmarks to train the 3DASM instead of manually annotated landmarks.

Evaluation on the PIE database: In our experiment, we choose 280 images from cameras c05, c29, c11, c37 and without expressions. This limited choice is because the face detector [13] works only for those sets. The ground truth for the landmarks comes from [5], where we have position of eyes and nose tip. As shown in Figure 4, the 3DASM gives better results. By visual inspection we have also noted that the performance is not only better on eyes and nose tip, but also on mouth corners and contour landmarks, which are needed for 3D face reconstruction.

6 Discussion and Perspectives

In this paper, we proposed a 3D Active Shape Model for 2D landmark location on nonfrontal face images. The novelty of our proposal is that we do not need to manually annotate the 2D landmarks on all training 2D images. Instead, we need to annotate manually only once the defined landmarks on the mean face of 3DMM. We use this 3DMM for learning a 3DPDM which describes the prior of intrinsic change caused by the characters of different persons in 3D space, and a 3DLTM which describes the prior of each landmark’s local texture characteristic in different pose separately. Our fitting framework for landmark location is simple and efficient. We mainly compare the performance with our previous CASM (trained with manually obtained landmarks). The results show that our proposed algorithm based on automatically generated training landmarks

gives better performances than CASM. Therefore we have validated the proposal of automatic landmark generation for training.

In this work, we focus on solving the pose variation problem during the landmark location, but expression can be handled by the same framework by increasing expression variability in 3DMM. That will be part of our future work. The proposed automated 2D landmark location will also be further validated in the generation of 3D facial reconstruction from 2D images.

References

- [1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In A. SIGGRAPH, editor, *In Proceedings of SIGGRAPH 99*, pages 187–194, August 1999.
- [2] P. Breuer, K. I. Kim, W. Kienzle, B. Scholkopf, and V. Blanz. Automatic 3d face reconstruction from single images or video. In *FG*, 2008.
- [3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models—their training and application. In *Computer Vision and Image Understanding*. Elsevier Science Inc., 1995.
- [4] B. Efraty, E. Ismailov, I. Kakadiaris, and S. Shah. Towards 3d-aided profile-based face recognition. In *in Proc. BTAS (3)*, 2009.
- [5] R. Gross. <http://ralphgross.com/facelabels>.
- [6] L. Gu and T. Kanade. 3d alignment of face in a single image. In *CVPR*, volume 1, pages 1305–1312, 17–22 June 2006.
- [7] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. zhang. Automatic 3d reconstruction for face recognition. In *FG*, pages 843–848, 17–19 May 2004.
- [8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [9] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. *ECCV*, 2008.
- [10] H. Ren, K. ah Sohn, and S. Kee. Automatic 3d face modeling with single image and perspective projection model. In *IAPR Conference on Machine Vision Applications*, 2005.
- [11] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *FG*, 2002.
- [12] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003.
- [13] P. Viola and M. Jones. Robust real-time object detection. In *International Journal of Computer Vision*, 2001.
- [14] L. Zhang and H. Ai. Multi-view active shape model with robust parameter estimation. In *ICPR (4)*, pages 469–468, 2006.
- [15] D. Zhou, D. Petrovska-Delacrétaz, and B. Dorizzi. Automatic landmark location with a combined active shape model. In *in Proc. BTAS (3)*, 2009.