Applications of 3D morphable models for faces with expressions

B. Chu[1,2], S. Romdhani[1] et L. Chen[2]

[1] Morpho, SAFRAN Group
11 boulevard Galliéni 92130 Issy-Les-Moulineaux - France
{baptiste.chu, sami.romdhani}@morpho.com

[2]Université de Lyon, CNRS
Ecole Centrale de Lyon, LIRIS UMR5205, F-69134
Lyon, France
{baptiste.chu, liming.chen}@ec-lyon.fr

Abstract

In this paper, we present a framework to represent the face of any individual, dealing with identity and expression variation and some applications of this model. A 3D morphable model (3DMM) is a generative method capable to reconstruct the 3D shape of human faces from a small set of coefficients. It is used in many applications such as identity or expression recognition, 3D scans processing or face animation. Such a 3D face generative model is used to build any face with expression as linear combination of deformations. In order to determine these basis deformations and with the hypothesis of Gaussian distribution of the 3D human faces space, a PCA is computed on a registered dataset. The registration of training 3D faces is first achieved so that it is robust to expression. To build the 3DMM with identities variations separated from expressions variations, two PCA are computed. Using this morphable model, a new face can be represented as a linear combination of these principal modes of variations. We show in this paper that such a morphable model can be used as shape prior in many applications like denoising or occlusion recuperation tool.

Keywords

3D faces, Facial modelling, Expression, Morphable Model, Registration, Denoising, Occlusion

1 Introduction

Owning to its numerous applications, modelling human faces is a popular research topic in the Computer Graphics community. Various methods have been proposed [1] [2] [3]. Some of these methods are based on 3D morphable model (3DMM) techniques [4] [5]. Learned from 3D scans, a 3DMM can approximate the face of any individual with a linear combination of deformations. As well as modelling the identity variations [4], morphable models can be extended by adding new modes of variations [6] : the expression.

There exist different methods to build a morphable model with expression. The first one, based on two PCA, has been introduced in [6]. The dataset is separated in two sets : the first with neutral scans and the second with expressive scans. A PCA is computed on each set to find the principal modes of variation of identity and of expression. In [7], Vlasic et al. introduce a second method based on tensor where interactions between identity and expression are taken into account. To estimate theses interactions, a large and representative ensemble of individuals must be acquired.
Given the complexity of getting such a database, we use in this paper the 2PCA method : Two PCA are computed on a registered training dataset to extract the principal modes of variation of identity and expression.
We describe in this paper all steps necessary to build this morphable model, in particular the registration of 3D face scans, robust to expression and to partial scans. Applied to the training dataset, registration transforms every scans to share the same geometrical topology.

As a result, the 3D morphable model is able to generate any individual face with expression. In this work, we present applications using the 3DMM as a shape prior to reconstruct occluded parts.

We start with a description of the 3D scans database. Then, we present the registration framework, based on a smoothness energy minimization and the construction of the 3D morphable model. Finally, some applications using the 3DMM as shape prior are presented like denoising or occlusion recuperation.

2 Data

In order to build an expression and identity morphable model, a representative training dataset of 3D face scans with variations in expression and identity is needed. There exist in the literature several public 3D face datasets with expressions, e.g., BU-3DFE [8], Bosphorus [9], D3DFacs [10].
With only 10 individuals, the D3DFacs dataset is too small to train a statistical model. Bosphorus, however, contains significant variations of expressions and subjects but their 3D scans contain only cropped facial region. To get an accurate model of human faces, the most complete possible scans are needed to analyse as many facial deformations as we can. Due to their cropped scans, Bosphorus is not used as training dataset.
BU-3DFE provides a significant set of expressive 3D face scans of 100 subjects with 4 intensity levels. Theses scans, simultaneously acquired from two viewpoints, are ranging from one ear to the other. Thus, we use the BU-3DFE database as training dataset.
In this database, each subject is asked to perform seven universal expressions (Neutral, Anger, Disgust, Fear, Happy, Sad, Surprise). For the non-neutral expressions, four stages of intensity are recorded. In our work, we use the neutral and the six highest intensity expression scan. Thus, the training set contains 700 scans.

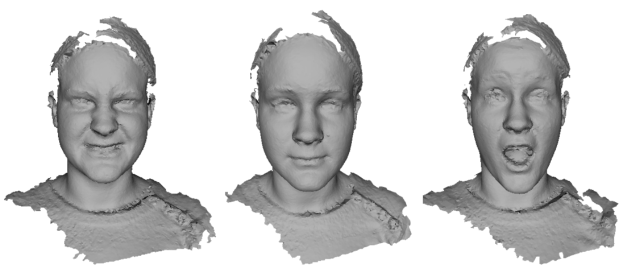

Figure 1 – *Angry, neutral and surprise scans*

3 Registration framework

Before performing a statistical analysis, each scan of the dataset must be brought into full correspondence with others by the registration framework presented in this section. Amberg in [11] proposed a registration method which allows a linear resolution.

3.1 Method

The correspondence between two scans is given by fitting a template surface, *T*, to each of the scanned surfaces, *S*. To get a good fitting between template and scan, the distance between each template's vertex and the surface of the scan is minimized.

The data error measures the matching quality between the template and the scan. It is computed as the sum of the squared distance between each template's vertex and the closest point on the scan surface :

$$E_{data} = \sum_{i=1}^{n_v} dist^2(T_i v_i, S) \qquad (1)$$

where n_v is the number of vertex in the template, T_i is the transformation applied to the template vertex v_i and $dist$ gives the distance between a vertex and the closest point of the scanned surface *S*.

When only data error is used, the system is under-constrained : As scans are incomplete, some vertices of the template do not have correspondence. Moreover, we want to keep vertex labelling : the tip of the nose of the template must remain the tip of the nose after registration. For these two reasons, a regularisation term is required. Some regularisation techniques are based on surface or volume preservation. They can be used when finding the deformation of a real object. In our framework, such a method cannot be used because of the variability between shapes of different heads of individual. Therefore, a regularization energy, as proposed by Popovic et al. [12] (minimization of the variation between two neighbour transformations) or Amberg [11] (minimization of the Laplacian of the surface) is used. In our framework, we used the smoothness energy proposed by Amberg : the smoothness error is defined as the norm of the Laplacian of the surface :

$$E_s = \int_S \left|\left|\nabla^2 T(x)\right|\right|^2 dx \qquad (2)$$

where $T(x)$ is the transformation function applied to the 3D coordinates of vertex x of the surface S:

$$T : \begin{cases} \mathbb{R}^3 \to \mathbb{R}^3 \\ x \mapsto Ax + t \end{cases} \qquad (3)$$

with A the affine part and t the translation part of the deformation.

Then, the global energy equation is defined as :

$$E(v) = E_S(v) + \alpha E_{data}(v) \qquad (4)$$

$$E_S(v) = \left\| D \begin{bmatrix} v \\ n(v) \end{bmatrix} \right\|^2 \qquad (5)$$

$$E_{data}(v) = \|Cv - c\|^2 \qquad (6)$$

where v and $n(v)$ are the vertices and the normals of the facets of the template, c refers to the coordinates of the points of the surface S in correspondence with template's vertex and C and D are suitable sparse matrices that only depend on the template initial shape and topology.

To improve the registration quality, some constraints are added referring to the data error. First, in order to align every scans at the neck, vertices of the neck boundary are projected onto a plane which is computed after a first rigid registration.
The scans in the dataset are acquired by a frontal device. Thus, the lack of information in the back of the head leads to poor deformations of this part of the template. We add a hard shape prior, with a back shell positioned after rigid registration, to get a well-deformed template.
Because we work with expressive scans, annotated mouth boundaries (with a Bezier curve) are used to improve the expression robustness of the registration process.
For each mouth vertex in the template, the corresponding point on the Bezier curve in the scan is found on the plane normal to the template's mouth boundary.

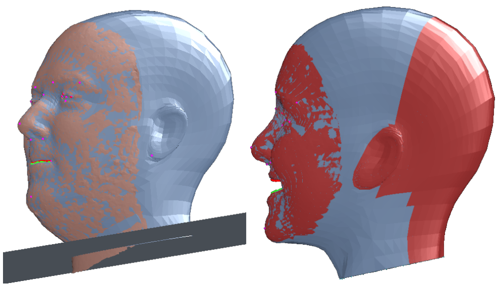

Figure 2 – *Additional constraints*

3.2 Pre-processing

To improve the quality of the registration process, we apply some pre-processings on the dataset. Some of the scans may contain non-informative parts (like clothes, hair or interior of the mouth). These parts could lead to registration errors due to wrong correspondences between template's vertex and point on the surface of the scan. In order to only keep the parts of the scans represented in the template, we clean the 3D scans by removing the non-informative parts. Wrong correspondences could also exist in the area of the mouth. To prevent these errors, we annotate the mouth boundary on the 3D scan.

At the begining of the registration, a first rigid transformation is computed to align the template and the scan. Some fiducial points are used to process this initialization. We manually annotate these points (like the nose tip, the corner of the mouth and the eye center) during the pre-processing.

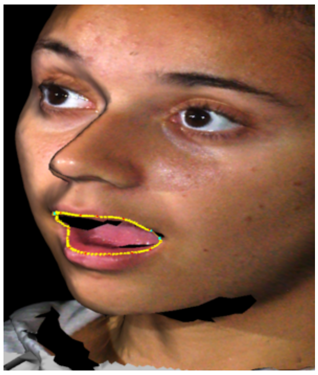

Figure 3 – *Manual annotations*

3.3 Minimization

Finally, the global energy function given in equation (4) is minimized. In this equation, α is weighting coefficient used to tune the optimization.

In equation (5), template's normals $n(v)$ are needed at each iteration. As (4) is a nonlinear equation, Amberg [11] introduces a new set of variables $\overline{n}$ for the normals and a new cost which makes $\overline{n}$ and the last iteration normals $n(v^{t-1})$ as close as possible.

$$E(v,\overline{n}) = E_S(v,\overline{n}) + \alpha E_{data}(v,\overline{n}) + \beta E_{normal}(v,\overline{n}) \quad (7)$$

$$E_S(v,\overline{n}) = \left\| D \begin{bmatrix} v \\ \overline{n} \end{bmatrix} \right\|^2 \quad (8)$$

$$E_{data}(v,\overline{n}) = \|Cv - c\|^2 \quad (9)$$

$$E_{normal}(v,\overline{n}) = \left\|\overline{n} - n(v^{t-1})\right\|^2 \quad (10)$$

With this approximation, equation (7) is linear. Given this linearity, we use a Gauss-Newton algorithm to solve iteratively the system. Hence this optimisation has a single global optimum and it can be found efficiently.

3.4 Results

Some registration results are shown in the following table. The first row refers to a neutral scan, the second and the third to disgust and surprise scans. Error between feature points on the template and on the scan and error of the distance between template and scan surface are used as quality metrics. The distance are given in percentage of inter-eye distance.

3D input scan	Feature point error	Point-Surface error
Neutral scan	0.017 %	0.061 %
Disgust scan	0.016 %	0.089 %
Surprise scan	0.007 %	0.052 %

These results show that the surface of the scan is well approximated by the template during registration. Also, they show the robustness of our method to expressive scans. Furthermore, given the linearity of the system, the computational time is quite low (about 2 minutes per scan with an Intel Core i5 processor at 2.5GHz).

As in previous work [13], we also judge the quality of the registration by analysing caricatures of registered scans. Registration artifacts can occur on caricatures when registration problems exists (e.g. nose registered to lips or neck to chin). Drawing a caricature consists of exaggerating the displacement of the vertex of the registered scan. Vertices v_{caric} of the caricature are computed as :
$v_{caric} = v_{template} + coeff_{caric}(v_{registered} - v_{template})$

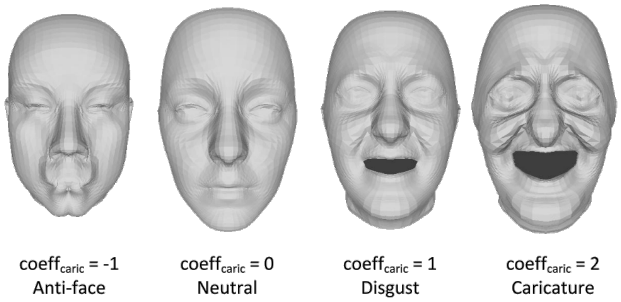

Figure 4 – *Caricatures computed from disgust scans : Anti-face, neutral, disgust and caricature*

4 3D morphable model

The morphable model is based on a training dataset which is BU-3DFE in this work. Thanks to the registration step described previously, a principal component analysis (PCA) can be applied to those registered 3D face scans to extract the principal modes of variation.

Each 3D face shape is represented by $S_{i,e} = (X_1, Y_1, Z_1, X_2, ..., Y_{n_{vert}}, Z_{n_{vert}}) : 3n_{vert} \times 1$, where e refers to the expression (0 for neutral and 1-6 for expression) and i in $1...n_{id}$ is the identity index. First, each face is centred by removing the mean shape $\overline{S}$. Separating the dataset in two sets (one for neutral faces and the second for expressive scans), two PCA are computed (one on each set).

The first one, computed on neutral scans $S_{i,0}$, gives an identity morphable model. New neutral faces can be generated by varying the identity coefficients vector $\alpha_{identity} : (n_{id} - 1) \times 1$:

$$S = \overline{S} + A_{ide}\alpha_{id} \tag{11}$$

The second PCA models the deformations due to expressions. A PCA is thus performed on the offsets between expressive scans and neutral scans : $\Delta S_{i,e} = S_{i,e} - S_{i,0}$ for $e = 1...6$. Face deformations due to expression can hence be generated by varying the expression coefficients vector $\alpha_{expression} : (6n_{id} - 1) \times 1$:

$$\Delta S_{expression} = A_{exp}\alpha_{exp} \tag{12}$$

Combining equation (11) and equation (12), we can generate any face with identity and expression variations :

$$S = \underbrace{\overline{S} + A_{id}\alpha_{id}}_{identity} + \underbrace{A_{exp}\alpha_{exp}}_{expression} \tag{13}$$

$$S = \overline{S} + \begin{bmatrix} A_{id} & A_{exp} \end{bmatrix} \begin{bmatrix} \alpha_{id} \\ \alpha_{exp} \end{bmatrix} \tag{14}$$

A_{id} and A_{exp} are truncated to keep only the most significant expression and identity vectors. Keeping the first 40 identity vectors and the first 40 expression vectors of the 3DMM allow a compact and accurate representation of the face space. In the following, we use this truncated model.

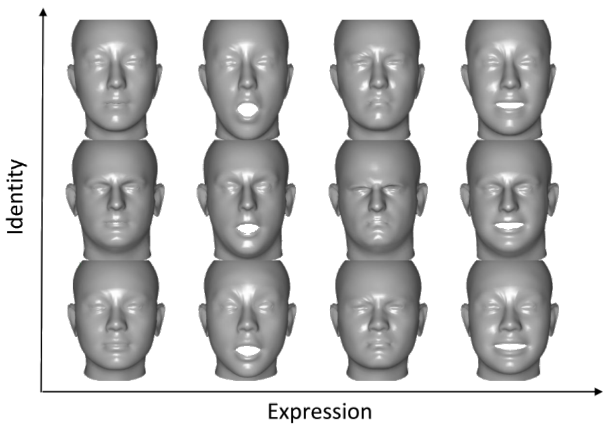

Figure 5 – *Faces generated by the 3DMM with identity and expression variations.*

5 Applications of 3D morphable model

The morphable model presented in the previous section can be used as a generator of realistic faces. Because we have expression coefficients, it can also be used to animate synthetic faces by changing expression [14].

3D model acquisition technologies often produce noisy or occluded data. The registration process and the 3DMM can be used to denoise or reconstruct scans.

5.1 Denoising

3D scans can be obtain with different technologies, which often produce noisy data. In the most of 3D mesh analysis, denoising is the first preprocessing step.

During the registration, the deformation energy, based on the Laplacian of the surface, is minimized to get a smooth registered scan. When applied to a noisy scan, registration can work as a denoising filter. An example of noisy mesh registration is shown on Figure 6 .

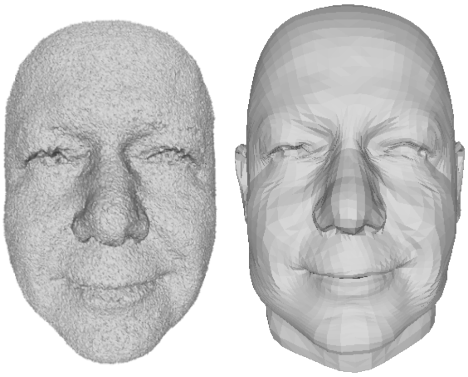

Figure 6 – *Scan denoising*

Given the noisy frontal 3D face on the left, the registration process gives a smooth registered scans presented in the right. The use of registration as denoising tool is a example of the current framework's application.

5.2 Reconstruction of occluded parts

During 3D acquisition, some occlusions due to hair, hand or eyeglasses can occur. To reconstruct the occluded parts of the scan, the morphable model can be used as shape prior. We use scans from the Bosphorus database [9] which some occlusions.

To reconstruct the occluded parts of the scan, we apply some preprocessing. First, the non-informative parts (hair, clothes, ...) of the scans are removed and the registration specific preprocessing (feature points and mouth boundary annotations) are carried out.

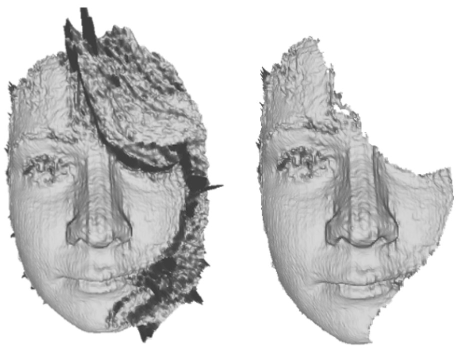

Figure 7 – *Occluded scan pre-processing*

After pre-processing, we applied the registration process, described in previous section, to the cleaned scan. In addition to give a registered scan, the registration process gives the list, $v_{corresp}$, of templates's vertices in correspondence. Thus, we know which part of the template refers to an occluded part of the scan. Using this information, the 3DMM is fitted to the registered scan using only the vertices in correspondence during the registration :

Coefficients α_{id} and α_{exp} of the morphable model

$$S = \overline{S} + \left[\begin{array}{cc} A_{id} & A_{exp} \end{array} \right] \left[\begin{array}{c} \alpha_{id} \\ \alpha_{exp} \end{array} \right]$$

are determined to minimize the distance between the template and the scan, defined by :

$$\sum_{v \in v_{corresp}} dist(v, S)$$

where $dist$ gives the distance between the vertex v of the template and the closest point on the surface of the scan.

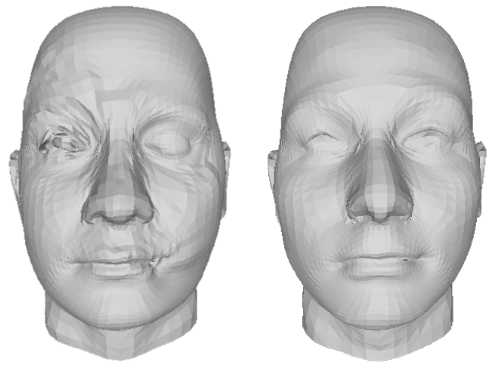

Figure 8 – *Registered scan and 3DMM fitting*

6 Conclusion

In this paper, a complete framework is presented to build a 3D morphable model able to generate 3D faces with identity and expression variations. To build the training database, we use a registration process robust to expression. Then, two PCA are computed on neutral and expressive scans. As a result, the morphable model is able to generate any individual face with expression. Finally, denoising and occlusion reconstruction are presented. The first one shows that the registration process, giving smooth meshes, can also be used as a denoising tool. The second one, an example of using the morphable model as shape prior, deals with reconstruction of 3D scans with occlusions. With a larger dataset, the morphable model can also be used as a compression method where faces are represented by a set of coefficients.

This process is only based on manual annotations.

Given the lack of facial markup, our morphable model framework can be extended to also deal with texture maps. The robustness of our model, specially in the expression area, can be improved by using a new template's meshing as proposed in [15] but with an higher registration computing time.

The accuracy of the morphable can also be improved by using a larger 3D faces databases. Currently, training dataset contains only expressions limited to the prototypic expressions space of 100 individuals. Furthermore, in the dataset used in this paper, scans cover only a partial view of the full 3D face. Despite the back-shell used during the registration, a better morphable model can be obtained with full 3D scans.

References

[1] Douglas DeCarlo, Dimitris Metaxas, and Matthew Stone. An anthropometric face model using variational techniques. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, page 67–74, 1998.

[2] Frédéric Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David H Salesin. Synthesizing realistic facial expressions from photographs. In *ACM SIGGRAPH 2006 Courses*, page 19, 2006.

[3] Gareth J Edwards, Andreas Lanitis, Christopher J Taylor, and Timothy F Cootes. Modelling the variability in face images. In *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, page 328–333, 1996.

[4] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, page 187–194, 1999.

[5] Brian Amberg, Pascal Paysan, and Thomas Vetter. Weight, sex, and facial expressions: On the manipulation of attributes in generative 3D face models. *Advances in Visual Computing*, page 875–885, 2009.

[6] Volker Blanz, Curzio Basso, Tomaso Poggio, and Thomas Vetter. Reanimating faces in images and video. In *Computer graphics forum*, volume 22, pages 641–650. Wiley Online Library, 2003.

[7] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. Face transfer with multilinear models. In *ACM Transactions on Graphics (TOG)*, volume 24, page 426–433, 2005.

[8] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale. A high-resolution 3D dynamic facial expression database. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, page 1–6, 2008.

[9] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Bosphorus database for 3D face analysis. In *Biometrics and Identity Management*, page 47–56. Springer, 2008.

[10] Darren Cosker, Eva Krumhuber, and Adrian Hilton. A facs valid 3d dynamic action unit database with applications to 3d dynamic morphable facial modeling. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2296–2303. IEEE, 2011.

[11] Brian Amberg. *Editing faces in videos*. PhD thesis, University of Basel, 2011.

[12] Brett Allen, Brian Curless, and Zoran Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM Transactions on Graphics (TOG)*, volume 22, page 587–594, 2003.

[13] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, page 1–8, 2007.

[14] Tony Ezzat, Gadi Geiger, and Tomaso Poggio. *Trainable videorealistic speech animation*, volume 21. ACM, 2002.

[15] Stefano Pasquariello and Catherine Pelachaud. Greta: A simple facial animation engine. In *Soft Computing and Industry*, page 511–525. Springer, 2002.