

3D face computational photography using PCA spaces

Jesús P. Mena-Chalco · Ives Macêdo · Luiz Velho · Roberto M. Cesar Jr.

Published online: 3 June 2009
© Springer-Verlag 2009

Abstract In this paper, we present a 3D face photography system based on a facial expression training dataset, composed of both facial range images (3D geometry) and facial texture (2D photography). The proposed system allows one to obtain a 3D geometry representation of a given face provided as a 2D photography, which undergoes a series of transformations through the texture and geometry spaces estimated. In the training phase of the system, the facial landmarks are obtained by an active shape model (ASM) extracted from the 2D gray-level photography. Principal components analysis (PCA) is then used to represent the face dataset, thus defining an orthonormal basis of texture and another of geometry. In the reconstruction phase, an input is given by a face image to which the ASM is matched. The extracted facial landmarks and the face image are fed to the PCA basis transform, and a 3D version of the 2D input image is built. Experimental tests using a new dataset of 70 facial expressions belonging to ten subjects as training set show rapid reconstructed 3D faces which maintain spatial coherence similar to the human perception, thus corroborating the efficiency and the applicability of the proposed system.

Keywords 3D face reconstruction · Principal components analysis · Computer vision · Computational photography

1 Introduction

The process of building 3D facial models is a relevant topic in computer vision which has recently received attention within the research community. This is an example of the so-called computational photography, where computer vision and graphics methods are used to solve a given problem. Modeling facial data depends on the nature of the considered problem. Usually, models with accurate geometry are preferred for face recognition, and simpler models are more suitable in applications where the speed of the process is a critical factor [7], e.g., face transmission or augmented reality.

Face images play a central role in different applications of computer vision and graphics. Different methods for 3D face detection, tracking, and representation have been developed to address applications such as face recognition [1, 12, 20, 26], facial expression analysis [22, 27], face synthesis [15, 28], and video puppeteering [3, 11]. As far as face synthesis is concerned, most 3D face reconstruction methods proposed so far are based on artificial 3D face models such as public available avatars [3]. Jiang et al. [12] explored an alignment and facial feature extraction algorithm for automatic 3D face reconstruction. Their algorithm subsequently applies principal component analysis on the shape to compute 3D shape coefficients. However, a frontal face image of a subject with normal illumination and neutral expression is required.

J.P. Mena-Chalco (✉) · R.M. Cesar Jr.
IME—Universidade de São Paulo, Rua do Matão, 1010, São Paulo, SP 05508-090, Brazil
e-mail: jmena@vision.ime.usp.br

R.M. Cesar Jr.
e-mail: cesar@vision.ime.usp.br

I. Macêdo · L. Velho
Instituto Matemática Pura e Aplicada, Estrada Dona Castorina, 110, Rio de Janeiro, RJ 22460-320, Brazil

I. Macêdo
e-mail: ijamj@impa.br

L. Velho
e-mail: lvelho@impa.br

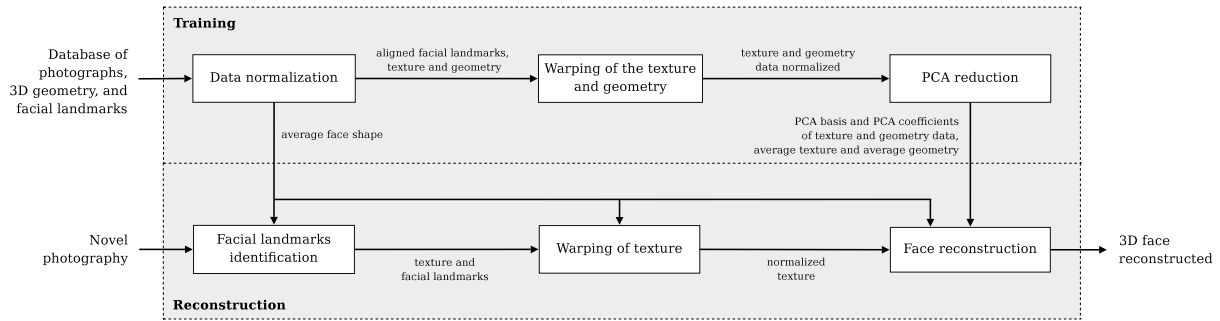


Fig. 1 Schematic data flow diagram of the 3D facial reconstruction system (dotted lines). Each block (solid lines) represents a process, while each arrow represents the information flow between processes

Despite more than three decades of research [13, 14], there are still some important 3D face photography open problems. This paper presents a new approach for 3D face computational photography using real-data based models. The main contributions of this paper rely on the new approach itself which, because of being based on real geometry data, produces more realistic 3D reconstructed faces.¹ The system works with few training samples and relies on standard vision and graphics algorithms and representations, thus leaving space for different improvements in the future.

Starting from the work of Vlasic et al. [24] and Macêdo et al. [18], an automatic system for 3D face reconstruction from 2D color images using a small training dataset of range images (registered texture and geometry data) has been created. It is worth noting that these previous works [18, 24] do not explore 3D data. The training set is composed by a small set of range images corresponding to seven different facial expressions of ten subjects. Our approach employs principal components analysis (PCA) to represent the face model (texture and geometry separately). In the system, the PCA face model is composed by two separate orthonormal basis which represent texture and geometry, respectively.

Given an input frontal face image to be 3D reconstructed, an active shape model (ASM) is used to extract the 2D facial feature points (landmarks) on 2D gray-level images. The set of feature points is used to normalize the input texture. The normalization is obtained by cropping the texture and geometry data and warping to the average face shape. The 3D facial geometry is produced by projecting the normalized texture onto the geometry space (obtained in the training procedure). The projection is produced by using a PCA vectorial basis and a linear optimization function to relate 2D texture and 3D geometry information. Finally, the 3D reconstruction is obtained by directly mapping the normalized texture onto the geometry. Figure 1 summarizes the proposed system architecture.

¹Preliminary results have been described in a conference [19].

This paper is organized as follows. The acquisition protocol used to create the dataset is described in Sect. 2. An overview of the proposed mathematical scheme is presented in Sect. 3. The introduced system is described in Sect. 4. Experimental results and discussions are shown in Sect. 5. The paper is concluded with some comments on our ongoing work in Sect. 6.

2 Acquisition protocol and face model

3D face data is acquired using a non-contact 3D scanner KONICA MINOLTA VIVID 910. The scanner is composed by a laser distance sensor and a digital camera. The scan volume specifications are: $111 \times 84 \times 40$ mm (min) to $1200 \times 903 \times 400$ mm (max) (width \times depth \times height, respectively). Texture images have been acquired with 640×480 pixels (24 bits). The acquisition time was of 2.5 seconds for the geometry data. The 3D geometry associated to each texture image contains approximately 50000 points.

The data is hence composed by registered texture and geometry data. Devices that provided controlled lighting conditions were used. We considered four halogen reflector lamps as shown in Fig. 2. The digital camera had the white balance properly adjusted.

We focused our study on the analysis of different male and female subjects, each one assuming one neutral pose and six distinct facial expressions (joy, sadness, surprise, anger, disgust, and fear). During acquisition, subjects were in frontal position with eyes open. They were not allowed to use eyeglasses or other objects that may modify the face appearance, but there were no restrictions on clothing, hairstyle or haircut. Figure 3 shows an example of a subject belonging to the dataset. The dataset also includes another 5 strongly differentiated facial expressions that may be explored in future applications: mouth and eyes open, mouth and eyes close, kiss expression, and wink expression.

The range images have been pre-processed to fill data gaps obtained in the data acquisition using a Laplacian algo-

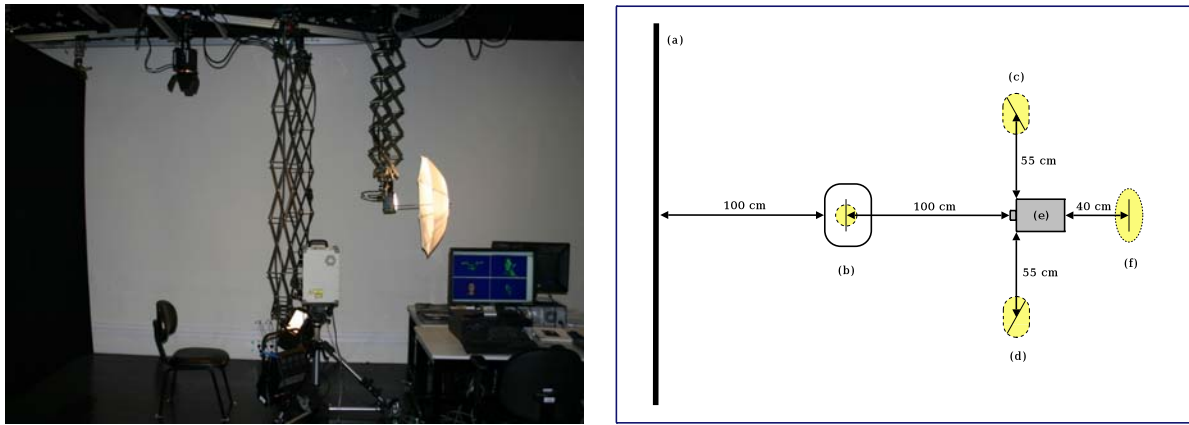


Fig. 2 Equipment used for data acquisition. Equipment setup (*left*) and corresponding illustrative diagram (*right*): (a) black canvas, (b) 600-watt halogen reflector lamp, (c)–(d) 100-watt halogen reflector lamp, (e) 3D scanner, (f) 1000-watt halogen reflector lamp



Fig. 3 Example of acquired data using the 3D scanner: (a) frontal faces with different expressions of one subject, (b) corresponding face geometry

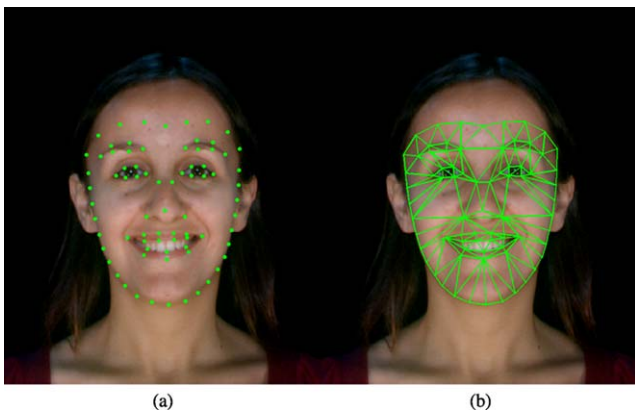


Fig. 4 Example of (a) facial landmarks and (b) triangulation considered in our approach

rithm. Therefore a depth value is associated with each pixel of the texture images.

Once the training images have been acquired and pre-processed, facial landmarks are manually placed over the texture images and aligned to the corresponding range images. We adopted a model with $K = 77$ landmarks and a triangulation that contains 120 elements (triangles). Fig-

ure 4 shows an example of facial landmarks and the considered triangulation.

3 Mathematical model overview

The list below summarizes the symbols used in the current paper, being presented to help the reader:

- l_i : landmarks of the i th face
- \bar{l} : average face shape
- x_i^t, x_i^g : the i th texture face and corresponding geometry belonging to the dataset
- $\mathbf{x}^t, \mathbf{x}^g$: normalized input texture face and the corresponding output reconstructed geometry, respectively
- L_i^t, L_i^g : texture and geometry of the i th face, cropped and warped to \bar{l} , respectively
- E^t, E^g : texture and geometry PCA basis
- α^t, α^g : texture and geometry coefficients expressed in terms of E^t, E^g , respectively
- s_x : weighting coefficients of \mathbf{x}^t in terms of the training samples

The proposed approach is based on learning a 3D face model using texture and geometry of a training face for some

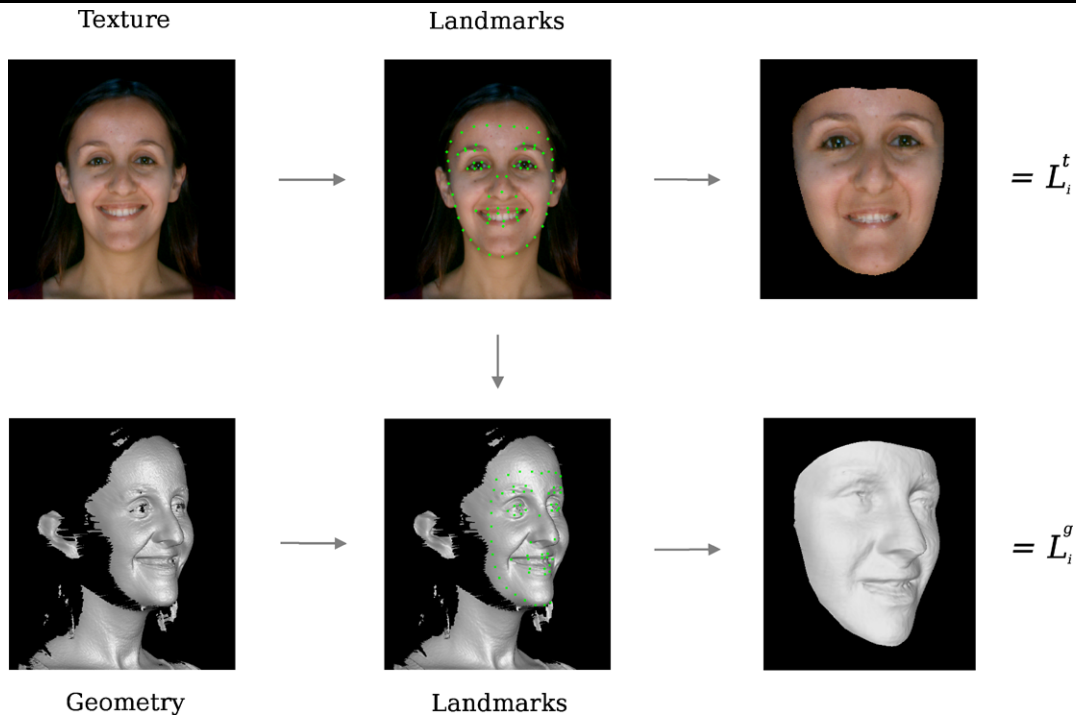


Fig. 5 Training set formation: texture and geometry of the i th sample. The same facial landmarks are used in the texture and geometry of the subject

different facial expressions. An input 2D face image (i.e., only texture) is then reconstructed by projecting it on the trained 2D texture space, decomposing it as weights of the training samples. The obtained weighting coefficients are then used to build a 3D model from the 3D training geometry samples.

The training dataset is composed by pairs of texture and geometry data from subjects with some different facial expressions. A set of landmarks $\{l^1, l^2, \dots, l^K\}$ is placed on the texture image and used to represent the input texture information. Therefore, each facial texture is represented by L^t composed by the landmarks information. Because texture and geometry data are registered, the texture landmarks have corresponding geometry counterparts, which are used to define the facial geometry L^g . This initial landmarks representation scheme is illustrated in Fig. 5.

The training phase consists in defining good texture and geometry space representations based on a given set of facial expression samples from the subjects. Therefore, texture and geometry are obtained for different facial expressions of N subjects, being denoted as $\{L_1^t, L_2^t, \dots, L_N^t\}$ and $\{L_1^g, L_2^g, \dots, L_N^g\}$, respectively. This data helps to define the initial texture and geometry spaces, as illustrated in Figs. 6(a) and 6(e). In order to have a more efficient and statistically optimized representation, both texture and geometry spaces are PCA-transformed (Figs. 6(b) and 6(d), respectively). Each training sample represents a vector expressed in these spaces.

The main goal of the 3D photography system is to obtain a geometry representation of a given face \mathbf{x} , provided as a texture image. A set of facial landmarks is automatically identified, and a normalized input texture \mathbf{x}^t is extracted from such input image and undergoes a series of transformations through the texture and geometry spaces, as illustrated in Fig. 6. The final result is the reconstructed geometry of the input face \mathbf{x}^g , i.e., a point in the geometry space.

4 The system

4.1 Training steps

The training procedure is composed by three phases. Firstly, the input landmarks are normalized by Procrustes analysis [6, 9], thus resulting in a dataset with landmarks aligned \bar{l} , in a common coordinate system. Procrustes analysis allows one to remove the translational, rotational, and scaling components of the shapes, defined by the landmarks, in order to find the best fit of all landmarked shapes. This analysis aligns iteratively all shapes so that the sum of the distances of each shape to the average is minimized, i.e., it minimizes the distance $\sum_{i=1}^N |l_i - \bar{l}|^2$ [9]. The facial landmarks may be aligned for the different input images because of homology among the individual representations. This fact allows one to map each input image by warping it onto the average face

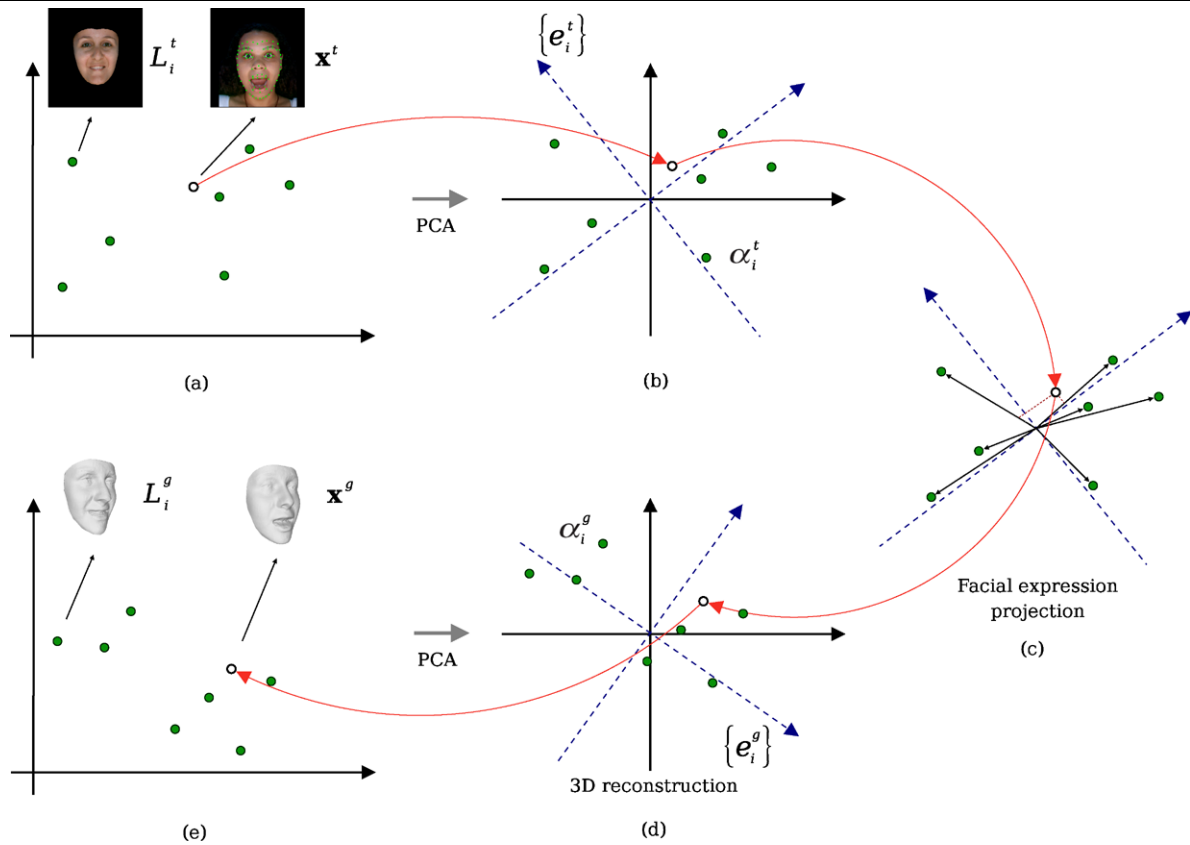


Fig. 6 Texture and geometry spaces: \mathbf{x}^t is an input texture face which undergoes a series of transformations through texture and geometry spaces until its geometry \mathbf{x}^g is built

data using the previously defined triangulation. The geometry data is also mapped onto the average face, produced by the Procrustes analysis.

Two PCA procedures are carried out separately for the geometry L_i^g and for the texture L_i^t data. This analysis leads to:

- An average texture model (t_0), an orthonormal basis ($E^t = \{e_i^t\}$) for the facial texture space, and the coefficients ($\{\alpha_i^t\}$) for each texture image in the training dataset expressed w.r.t. $\{e_i^t\}$. See Fig. 8(a) for an example.
- An average geometry model (g_0), an orthonormal basis ($E^g = \{e_i^g\}$) for the facial geometry space, and the coefficients ($\{\alpha_i^g\}$) for each 3D geometry data in the training dataset expressed w.r.t. $\{e_i^g\}$. See Fig. 8(b) for an example.

In order to work with the same number of principal components in the aforementioned spaces, we use the minimum amount of components representing a pre-defined amount of total variance kept by both basis. The results shown in this paper were drawn from those in which principal components kept at least 95% of the total variance.

The following algorithm summarizes the training procedure:

Face3D-Training($X^t = \{x_1^t, \dots, x_N^t\}$, $X^g = \{x_1^g, \dots, x_N^g\}$, landmarks = $\{l_1, \dots, l_N\}$, triangulation)

1. $\bar{l} \leftarrow$ PROCUSTES-ANALYSIS(landmarks)
2. $L^t \leftarrow$ WARP(X^t , landmarks, \bar{l} , triangulation)
3. $L^g \leftarrow$ WARP(X^g , landmarks, \bar{l} , triangulation)
4. $\{t_0, g_0\} \leftarrow$ MEAN(L^t, L^g)
5. $\{E^t, E^g\} \leftarrow$ PRINCIPAL-COMPONENT-ANALYSIS(L^t, L^g)

The procedure WARP allows one to normalize the textures X^t and geometries X^g of the dataset, mapping them into the average face shape using a defined triangulation. The procedure MEAN calculates the average texture model t_0 and the average geometry model g_0 , based on L^t and L^g , respectively.

4.2 Face reconstruction

The input to the system is a 2D photography (frontal face image) to which the ASM is applied in order to automatically detect the facial landmarks. The ASM landmarks are extracted from the gray-scale input image. ASM has been

proposed in [4] and allows the alignment and representation of image data using a statistical model of the target object obtained from the training data. A face model is represented by a set of landmarks manually placed over training face images (not necessarily those obtained for the 3D face model). The sets of landmarks for the training images are aligned in order to minimize the distance between corresponding points (i.e., homologous points). A point distribution model (PDM) is obtained from the variance of the distances among the different points. The PDM is used to constraint the shape variation in the ASM matching process.

The facial landmarks are aligned to the average face shape obtained in the training process. Thus, the texture is warped to the average shape (similar to the process done in training phase). This process allows one to normalize the texture of the input image.

Let \mathbf{x}^t be the warped texture of the input image \mathbf{x} , and t_0 the normalized average texture obtained in training process. The texture coefficients α_x^t , are calculated by projecting $(\mathbf{x}^t - t_0)$ onto the respective orthonormal basis $(\{e_i^t\})$:

$$\alpha_x^t = E^t \cdot (\mathbf{x}^t - t_0) \quad (1)$$

where E^t is a transformation matrix defined by the orthonormal basis for the texture space learned in the training process.

Once the texture coefficients α_x^t are obtained, the texture coefficients α^t of all images considered in the training process are used to calculate the coefficients s_x , defined as

$$\alpha^t \cdot s_x = \alpha_x^t \quad (2)$$

where α^t is the matrix defined by the coefficients for each texture image in the training dataset. Intuitively, s_x represents weighting coefficients obtained by projecting α_x^t onto α^t (Fig. 6(c)). It is important to recall that each sample represented in α^t is associated to a different facial expression. Therefore, s_x represents a decomposition of \mathbf{x}^t in terms of the different facial expressions learnt by the system (e.g., as we would say that \mathbf{x}^t is $a\%$ happy, $b\%$ angry, $c\%$ neutral, etc.).

The geometry coefficients α_x^g of \mathbf{x}^t are then calculated using the geometry coefficients of all training geometry samples α^g :

$$\alpha_x^g = \alpha^g \cdot s_x \quad (3)$$

The normalized geometry \mathbf{x}^g of the test face image \mathbf{x} is then reconstructed by

$$\mathbf{x}^g = (E^g \cdot \alpha_x^g) + g_0 \quad (4)$$

where E^g is a transformation matrix defined by the orthonormal basis for the geometry space learnt in the training

process. In order to reduce the noise on the reconstructed facial geometry (surface), the improved technique for smoothing polygonal surface meshes proposed by J. Vollmer et al. [25] has been applied. This smoothing technique was selected because of being robust and efficient to smooth a mesh which is not necessarily topologically correct. Finally, the input texture warped to the average shape face is directly mapped onto the resulting 3D smooth geometry. The missing blank areas are filled by interpolation of adjacent 3D points.

The following algorithm summarizes the proposed approach:

Face3D-Reconstruction(\mathbf{x})

1. $l_x \leftarrow \text{ASM}(\mathbf{x}) \triangleright$ Landmarks of the input photography \mathbf{x}
2. $\mathbf{x}^t \leftarrow \text{WARP}(\mathbf{x}, l_x, \bar{l})$
3. $\alpha_x^t \leftarrow E^t \cdot (\mathbf{x}^t - t_0)$
4. $s_x \leftarrow \text{PROJECT}(\alpha_x^t, \alpha^t)$
5. $\alpha_x^g \leftarrow \alpha^g \cdot s_x$
6. $\mathbf{x}^g \leftarrow (E^g \cdot \alpha_x^g) + g_0$
7. **return** \mathbf{x}^g

The procedure WARP normalizes the input texture \mathbf{x} using the corresponding facial landmarks l_x and the average face shape \bar{l} obtained in training phase. The procedure PROJECT allows one to represent α_x^t in terms of α^t as indicated in (2).

5 Results and discussion

The system was implemented as described in the previous section by using MATLAB. The experiments presented in this paper were conducted on a 1.5 GHz Intel Core 2 Duo with 2-Gb memory running the Linux operating system.

The system has been tested using real data. All experiments were performed with seven poses of ten subjects (five female and five male subjects), consisting in a total of 70 training images (texture and geometry). Figure 7 shows the training images used in this work. We used a 3D scanner KONICA MINOLTA VIVID 910 for the data acquisition as mentioned in Sect. 2. In all experiments, for a given input texture image, a face geometry of 26651 points has been obtained in at most 4 seconds. Eight principal components were required to represent the total variance of the dataset. The average face of the generated dataset is show in Fig. 8.

Three different types of experiments have been performed to evaluate the 3D facial reconstruction quality of the system. One experiment was considered to investigate the 3D reconstruction taking the seven facial expressions of one person seen in the training process as input images. The 2D photographs shown in Fig. 9(a) have been used as input to the system. In this case, the subject belongs to the training data. Figure 9(b) shows the corresponding reconstructed



Fig. 7 Facial expressions of the dataset. The complete dataset is composed of 70 2D photographs and 70 range images corresponding to 7 facial expressions of 10 subjects

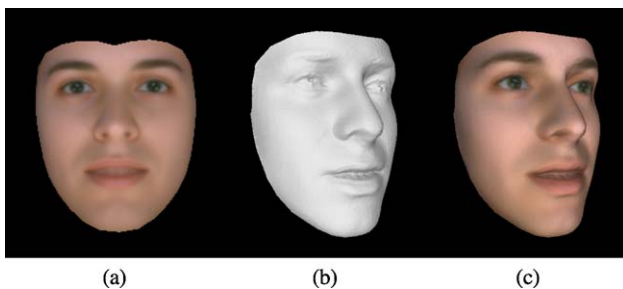


Fig. 8 Average face obtained in the training phase: (a) texture (t_0), (b) geometry (g_0), and (c) geometry with mapped texture

3D facial geometry. Figure 9(c) shows the facial geometry of the input image with the texture mapped onto it. As it can be seen, the system was able to successfully reconstruct the 3D face structure of the subject maintaining the facial spatial coherence.

Note that the mouth is slightly open in the 3D reconstructed faces for the neutral pose and the sad expression. This is because the obtained geometry coefficients were not higher enough to modify the slightly open mouth of the average geometry g_0 calculated in the training phase.

A second experiment was performed for 3D reconstruction using seven facial expressions of a different person from those present in the training phase. Photographs of differ-

ent facial expressions have been obtained in a controlled environment (the obtained texture information has the same technical specifications from those used to train the system). In this second case, the facial landmarks have been identified using the ASM approach described above. This input data has been fed to the system in order to reconstruct a 3D geometry from photographs.

Figure 10 shows the 3D face reconstruction for some facial expressions. The calculated 3D reconstructions are shown in Fig. 10(b). Figure 10(c) shows the 3D geometry where the 2D photography was mapped. In the reconstructed 3D faces, the facial spatial coherence is maintained. Note that, for the facial expression of anger, the obtained geometry presents a closed mouth face. This is because several expressions of anger in the dataset were expressed with a closed mouth, as shown in anger expression of Fig. 9. In general, the facial spatial coherence is maintained in all reconstructed 3D facial expressions. Results qualitatively similar were obtained in our tests, regardless of the subject being or not present in the training dataset.

Finally, considering the two subjects shown above, a third experiment has been devised to test the system using two facial expressions that are not present in the training phase: mouth and eyes open, and kiss expression with open eyes.

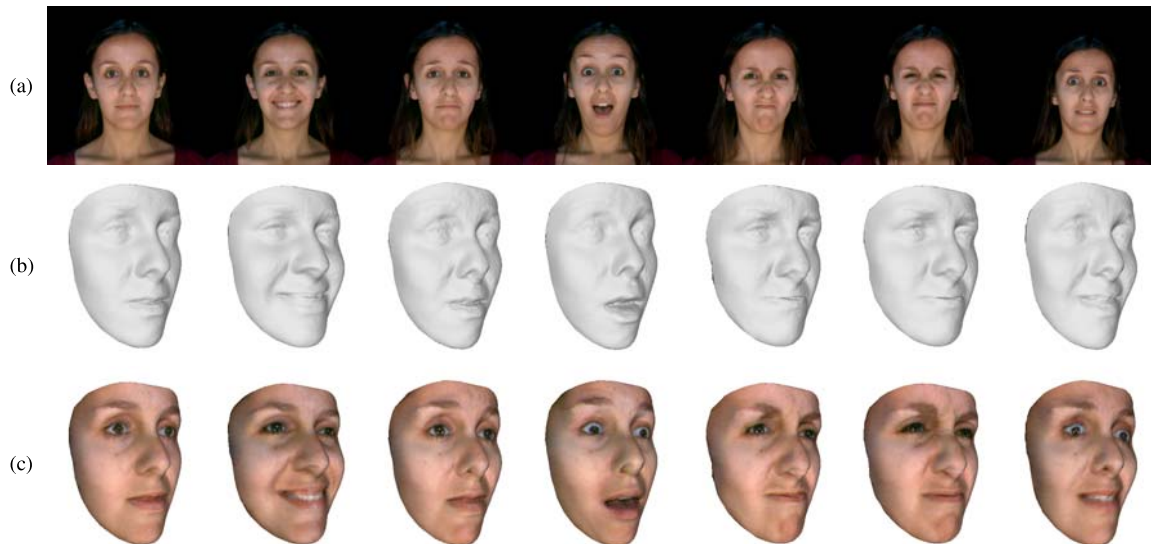


Fig. 9 3D reconstruction of a subject present in the dataset: **(a)** facial expressions showing neutral, joy, sadness, surprise, anger, disgust, and fear, respectively; **(b)** reconstructed 3D face geometry; and **(c)** 3D geometry with the mapped texture. In the system, the 2D photographs

with several facial expressions of a subject present in the dataset were used to be 3D reconstructed. Note that the reconstructed geometry presents spatial coherence regarding each facial expression

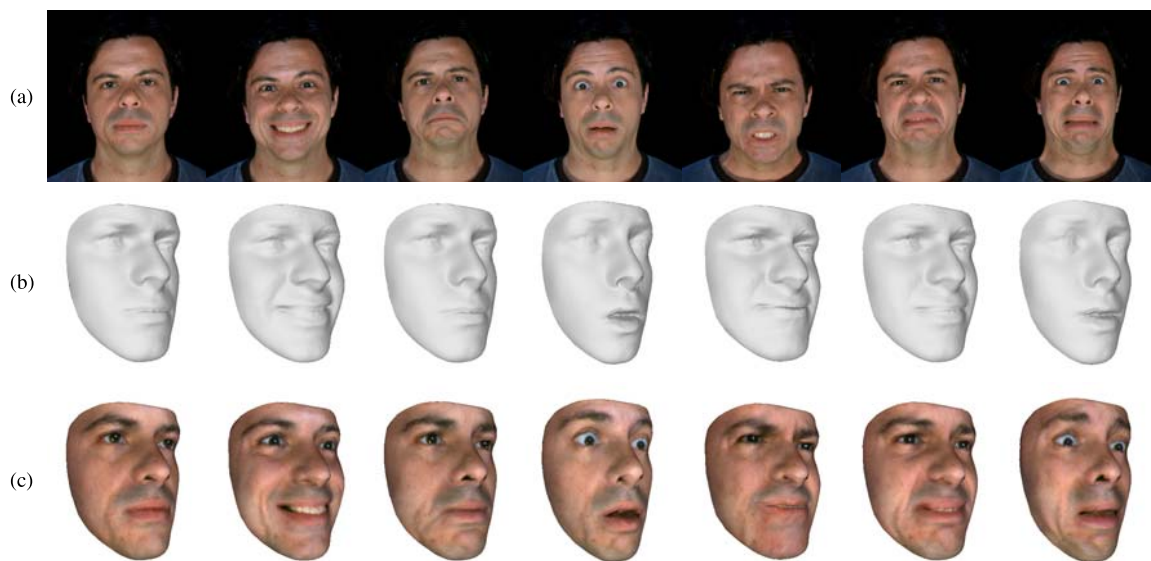


Fig. 10 3D reconstruction of a subject not present in the dataset: **(a)** facial expressions showing neutral, joy, sadness, surprise, anger, disgust, and fear, respectively; **(b)** reconstructed 3D face geometry; and **(c)** 3D geometry with the mapped texture. In the system, the 2D

photographs with several facial expressions of a subject present in the dataset were used to be 3D reconstructed. Note that the reconstructed geometry presents spatial coherence regarding each facial expression

The textures shown in Fig. 11(a) have been used as input to the system. ASM has been applied to calculate the facial landmarks and used in the face reconstruction system. As it can be seen in Fig. 11(b), although the overall structure is recovered, details, such as the open mouth, are not completely reconstructed as expected. The obtained expressions are not realistic when compared to the input texture because the dataset does not include any samples with similar facial

expressions. This limitation indicates that there is room for improving the method, as discussed in the next section.

6 Conclusions and future works

This paper describes a proof-of-concept system for reconstructing the 3D geometry of human faces from 2D monoc-



Fig. 11 3D reconstruction of facial expressions not present in the training phase: (a) facial expressions of one subject present in the dataset (*first two photographs*) and one subject not present in the dataset (*last two photographs*); (b) reconstructed 3D face geometry; and (c) 3D geometry with the mapped texture

ular color images. This task is accomplished through a “resynthesis by analysis” approach, in which models for human facial texture and geometry spaces are built from acquired real-world data, being subsequently used to perform 3D face reconstruction.

The mathematical model behind the system is based on building suitable texture and geometry spaces from the training set and transformations between such spaces. Face reconstruction from an input texture image is then carried out by transforming the input data through these spaces until a geometry model is created. The experimental results have shown that the proposed method may be applied to 3D reconstruction of faces from subjects present or not in the training set.

There are three main phases in our system: (1) data acquisition and processing; (2) data analysis and model construction; and (3) resynthesis and face reconstruction. By inspecting them, we identify a number of future research directions which can be exploited in order to improve the quality of our prototype system.

Data acquisition and processing The main issue in this first phase regards fitting a template face model to an RGBZ image (i.e., a color photograph plus a range image). This is a crucial step, and its result induces a correspondence in the example data which is exploited in the following phases.

For our prototype, fitting is accomplished by hand. This limits the system to coarse face templates and just a handful of training subjects. Thus, for improved scalability and quality, an investigation of (semi-)automatic methods to perform this template fitting on medium-sized datasets with fine

face templates must be considered. As examples, we may mention works on fitting AAM’s to low-resolution color images [5], the semi-automatic subdivision template matching approach used in [8], and the variational surface matching method of [17]. Each of these may be adapted to our needs and evaluated.

Data analysis and model construction With the training examples in proper correspondence, we are able to build mathematical models for both face texture and geometry spaces, describing a span of facial appearance and shape configurations.

In our system, a PCA-based affine model was employed mainly for its simplicity and well succeeded results of related affine models (e.g., morphable and active shape models). However, for richer representations of facial appearances and shapes, more elaborate mathematical models are needed and should be investigated. We have been considering simple, yet powerful, nonlinear models like tensor-based multilinear spaces [18, 24] and Kernel PCA-based spaces [21, 23].

Resynthesis and 3D face reconstruction The last phase of our system is responsible for reconstructing a 3D face model from a given 2D monocular color image. This is accomplished in two steps: texture registration and texture-geometry mapping.

In the first step, a 2D template face model is fitted to the input photograph to register the user’s face image with those from the dataset. This task is very similar to that in the first phase and can be improved with analogous techniques.

The second step is central in this phase. The registered texture is projected onto the texture space built at the second phase. Its representation is used to retrieve a description of the corresponding geometry with respect to the associated geometry space, also built at the previous phase. In this step, the most important tasks comprise projection onto the texture space, texture-geometry representation mapping, and geometry reconstruction from its representation. The first and last tasks are loosely coupled and depend mostly on the models chosen for the texture and geometry spaces. The second task is implemented by a mathematical model representing the mapping of appearance to geometry. Our system implements a very simple linear mapping between PCA coefficients and yet yields quite nice results. An interesting investigation concerns to study more elaborate mappings constructed from training data (e.g., nonlinear regression models as kernel methods and neural networks [10]).

Our experiments with the system have shown that a key point for the performance of the system with faces not present in the training phase relies on the identification of the facial landmarks. The system is sensitive to the identification of the facial landmarks considered as the first step

of the reconstruction, i.e., $\{l_1, l_2, \dots, l_K\}$. Unfortunately, the ASM has not always performed as would be desirable, thus leading eventually to bad reconstruction results. Thus, a better method for automatic landmark identification would certainly help in improving the performance obtained. This is one of the main topics of our ongoing work.

Our plans for future research also include a more comprehensive treatment of different applications which might benefit from our 3D face reconstruction system (e.g., video rewrite [2], puppetry [24], and conferencing). Other interesting investigations involve incorporating data-driven models to add geometric details to the reconstructed 3D faces [8] and enhancing facial attractiveness [16].

Acknowledgements Financial support for this research has been provided by CAPES, CNPq, FAPESP, and FINEP. The authors are grateful to anonymous reviewers for the critical comments and valuable suggestions.

References

- Blanz, V.: Face recognition based on a 3D morphable model. In: International Conference on Automatic Face and Gesture Recognition, pp. 617–624 (2006)
- Bregler, C., Covell, M., Slaney, M.: Video rewrite: driving visual speech with audio. In: SIGGRAPH '97: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 353–360. ACM/Addison–Wesley, New York/Reading (1997)
- Chai, J.-X., Xiao, J., Hodgins, J.: Vision-based control of 3D facial animation. In: Breen, D., Lin, M. (eds.) Eurographics/SIGGRAPH Symposium on Computer Animation, pp. 193–206. Eurographics Association, San Diego (2003)
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models: Their training and application. *Comput. Vis. Image Underst.* **61**(1), 38–59 (1995)
- Dedeoglu, G., Baker, S., Kanade, T.: Resolution-aware fitting of active appearance models to low resolution images. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV (2). Lecture Notes in Computer Science, vol. 3952, pp. 83–97. Springer, Berlin (2006)
- Dryden, I.L., Mardia, K.V. (eds.): *Statistical Shape Analysis*. Wiley, Chichester (1998)
- Elyan, E., Ugail, H.: Reconstruction of 3D human facial images using partial differential equations. *J. Comput.* **2**(8), 1–8 (2007)
- Golovinskiy, A., Matusik, W., Pfister, H., Rusinkiewicz, S., Funkhouser, T.: A statistical model for synthesis of detailed facial geometry. *ACM Trans. Graph.* **25**(3), 1025–1034 (2006)
- Goodall, C.: Procrustes methods in the statistical analysis of shape. *J. R. Stat. Soc. Ser. B* **53**(2), 285–339 (1991)
- Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning*. Springer, Berlin (2003)
- Hong, P., Wen, Z., Huang, T.S., Shum, H.Y.: Real-time speech-driven 3D face animation. In: 3D Data Processing Visualization and Transmission, pp. 713–716 (2002)
- Jiang, D.L., Hu, Y.X., Yan, S.C., Zhang, L., Zhang, H.J., Gao, W.: Efficient 3D reconstruction for face recognition. *Pattern Recognit.* **38**(6), 787–798 (2005)
- Kanade, T.: *Picture processing system by computer complex and recognition of human faces*. Doctoral Dissertation, Kyoto University, November 1973
- Kittler, J.V., Hilton, A., Hamouz, M., Illingworth, J.: 3D assisted face recognition: A survey of 3D imaging, modelling and recognition approaches. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 3, p. 114 (2005)
- Lee, M.W., Ranganath, S.: 3D deformable face model for pose determination and face synthesis. In: International Conference on Image Analysis and Processing, pp. 260–265 (1999)
- Leyvand, T., Cohen-Or, D., Dror, G., Lischinski, D.: Data-driven enhancement of facial attractiveness. *ACM Trans. Graph.* **27**(3), 1–9 (2008)
- Litke, N., Droske, M., Rumpf, M., Schröder, P.: An image processing approach to surface matching. In: SGP '05: Proceedings of the Third Eurographics Symposium on Geometry Processing, p. 207. Eurographics Association, Aire-la-Ville (2005)
- Macêdo, I., Brazil, E.V., Velho, L.: Expression transfer between photographs through multilinear AAM's. In: SIBGRAPI, pp. 239–246. IEEE Computer Society, Los Alamitos (2006)
- Mena-Chalco, J.P., Macêdo, I., Velho, L., Cesar, R.M. Jr.: PCA-based 3D face photography. In: Brazilian Symposium on Computer Graphics and Image Processing, pp. 313–320. IEEE Computer Society, Los Alamitos (2008)
- Onofrio, D., Tubaro, S.: A model based energy minimization method for 3D face reconstruction. In: ICME, pp. 1274–1277. IEEE Press, New York (2005)
- Schölkopf, B., Smola, A., Müller, K.-R.: Kernel principal component analysis. In: Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-D. (eds.) International Conference on Artificial Neural Networks—ICANN. Lecture Notes in Computer Science, vol. 1327, pp. 583–588. Springer, Berlin (1997)
- Soyel, H., Demirel, H.: Facial expression recognition using 3D facial feature distances. In: International Conference on Image Analysis and Recognition, pp. 831–838 (2007)
- Twining, C.J., Taylor, C.J.: Kernel principal component analysis and the construction of non-linear active shape models. In: Cootes, T.F., Taylor, C.J. (eds.) British Machine Vision Conference. British Machine Vision Association (2001)
- Vlasic, D., Brand, M., Pfister, H., Popović, J.: Face transfer with multilinear models. *ACM Trans. Graph.* **24**(3), 426–433 (2005)
- Vollmer, J., Mencl, R., Müller, H.: Improved Laplacian smoothing of noisy surface meshes. *Comput. Graph. Forum* **18**(3), 131–138 (1999)
- Wang, Y.J., Chua, C.S.: Face recognition from 2D and 3D images using 3D Gabor filters. *Image Vis. Comput.* **23**(11), 1018–1028 (2005)
- Yabui, T., Kenmochi, Y., Kotani, K.: Facial expression analysis from 3D range images; comparison with the analysis from 2D images and their integration. In: International Conference on Image Processing, pp. 879–882 (2003)
- Zhang, Y., Xu, S.: Data-driven feature-based 3D face synthesis. In: 3-D Digital Imaging and Modeling, pp. 39–46 (2007)



Jesús P. Mena-Chalco received the Eng. degree in Systems Engineering from the National University of San Agustín, Peru. He received the M.S. degree in Computer Science from University of São Paulo, Brazil. Currently, he is a Ph.D. student in the Department of Computer Science at the University of São Paulo, Brazil. His research interests are mainly in the fields of pattern recognition, computer vision, and bioinformatics.



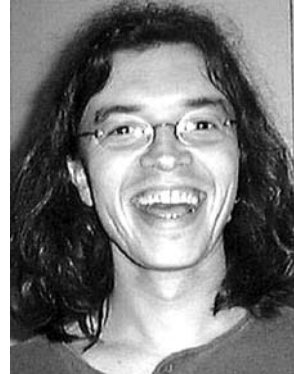
Ives Macêdo is a D.Sc. student of Mathematics (with a major in Computer Graphics) at the Vision and Graphics Laboratory (Visgraf Lab) of the National Institute of Pure and Applied Mathematics (IMPA), Rio de Janeiro, Brazil. He also received the M.S. degree in Mathematics from IMPA and the B.Sc. degree in Computer Science from the Informatics' Center (CIn) of the Federal University of Pernambuco (UFPE), Brazil. Among other things, his research interests include facial animation, machine learning,

and physically-based modeling and animation, his current research focus.



Luiz Velho is a Full Researcher/Professor at IMPA—Instituto de Matemática Pura e Aplicada of CNPq and the leading scientist of VISGRAF Laboratory. He received the B.E. degree in Industrial Design from ESDI/UERJ in 1979, the M.S. in Computer Graphics from the MIT/Media Lab in 1985, and a Ph.D. in Computer Science in 1994 from the University of Toronto under the Graphics and Vision groups. His experience in computer graphics spans the fields of modeling, render-

ing, imaging, and animation. He has published extensively in conferences and journals of the area. He is the author of several books and has taught many courses on graphics-related topics. He is a member of the editorial board of various technical publications. He has also served on numerous conference program committees.



Roberto M. Cesar Jr. received the B.S. degree in Computer Science from the São Paulo State University, Brazil, the M.S. degree in Electrical Engineering from the State University of Campinas, Brazil, and the Ph.D. degree in Computational Physics from the Institute of Physics, University of São Paulo (USP), São Paulo, Brazil. He is currently a Full Professor with the Department of Computer Science, IME, USP. His main research topics include shape analysis, computer vision, and structural pattern recognition.