3D Linear Facial Animation Based on Real Data

Andréa Britto Mattos
IME - Universidade de São Paulo
amattos@ime.usp.br

Jesús P. Mena-Chalco
IME - Universidade de São Paulo
jmena@ime.usp.br

Roberto M. Cesar Jr
IME - Universidade de São Paulo
cesar@ime.usp.br

Luiz Velho
Instituto Nacional de Matemática Pura e Aplicada
lvelho@impa.br

***Abstract*—In this paper we introduce a Facial Animation system using real three-dimensional models of people, acquired by a 3D scanner. We consider a dataset composed by models displaying different facial expressions and a linear interpolation technique is used to produce a smooth transition between them. One-to-one correspondences between the meshes of each facial expression are required in order to apply the interpolation process. Instead of focusing in the computation of dense correspondence, some points are selected and a triangulation is defined, being refined by consecutive subdivisions, that compute the matchings of intermediate points. We are able to animate any model of the dataset, given its texture information for the neutral face and the geometry information for all the expressions along with the neutral face. This is made by computing matrices with the variations of every vertex when changing from the neutral face to the other expressions. The knowledge of the matrices obtained in this process makes it possible to animate other models given only the texture and geometry information of the neutral face. Furthermore, the system uses 3D reconstructed models, being capable of generating a three-dimensional facial animation from a single 2D image of a person. Also, as an extension of the system, we use artificial models that contain expressions of visemes, that are not part of the expressions of the dataset, and their displacements are applied to the real models. This allows these models to be given as input to a speech synthesis application in which the face is able to speak phrases typed by the user. Finally, we generate an average face and increase the displacements between a subject from the dataset and the average face, creating, automatically, a caricature of the subject.**

***Keywords*-Computer Graphics, Facial Animation, 3D Reconstruction.**

I. Introduction

Facial animation is a topic that attracts much interest among researchers, being widely studied and having many applications, especially in the entertainment field. Obtaining a realistic facial animation is a difficult task due to the very complex structure of the face. Also, humans are extremely familiar with face movements, and therefore, they can easily identify small details that are unnatural or inconsistent in a facial animation. When trying to achieve realistic results, the biggest challenges in producing computer facial animations are obtaining a consistent representation of the face and modeling the animation itself. The most realistic results in modeling faces usually come from obtaining information of real humans, by 3D scans or reconstructed models from photographs [1], [2]. To increase realism, statistical techniques are also used in surfaces that are too smooth to produce wrinkles and pores [3]. In turn, to accomplish a feasible animation, there are at least four fundamental approaches that are mainly used in facial animation: direct parametrization, muscle-based, performance-driven and interpolation.

In direct parametrization, the face is represented by a polygonal mesh and it moves through a set of parameters that modifies characteristic points of the face. Thus, this approach aims to build a model that can generate a wide range of faces and facial expressions based on this set of control parameters, that should be as small as possible. One disadvantage of the process is that the parameters depend on the topology of the face, and must be rewritten to be used in other models. Seeking to overcome this problem, the MPEG-4 standard was created, intending to be an international standard to the animation of parametric faces, and being used in some projects [4], [5], [6]. The goal of the direct parametrization is defining a model that allows the specification of all possible expressions. However, this ideal model has not been achieved so far.

The muscle-based approach, in turn, seeks to simulate the geometry of bones, muscles and tissues of the face. In this technique, the vertices of the model are not altered directly and must follow the interactions of an internal structure as it is modified. The pseudo-muscle based approach is a variant for the method, which does not simulate exactly the detailed anatomy of the face. Instead, it aims to select simplified models that can be controlled in the same way. Both approaches have been explored in some studies [7], [8], [9] and have, as limitations, the high complexity and computational cost used in the process of the animation. Also, there are difficulties in reproducing subtle movements of the surface of the skin, such as wrinkles.

The performance-based process of animation acquires real human actions and reproduce it to synthetic models, usually requiring some appropriate equipment for the movement

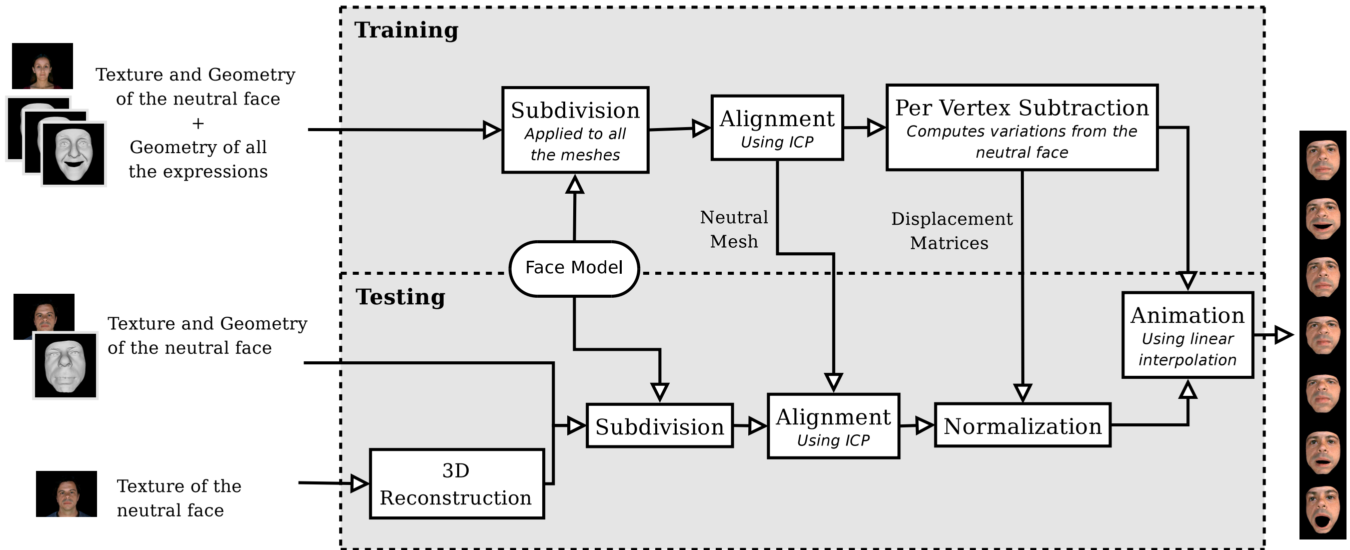


Figure 1. Schematic data flow of the animation system (dotted lines). Each block (solid lines) represents a process while each arrow represents the information flow between process. The system input, in the testing phase, is a 3D facial geometry fitted with texture or a 2D photograph.

extraction. In facial animation, the strategy commonly used is to provide an initial map of the human's neutral face and the target's neutral face, and use an appropriate function to update the correspondences between them as the real face moves. The most common example of performance-based animation is motion capture, which traditionally is made through special markers applied to the human's face [10]. However, some good results have been achieved using markerless approaches [11], [12], [13], although special markers can still be used to emphasize useful information like wrinkles [14].

The last method is the oldest and most simple, but it is probably also the one most used until now: the interpolation of key frames, represented by different facial expressions. The strategy provides a smooth transition into two different poses by the computation of the in-between points, and if the poses to be animated have the same topology, the interpolation only changes the positions of the vertices. The interpolation may be linear or non-linear considering that the physical movements of the face are usually performed in a non-linear way. It is also possible to apply the interpolation in smaller regions of the face at a time, as it is suggested in [15]. Linear interpolation is used in a variety of studies being applied alone or combined with the other techniques as mentioned [1], [5], [6], [16], [17], [18], [19].

The limitation of the interpolation method is that the key facial expressions should be provided. They can either be acquired by real humans, manually sculpted or generated by statistically-based techniques [20]. However, once the expressions are available, the results of the interpolation method can overcome the achievements of other techniques, which might seem artificial, especially when not taking into account movements of every part of the face. In this work, we explore linear interpolation applied to a realistic dataset of models, acquired by scanning real people's faces. The used dataset contains models displaying the six universal expressions proposed by Paul Ekman's Facial Action Coding System (joy, sadness, fear, anger, surprise and disgust) [21] and these expressions, along with the neutral face, represent the keyframes that should be interpolated in the animation process. The data flow of the system is shown in Figure 1.

In the considered process, the face models to be interpolated must have the same number of vertices, which is not guaranteed when dealing with real data from 3D scans. Therefore, it is necessary to manipulate the raw data obtained from the 3D scanner in order to generate models that satisfy this restriction. In this context, previous works have investigated strategies for the computation of one-to-one correspondences in facial meshes. In [16], a strategy using an optic-flow-based algorithm was introduced, being used latter in [17], [22]. The alternative proposed by D. Vlasic *et al.*[19] searches the optimization of the overall similarity of the meshes, taking into account the proximity of some manually selected points and reference vertices. Also, in [18], some corresponding points are manually selected, and the matching of the remaining vertices of the dense surface is made by volume morphing, using Radial Basis Functions. However, all these approaches focus on finding dense correspondences between the meshes, which can lead to some problems, such as insufficient correspondences or bad matching in plain regions of the face. The main difference of our approach is that we are only concerned in establishing a few corresponding points that form an initial model, and the matching of intermediate points is made by a subdivision process. Section II describes the subdivision method applied in order to reproduce the equivalent meshes. After generating the subdivided models it is possible, for each pose, to compute the shift that each vertex suffers with respect to the neutral face. This information is kept in a set of matrices used to animate other models in which the geometry of other expressions are not available. This animation process is described in Section III.

Then, a 3D facial reconstruction system is used to achieve a more elaborated situation: using a single 2D image as input, an animated 3D model that displays the facial expressions contained in the dataset is generated.

Moreover, an additional application is considered, in which the same process is applied to artificial models with expressions that are not in the dataset, making it possible to transfer the information of the displacement matrix computed to the models of the dataset, and thus, generating new expressions.

One last experiment was made, in which we generate an average face using models from the dataset, and choose a subject to be automatically caricatured by increasing the distances from the subject and the dataset.

II. Model

3D face data was acquired using a non-contact 3D scanner Konica Minolta Vivid 910 and reflector lamps that provided controlled lighting conditions. The scanner is composed by a laser sensor and a digital camera video. The data is hence composed by registered texture and geometry data. The output mesh contains approximately 60k vertices. Texture images have been acquired with 640×480 pixels (24 bits).

The model considered is similar to the one proposed by the MPEG-4 facial animation standard. We use an initial model of 48 vertices and 74 triangles as shown in Figure 2. These points were manually located on each acquired face. The model used does not consider the inner mouth region because the geometry of the tongue and the upper and lower teeth was not available in all the expressions, since these structures are not obtained by the 3D scanner. Besides, the image of neutral face is used to map the texture in which this region is not shown. Thus, it is possible to animate other face models using only the information of the neutral face.

As mentioned previously, it is necessary to process the raw data obtained by the 3D scanner in order to generate models with equivalent geometry. This is made by following the approach of Guskov *et. al* [23] and Golovinskiy *et. al* [3] in which the initial model of 48 landmarks is subdivided and re-projected successively, i.e. the raw mesh is refined to 9648 vertices applying two linear subdivisions followed by two Loop subdivisions [24]. This subdivision scheme divides each triangle of the initial model into four smaller triangles, adding new vertices to the middle of each edge. After each subdivision, the new vertices are projected onto

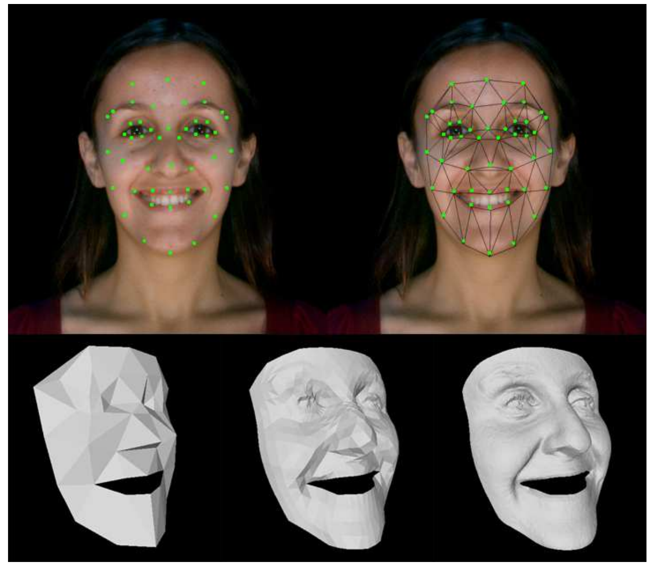


Figure 2. Model used, with 48 points and 74 triangles (top), and the results of the subdivision process after 0, 2 and 4 iterations (bottom).

Table I
STRATEGY OF SUBDIVISIONS AND PROJECTIONS USING THE INITIAL FACE MODEL.

Description	**Vertices**	**Triangles**
Initial face model (base mesh)	48	74
Linear subdivision and projection	170	296
Linear subdivision and projection	636	1184
Loop subdivision and projection	2456	4736
Loop subdivision and projection	9648	18944

the raw mesh, using the directions of its corresponding normal vectors.

In the dataset, the subdivision and projection process is performed using the same initial model with the same number of subdivision steps, i.e. the resulting meshes keep the same number of vertices, since every initial triangle is divided considering the same strategy. Therefore, the process of subdivisions and projections allows to accomplish the resulting meshes into per-vertex correspondences yielding, in each iteration, face models with refined approximations of the raw mesh. Note that each mesh is processed independently. Then, the problem of matching plain facial regions, mentioned in Section I, is avoided. See in Table I the strategy of subdivisions and projections considered in the 3D facial animation system.

III. Animation System

Once the models are subdivided, for every vertex that is part of the neutral model, it is possible to locate the corresponding vertex in all the expressions, since they all have the same index given in the subdivision process.

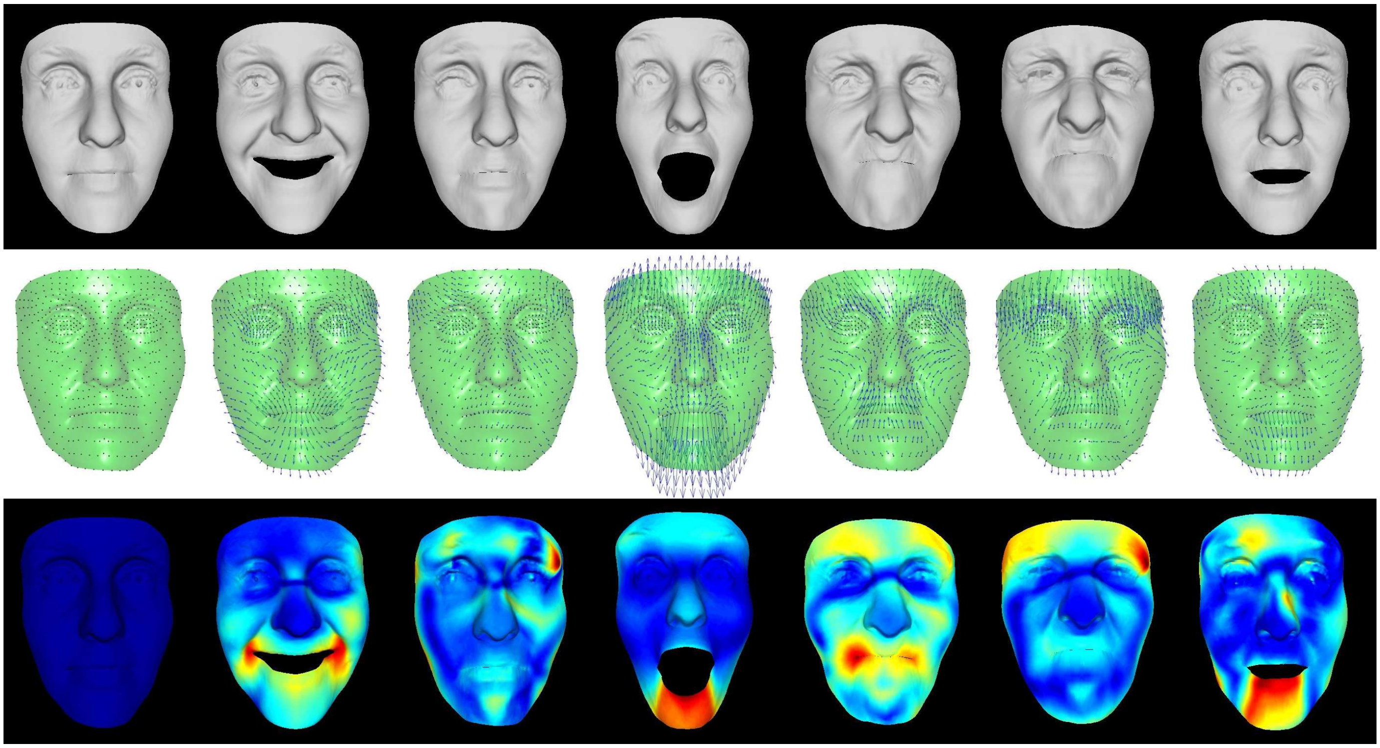


Figure 3. Geometry of the expressions contained in the dataset (top), displacements computed from the neutral face represented by vectors that determine the directions of the vertices of the neutral face in each expression (middle) and a color map for the applied displacements (bottom), where cold-colors indicate regions of the face that are slightly modified, and hot-colors indicates regions of high changes with respect to the neutral face.

Therefore, it is trivial to compute the variation of the position of a vertex with respect to the neutral face. Considering that there are m models to be interpolated, each one containing n vertices, we denote the models by

$$P_k, \quad k = 1, \ldots, m$$

where P_k represents the facial expressions (P_1 is the neutral pose), and the vertices position on the three axis of a pose P_k by

$$v_{xi}^k, \quad v_{yi}^k, \quad v_{zi}^k, \quad i = 1, \ldots, n$$

Thus, the linear difference between every vertex of a given pose and its corresponding vertex at the neutral face is represented by

$$\Delta x_i^k = v_{xi}^k - v_{xi}^1 \tag{1}$$

$$\Delta y_i^k = v_{yi}^k - v_{yi}^1 \tag{2}$$

$$\Delta z_i^k = v_{zi}^k - v_{zi}^1 \tag{3}$$

See in Figure 3 the difference between the neutral face and the six expressions.

In face scans, it is common that the person's face does not remain static at the time of the acquisition of the facial expressions. Therefore, before computing the difference from the neutral face, it is necessary to align all models to be interpolated. The ICP method [25] is applied, thus aligning the models with respect to translation and rotation.

After computing the shifts for all vertices, it is possible to build a matrix $M_{n\times 3}^k$ for each of the k poses containing the translations that should be applied to the vertices of the neutral face. The interpolation process gradually increments this variation to its final value. It is also possible to combine multiple poses into one single expression, by applying amounts of the displacements of each pose that should be combined. There is a variety of combinations of how the six universal expressions can be combined generating new ones, which can be applied to our system generating expressions that are not originally part of the dataset.

At the end of the process, an animation is obtained given the geometry of all the expressions of a single person of the dataset. The knowledge obtained in this process can be used to animate other people from the dataset given only the geometry of the neutral expression. Consider that all the previously described steps were applied to an arbitrary person denoted person *base*. If the same subdivision process is applied to the neutral face of another person of the dataset, the resulting mesh will be equivalent to the neutral mesh previously subdivided of the base.

Once the set of matrices M^k is computed for the base, its values can be used to transfer the animation from the base to another model. However, it must be taken into account that the dimensions of the model can be significantly different from the dimensions of the base. The influence of the transferred displacements could be too small, in case the model is much bigger then the base, or too large, otherwise. Therefore, a normalization is applied, by computing the dimensions of the bounding boxes of the model and the base. The displacements made by the i vertices of the model considering the pose k are computed as:

$$\Delta x_i^k = v_{xi(base)}^k \times \frac{X_{base}}{X_{model}} \tag{4}$$

$$\Delta y_i^k = v_{yi(base)}^k \times \frac{Y_{base}}{Y_{model}} \tag{5}$$

$$\Delta z_i^k = v_{zi(base)}^k \times \frac{Z_{base}}{Z_{model}} \tag{6}$$

where X, Y and Z represent the dimensions of the bounding boxes in each axis, for the base and the model. Also, it must be considered that, besides the dimensions, the position and rotation of the model and the base can be different. The ICP method is applied to align the model with respect to the base from which the animations will be transferred.

In fact, it is not relevant if the new model to be animated is part of the dataset or not. We can simply consider any model that contains the information of geometry for the neutral face to produce a similar animation. This fact allows to include two additional functionalities to the system: to use and animate the result of reconstructed models and to generate new expressions to the models of real people considered in the dataset by computing the variation matrices of other models.

A. 3D Facial Reconstruction System

A PCA-based 3D face reconstruction system is used in order to obtain a 3D version of a 2D input image [2]. Given an input frontal face image to be 3D reconstructed, a set of facial landmarks is manually selected on the 2D image. The set of facial landmarks (feature points) is used to normalize the input texture. 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 vector basis and a linear optimization function to relate 2D texture and 3D geometry information as described in [26]. Finally, the 3D reconstruction is obtained by directly mapping the normalized texture onto the geometry.

In the reconstruction system the same 3D face dataset considered in [26] was used, consisting of the neutral face and the six poses of ten subjects (5 female and 5 male subjects). Therefore, the training procedure was performed in a total of 70 facial expressions.

It is important to note that, the reconstructed face presents a facial model different to the model considered in our work. For this reason, in the current version of our system, the 48 vertices of the initial face model (see Figure 2) were manually located onto the reconstructed mesh. Thus, the reconstructed face with the same model used in the animation system has been obtained performing the strategy of subdivisions and projections in the directions of its normals.

B. Speech Synthesis System

The dataset of real people used in our work contained the six expressions displayed in Figure 3. In order to add more expressions to these models, the same process was applied to another set of three-dimensional meshes which contained the geometry of other facial expressions not considered in the dataset. This set of models was created by the software FaceGen Modeler (www.facegen.com), which includes additional facial expressions besides the six adopted, such as expressions of visemes, which are the visual correspondent of phonemes. These are artificial models, whose geometry differs extensively from the real models used. The neutral face created by the software was exported, together with the expressions of visemes and the expression of the model with eyes closed, and all these meshes were given as input to same procedure described in Section III, computing the displacement matrices of each pose. Since, after the subdivision process, the neutral model of the artificial set is equivalent to the subdivided neutral model of the dataset, the matrices could be used to transform the real models.

The models were then used by a speech synthesis application which manipulates the face, synchronizing its lips movement with synthesized text that is typed by the user. The developed system uses a text-to-speech library to recognize the phonemes contained in the text, that can be spoken in Portuguese or English, and a callback is registered to notify the system whenever a phoneme event occurs. Then, the related viseme is displayed, according to a map previously defined of phonemes and visemes according to the MPEG-4 standard for the English language, which consider 15 phonemes. Because there are no standards of visemes mappings in Portuguese, an approximation is performed, considering a subset of MPEG-4 visemes that correspond to phonemes present in the Portuguese language. During human's speech, the current viseme is affected by the previously displayed viseme, and also by the face expression at the time [27]. In our system, these details are taken into consideration, using the fact that multiple poses can be easily combined. The expression having the models with eyes closed is used as a blinking movement to provide greater realism to the animation.

C. Caricature Auto-generation

A caricature is a face representation in which the shape of some facial attributes appear highly exaggerated. Commonly, this facial attributes refer to parts of the face in which there is already a bit disparity in comparison to an average face. The ability of performing interpolations by computing linear distances is used to generate automatic caricatures. The idea of the approach is suggested by V. Blanz and T. Vetter [16] and consists of caricaturing individual faces by increasing their distance with respect to an average face. In our work, the average face has been obtained computing the mean of each vertex of neutral faces belonging to the facial reconstruction system.

The face models are aligned and the linear difference between each vertex of the model and its related vertex on the average face is computed. The knowledge obtained with these displacements is applied on the face model in order to create caricatures in a simple way. By increasing its distance with respect to the average face it is possible to expand the disparity of attributes in which there were already contrasts with respect to the average face. It is important to note that the reverse displacements, applied in the face model, allows to obtain a deformation that approximates the model from the average face.

IV. Results

The system has been tested using real data. All experiments were performed with the animation knowledge obtained by the displacement matrices computed using the neutral face and the six facial expressions of a person of the dataset. Figure 4 shows the obtained interpolation given the geometry of all the expressions of this person. This data was considered in the training phase for the computation of the matrices applied in the experiments described next.

Some experiments have been performed to evaluate the system, in which we explore transferring the animations from the base model to another person of the dataset. One experiment was considered to investigate the expressions transfer to another model, given the texture and geometry information of the neutral face. The results are presented in Figure 5(a). A second experiment was performed to animate a 3D reconstructed face taking as input only a 2D photograph, and the results are shown Figure 5(b).

In another experiment (see Figure 6), it is shown how the expressions of artificial models were applied to the real models we were working with, allowing them to be part of the speech synthesis application developed. The artificial model contains approximately 6.000 vertices, having a topology that differs significantly from the one of the real models. Note that not every method of expression transfer works when models have topologies that are too different. Interpolation methods that rely on dense correspondence would fail in such situation.

Finally, in the last experiment, automatic caricatures of two subjects were generated. The average face was created by computing the mean of each vertex of 10 different faces from the dataset, presented in Figure 7(a). The resulting average face can be seen in Figure 7(b). Figure 8 shows

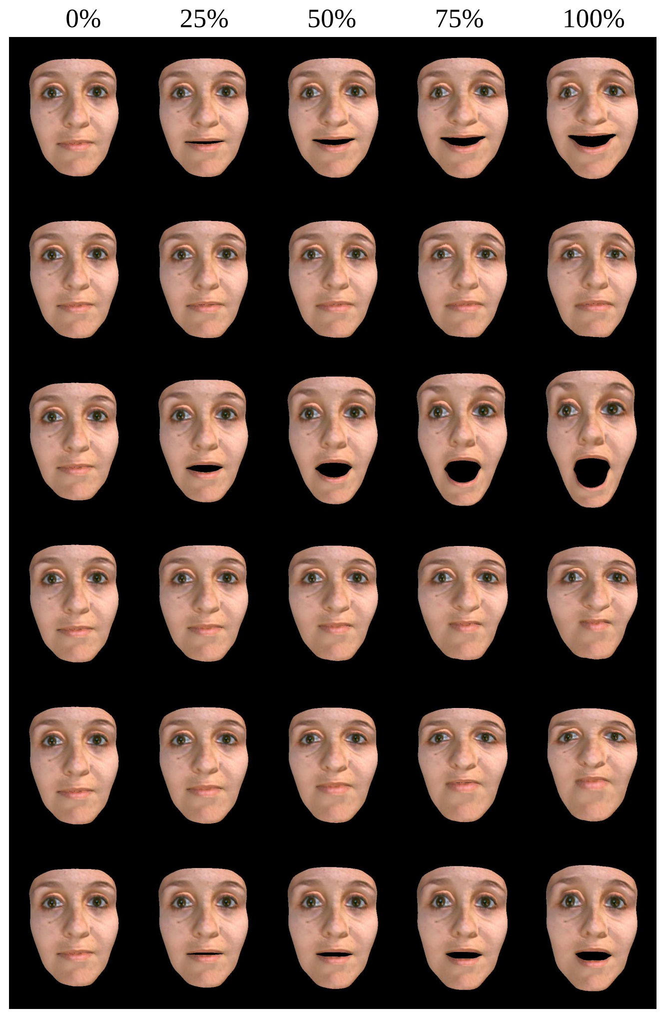


Figure 4. Results given the geometry of all the expressions and the texture of the neutral face. The interpolations of the six expressions are presented in each line of the image, at increasing stages.

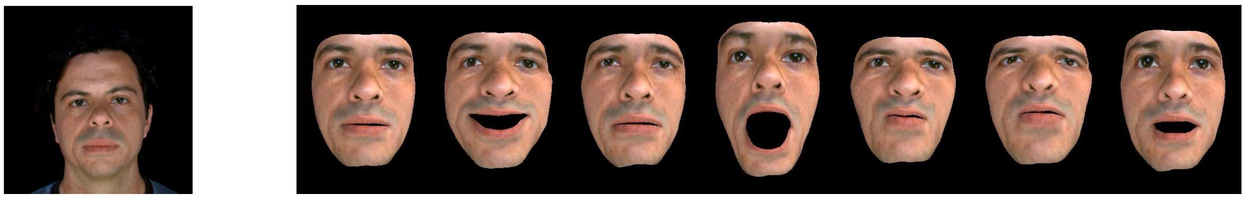
the effects of increasing and decreasing the displacements of the two models with respect to the average face. Figure 8(a) shows the caricature of a person who is also part of the computation of the average face and the caricature shown in Figure 8(b) considered a subject who is not part of this computation.

Videos of all the experiments and of the speech synthesis system using the real models are available at www.vision.ime.usp.br/~dedea/sibgrapi2010.

V. Conclusion

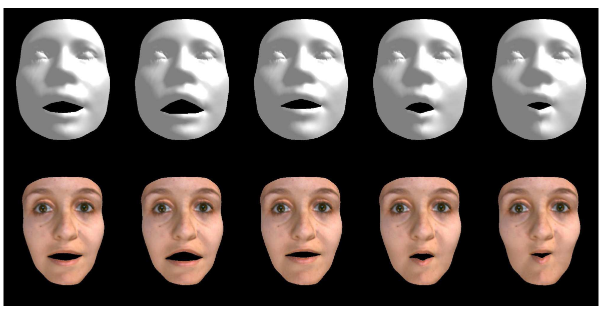
This paper describes a system of 3D facial animation using real human faces. The animation is performed through linear interpolations which require the computation of one-to-one correspondences between the faces. We show a strategy for computing the correspondences based on a few manually selected points followed by a subdivision process,

(a)



(b)

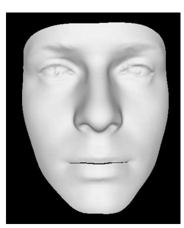
Figure 5. Experimental results for different types of input: (a) texture and geometry of the neutral face, (b) only the texture of the neutral face. Each image is associated to an facial expression with 100% of displacement.



/a/ /e/ /i/ /o/ /u/

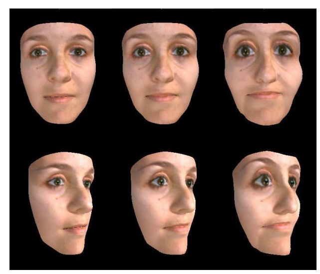
Figure 6. Visemes expressions for the vowels, generated in our model from the matrix previously computed by artificial models.

(a)

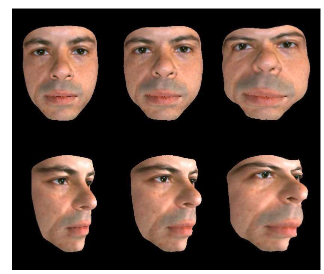


(b)

Figure 7. Faces considered in the computation of the average face on the auto-generation of caricatures: (a) 3D facial geometry of 10 subjects, (b) neutral average face.



(a)



(b)

Figure 8. Results of the caricature generation. The original model (middle), the transformed model by decreasing the displacements from the average face (left), and the caricaturization by increasing the displacements (right).

which avoids the problems of dense matching approaches. Displacement matrices are calculated from real-world data, being subsequently used to perform simple facial animation. The experimental results have shown that the proposed system may be applied to animate faces from subjects not present in the training phase.

We have shown that the ability to interpolate models of a realistic dataset can lead us to interesting results, but there is still room for improvements. It has been identified a number of future research directions which can be exploited in order to improve the quality of the animation system. It is now considered a simple method of animation, but more realistic results might be obtained considering non-linear movements. It may also be considered the possibility of acquiring the viseme information from scanning real people, instead of obtaining this information from artificial models, an thus, extending the used dataset. So far, in all the experiments reported in Section IV, the base model was represented by an arbitrary person of the dataset. We intend to investigate if, for each one of the six considered poses, combining the facial expressions of every person of the dataset generating an average expression and using the average expression as a base could lead us to better results. Finally, the subdivision strategy requires that the initial points are manually placed in the models. Naturally, this approach leads to the problem that marking several images may be an exhausting process. We aim to explore an automatic method for finding correspondences between the raw data obtained by the scanner, so meshes with the same topology can be generated and easily interpolated.

Acknowledgments

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

References

[1] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing realistic facial expressions from photographs," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*. ACM, 2006.

[2] J. P. Mena-Chalco, I. Macêdo, L. Velho, and R. M. Cesar-Jr., "PCA-based 3D face photography," in *Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing, 21 (SIBGRAPI)*. IEEE Computer Society, 2008, pp. 313–320.

[3] A. Golovinskiy, W. Matusik, H. Pfister, S. Rusinkiewicz, and T. Funkhouser, "A statistical model for synthesis of detailed facial geometry," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 1025–1034, 2006.

[4] Z. Wu, P. S. Aleksic, and A. K. Katsaggelos, "Lip tracking for MPEG-4 facial animation," in *ICMI '02: Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*. IEEE Computer Society, 2002.

[5] K. Balci, E. Not, M. Zancanaro, and F. Pianesi, "Xface open source project and SMIL-agent scripting language for creating and animating embodied conversational agents," in *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*. ACM, 2007.

[6] B. cai Yin, C. zhang Wang, Q. Shi, and Y. feng Sun, "MPEG-4 compatible 3D facial animation based on morphable model," in *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*. IEEE Computer Society, 2005.

[7] E. Sifakis, I. Neverov, and R. Fedkiw, "Automatic determination of facial muscle activations from sparse motion capture marker data," in *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*. ACM, 2005.

[8] E. Sifakis, A. Selle, A. Robinson-Mosher, and R. Fedkiw, "Simulating speech with a physics-based facial muscle model," in *SCA '06: Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Eurographics Association, 2006.

[9] B. Choe and H.-S. Ko, "Analysis and synthesis of facial expressions with hand-generated muscle actuation basis," in *SIGGRAPH '05: ACM SIGGRAPH 2005 Courses*. ACM, 2005.

[10] L. Dutreve, A. Meyer, and S. Bouakaz, "Feature points based facial animation retargeting," in *VRST '08: Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. ACM, 2008.

[11] T. Weise, H. Li, L. Van Gool, and M. Pauly, "Face/off: live facial puppetry," in *SCA '09: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, 2009.

[12] J.-x. Chai, J. Xiao, and J. Hodgins, "Vision-based control of 3D facial animation," in *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Eurographics Association, 2003.

[13] D. Vlasic, M. Brand, H. Pfister, and J. Popovic, "Face transfer with multilinear models," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*. ACM, 2006.

[14] B. Bickel, M. Botsch, R. Angst, W. Matusik, M. Otaduy, H. Pfister, and M. Gross, "Multi-scale capture of facial geometry and motion," in *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*. ACM, 2007.

[15] P. Joshi, W. C. Tien, M. Desbrun, and F. Pighin, "Learning controls for blend shape based realistic facial animation," in *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Eurographics Association, 2003.

[16] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999.

[17] V. Blanz, C. Basso, T. Poggio, and T. Vetter, "Reanimating faces in images and video," in *Proceedings of EUROGRAPHICS*, 2003.

[18] J.-y. Noh and U. Neumann, "Expression cloning," in *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001.

[19] D. Vlasic, M. Brand, H. Pfister, and J. Popovic, "Face transfer with multilinear models," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*. ACM, 2006.

[20] I.-J. Kim and H.-S. Ko, "Intuitive quasi-eigen faces," in *GRAPHITE '07: Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*. ACM, 2007.

[21] P. Ekman and W. V. Friesen, "The facial action coding system: A technique for the measurement of facial movement," in *Consulting Psychologists*, Palo Alto, California, 1978.

[22] C. Basso, P. Paysan, and T. Vetter, "Registration of expressions data using a 3D morphable model," in *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 2006.

[23] I. Guskov, P. Schröder, K. Vidimče, and W. Sweldens, "Normal meshes," in *Proceedings of the Computer Graphics Conference 2000 (SIGGRAPH-00)*, S. Hoffmeyer, Ed. New York: ACMPress, Jul. 2000, pp. 95–102.

[24] C. Loop, "Smooth subdivision surfaces based on triangles," Master's thesis, University of Utah, Department of Mathematics, 1987.

[25] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–258, Feb. 1992.

[26] J. P. Mena-Chalco, I. Macêdo, L. Velho, and R. M. Cesar-Jr., "3D face computational photography using PCA spaces," *The Visual Computer*, vol. 25, no. 10, pp. 899–909, 2009.

[27] P. S. L. Rodrigues, "Um sistema de geração de expressões faciais dinâmicas em animações faciais 3D com processamento de fala," Ph.D. dissertation, PUC-Rio, Dec. 2007.