Realistic Facial Animation Generation Based on Facial Expression Mapping

Hui Yu[1*], Oliver Garrod[2], Rachael Jack[2], Phlippe Schyns[2]
[1]University of Portsmouth, [2]University of Glasgow

ABSTRACT

Facial expressions reflect internal emotional states of a character or in response to social communications. Though much effort has been taken to generate realistic facial expressions, it still remains a challenging topic due to human being's sensitivity to subtle facial movements. In this paper, we present a method for facial animation generation, which reflects true facial muscle movements with high fidelity. An intermediate model space is introduced to transfer captured static AU peak frames based on FACS to the conformed target face. And then dynamic parameters derived using a psychophysics method is integrated to generate facial animation, which is assumed to represent natural correlation of multiple AUs. Finally, the animation sequence in the intermediate model space is mapped to the target face to produce final animation.

Keywords: Facial animation, FACS, deviation transfer, facial expression mapping, 3D face, action unit (AU).

1. INTRODUCTION

Facial animation has been the interest in computer graphics and animation communities [1], [2]. A great deal of effort has been taken to generate facial animation. Since the performance-driven animation method, also known as expression mapping, was introduced by Williams [9], it has been widely investigated [3], [4], [6], [7], [8], especially with the development of advances in 3D scanning and motion capture technologies [10]. Readers are referred to the book by Parke and Waters [5] for an excellent survey about facial animation.

Due to the complexity of facial muscles and subtle movements involved in conveying emotional information, it is nontrivial to model realistic facial expression using muscle simulation or modeling methods [24]. Thus, how to efficiently reuse existing facial expression data to animate various new face models but avoid time-consuming and tedious manual intervention has been a challenging topic.

In this paper, we present a facial expression mapping method, which transfers facial expressions from a source model to the target 3D face through encoding deformation information in local geometric shape. Through recording AUs according to Facial Animation Coding System (FACS) from real human beings and mapping to different face models, we demonstrate that the proposed method can easily generate facial animation reflecting true muscle movements.

2. RELATED WORK

Since the pioneer work conducted by Parke [19], research on facial animation has been widely studied. Enormous research has been conducted on improving fidelity of facial expression as well as on reducing time-consuming and tedious intervention from users.

Facial expressions are actually driven by facial muscles beneath the skin. Physically-based methods try to simulate facial muscles and skin to animate face models [11], [12], [15]. Those methods require high computation and nontrivial skills to control the parameters simulating muscle forces.

Geometry-based methods tend to animate face models using geometric information involving using feature points or controlling geometric parameters. Zhang et al. [17] developed a geometry-driven method for facial expression synthesis using feature points and face region subdivision, which was capable to infer missing feature motions to generate expression detail. Pighin et al. [16] proposed an image-based method for generating photo-realistic 3D facial expressions from 2D images. Parke et al. [26] presented a parametric method for generating facial expressions using parameter vector as representation of the motion of a set of vertices. However, it is still time-consuming for reusing those two methods in application where various new faces are to be animated.

Summer et al. [18] proposed a method for deformation transfer from the source mesh model to the target mesh based on triangle transformation. This method is barely feasible if the source and target meshes have different mesh topology or less shape resemblance. Darren et al [23] described a method for performance driven facial animation by re-mapping animation parameters between multiple types of facial models, which was able to map facial expressions through extracting meaningful facial actions parameters from video performance. Curio et al [4] presented a system for 3D motion retargeting combining a laser scanner and motion capture system. Their system produced a set of vectors of morph weights using optimization system applied on motion capture markers. Asthana et al. [14] proposed a real-time facial performance transfer method through modeling parametric correspondence for 2D images. Noh et al [6] introduced a method for reusing animation data for motion retargeting. Movements of the source face are represented by motion vectors which were then transferred to the target face model. However, that method only works well for cloning facial expressions for highly resemble face shapes.

3. DEFORMATION TRANSFER

3.1 Find Correspondence

Due to variable number of vertices and mesh types across different individuals, we need to establish correspondence between the source face and the target face model. Since it is a nonlinear problem to fit a template mesh to a target model data with different vertex number and topologies, the optimization could be trapped in local minima leading to fake mesh fitting. We used a method similar to the one proposed by Allen et al [28]. The registration process consists of two stages: rigid registration and nonrigid registration. One result of registration example is demonstrated in Fig. 1.

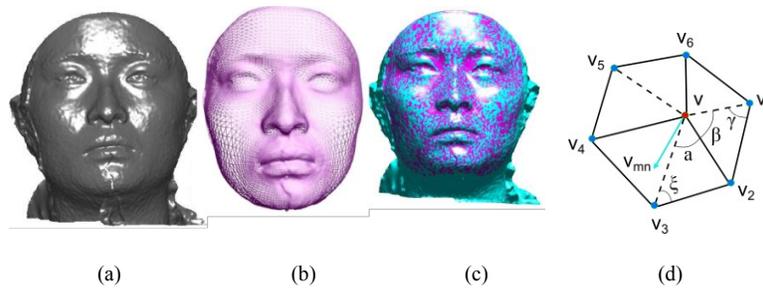(a)   (b)   (c)   (d)

Figure 1. An example of registration result. (a) Geometry of original face models; (b) conformed mesh models; (c) overlap of conformed face models on original models illustrating the error of registration; (d) Illustration of 1-ring vector

3.2 Deviation Transfer

Sorkine et al. [22] proposed a method for Laplacian surface editing, which is invariant to rotation and scaling. The sensitivity to linear transformation of Laplacian coordinate is overcome by applying a transformation T_i for each vertex v_i. As a consequence, $T_i(V')$ is a function of V', so the reconstructed shape is obtained by minimizing the following error function:

$$E(\widetilde{V}) = \sum_{i=1}^{n} \| T_i(\widetilde{V})\delta_i - L(\widetilde{v}_i) \|^2 + \sum \| \widetilde{v}_i - u_i \|^2 \qquad (1)$$

Inspired by the Laplacian coordinate method, we propose an approach based on local geometry encoding, which is based on the relationship of 1-ring vectors between the original shape and the deformed shape. Laplacian coordinate encodes vertex v_i in 1-ring neighbours to one vector δ_i, while we try to explicitly use these 1-ring neighbours. Mean curvature normal as an approximation of Laplacian coordinates in 1-ring neighbourhood has the similar characteristic of capturing the local geometric information with Laplacian coordinates, which is demonstrated in Fig. 1 (d).

In this paper, we propose to use mean curvature normal [29] as the extra vector of 1-ring vector for encoding the local deformation information which is to be introduced as follows.

We use 1-ring vector to capture local mesh detail. Let V be 1-ring vector, V=[v_i-v_{i1} v_i-v_{i2},...v_i-v_{im}], where are 1-ring neighbours of v_i. We further take the mean curvature into account in 1-ring vector to capture more local surface information in 1-ring neighbourhood. Therefore 1-ring vector V=[v_i-v_{i1} v_i-v_{i2},...v_i-v_{im}, v_n],   where v_n is the mean curvature normal vector at vertex v_i.

We suppose that there exists a transformation matrix T_i, which relates the 1-ring vector V of the original shape to the deformed shape. Thus the error function can be formulated as follows:

$$E(\widetilde{V}) = \sum_{i=1}^{n} \| T_i (V_i) V_i - \widetilde{V}_i \|^2 \qquad (2)$$

Where $\widetilde{V}_i$ is the 1-ring vector of $\widetilde{v}_i$ on the deformed shape. Specifically, for vertex v_i, this relationship can be expressed as $T V = \widetilde{V}$.

We notice that in (2), for a given shape M and its deformed shape $\widetilde{M}$ and 1-ring vector $\widetilde{V}_i$ we can simply work out the transformation T for vertex v_i: $T = \widetilde{V}V'(VV')^{-1}$.

Specifically, our goal is to transfer this deviation from the source face to a different neutral face model. The basic idea is that we expect the transformation matrix T to encode the deformation information for each 1-ring vector, which is transferable to any face models. Then, we can apply transformation matrix T to the target face model to obtain the deformation. Ideally, by applying the transformation matrix T to the neutral face model, we can obtain its deformed expression model. Thus, the minimization of error function in (2) can be reformulated as: $\arg\min_{T_i} \left(\| T_i V_i^t - \widetilde{V}_i^t \|^2 \right)$ where V_i^t and $\widetilde{V}_i^t$ are the 1-ring vector for the target neutral and unknown deformed face model.

To prevent the edge length of 1-ring neighborhood from changing in any cases, we apply the following edge energy term presented in [25]: $\sum \| (v_i - v_j) - r(v_i - v_j) \|^2$.

Thus, deformation can be achieved by minimizing the following energy:

$$E = \| T_i V_i^t - \widetilde{V}_i^t \|^2 + \| HV - r(V) \|^2 \qquad (3)$$

The solution to the minimization problem of the above equation is equivalent to solving a system of linear equation, which can be represented as the following form:

$$\arg\min_{\widetilde{v}_1...\widetilde{v}_n} \| A\widetilde{X} - b \|_2^2 \qquad (4)$$

where $\widetilde{X} = [\widetilde{v}_1, \widetilde{v}_2,...\widetilde{v}_n, \widetilde{v}_{m1}, \widetilde{v}_{m2},...\widetilde{v}_{mn}]'$ is a vector consisting of $3xn$ unknown entries of the deformed face model: vertex $\widetilde{v}_i$ and mean curvature normal vector $\widetilde{v}_{mi}$. And b contains information from the source deformation and given target neutral face vertices. Each entry of b is computed as $b_i = [T_i'T_i]^{-1}T_iV$.

The dimension of large and sparse matrix A is [3xn, 3xn], which associates unknown $\widetilde{X}$ with b. The linear system can be solved using the normal equation:

$$\widetilde{X} = (A'A)^{-1}A'b. \qquad (5)$$

It should be noted that by solving (5), we obtain both vertex and mean curvature normal of the deformed face model. The first n entries of $\widetilde{X}$ correspond to vertices of the new transformed face model. Since the model is in 3D space, each entry actually consists of 3 components (x, y, z).

It is expensive to solve the linear system directly. It takes more time to solve with a dense model. Therefore, it is more efficient to decompose A'A by LU factorization in advance for computation. Then component x, y and z are solved separately in (5), whereas we only need to decompose A'A once. Fig. 2 (a) demonstrates the examples of facial expression mapping.

After deforming the template mesh to the target face model, the deformed template mesh is nicely fitted to the target face. We then parameterize the target face in the space of the source model. More precisely, we concern only those vertices falling in the surface of the source model, since those vertices are possible to move driven by the conformed template mesh. Here, two steps are conducted to achieve this movement mapping: vertex projection and barycentric coordinate calculation. Fig. 2 (b) shows examples of the movement mapping result for a face model of point cloud data.

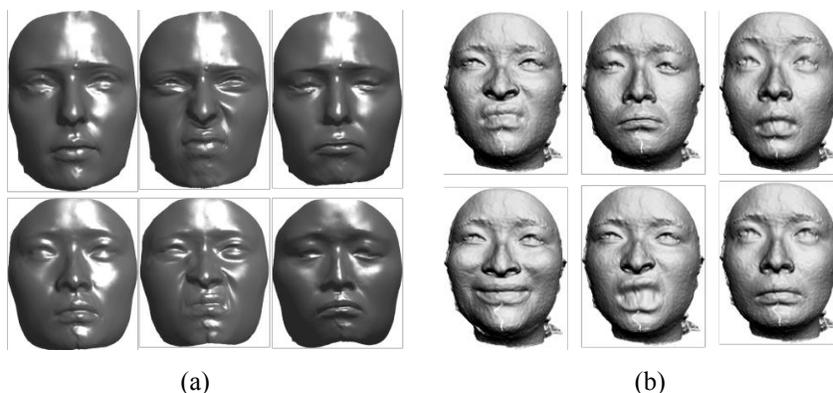(a)   (b)

Figure 2. (a) Facial expressions are transferred from the source face models (top row) to the target face model (bottom row). Column 1 is neutral face model. Expressions from column 2 to 3 are disgusting, sad and surprise. (b) Illustration of movement mapping result of female face (point cloud): disgust, sad surprise, happiness, anger and fear.

4. ANIMATION GENERATION

With the proposed method, it is easy to automatically animate the target face model. In this paper, we use the method developed by Yu etc. [27] to obtain the animation curve which is controlled by six parameters. Fig. 3 shows some animation examples.
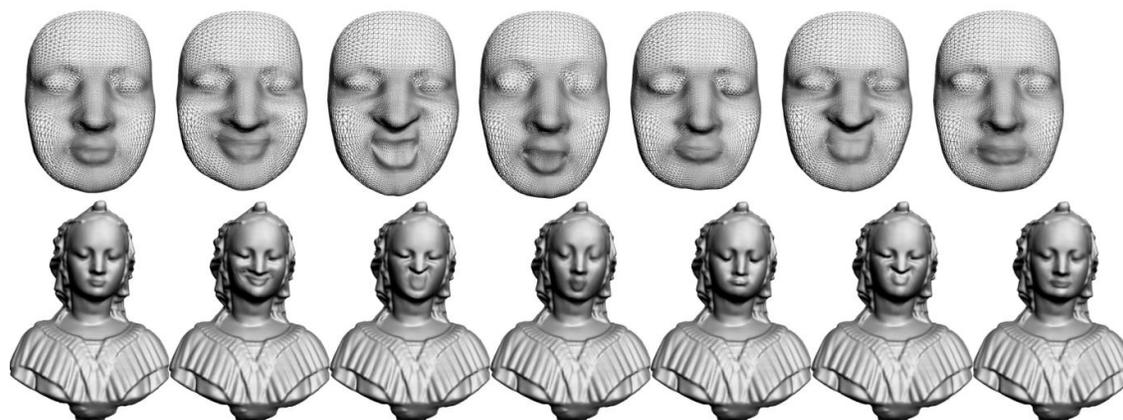

Figure 3. Example of facial expression generation for a bust model. Top row: peak frame of facial expressions of the conformed mesh; bottom row: peak frame of the final facial expressions.

5. CONCLUSION

We have presented a 3D facial expression mapping system for automatic facial animation generation. The facial expression deviation is transferred based on local encoded geometric information of face models. We have demonstrated that the proposed method works for different types of face models, such as point cloud model, triangular face model and quadrilateral mesh models. Apart from manually labeling a few feature landmarks on the face, all the rest procedure is automatic, which enables the method work very efficient.

To improve the current method, we would like to make it as a fully automatic procedure by developing an automatic and precise feature point detection method on 3D faces. Furthermore, we plan to improve the computation of the expression mapping method to make the animation generation more efficiently.

REFERENCES

[1]   N. Ersotelos, F. Dong, Building Highly Realistic Facial Animation: A Survey, *The Visual Computer*, Vol.24, Iss.1, Pp.13,2008

[2]   Z. Deng And J. Y. Noh, Computer Facial Animation: A Survey, , Pages 1-28, *Springer Press, Data-Driven 3d Facial Animation*, 2007

[3]   L. Zhang, N. Snavely, B. Curless, S.M. Seitz, Spacetime Faces: High Resolution Capture For Modeling And Animation. *ACM Transactions On Graphics* 23, 3 548-558,2004

[4]   C. Curio, M. Breidt, M. Kleiner, Q. Vuong, M. Giess, H. Bulthoff. Semantic 3D Motion Retargeting for Facial Animation.. In: Proceedings of the ACMAPGV; 2007. p.59–64.

[5]   F. I. Parke And K. Waters. Computer Facial Animation. A K Peters, Wellesley, Massachusetts, 1996.

[6]   J. Noh, U. Neumann: 'Expression Cloning'. *Proc. Siggraph*, 2001, Pp. 277–288

[7]   V. Blanz, C. Basso, T. Poggio, T. Vetter: 'Reanimating Faces In Images And Video'. *Proc. Eurographics*, 2003, Pp. 641–650

[8]   Q. Zhang, Z. Liu, B. Guo, D. Terzopoulos, H. Shum: 'Geometry-Driven Photorealistic Facial Expression Synthesis', *IEEE Trans. Vis. Comput. Graphics*, 2006, 12, (1), Pp. 48–60

[9]   L. Williams, 1990. Performance-Driven Facial Animation. In *Siggraph '90, Acm Press*, 235–242.

[10]   L.Zhang, N.Snavely, B. Curless And S. M. Seitz, Spacetime Faces: High Resolution Capture For Modeling And Animation, *Acm Transaction On Graphics*, Vol. 23 No.3 Pp.548-558, 2004

[11]   K.Kahler J.Haber H. Y.H.S., Head Shop: Generating Animated Head Models With Anatomical Structure. *Acm Siggraph/Eurographics Symposium On Computer Animation*, Pp.55-63

[12]   E. Sifakis, A. Selle, A. Robinson-Mosher, R. Fedkiw. Simulating Speech With A Physics-Based Facial Muscle Model. *Eurographics* 2006, 261-270.

[13]   K. K¨ahler, J. Haber, and H. P. Seidel, "Geometry-Based Muscle Modeling For Facial Animation," in Proc. Graphics Interface, 2001, pp. 37–46.

[14]   A. Asthana, M. de la Hunty, A. Dhall, & R. Goecke, Facial performance transfer via deformable models and parametric correspondence. IEEE Transactions on Visualization and Computer Graphics, 18(9), 1511-1519. 2012.

[15]   D. Terzopoulos And K. Waters. Physically-Based Facial Modelling, Analysis, And Animation. The Journal Of Visualization And Computer Animation, 1:73–80, 1990.

[16]   F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, And D. H. Salesin, "Synthesizing Realistic Facial Expressions From Photographs," In Siggraph'98: Proc. Of The 25th Annual Conference On Computer Graphics And Interactive Techniques. New York, Ny, Usa: Acm Press, 1998, Pp.75–84.

[17]   Q. Zhang, B. G. Z. Liu, And H. Shum, "Geometry-Driven Photorealistic Facial Expression Synthesis," In Proc. Eurography/Siggraph Symposium On Computer Animation, 2003, Pp. 177–187.

[18]   R. Sumner, P. J. Deformation Transfer For Triangle Meshes. In: Proceedings Of The Acm Siggraph 04; 2004. P. 399–405.

[19]   F. Parke, Computer Generated Animation of Faces, In Acm National Conferences, Vol. 1. Acm Press, 1972, Pp.451-457.

[20]   R. Hartley. A. Zisserman. Multiple View Geometry In Computer Vision. 2nd Ed. Cambridge University Press, 2004.

[21]   J. 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, Pages 193–206, Airela-Ville, Switzerland, 2003. Eurographics Association.

[22]   O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. R¨Ossl, and H.-P. Seidel. Laplacian Surface Editing. In Proc. Of Eurographics Symposium On Geometry Processing 04, Pages 179–188, 2004.

[23]   D. Cosker, R. Borkett, D. Marshall And P. L. Rosin, "Towards Automatic Performance Driven Animation Between Multiple Types Of Facial Model", Iet Computer Vision, Vol. 2, No. 3, Pages 129-141. 2008.

[24]   B. Choe, H. Lee, H. Ko: Performance-Driven Muscle-Based Facial Animation. J. Vis. Comput. Animat. 12, 67–79 (2001)

[25]   Y. Weng, W. Xu, Y. Wu, K. Zhou, B. N. Guo. 2D Shape Deformation Using Nonlinear Least Squares Optimization The Visual Computer. Volume 22 Issue 9, September 2006

[26]   F. I. Parke,. Parameterized models for facial animation. IEEE computer graphics and applications. Vol.2, No 9, pp 61-68.1982

[27]   H. Yu, Oliver Garrod and Philippe Schyns. Perception-Driven Facial Expression Synthesis, *Computer & Graphics*, Vol. 36, Issue 3, May 2012, pp152–162.

[28]   Allen, B., Curless, B., and Popovic, Z. The space of human body shapes: reconstruction and parameterization from range scans. ACM Trans. Graph., 22(3):587–594, 2003

[29]   M. Meyer, M. Desbrun., P. Schröder, A. H. Barr: Discrete Differential-Geometry Operators For Triangulated 2-Manifolds. In: Proc. Of Visualization And Mathematics, Pp. 35–57 (2003)