REALISTIC FACIAL ANIMATION SYNTHESIS AND TRANSFER BASED ON FLEXIBLE EXPRESSION RATIO IMAGE

Pith Xie[1,2], Yiqiang Chen[1], Junfa Liu[1]

[1](Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)
[2](Shanghai Institute of Computing Technology, Chinese Academy of Sciences, Shanghai 201203, China)
pithxie@yahoo.com,yqchen@ict.ac.cn,jfliu@ict.ac.cn

Abstract

Geometry-controlled image warping performs well in exhibiting shape variations but bad in exhibiting wrinkle such as fossette. Successfully applied in expression cloning, Expression Ratio Image(ERI) provides a way to quantize transferable common wrinkle. However, the warping used in ERI makes it difficult to restrain alignment error, which however prevents further analysis such as PCA. In this paper, Flexible Expression Ratio Image(FERI) is defined and further compressed as Eigen FERI with PCA to quantize compactly common wrinkle. Based on FERI, a geometry-mapped mechanism consisting of shape-varying mechanism and wrinkling mechanism is constructed to synthesize and transfer realistic facial animation. Given the geometric parameter Facial Animation Parameter(FAP), the shape variation deriving from shape-varying mechanism and the wrinkle deriving from wrinkling mechanism combine to generate realistic facial animation. Having no need for any example photo gallery but only two frontal neutral photos respective captured from source face and target face, as demonstrated in the experiment, the geometry-mapped mechanism based on FERI can map FAP series to the synchronous realistic source animation and target animation.

1. INTRODUCTION

Realistic facial expression synthesis is one interesting yet difficult problem in computer graphics, which obtains increasingly application in many fields, such as entertainment industry.

Geometry-controlled image warping[1,2] can synthesize shape variation efficiently. However, it does not generate wrinkle caused by muscle contraction. Example-based methods[3,4], interpolating and extrapolating on example images, can exhibit lifelike shape variation as well as wrinkle and has been applied to synthesize realistic facial animation. For any novel target, however, they need a photo gallery containing abundant expressions yet difficult to build in practice.

Du et al.[5] approximates a function with machine learning algorithm to fit the mapping from shape variation to dynamical texture, and then imports flexible model to synthesize realistic facial animation. Their method enhances the generality of synthesis mechanism by machine learning algorithm. However, dynamic texture reflects more personalized texture than common wrinkle, which can not be transferred between different targets.

All above researches propel the research on realistic facial animation synthesis. However, they are difficult to be used to transfer realistic facial animation, which limits their spreading in application. At this background, the research on facial animation transfer attracts more and more attention.

Noh and Neumann[6] propose an expression transfer technology based on 3D morphing. It firstly computes the vertex motion vectors, then transforms them to morph the target model to exhibit various expressions. This technology elaborates on transferring expression between models with different topology.

Pyun[7] et al. propose an example-based expression transfer technology. It constructs key-models of both source model and target model, then evaluate the weight values for the target key-models and output a expressive model by blending the target key-models with those weights. This technology can transfer facial animation exactly, however, it can not get rid of the key-model example and transfers facial animation to any novel target without key-models.

Liu[8] et al. align the expressive face image to the reference image by warping, and then extract ERI to represent wrinkle. Independent of personalized texture, ERI can be successfully applied in expression cloning. However, the warping used in ERI makes it difficult to restrain alignment error, which however prevents further analysis such as PCA.

To sum up, the researches have made a great progress respectively on realistic facial animation synthesis or transfer. However, the researches related to both fields are relatively deficient. In this paper, Flexible Expression Ratio Image(FERI) is defined and then compressed as Eigen FERI with PCA to quantize compactly transferable common wrinkle. Then, we construct a geometry-mapped mechanism consisting of shape-varying mechanism and wrinkling mechanism; Finally, given geometric parameter FAP, shape-varying mechanism generates shape variation while wrinkling mechanism generates wrinkle, both them combine to output the corresponding realistic facial animation. Having no need for any example photo gallery but only two frontal neutral photos respective captured from the source face and target face, the geometry-mapped mechanism based on FERI can map FAP series to the synchronous realistic source animation and target animation.

2. FLEXIBLE EXPRESSIVE RATIO IMAGE

Realistic facial animation requires exhibiting wrinkle resulting from muscle contraction. Muscle contraction inspires not only shape variation, but also wrinkle. For mathematic modeling, the physical feature should be quantized as parameters.

2.1 Physical Feature Quantization

The physical features for modeling facial animation includes muscle contraction, shape variation, and wrinkle. FAP provided by MPEG-4 has high versatility and low calculation complexity and thus is used here to quantize muscle contraction. Shape variation is usually quantized as a colorful image resulting from warping.

ERI quantizes transferable common wrinkle as a ratio matrix, which is constructed by 3 steps: marking a few feature points along facial edge; aligning the expressive image to reference image through warping; computing the ratio image. By warping, various expressive images is aligned with the same mesh. So in essence, ERI behave as a rigid ratio image.

The warping used in ERI may be done by pixel value interpolation, e.g. triangulation or thin plate spline method[9]. However, aligning edge exactly is so difficult that the alignment error is reserved in ERI. When PCA is imported to compress ERI, the partial alignment error around feature edge will increase the reconstruction error. Taking partial occlusion as example, partial error will be converted into global noise in the PCA reconstruction[10]. In the AAM(Active Appearance Models), the PCA (Principal Component Analysis) reconstruction error is used as a distance measure for the evaluation of alignment quality[11]. Thus, it is better to reduce partial alignment error to do PCA.

2.2 Flexible Expressive Ratio Image

In this paper, Flexible Expressive Ratio Image(FERI) is defined to quantize common wrinkle.

Above all, we divide pixels in face image into 7 subset just like Fig.1(a). Those pixels closed by the outmost edge curve compose face set F. Those pixels closed by other 6 edge curves compose feature subset S. The difference set F−S is wrinkle subset W. As the key to character recognition, Feature subset S mainly reflects personalized information. On the contrary, wrinkle subset W mainly reflects transferable common wrinkle information.

FERI behaving as a ratio matrix, in which each ratio derives from the following arithmetic and logic operation:

$$\begin{cases} R(i,j) = I_E(i,j)/I_R(i,j) & (i,j) \in W_E \cap W_R \\ R(i,j) = 1.0 & (i,j) \notin W_E \cap W_R \end{cases} \tag{2.1}$$

In the above formula, $R(i,j)$ represents the intensity in the coordinate (i,j). Suffix R and E represent reference face and expression face.

FERI extraction algorithm includes 3 procedures as follows:

(a) (b) (c)

Fig.1. FERI calculation. (a) Edge curves(green) in reference face;(b) Edge curves(green) in expressive face;(c) FERI

1. Mark dense edge data points along the edge curves respectively surrounding the feature subset and the wrinkle subset in the initial frame. Due to the salient gray of the edge, optical flow can be used to track those edge data points frame by frame;
2. Use polar-coordinate spline interpolation with periodic boundary conditions to fit the closed edge curves, such as Fig.1(a), Fig.1(b);
3. Calculate FERI such as Fig.1(c) according to formula (2.1).

The pixel set out of the $W_E \cap W_R$ can be divided into 2 portions. One portion is the union of feature subset mainly exhibiting shape variation, where intensity ratio varies so intensive that the ratio in wrinkle subset is hardly reflected after PCA. Thus, the intensity ratio of this portion is screened by constant 1.0. The other portion mainly exhibiting the background does not deserve attention, whose intensity ratio is also screened by constant 1.0.

Fig.1(c) intuitively shows a FERI corresponding to Fig.1(b). In step 1, the optical flow tracking error is checked by eye and corrected manually. At last, the approximation error of spline function can be restricted to a fine order of magnitude. The FERI extraction algorithm can effectively restrict the alignment error and helps to the later PCA.

FERI behaves as a pixel matrix with high redundancy. So it has to be compressed as Eigen FERI by PCA to act as wrinkle parameter.

3. GEOMETRY-MAPPED MECHANISM BASED ON FERI

After physical feature quantization, the geometry-mapped mechanism based on FERI is modeled as the running flow in Fig.2.

The shape-varying mechanism takes the triangular facial mesh as template and then generates shape variation by affine transformation[2]. The wrinkling mechanism contains 3 components. The first trains a SVM on sample set to predict the mapping from FAP to Eigen FERI. The second reconstructs FERI with inverse PCA on Eigen FERI; The third smoothes the global noise with Gauss filter.

3.1 Shape-Varying Mechanism

In this paper, vertex of triangular facial mesh is classified to active vertex and passive vertex. Active vertex is directly controlled by FAP while passive vertex is only influenced by

FAP. As shown in Fig.3, the black point is active vertex while the red point is passive vertex, all them connect one another to construct a triangular facial mesh.

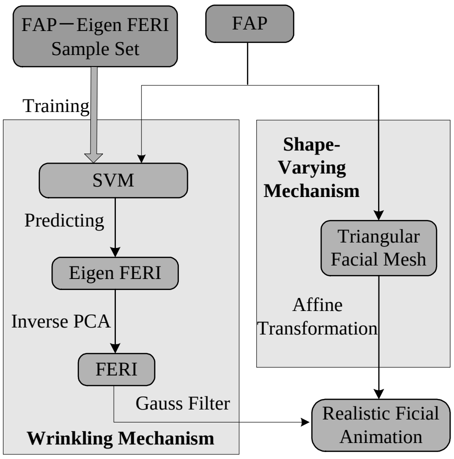
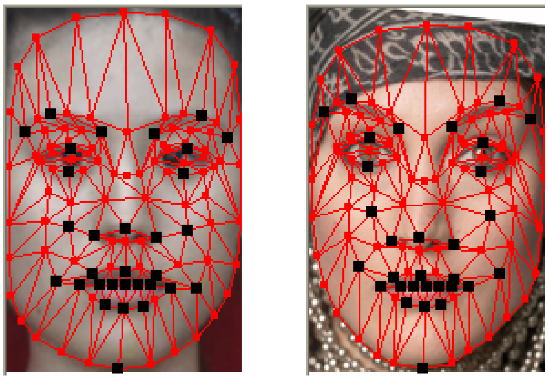


Fig.2 The running flow of the geometry-mapped mechanism based on FERI



(a) (b)

Fig.3 Triangular facial mesh. (a)Source mesh (b)Target mesh

FAP value does not determine the displacement of passive vertex but that of the active vertex. Thus, thin plate spline is used to evaluate the displacement of passive vertex. Then, those patchs in 2D triangular mesh are warped by affine transformation to generate a novel frame. Affine transformation can excellently exhibit shape variation without wrinkle.

3.2 Wrinkling Mechanism

According to physiological constraints, muscle contraction inspires wrinkle. The causality between them can be abstracted as a nonlinear mapping function from FAP to Eigen FERI. When accepting little error, the mapping function can be approximated by machine learning algorithm.

3.2.1 FAP-Eigen FERI Sample Set

Lucas-Kanade optical flow tracking algorithm[12] is used to track active vertex, as well as edge data points in the edge curves. Then synchronous FAP and Eigen FERI compose a sample set.

According to MPEG-4, 35 active vertex is marked on the performer's face with yellowy watercolor before video recording, which helps to optical flow tracking frame by frame while reserving the original wrinkle. Fig.4 shows those tracked points with greeness. Then FAP value of each frame can be calculated.

As to edge curves, whose gray is so salient that those edge data points in them are easy to be localized in the initial frame and tracked in the following frames. Fig.5 shows those tracked edge data points with green and those edge curves fitted by spline function with blueness. Then Eigen FERI is extracted according to (2.2) and the synchronous FAP-Eigen FERI series compose a sample set:

Fig.4 Tracked active vertex

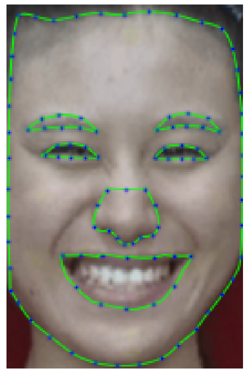


Fig.5 Edge closed curve fitted by edge data points

$$\{(x_1,y_1),(x_2,y_2),\cdots\cdots,(x_l,y_l)\}\subset X\times R \quad (3.1)$$

In the expression (3.1), X represents FAP while Y represents Eigen FERI.

3.2.2 SVM Training And Predicting

SVM proposed by Vapnik[13] et al. is a novel machine learning method aiming at data classfication and regression. Based on statistical learning theory and structure risk minimization criterion, SVM find a tradeoff between the complexity and learning ability so as to enhance the generality.

The regression function mapping FAP to Eigen FERI can be abstracted as the following expression:

$$y=<w,x>+b \qquad w\in x, b\in R \quad (3.2)$$

$<\bullet,\bullet>$ represents the dot metrix. In SVM, the aim is to solve a quadratic programming problem:

$$\min_{w,b,\xi,\xi^*}\frac{1}{2}w^T w+c\sum_{i=1}^{l}(\xi_i+\xi_i^*) \quad (3.3)$$

$$S.T.\ \ w^T\Phi(x_i)+b-S_i\le c+\xi_i \quad (3.4)$$

$$S_i-w^T\Phi(x_i)-b\le c+\xi_i^* \quad (3.5)$$

$$\xi_i,\xi_i^*\ge 0, i=1,\cdots\cdots,l \quad (3.6)$$

Accordingly, formula (3.2) can be converted into:

$$y=\sum_{i=1}^{l}(-\alpha_i+\alpha)k(x_i,x)+b \quad (3.7)$$

Those parameters in (3.7) can be solved by SVM training algorithm. Finally, the trained SVM is used to predict Eigen

FERI. In FERI reconstruction, Gauss filter is imported to filter the noise deriving from SVM prediction error.

4. Experiment Analysis

In this paper, a facial motion video including 950 frames is captured with the resolution 130×200, which covers basic expressions.

First, 68-dimension FAP and 23-dimension Eigen FERI are extracted respectively. Second, a SVM is trained on 950 samples to approximate the mapping from FAP to Eigen FERI. Third, source facial mesh and target facial mesh are respectively constructed as Fig.3. Fourth, the geometry-mapped mechanism based on FERI maps the same FAP series to 2 realistic facial animations as Fig.6 and Fig.7.

By comparing Fig.6 and Fig.7, the synthesized animation and the transferred animation are both realistic. In addition, the synthesized animation and transferred animation corresponds to each other so excellently that they are played synchronously.

Fig.6 Geometry-mapped realistic facial animation synthesis based on FERI

Fig.7 Geometry-mapped realistic facial animation transfer based on FERI

The experiment demonstrates that the geometry-mapped mechanism based on FERI can not only synthesize realistic facial animation, but also transfer realistic facial animation.

5. Conclusion

The FERI extraction algorithm restricts the alignment error while quantizing transferable common wrinkle. Due to the lower computation complexity, the shape-varying mechanism and the wrinkling mechanism generate shape variation and wrinkle efficiently.

Having no need for any example photo gallery but only two frontal neutral photos respectively captured from source and target face, the geometry-mapped mechanism based on FERI can synthesize and transfer realistic facial animation. Due to the loose precondition, it can be applied in complex and uncontrolled environment, such as internet interaction.

Although FERI in this paper is used to animate 2D facial model, it can be converted into normal maps[14] to animate 3D facial model.

6. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 60303018, 60575032), Shanghai Qingmingxing Project (05QMX1447) and Beijing Multimedia and Intelligent Software Technology Laboratory.

7. REFERENCE

[1] T. Beier, S. Neely, "Feature-based image metamorphosis," In Computer Graphics, pages 35-42. Siggraph, July 1992.

[2] G. Wolberg, Digital Image Warping. IEEE Computer Society Press, 1990.

[3] Moiza. G, Tal. A, Shimshoni. I, Barnett. D, Moses. Y, "Reconstruction of movies of facial expressions," IEEE ICCV 2001, Page(s):8 – 15, 2001

[4] Qingshan Zhang, Liu Z., Gaining Quo, Terzopoulos D., Heung-Yeung Shum, "Geometry-driven photorealistic facial expression synthesis," Visualization and Computer Graphics, IEEE Transactions on Volume 12, Issue 1, Jan.-Feb. 2006 Page(s): 48 – 60.

[5] Du YZ, Lin XY, "Dynamic facial texture generation based on shape-appearance dependence mapping strategy," Journal of Software, 2004,15(2):308~316.

[6] J.Y. Noh and U. Neumann, "Expression Cloning," Proc. ACM SIGGRAPH Conf., pp. 277-288, 2001.

[7] H. Pyun, Y. Kim, W. Chae, H.W. Kang, and S.Y. Shin, "An Example-Based Approach for Facial Expression Cloning, " Proc. 2003 ACM SIGGRAPH/Eurographics Symp. Computer Animation, pp. 167-176, 2003.

[8] Liu Z, Shan Y, Zhang Z, "Expressive expression mapping with ratio images, " In Proc. SIGGRAPH'01, Los Angeles, CA, USA, 2001, pp.271-276.

[9] Stan Z. Li, Ying Zheng, Zhen Lei, ZengFu Wang, "Shape and Texture Based Deformable Models for Image Analysis, " In Deformable Models - Theory & Biomaterial Applications, Suri, Jasjit S. & Farag, Aly (Eds.). Springer. ISBN: 0-387-31204-8. May 2006

[10] C. Hu, R. Feris, and M. Turk, "Active wavelet networks for face alignment, " In British Machine Vision Conference, East Eaglia, Norwich, UK, 2003.

[11] Xiangsheng Huang, Stan Z. Li, Yangsheng Wang, "Evaluation of Face Alignment Solutions Using Statistical Learning, " FGR 2004: 213-218

[12] B. Lucas , T. Kanade, " An iterative image registration technique with an application to stereo vision, " IJCAI’81, pages 674–679, 1981.

[13] V.Vapnik, Statistical Learning Theory, Wiley, New York, NY, 1998.

[14] Pei-Hsuan Tu, I-Chen Lin, Jeng-Sheng Yeh, Rung-huei Liang, Ming Ouhyung, "Expression detail for realistic facial animation, " Proc. CAD/ Graphics 2003, pp. 20-25, Macau, Oct 28-30, 2003.