Intuitive Quasi-Eigen Faces

Ig-Jae Kim*
Graphics and Media Lab, Seoul National University
Korea Institute of Science and Technology

Hyeong-Seok Ko†
Graphics and Media Lab, Seoul National University

Abstract

In blendshape-based facial animation, two main approaches are used to create the key expressions: manual sculpting and statistically-based techniques. Hand-generated expressions have the advantage of being intuitively recognizable, thus allowing animators to use conventional keyframe control. However, they may cover only a fraction of the expression space, resulting in large reproduction/animation errors. On the other hand, statistically-based techniques produce eigenfaces that give minimal reproduction errors but are visually non-intuitive. In this paper we propose a technique to convert a given set of hand-generated key expressions into another set of so-called *quasi-eigen* faces. The resulting expressions resemble the original hand-generated expressions, but have expression space coverages more like those of statistically generated expression bases. The effectiveness of the proposed technique is demonstrated by applying it to hand-generated expressions.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling —Curve, surface, solid and object modeling; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism —Animation;

Keywords: facial animation, blendshape interpolation, principal component analysis

1 Introduction

The facial expressions of animated characters play a central role in delivering the story. For an animation studio, therefore, the ability to generate expressive and plausible facial expressions is a critical skill. Nevertheless, as yet no standard procedure for generating facial expressions has been established; when facial animations are required, a whole gamut of approaches are mobilized, ranging from labor-intensive production work to state-of-the-art technical supports. This paper proposes a small but very useful innovation in the area of 3D facial animation, which can be adopted in a wide range of facial animation productions.

Probably, the most popular approach currently used in facial animation productions is the so-called *blendshape* technique, which synthesizes expressions by taking a linear combination of a set of pre-modeled expressions. We call this expression set the *expression basis*. Many commercial animation packages such as Maya and Softimage support blendshape-based facial animation. The technique we develop in this paper is for systems of this type.

*e-mail: ijkim@graphics.snu.ac.kr
†e-mail: ko@graphics.snu.ac.kr

GRAPHITE 2007, Perth, Western Australia, December 1–4, 2007.
© 2007 ACM 978-1-59593-912-8/07/0012 $5.00

A fundamental question in developing a blendshape-based facial animation system is how to control the expressions. One approach is to let the animators manually control the weights assigned to each member of the expression basis set in order to produce the desired expression sequences. Another popular approach that can be taken provided facial motion capture is available, is to set up the system so that the facial animation is driven by a human performance. In this approach, if the basis is taken from the human subject, in principle the original performance can be reproduced. Although such reproduction may not be needed in an animation production, it has theoretical significance to developers because it can utilized as a benchmark of a blendshape technique: if a method can accurately reproduce the original performance, it can produce other facial animations accurately. The present work assumes that the facial animation system is operated by performance-driven control, but also assumes that manual control can be added whenever the results need to be edited.

Another fundamental issue that must be resolved when developing a blendshape technique is how to form the expression basis. The present work is related to this issue. A casual approach practiced by many animation studios is to use an expression basis comprised of manually modeled, intuitively recognizable key expressions. The basis[1] should contain sufficient elements span the desired range of expressions. An advantage of using a hand-generated basis is that the combinations of basis elements produce somewhat predictable results. A disadvantage of this approach is that the linear combinations may cover only a portion of full range of facial expressions. When the system is used to reproduce a human performance, the lack of coverage manifests as reproduction errors. In the context of blendshape-based reproduction of human performances, another well-established approach to obtain the expression basis is to use principal component analysis (PCA). In this method, a set of mutually-orthogonal principal components that spans the expression space is generated by statistical analysis of performance data. Because this technique gives quantitative information on the coverage of each component, by selecting the dominant components we can form an expression basis whose coverage is predictable and greater than that of manually generated bases, resulting in more accurate reproduction of the original performance. A drawback of this approach is that the expressions corresponding to the principal components are visually non-intuitive. Hence animators cannot predict the expression that will be produced by a particular linear combination.

Here we propose a new approach to basis generation that gives coverages comparable to those of statistically generated bases while at the same time having basis elements with meaningful shapes. This approach is based on our observation that a hand-generated expression can be modified such that the resulting expression remains visually close to the original one but its coverage over the expression space increases. It is also based on the relaxation we take that the basis elements do not need to be strictly orthogonal to each other; they can still span the expression space. The remainder of this paper is organized as follows. Section 2 reviews the related work in facial animation. Section 3 gives the mathematical description of the

[1]The term "basis" is usually reserved for an independent set of elements that spans an entire space. In this paper, however, we use the term to loosely mean a set of expressions from which linear combinations are taken.

problem solved in this paper. Section 4 presents the procedure for obtaining the quasi-eigen faces. Section 5 reports the experimental results, and Section 6 concludes the paper.

2 Related work

A large number of techniques for synthesizing human expressions have been proposed since the pioneering work of [Parke 1972]. Facial expression can be viewed as resulting from the coordination of (mechanical) components such as the jaw, muscles, and skin. Various researchers have explored physically based techniques for synthesizing facial expressions[Waters 1987; Terzopoulos and Waters 1990; Essa and Pentland 1997; Kahler et al. 2001; Choe et al. 2001; Sifakis et al. 2005]. The present work takes a different approach: expressions are synthesized by taking linear combinations of several key expressions. Thus, instead of looking into the physics of facial components, we utilize facial capture data to obtain realistic results. In this section, we review previous work, with a focus on blendshape techniques and performance-driven facial animation techniques.

The blendshape technique has been widely used for expression synthesis. To generate human expressions in real-time, [Kouadio et al. 1998] used linear combinations of a set of key expressions, where the weight assigned to each expressioin was determined from live capture data. [Pighin et al.1998] created a set of photorealistic textured 3D expressions from photographs of a human subject, and used the blendshape technique to create smooth transitions between those expressions. [Blanz and Vetter 1999] introduced a morphable model that could generate a 3D face from a 2D photograph by taking a linear combination of faces in a 3D example database. To increase the covering range of the key expressions, [Choe and Ko 2001] let animators sculpt expressions corresponding to the isolated actuation of individual muscles and then synthesized new expressions by taking linear combinations of them.

A critical determinant of the quality of the expression generated by blendshape-based synthesis is the covering range of the key expressions being used. [Chuang 2002] used a PCA-based procedure to identify a set of key expressions that guarantees a certain coverage. However, the resulting principal components did not correspond intuitively meaningful human expressions. [Chao et al. 2003] proposed another basis generation technique based on independent component analysis. In the key expression set produced using this approach, the differences among the elements were more recognizable than those generated by [Chuang 2002]; however, the individual elements in the set still did not accurately represent familiar/vivid human expressions. As a result, conventional keyframe control is not easy using this approach. To enable separate modifications of specific parts of the face in a blendshape-based system, [Joshi et al. 2003] proposed automatic segmentation of each key expression into meaningful blend regions.

[Williams 1990] introduced a performance-driven approach to synthesize human expressions. This approach utilizes the human ability to make faces and has been shown to be quite effective for controlling high-DOF facial movements. The uses of this approach for blendshape-based reproduction of facial performances were introduced above. [Noh and Neumann 2001; Pyun et al. 2003; Na and Jung 2004; Wang et al. 2004] proposed techniques to retarget performance data to synthesize the expressions of other characters. Recently, [Vlasic et al. 2005] developed a *multilinear* model that can transfer expressions/speech of one face to other faces. Another class of performance-driven facial animation techniques is the speech-driven techniques. [Bregler et al. 1997; Brand1999; Ezzat et al. 2002; Chao et al. 2004; Chang and Ezzat. 2005; Deng et al. 2005] are several representative works exploring this research direction.

3 Problem Description

Let $\mathbf{v} = \mathbf{v}(t) = [\mathbf{v}_1^T, \ldots, \mathbf{v}_N^T]^T$ represent the dynamic shape of the 3D face model at time t. It is a triangular mesh consisting of N vertices, where $\mathbf{v}_i$ represents the 3D position of the i-th vertex. We assume that the geometry $\mathbf{v}^0 = [(\mathbf{v}_1^0)^T, \ldots, (\mathbf{v}_N^0)^T]^T$ of the neutral face is given. We also assume that motion capture data are given in a $3N \times L$ matrix $\Xi = [\mathbf{v}(1), \ldots, \mathbf{v}(L)]$, where L is the duration of the motion capture in number of frames. We are interested in finding a set of facial expressions, linear combinations of which span Ξ.

Let $\hat{E}^H = \{\hat{\mathbf{e}}_1^H, \ldots, \hat{\mathbf{e}}_n^H\}$ be the hand-generated expression basis that is given by the animator. Here, n is the number of elements and $\hat{\mathbf{e}}_i^H$ is the geometry of the i-th element. Let $\mathbf{e}_i^H$ represent the displacement of $\hat{\mathbf{e}}_i^H$ from the neutral face, i.e., $\mathbf{e}_i^H = \hat{\mathbf{e}}_i^H - \mathbf{v}^0$. In this paper, we call the set of displacements such as $E^H = \{\mathbf{e}_1^H, \ldots, \mathbf{e}_n^H\}$ also the (hand-generated) expression basis if it does not cause any confusion. When the weights w_i^H are given, we synthesize the expression $\mathbf{v}$ by

$$\mathbf{v} = \mathbf{v}^0 + \sum_{i=1}^{n} w_i^H \mathbf{e}_i^H. \tag{1}$$

A potential problem of the hand-generated expression basis E^H is that linear combinations of the basis elements may not span Ξ. The goal of this paper is to develop a procedure to convert E^H into another basis $E^{QE} = \{\mathbf{e}_1^{QE}, \ldots, \mathbf{e}_n^{QE}\}$, such that the new basis spans Ξ and each element $\mathbf{e}_i^{QE}$ visually resembles the corresponding element $\mathbf{e}_i^H$ in E^H. We call the elements in the new basis *quasi-eigen faces*[2].

4 Obtaining Quasi-Eigen Faces

If the facial vertices $\mathbf{v}_1, \ldots, \mathbf{v}_N$ are allowed to freely move in 3D space, then $\mathbf{v}$ will form a $3N$-dimensional vector space. Let us call this space the *mathematical expression space* $\mathbf{E}$. However, normal human expressions involve a narrower range of deformation. If we plot each expression in Ξ as a point in $3N$-dimensional space, the point cloud forms an approximate hyperplane. The PCA is designed to identify the orthogonal axes that spans the hyperplane. The analogical situation is shown in Figure 1. The 3D coordinate system can be viewed as $\mathbf{E}$, the dots as forming the Ξ-hyperplane, and the solid perpendicular axes as the principal components.

The procedure for obtaining the quasi-eigen faces is based on the principal components. Finding the principal components requires the point cloud to be centered at the origin. Let $\mu = [\mu_1^T, \ldots, \mu_N^T]^T$ be the mean of $\mathbf{v}$ where the summation is taken over the entire motion capture data Ξ. Then, we can obtain a centered point cloud $\tilde{\mathbf{D}} = [\tilde{\mathbf{v}}(1)^T, \ldots, \tilde{\mathbf{v}}(L)^T]^T$, where $\tilde{\mathbf{v}}(i) = \mathbf{v}(i) - \mu$. Now we construct the covariance matrix $\mathbf{C}$ using

$$\mathbf{C} = \frac{1}{L} \tilde{\mathbf{D}} \tilde{\mathbf{D}}^T. \tag{2}$$

$\mathbf{C}$ is a symmetric positive-definite matrix, and hence has positive eigenvalues. Let $\lambda_1, \ldots, \lambda_{3N}$ be the eigenvalues of $\mathbf{C}$ in order of magnitude, with λ_1 being the largest. The m eigenvectors

[2]The term "quasi-eigen faces" was chosen to indicate that its coverage of the motion capture data amounts to the eigenfaces.

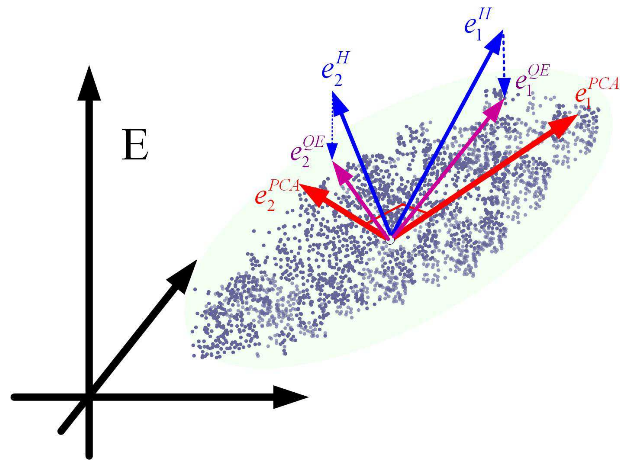

Figure 1: Analogical drawing of the motion capture data and the bases discussed in the paper (red arrows: PCA basis, blue arrows: hand-generated basis, purple arrows: quasi-eigen basis)

$E^{PCA} = \{\mathbf{e}_1^{PCA}, \ldots, \mathbf{e}_m^{PCA}\}$ corresponding to $\{\lambda_1, \ldots, \lambda_m\}$ are the principal axes we are looking for, the coverage of which is given by $\sum_{i=1}^{m} \lambda_i / \sum_{i=1}^{3N} \lambda_i$. The facial expressions in E^{PCA} are called the *eigenfaces*. Since the coverage is usually very close to unity even for small m (e.g., in the case of the motion capture data used in this paper, $m = 18$ covers 99.5% of Ξ), the above procedure provides a powerful means of generating an expression basis that covers a given set Ξ of expressions. A problem of this approach is that, even though the eigenfaces have mathematical significance, they do not represent recognizable human expressions.

In the context of generating the eigenfaces, we can now describe the method to convert the hand-generated expression basis into the quasi-eigen basis (i.e., the set of quasi-eigen faces). This method is based on our observation that the hand-generated elements may lie out of the hyperplane. In the analogical situation drawn in Figure 1, let's consider two hand-generated expressions (two 3D vectors in the figure) that do not lie on the hyperplane. Although any linear combination of the two expressions should be allowed in principle, if we stipulate that the result must lie on the hyperplane to form a valid expression, then the ratio between the weights assigned to the two expressions must be fixed. This means that, the two expressions, rather than spanning a two dimensional range of expressions, in fact only cover a one dimensional range, resulting in significant coverage loss. Linear combinations that are formed disregarding this constraint will be positioned out of the hyperplane, which explains why reproduction of an original performance generated using a hand-generated bases usually contain large errors.

A simple fix to the above problem would be to project the hand-generated elements onto the hyperplane; the quasi-eigen faces we are looking for in this paper are, in fact, the projections of the hand-generated basis elements. To find the projection of a hand-generated element onto each principal axis, we first compute

$$w_{ij}^{\text{PCA-to-QE}} = \mathbf{e}_j^{PCA} \cdot (\mathbf{e}_i^H - \mu), \tag{3}$$

where i ranges over all the hand-generated elements, and j ranges over all the principal axes. Now, we can obtain the quasi-eigen faces by

$$\mathbf{e}_i^{QE} = \mu + \sum_{j=1}^{m} w_{ij}^{\text{PCA-to-QE}} \mathbf{e}_j^{PCA}. \tag{4}$$

With the quasi-eigen basis, a general expression is synthesized by the linear combination $\sum_{i=1}^{n} w_i^{QE} \mathbf{e}_i^{QE}$. We would note that in most blendshape-based facial animation systems, the weights are always positive and, in some cases, are further constrained to lie within the range $[0, 1]$ in order to prevent extrapolations. When the eigenfaces are used, however, the weights w_i^{QE} is supposed to take on both positive and negative values. The weights of the quasi-eigen basis should be treated like the eigenfaces: even though they are not orthogonal, their ingredients are from an orthogonal basis. Allowing negative weights obviously increases the accuracy of the reproduction of a performance. Although keyframe-animators would not be familiar with negative weights, allowing weights to take on negative values can significantly extend the range of allowed expressions.

The projection steps of Equations 3 and 4 will modify to the hand-generated elements. We need to assess whether the new expressions are visually close to the original ones. If a hand-generated expression lies on the hyperplane (or, is contained in the motion capture data), then it will not be modified by the projection process. When a hand-generated expression is out of the hyperplane, however, the projection will introduce a minimal Euclidean modification to it. Although the scale for visual differences is not the same as that of Euclidean distance, small Euclidean distances usually correspond to small visual changes.

Another aspect that must be checked is the coverage of E^{QE}. In the analogical case shown in Figure 1, when there are two 3D vectors that do not coincide, it is highly likely that the projections of those vectors span the hyperplane. Similarly, if the number of hand-generated expressions is equal to or larger than m, it is highly probable that the projections of those expression will cover the hyperplane Ξ. Below we introduce several measures that can help avoid potential (but very rare) degenerate cases.

Preventive Treatments: We can guide the sculpting work of the animator so as to avoid overlap among the hand-generated expressions. For example, we can take $\hat{\mathbf{e}}_i^H$ to represent the full actuation of a single expression muscle with other muscles left relaxed, which intrinsically rules out the possibility of two hand-generated elements having almost identical shapes [Choe and Ko 2001]. For this purpose, animators can refer to reference book showing drawings of the expressions corresponding to isolated actuation of individual muscles [Faigin 1990]. The facial action coding system [Ekman and Friesen 1978] can also be of great assistance constructing non-overlapping hand-generated expression bases.

Post-Treatments: In spite of the above preventive treatments, the quasi-eigen basis may leave out a PCA-axis. Situations of this type can be identified by looking at the matrix $\mathbf{W}^{\text{PCA-to-QE}} = (w_{ij}^{\text{PCA-to-QE}})$. If $\sum_{i=1}^{n} |w_{ij}^{\text{PCA-to-QE}}|$ is less than a threshold ε, we conclude that $\mathbf{e}_j^{PCA}$ is missing in the quasi-eigen basis. In such a case, we can simply augment the basis with $\mathbf{e}_j^{PCA}$, or, can explicitly notify the animator regarding the missing eigenface $\mathbf{e}_j^{PCA}$ and let him/her make (minimal) modification to it so that its projection can be added to the keyframing basis as well as the quasi-eigen basis.

5 Experiments

To test the proposed method, we obtained a set of facial capture data, and modeled a hand-generated expression basis, based on the actuation of the expression muscles. We followed the procedure described in the previous section and produced the quasi-eigen basis from the hand-generated expression basis.

5.1 Capturing the Facial Model and Performance

We captured the performance of an actress using a Vicon optical system. Eight cameras tracked 66 markers attached to her face, and an additional 7 markers that were attached to her head to track the gross motion, at a rate of 120 frames per second. The total duration of the motion capture was $L = 35,000$ frames. We constructed the 3D facial model using a *Cyberware* 3D scanner. We established the correspondence between the 3D marker positions and the geometrical model of the face using the technique that was introduced by Pighin et al. [1998].

5.2 Preparing the Training Data Ξ

The motion capture data Ξ is a sequence of facial geometries. We convert the marker positions of each frame obtained into a facial mesh.[3] For this, we apply an interpolation technique that is based on the radial basis function. The technique gives the 3D displacements of the vertices that should be applied to the neutral face.

5.3 Preparing the Hand-Generated Expression Basis

We performed PCA on the data obtained in Section 5.2. Covering 99.5% of Ξ corresponded to taking the first $m = 18$ principal components. We asked animators to sculpt a hand-generated expression basis $\hat{E}^H = \{\hat{\mathbf{e}}_1^H, \ldots, \hat{\mathbf{e}}_n^H\}$ consisting of $n = 18$ elements. If the elements are clustered in E, then their projections will also be clustered in the hyperplane; this will result in poor coverage, requiring the hand-generation of additional basis elements. To reduce the hand-work of the animators, we guided the sculpting work by considering the size and location of the expression muscles, so that each basis element corresponds to the facial shape when a single expression muscle is fully actuated and all other muscles relaxed. In our experiment, we made 18 hand-generated expressions. Six elements are for the actuation of muscles in the upper region, 12 are for muscles in the lower region.

5.4 Obtaining the Quasi-Eigen Faces

Starting from the given hand-generated basis, we followed the steps described in Section 4 to obtain the quasi-eigen faces. A selection of the quasi-eigen expressions are shown in Figure 2 along with the corresponding hand-generated expressions. Comparison of the quasi-eigen and hand-generated expressions verifies that although projecting a hand-generated expression onto the hyperplane may involve non-negligible geometry modifications, the original visual impression is preserved.

Running the preprocessing steps, which included the PCA on 6,000 frames of training data, took 158 minutes on a PC with an Intel Pentium 4 3.2GHz CPU and Nvidia geforce 6800 GPU. After the training was complete, we could create quasi-eigen faces in realtime.

[3]The proposed technique can also work with the marker data directly. However, in this section we show how the proposed technique works when the facial motion is given as mesh data. The two approaches should yield the same results.

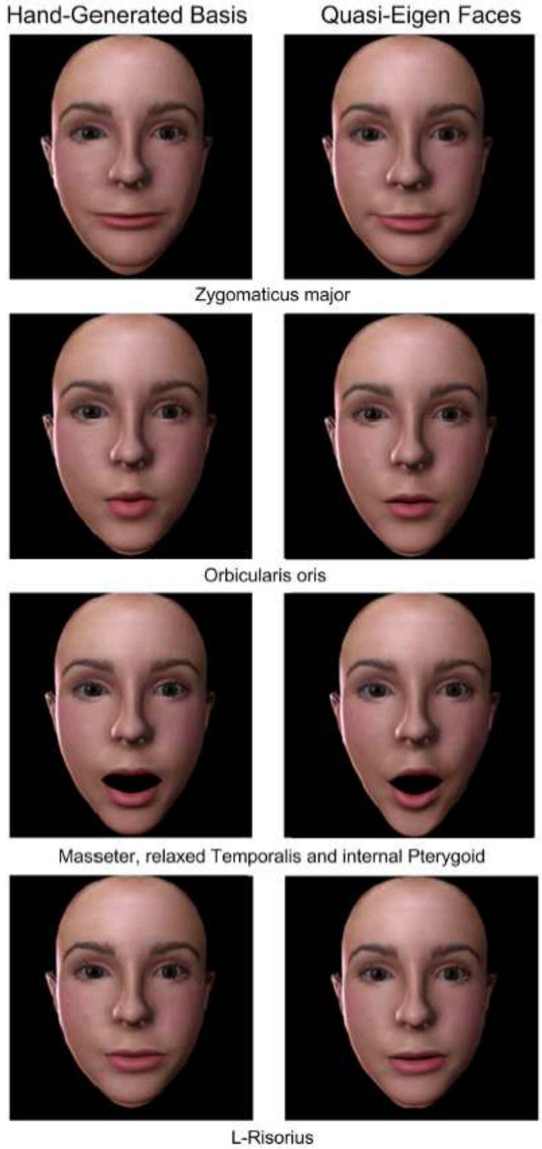

Figure 2: Side-by-side comparison of the hand-generated basis and the quasi-eigen basis in four selected elements

5.5 Analysis

Now, we approximate each frame of Ξ with a linear combination of the quasi-eigen faces. Let $\mathbf{v} = \mathbf{v}^0 + \sum_{i=1}^{n} w_i^{QE} \mathbf{e}_i^{QE}$ be the reconstruction of a frame, and let $\mathbf{v}^* = \mathbf{v}^0 + \mathbf{d}^*$ be the original expression of Ξ where $\mathbf{d}^*$ is the $3N$-dimensional displacement vector from the neutral expression. We find the n-dimensional weight vector $\mathbf{w}^{QE} = (w_1^{QE}, \ldots, w_n^{QE})$ by minimizing

$$|\mathbf{v}^* - \mathbf{v}|^2 = \sum_{j=1}^{N} \left| \mathbf{d}_j^* - \sum_{i=1}^{n} w_i^{QE} \mathbf{e}_{ij}^{QE} \right|^2, \quad (5)$$

where $\mathbf{d}_j^*$ and $\mathbf{e}_{ij}^{QE}$ are the displacements of the j-th vertex of $\mathbf{v}^*$ and $\mathbf{e}_i^{QE}$, respectively, from $\mathbf{v}^0$. We solve equation 5 using the quadratic programming, which required about 0.007 second per frame. To evaluate the accuracy of the reproduction, we used the following error metric:

$$\alpha[\%] = 100 \times \frac{\sqrt{\sum_{j=1}^{N} \left| \mathbf{v}_j^* - \mathbf{v}_j \right|^2}}{\sqrt{\sum_{j=1}^{N} \left| \mathbf{v}_j^* \right|^2}} \tag{6}$$

For comparison, the above analysis was also performed using the bases E^H and E^{PCA}. The α values obtained using the three bases were $\alpha^{QE} = 0.72\%$, $\alpha^H = 5.2\%$, and $\alpha^{PCA} = 0.62\%$. The results thus indicate that, in terms of coverage, E^{QE} is slightly inferior to E^{PCA} and far better than E^H.

Qualitative comparison of the reconstructions can be made in the accompanying video. Figure 3 shows still images extracted from the video. The three images in the left column are taken from the original motion capture. The middle and right columns show the reconstruction of those frames with E^H and E^{QE}, respectively. Figure 4 visualizes the errors introduced during the reconstruction, for the frame shown on the top of Figure 3. The red dots represent the captured marker positions whereas the blue dots represent their positions in the reconstructed result. We can clearly see that reconstruction with the quasi-eigen faces (shown on the right) makes a less amount of error than the reconstruction with the hand-generated basis (shown on the left).

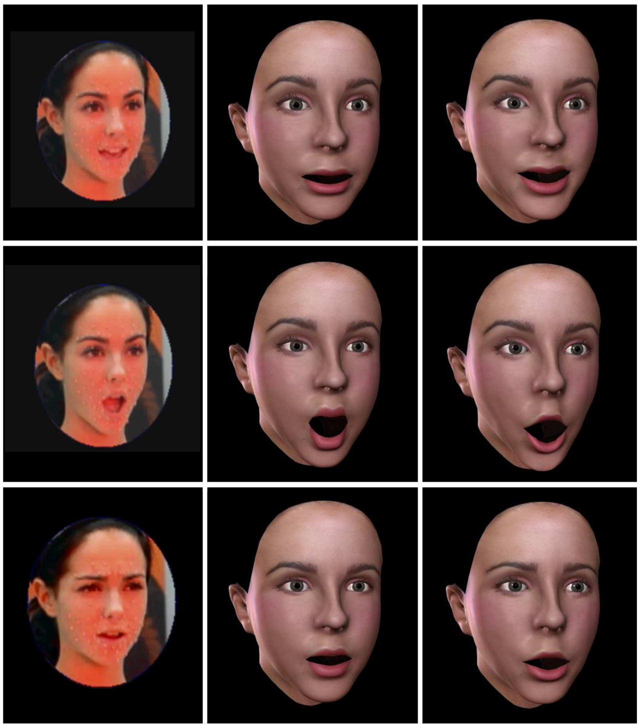

Figure 3: Reconstruction of the original performance (left: motion capture, middle: with E^H, right: with E^{QE})

6 Conclusion

In this paper, we have presented a new method for generating expression bases for blendshape-based facial animation systems. Animation studios commonly generate such bases by manually modeling a set of key expressions. However, hand-generated expressions may contain components that are not part of human expressions, and reconstruction/animation by taking linear combinations of these expressions may produce reconstruction errors or unrealistic results. On the other hand, statistically-based techniques can produce high-fidelity expression bases, but the basis elements are not intuitively recognizable. Here we have proposed a method for generating so-called quasi-eigen faces, which have intuitively recognizable shapes but significantly reduced reconstruction errors compared to hand-generated bases.

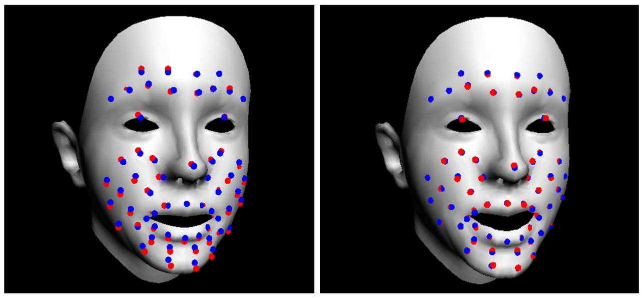

Figure 4: Comparison of the reconstruction errors (left: for E^H, right: for E^{QE})

In this paper we have focused on the reproduction of captured performances. This approach was taken based on our experience in facial animation that, in most cases, technically critical problems reside in the analysis part rather than in the synthesis part. If the analysis is performed accurately, then expression synthesis, whether it be reproduction or animation of other characters, will be accurate. The experiments performed in the present work showed that the proposed technique produces basis elements that are visually recognizable as typical human expressions and can significantly reduce the reconstruction error. Even though we did not demonstrate in the paper, the proposed technique can be effectively used for synthesizing expressions of other characters than the captured subject.

The proposed technique is an animator-in-the-loop method whose results are sensitive to the hand-generated expressions provided by the animator. If the animator provides inadequate expressions, the projection will not improve the result. We have found that a muscle-based approach to the modeling of the hand-generated expressions, as used in Section 5, effectively extends the coverage of the basis. Application of the proposed projection to hand-generated elements of this type reduces the reconstruction error. The muscle-based approach is not, however, the only way to obtain non-overlapping hand-generated expressions. A better guidance may be developed in the future, which can help the animator sculpt intuitively meaningful but non-overlapping faces.

7 Acknowledgment

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (National Research Laboratory M10600000232-06J0000-23210), Ministry of Information and Communication, the Brain Korea 21 Project, and Automation and System Research Institute at Seoul National University.

References

BLANZ, V. AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of SIGGRAPH 1999*, ACM Press, 187–194.

BRAND, M. 1999. Voice puppetry. In *Proceedings of SIGGRAPH 1999*, ACM Press, 21–28.

BREGLER, C. 1997. Video Rewrite:driving visual speech with audio. In *Proceedings of SIGGRAPH 1997*, ACM Press, 353–360.

CHAI, J., XIAO, J. AND HODGINS,J. 2003. Vision-based control of 3D facial animation. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 193–206.

CHANG, Y.-J., AND EZZAT, T. 2005. Transferable Videorealistic Speech Animation. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 143–151.

CHAO, Y., FALOUTSOS, P. AND PIGHIN, F. 2003. Unsupervised learning for speech motion editing. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 225–231.

CHAO, Y., FALOUTSOS, P., KOHLER, E., AND PIGHIN, F. 2004. Real-time speech motion synthesis from recorded motions. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 347–355.

CHOE, B., LEE, H AND KO, H.-S. 2001. Performance-driven musclebased facial animation. In *Journal of Visualization and Computer Animation 1999*, 67–79.

CHOE, B. AND KO, H.-S. 2001. Analysis and synthesis of facial expressions with hand-generated muscle actuation basis. In *Proceedings of Computer Animation 2001*, 12–19.

CHUANG, E. 2002. *Performance driven facial animation using blendshape*. Stanford University Computer Science Technical Report, CS-TR-2002-02, 2002.

DENG, Z., LEWIS, J. P., AND NEUMANN, U. 2005. Synthesizing Speech Animation by Learning Compact Speech Co-Articulation Models. In *Proceedings of Computer graphics international.*, IEEE Computer Society Press, 19–25.

EKMAN, P. AND FRIESEN, W. V. 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Inc.

ESSA, I. AND PENTLAND, A. 1997. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transaction on Pattern Analysis and Machine Intelligence Vol. 19, No. 7, 757–763.

EZZAT, T., GEIGER, G., AND POGGIO, T. 2002. Trainable videorealistic speech animation. In *Proceedings of SIGGRAPH 2002*, ACM Press, 388–398.

FAIGIN G. 1990. The Artists Complete Guide to Facial Expression. Watson-Guptill Publications

GUENTER, B., GRIMM, C., WOOD, D. MALVAR, H. AND PIGHIN, F. 1998. Making faces. In *Proceedings of SIGGRAPH 1998*, ACM Press, 55–66.

JOSHI, P., TIEN, W., C., DESBRUN, M., AND PIGHIN, F. 2003. Learning controls for blend shape based realistic facial animation. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*

KAHLER, K., HABER, J., AND SEIDEL, H.-P. 2001. Geometry-based muscle modeling for facial animation. In *Proceedings of Graphics Interface 2001*, 37–46.

KOUADIO, C., POULIN, P., AND LACHAPELLE, P. 1998. Real-time facial animation based upon a bank of 3D facial expression. In *Proceedings of Computer Animation 1998*, IEEE Computer Society Press.

NA, K. AND JUNG, M. 2004. Hierarchical retargetting of fine facial motions. In *Proceedings of Eurographics 2004*, Vol. 23.

NOH, J. AND NEUMANN, U. 2001. Expression cloning In *Proceedings of SIGGRAPH 2001*, ACM Press, 277–288.

PARKE, F. I. 1972. Synthesizing realistic facial expressions from photographs. In *Proceedings of ACM Conference 1972*, ACM Press, 451–457.

PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 1998. Synthesizing realistic facial expressions from photographs. In *Proceedings of SIGGRAPH 1998*, ACM Press, 75–84.

PIGHIN, F., SZELISKI, R., AND SALESIN, D. H. 1999. Resynthesizing facial animation through 3D model-based tracking. In *Proceedings of International Conference on Computer Vision 1999*, 143–150.

PYUN, H., KIM, Y., CHAE, W., KANG, H. W., AND SHIN, S. Y. 2003. An example-based approach for facial expression cloning. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 167–176.

SIFAKIS, E., NEVEROV, I., AND FEDKIW, R. 2005. Automatic determination of facial muscle activations from sparse motion capture marker data. In *Proceedings of SIGGRAPH 2005*, ACM Press, 417–425.

TERZOPOULOS, D., AND WATERS, K. 1990. Physically-based facial modeling, analysis and animation. *The Journal of Visualization and Computer Animation*, 73–80.

VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIC, J. 2005. Face transfer with multilinear models. In *Proceedings of SIGGRAPH 2005*, ACM Press, 426–433.

WANG, Y., HUANG, X., LEE, C. S., ZHANG, S., LI, Z., SAMARAS, D., METAXAS, D., ELGAMMAL, A., AND HUANG, P. 2004. High resolution acquisition, learning and transfer of dynamic 3-D facial expressions. In *Proceedings of Eurographics 2004*, 677–686

WATERS, K. 1987. A muscle model for animating three-dimensional facial expressions. In *Proceedings of SIGGRAPH 1987*, ACM Press, 17–24.

WILLIAMS, L. 1990. Performance-driven facial animation. In *Proceedings of SIGGRAPH 1990*, ACM Press, 235–242.

ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. 2004. Spacetime faces: High resolution capture for modeling and animation. In *Proceedings of SIGGRAPH 2004*, ACM Press, 548–558.