Proceedings of Computer Animation 2001, pages 12–19, November 2001

Analysis and Synthesis of Facial Expressions with Hand-Generated Muscle Actuation Basis

Byoungwon Choe* Hyeong-Seok Ko

School of Electrical Engineering and Computer Science
Seoul National University, Seoul, KOREA

Abstract

We present a performance-driven facial animation system for analyzing captured expressions to find muscle actuation and synthesizing expressions with the actuation values. Significantly different approach of our work is that we let artists sculpt the initial draft of the actuation basis—the basic facial shapes corresponding to the isolated actuation of individual muscles, instead of calculating skin surface deformation entirely relying on the mathematical models such as finite element methods. We synthesize expressions by linear combinations of the basis elements, and analyze expressions by finding the weights for the combinations. Even though the hand-generated actuation basis represents the essence of the subject's characteristic expressions, it is not accurate enough to be used in the subsequent computational procedures. We also describe an iterative algorithm to increase the accuracy of the actuation basis. The experimental results suggest that our artist-in-the-loop method produces more predictable and controllable outcome than pure mathematical models, thus can be a quite useful tool in animation productions.

1. Introduction

Since Williams' pioneering work on performance-driven facial animation [24], applying facial expressions from human faces to computer-generated characters has been widely studied [9, 3, 17, 12, 20]. To control facial movement, facial expressions were analyzed into the position of feature points [9, 3, 17] or the weights for blending pre-modeled expressions [12, 20]. Another fascinating approach, which we took in this work, is finding muscle actuation parameters from facial expressions [22, 6, 1]. Expressions can be easily modified by editing muscle actuation curves [1], and the actuation values can be converted to other control parameters such as the values of Actuation Units in Facial Action Coding System [4] without much effort.

Terzopoulos *et al.* [22] and Essa *et al.* [6] analyzed the expressions recorded in video footage into muscle actuation values for *facial expression recognition*. They also synthesized facial expressions with the actuation values. The synthetic expressions, however, showed only conspicuous ones such as 'opening mouth' or 'raising eyebrows', which were not yet to be used in high quality animation production. Recently, Choe *et al.* [1] proposed an algorithm to find muscle actuation values from the trajectory of feature points generated by an optical capture system. They could reproduce delicate facial movements by extracting complicated set of muscle actuation values with a linear finite element model, and showed the possibility of practical use in character animation. Still, the heuristic muscle model could misinterpret the original expressions, and simplified finite element model occasionally produced unnatural artifacts in skin deformation.

In this work, instead of relying entirely on the mathematical models to compute the 3D facial shape, we include the artists' modeling capability as an integral part of the method. According to our previous tests, the result of pure mathematical modeling was usually distant from what was *expected*.[1] Such expectation cannot be quantitatively stated; we thought that an artist may be able to form the expected (or desired) facial shape. Thus we made the artists sculpt manually a set of expressions called the *muscle actuation basis*, and let the computer program synthesize expressions based on the basis elements. Each element of the actuation basis corresponds to the facial shape when a single expression muscle is fully actuated and the rest are left relaxed.

We can synthesize a facial expression by the linear combination of the basis elements on the same principle as the linear muscle model [1]. Then our algorithm is basically re-

*133-316, Seoul National University, Shillim-dong, Gwanak-gu, Seoul, 151-742, KOREA. choe@graphics.snu.ac.kr

[1] Some of the reasons might be that we could not calibrate the muscle size and layout of the computer model with those of the subject being captured, and we made too many simplifying assumptions to use mathematical models.

duced to the methods that synthesize expressions by blending pre-modeled expressions, which was experimented by Kouadio *et al.* [12] and Pighin *et al.* [19, 20]. Our method is different from theirs in the pre-modeled expression set: we use an artificial but functional set of expressions instead of using real human expression samples such as 'happiness' or 'sadness'. Using the actuation basis rather than real human expression samples has an important consequence. The elements in the actuation basis are *orthogonal* to each other, and form a meaningful basis for the facial expression space—the actuation basis can produce (or in mathematical terms, *span*) the complete set of human expressions. When real human expressions are used, on the other hand, the linear combinations of them cannot generally guarantee to produce the complete set of expressions.[2]

We can summarize our facial animation process into two major steps: **modeling** to set up the neutral face and the actuation basis of a subject and **analysis** to find muscle contractions from the subject's performance using the actuation basis. We can synthesize expressions by applying the analyzed muscle contractions to any computer model with an equivalent muscle structure.

In order to model the actuation basis, we first obtain the neutral face of the subject using a 3D scanning device. Starting from the neutral face, we let an artist sculpt the basis elements considering the human facial anatomy [2]. The work of Faigin [7], which illustrates the facial shape corresponding to the actuation of each individual muscle, serves a good guide for the job. It could be expected that the first hand-generated draft would not give a satisfactory result. Moreover, considering that the accuracy of the actuation basis greatly affects the result of the **analysis**, we need to develop a procedure for improving the basis. The improvement procedure (described in Section 4.2), in turn, refers to the result of the **analysis** on some trial data; the procedure takes the form of fixed point iteration between **modeling** and **analysis**.

Once the actuation basis of a subject is ready, we can start analyzing the expressions captured from the subject. We approximate each frame of the facial performance by a linear combination of the basis elements. Finding the best approximation can be formulated as a *constrained quadratic programming*, and the coefficients in the resulting solution are interpreted as the muscle contraction values.

The rest of this paper is organized as follows. Section 2 reviews related work in facial animation. Section 3 and Section 4 present the **modeling** and **analysis** procedures respectively. Section 5 shows the experimental results of our method, and Section 6 concludes the paper.

[2]There have been efforts to resolve the correlation among human expression samples and map the expressions into an orthogonal domain [14]. A popular domain studied first in psychology was a two-dimensional space represented by *pleasure* and *arousal* axes [8]. However, the quantitative use of the parameters (e.g., for expression synthesis) does not seem suitable since the dimension is quite limited and assigning the coordinate values is done in a subjective manner.

2. Background

This section reviews the state-of-the-art techniques on performance-driven facial animation and muscle-based facial modeling. More topics on facial modeling and animation can be found in [18].

Williams [24] introduced a performance-driven facial animation system which synthesized expressions by changing texture coordinates calculated from the position of feature points on the face. Guenter *et al.* [9] captured both the 3D geometry and shading information of a human face, and reproduced photorealistic expressions. Eisert and Girod [3] modeled a face with a triangular B-spline surface, and analyzed facial expressions by estimating the facial animation parameters of MPEG-4 standard. Pighin *et al.* [19] reconstructed the geometry and texture of an individual face from five photo images of the subject. With this method, they modeled basic expressions such as 'joy' or 'surprise', and synthesized novel expressions by blending them. The result was photo-realistic, showing detailed wrinkles and creases. Later, they proposed an algorithm to find the blending weights from the video recording of a performance [20]. Kouadio *et al.* [12] animated a synthetic character by the linear combination of previously modeled 3D facial expressions by extracting the interpolation weights from the feature points traced by an optical capture device.

Waters [23] introduced an anatomically based muscle model which was kinematically formulated. Terzopoulos *et al.* [21, 13] represented the mesh of the skin surface by a mass-spring model, and calculated skin deformation due to muscle actuation. Koch *et al.* predicted the geometry of skin surface due to the skull shape change using a finite-element model [11], and synthesized expressions by embedding expression muscles [10].

Terzopoulos and Waters [22] developed a method to extract muscle contractions from the expressions recorded in video footage based on a dynamic muscle model. Essa *et al.* [5, 6] developed a system to estimate muscle actuation corresponding to a given expression using feedback control theory. Choe *et al.* [1] calculated muscle actuation values based on the finite element skin model and linear muscle model.

3. Modeling Muscle Actuation Basis

Muscle actuation basis is a set of expressions $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_m\}$, each of which represents the 3D facial shape when a single expression muscle is fully actuated and the others are relaxed. Figure 1 shows an example of the actuation basis.

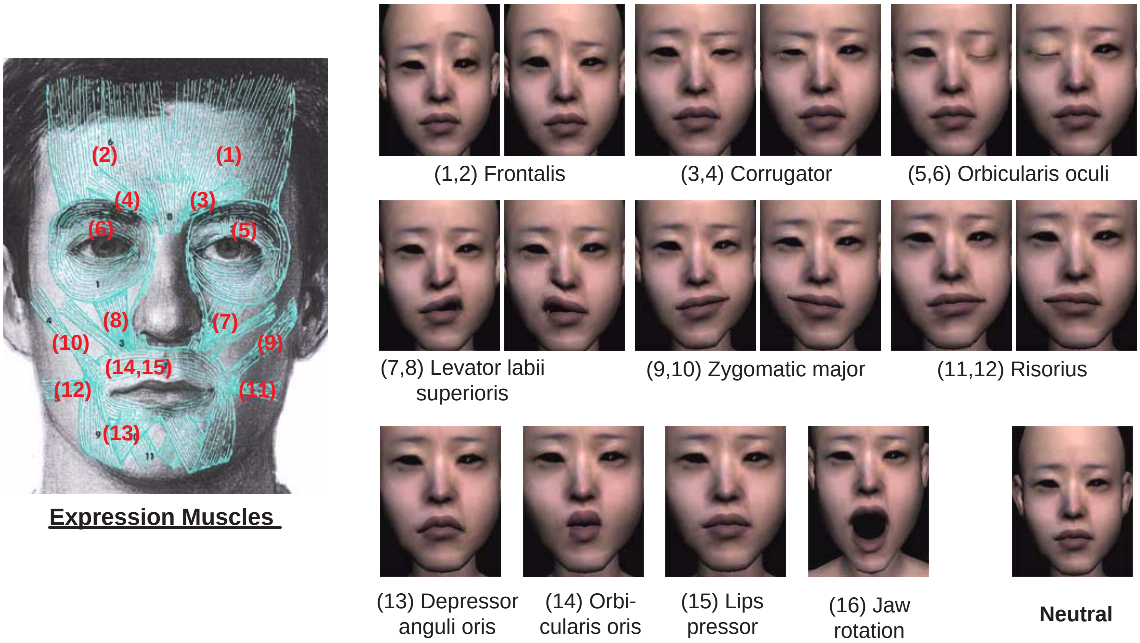

Figure 1. Expression muscles and the corresponding basis elements in the actuation basis.

Once we have the actuation basis, we can synthesize facial expressions by linear combinations of the basis elements if we assume the linear muscle model [1]. Let $\mathbf{E}_0$ denote the neutral expression—the position of about 1,500 vertices that constitute the facial surface. Let $\mathbf{e}_i = \mathbf{E}_i - \mathbf{E}_0$ $(i = 1, 2, \ldots, m)$ be the difference between the basis element and the neutral expression, where m is the number of the basis elements. When the muscle contractions $x_1, x_2, \ldots, x_m$ are given, we synthesize an expression $\mathbf{E}$ by

$$\mathbf{E} = \mathbf{E}_0 + \sum_{i=1}^{m} x_i \mathbf{e}_i. \tag{1}$$

We normally expect the muscle contractions have the value in $[0, 1]$ since each basis element embodies the full actuation of an expression muscle.

In this work, we used an actuation basis of 16 elements as shown in Figure 1: *six* for the muscles in the upper region around the eyebrows (Figure 1 (1)∼(6)), *ten* for the muscles in the lower region around the mouth (Figure 1 (7)∼(16)). We get the reduced set of 16 basis elements to represent the operation of not less than 26 expression muscles in the human face. The operation of several muscles can be merged into a single basis element if they are dependent on each other, and the operation of a single muscle should be represented by multiple basis elements if the actuation of the muscle can produce multiple distinct shapes:

- **Merging:** We merge the operation of the muscles into a single basis element if they usually actuate simultaneously. For example, we merge the three muscles *Levator labii superioris alaeque nasi*, *Levator labii superioris*, and *Zygomatic minor* which are known as the sneering muscles (see Figure 1 (7, 8)) into the single basis element *Levator labii superioris*. The basis elements *Corrugator* (Figure 1 (3, 4)), *Risorius* (Figure 1 (11, 12)), and *Depressor anguli oris* (Figure 1 (13)) are also the results of merging the operation of two or three muscles.

- **Mouth:** The operation of *Orbicularis oris* around the mouth is very complicated, and the full actuation of the muscle can generate many different shapes. In this work, we created two basis elements to represent the operation of the muscle: normal *Orbicularis oris* which corresponds to the mouth shape when pronouncing /u/ sound (Figure 1 (14)), and the *Lips pressor* which corresponds to the protruded (upset) mouth (Figure 1 (15)).[3] Gentle closing of the mouth is covered by the neutral expression $\mathbf{E}_0$.

- **Eyes:** *Orbicularis oculi*, the sphincter muscle at the eyes, consists of the palpebral and orbital parts. In this

[3] *Orbicularis oris* was an obvious choice for the basis, but the inclusion of *Lips pressor* was based upon our experiences: without the *Lips pressor*, we observed the elements *Risorius* and *Orbicularis oris* had to combine frequently to produce the shape of *Lips pressor*, which was quite unnatural in the operation of human expression muscles.

work, we implemented only the operation of the palpebral part (gentle closing of the eyes) as a basis element (Figure 1 (5, 6)). Therefore emphatic closing of the eyes cannot be generated.

We let artists model the basis elements considering the size and location of expression muscles [2]. Faigin [7] illustrated the facial expressions resulting from the actuation of a single or multiple expression muscles, which served an excellent guide to the modeling job. The actuation basis only used for expression *synthesis* does not need to come from a human subject. However, the actuation basis for expression *analysis* should accurately reflect the operation of expression muscles of the subject because it affects drastically the result of expression analysis (Section 4.1). Therefore artists were asked to watch carefully the video recording of the subject (or the *training data* in Section 4.2) where the subject was asked to make all kinds of expressions including the extreme actuation of each muscle.

It would be impractical to assume that the hand-generated actuation basis is accurate enough. Fortunately, there is a way to evaluate the given basis: we simply run the expression analysis procedure on the training data, then we can infer that the basis is not accurate when the resulting muscle contractions go far beyond the expected range $[0, 1]$. In such a case, we ask the artists to re-model the basis elements. We repeat the step until a reasonable basis is obtained. However, it would be still impractical to assume that the resulting basis is accurate enough to start our computational steps of expression analysis. We present an algorithm that improves further the actuation basis (at this time without the help of artists) by taking iterations between the expression analysis and basis modification procedures. The algorithm cannot be fully described until the expression analysis procedure is understood, so the description is deferred to the end of the next section.

4. Analysis of Facial Expressions

This section presents the procedures to extract muscle contractions from facial performances, and shows how the procedure can be used for improving the hand-generated actuation basis.

4.1. Extracting Muscle Contractions

We analyze the facial expressions by finding muscle contractions to reproduce optimally the marker trajectories generated by optical capture systems. We improved the algorithm proposed by Choe *et al.* [1].

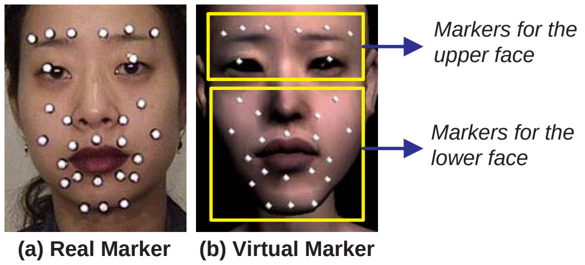

Figure 2. Real markers and corresponding virtual markers on the synthetic model.

4.1.1 Coordinate Alignment

While the 3D geometry of the synthetic face is resolved in its own local coordinate system $\{M\}$ (*model coordinate system*), the marker points in the performance data are resolved in another coordinate system $\{P\}$ (*performance coordinate system*). Before calculating the muscle contractions, we first have to transform the marker points from the performance coordinate system to the model coordinate system. We assume the transform from $\{P\}$ to $\{M\}$ is an affine (similarity) transform with scale s, rotation $\mathbf{R}$, and translation $\mathbf{t}$. We calculate the transform only once at the first frame of the performance data, and apply the same transform to all the other frames. In the following, we use the notation ${}^{Z}\mathbf{p}$ to denote the 3D coordinate of a point $\mathbf{p}$ resolved in the coordinate system $\{Z\}$

Let the position of the j-th marker be ${}^{P}\mathbf{p}_j$. We define the *virtual marker* ${}^{M}\mathbf{p}_j$ to be the corresponding point in $\{M\}$. Figure 2 shows the real and virtual markers. We want to find s, $\mathbf{R}$, and $\mathbf{t}$ that satisfy the following equations,

$$ {}^{M}\mathbf{p}_j = s\mathbf{R}\,{}^{P}\mathbf{p}_j + \mathbf{t}, \quad (j = 1, 2, \ldots, n) $$

where n is the number of markers. To initiate the computation, we first manually mark the virtual marker positions looking at both the first frame of the performance and the 3D model of the face. Then we solve the linear least square problems to find s, $\mathbf{t}$, and $\mathbf{R}$ sequentially, and repeat the procedure until the least square error is saturated.

The accuracy of s, $\mathbf{R}$, and $\mathbf{t}$ solved from the above procedure is at best limited to the accuracy of hand-marked position of the virtual markers. We note that, once we have s, $\mathbf{R}$, and $\mathbf{t}$, then the real markers can now be transformed to the model coordinate system. But the resulting points may not lie exactly on the surface of the 3D face model. By slightly adjusting the points along the normal directions, we can make the points lie on the surface, and get the next estimation of the (virtual) markers. We can further improve the accuracy of s, $\mathbf{R}$, and $\mathbf{t}$ by repeating the least square procedure with the new position of the markers.

The above assumes that the facial shape at the first frame of the performance is the same with the pre-modeled neutral face. Therefore we asked the actors to make a consistent neutral expression at the beginning of each performance capture.

4.1.2 Calculating Muscle Contractions

The final position of the virtual marker j in the above procedure will be a point within one of the triangles that constitute the facial geometry. Thus the marker point $\mathbf{p}_j$ can be encoded by the triangle index and the relative position within the triangle (barycentric coordinates), which do not depend on the subsequent deformation of the face. But the marker point will have different 3D position depending on the current shape of the face.

Let $\mathbf{d}_{ij}$ be the displacement of $\mathbf{p}_j$ at basis element $\mathbf{E}_i$ from $\mathbf{p}_j$ at neutral expression $\mathbf{E}_0$. From the synthesis equation (1), if muscle contractions x_i are given, the total displacement $\mathbf{d}_j$ of $\mathbf{p}_j$ is given by

$$\mathbf{d}_j = \sum_{i=1}^{m} x_i \mathbf{d}_{ij} \ .$$

We find the muscle contractions so that $\mathbf{d}_j$ is closest to the observed displacement $\hat{\mathbf{d}}_j$ from the performance by minimizing

$$\sum_{j=1}^{n} |\hat{\mathbf{d}}_j - \mathbf{d}_j|^2 = \sum_{j=1}^{n} |\hat{\mathbf{d}}_j - \sum_{i=1}^{m} x_i \mathbf{d}_{ij}|^2.$$

Because the muscle contractions should be lie in $[0, 1]$, we can find the contractions by solving the following optimization problem:

$$\begin{array}{rl} \textbf{minimize} & \displaystyle\sum_{j=1}^{n} |\hat{\mathbf{d}}_j - \sum_{i=1}^{m} x_i \mathbf{d}_{ij}|^2 \\ \textbf{subject to} & 0 \le x_i \le 1 \ (i = 1, 2, \ldots, m) \end{array} \ . \qquad (2)$$

The muscle contraction vector $\mathbf{x} = [x_1, x_2, \ldots, x_m]^T$ can be obtained by solving the constrained quadratic programming

$$\begin{array}{rl} \textbf{minimize} & \frac{1}{2}\mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{x}^T \mathbf{c} \\ \textbf{subject to} & 0 \le x_i \le 1 \ (i = 1, 2, \ldots, m) \end{array} \ , \qquad (3)$$

where

$$\mathbf{Q} = 2 \begin{pmatrix} \sum_{j=1}^{n} |\mathbf{d}_{1j}|^2 & \sum_{j=1}^{n} \mathbf{d}_{1j} \cdot \mathbf{d}_{2j} & \cdots & \sum_{j=1}^{n} \mathbf{d}_{1j} \cdot \mathbf{d}_{mj} \\ \sum_{j=1}^{n} \mathbf{d}_{2j} \cdot \mathbf{d}_{1j} & \sum_{j=1}^{n} |\mathbf{d}_{2j}|^2 & \cdots & \sum_{j=1}^{n} \mathbf{d}_{2j} \cdot \mathbf{d}_{mj} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^{n} \mathbf{d}_{mj} \cdot \mathbf{d}_{1j} & \sum_{j=1}^{n} \mathbf{d}_{mj} \cdot \mathbf{d}_{2j} & \cdots & \sum_{j=1}^{n} |\mathbf{d}_{mj}|^2 \end{pmatrix},$$

$$\mathbf{c} = 2 \begin{pmatrix} \sum_{j=1}^{n} \hat{\mathbf{d}}_j \cdot \mathbf{d}_{1j} \\ \sum_{j=1}^{n} \hat{\mathbf{d}}_j \cdot \mathbf{d}_{2j} \\ \vdots \\ \sum_{j=1}^{n} \hat{\mathbf{d}}_j \cdot \mathbf{d}_{mj} \end{pmatrix} .$$

We solve this problem using the active set method, and apply Lagrange method for the sub-problems derived from the active sets [15]. To make the optimization procedure more robust, we divided the face into the upper and lower regions. The contractions of *Frontalis*, *Corrugator*, and *Orbicularis oculi* were calculated using only the markers in the upper region, and contractions of the other muscles were calculated using only the markers in the lower region (Figure 2). A muscle contraction value larger than one can be thought of as an exaggerated expression. So, we set only $x_i \ge 0$ as the constraints if we need to allow the exaggeration.

4.2. Improving the Actuation Basis

The actuation basis only used for expression synthesis can be entirely depend on the craftsmanship of the artist. However, the actuation basis for the subject being captured needs to have a certain level of accuracy to get reliable expression analysis results. It is not likely that hand-generated basis has such an accuracy. Therefore we develop an iterative algorithm that increases the accuracy of an actuation basis.

The algorithm takes the form of a fixed point iteration between **modeling** and **analysis**—the result of **modeling** is used for the **analysis**, and the result of **analysis** is in turn used for improving the actuation basis. For the iteration, we collect a performance data called *training data* in which the actor is asked to make all sorts of expressions. We let the actor fully contract each individual muscles. Even though ordinary people cannot make isolated muscle actuation, the facial expressions generated in the process of trying to use only a single muscle contain important information about the operation of the muscles, and helps to find more optimal basis elements. The training data also includes a significant amount of ordinary expressions that involve compound actuation of multiple muscles.

We first calculate muscle contractions at all frames of the training data by solving (3). Then the following equations should be satisfied in ideal cases for the marker point $\mathbf{p}_j$:

$$\begin{array}{c} x_1^{(1)}\mathbf{d}_{1j} + x_2^{(1)}\mathbf{d}_{2j} + x_3^{(1)}\mathbf{d}_{3j} + \cdots + x_m^{(1)}\mathbf{d}_{mj} = \hat{\mathbf{d}}_j^{(1)} \\ x_1^{(2)}\mathbf{d}_{1j} + x_2^{(2)}\mathbf{d}_{2j} + x_3^{(2)}\mathbf{d}_{3j} + \cdots + x_m^{(2)}\mathbf{d}_{mj} = \hat{\mathbf{d}}_j^{(2)} \\ \vdots \\ x_1^{(N)}\mathbf{d}_{1j} + x_2^{(N)}\mathbf{d}_{2j} + x_3^{(N)}\mathbf{d}_{3j} + \cdots + x_m^{(N)}\mathbf{d}_{mj} = \hat{\mathbf{d}}_j^{(N)}, \end{array}$$

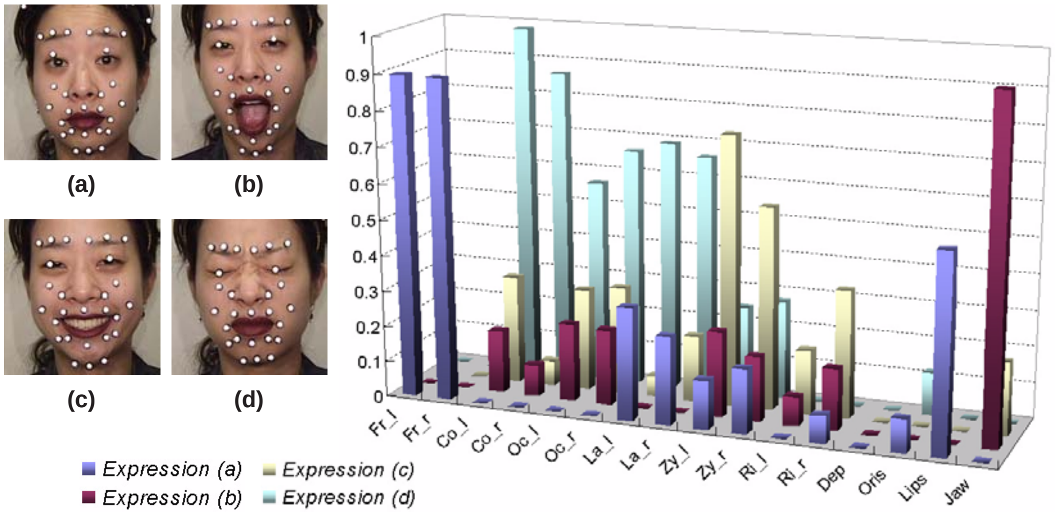

Figure 3. The four snapshots and plotting of the corresponding muscle contraction vectors

where $x_1^{(t)}, \ldots, x_m^{(t)}$ are analyzed muscle contractions at frame t, $\hat{\mathbf{d}}_j^{(t)}$ is the observed displacements at frame t, and N is the total number of frames in the training data. In reality, however, the equalities do not hold. But, if we solve the equations for $(\mathbf{d}_{1j}, \ldots, \mathbf{d}_{mj})$, the least square solution can provide us the improved position of the marker point $\mathbf{p}_j$ in each of the basis elements $\mathbf{E}_1, \ldots, \mathbf{E}_m$. If we perform the above steps for all the marker points $\mathbf{p}_j$ $(j = 1, \ldots, n)$, we can get a new (improved) actuation basis.

Thus we get an improved actuation basis from the initial draft: (1) calculating muscle contractions from the initial draft, (2) finding new $\mathbf{d}_{ij}$ $(i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$ with the muscle contractions. We repeat the cycle until the *total analysis error* $\sum_t \sum_j |\hat{\mathbf{d}}_j^{(t)} - \mathbf{d}_j^{(t)}|^2$ is saturated, and finally get the optimized displacements $\bar{\mathbf{d}}_{ij}$ $(i = 1, 2, \ldots, m, j = 1, 2, \ldots, n)$. Finally, using the scattered data interpolation method with a radial basis function [16], we can form an optimized actuation basis from $\bar{\mathbf{d}}_{ij}$ and the initial draft of actuation basis.

5. Experiments

We implemented our algorithms on a PC platform. For the experiment, we also developed an optical capture system with five video cameras [1], which generated 3D trajectory of the markers and gross motion of the head at 30 frames per second. The experimental results (demo clips) described in this Section is available at `http://graphics.snu.ac.kr/research/basis/`.

First, we 3D-scanned the face of an actor, let an artist model the actuation basis of it, ran the improvement algorithm described in Section 4.2, and got the final actuation basis. Then we captured performances and analyzed them into muscle actuation values. Figure 3 shows four expression snapshots during a performance: (a) raising eyebrows, (b) jaw rotation, (c) smiling, and (d) frowning. The graph in the figure plots the muscle contractions at the snapshots which were analyzed by the algorithm described in Section 4.1. The result agrees well with our anticipation:

- **Expression (a):** The contractions (about 0.9) of the left and right *Frontalis* are dominant in raising eyebrows.
- **Expression (b):** We can see the jaw rotation is conspicuous.
- **Expression (c):** The two dominant contractions in the middle correspond to the left and right *Zygomatic majors*, which matches well with the muscle actuation in real smiling.
- **Expression (d):** We can see the six muscles are dominant in the last row: the pairs of *Corrugator*, *Orbicularis oculi*, and *Levator labii superioris*. The contraction of *Corrugator* and *Levator labii superioris* matches well with the muscle actuation in real frowning. *Orbicularis oculi* resulted from the close of the eyes are not directly related to this expression.

Figure 4 shows the result of applying the contractions of the expressions to the computer model of the actor and other two cartoon-like characters.

Figure 5 plots the contractions of left *Frontalis* and *Jaw rotation* during the performance of "Tulip Season" contained in the demo clip. The x-axis represents the frame

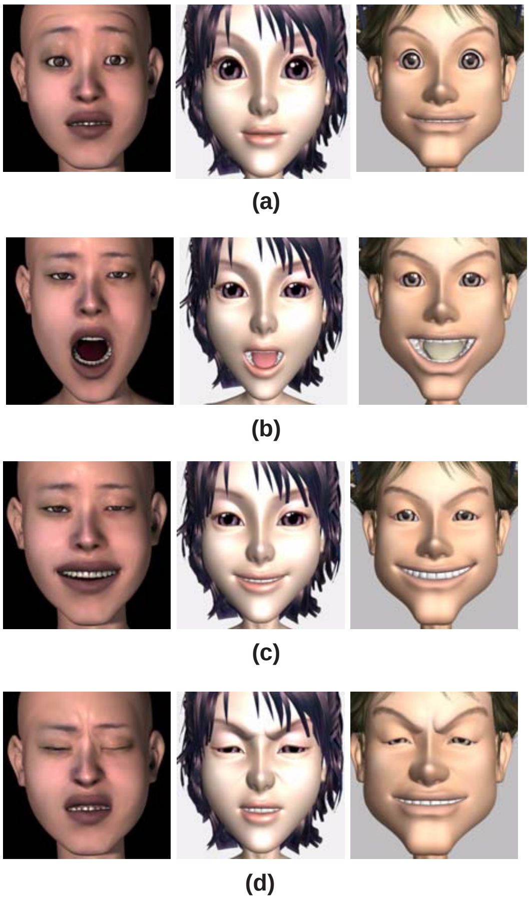

Figure 4. Results of applying the muscle contractions in Figure 3 to different 3D models.

number and the y-axis represents the muscle contractions in $[0, 1]$. The figure also shows *per-frame marker error* $(\sum_{j=1}^{n} |\hat{\mathbf{d}}_j - \mathbf{d}_j|)/n$, which is measured in centimeters. The error was computed separately for the upper and lower regions of the face. The error is bigger in the lower region due to the nonlinear and irregular movement around the mouth, which is mainly caused by *Orbicularis oris* muscle.

6. Conclusion

In this paper, we presented a new muscle-based facial animation technique that uses the actuation basis, a set of 3D facial shapes corresponding to the full actuation of individual muscles. Instead of completely relying on a mathematical method, we let artists manually sculpt (the initial draft of) the basis elements so that we could get more predictable deformation of the face. To increase the accuracy of the actuation basis, we developed an iterative algorithm that refined the actuation basis. Once an actuation basis was ready, a performance could be analyzed quite accurately, and the result could be applied to any 3D models with equivalent muscle structures.

An interesting contribution of this paper is that it proposed a technique that includes the artists' modeling capability as an integral part of the algorithm. The manual shaping of the basis elements complemented the pure mathematical approach which produced unexpected results occasionally. The proposed method is robust, and we believe that this artist-in-the-loop method can be quite useful in animation productions until the mathematical models can accurately simulate the operation of the muscles and concomitant movements in the facial tissue and skin.

Acknowledgment

This work was supported by the Brain Korea 21 Project. This work was also supported by ASRI (The Automation and Systems Research Institute) at Seoul National University. We thank Eunjeong Kang very much for modeling all the computer models and basis elements of them used in the experiments.

References

[1] B. Choe, H. Lee, and H.-S. Ko. Performance-driven muscle-based facial animation. *The Journal of Visualization and Computer Animation*, 12(2):67–79, May 2001.

[2] C. D. Clemente. *Anatomy: A Regional Atlas of the Human Body, 2nd edition*. Urban and Schwarzenberg, 1981.

[3] P. Eisert and B. Girod. Analyzing facial expression for virtual conferencing. *IEEE Computer Graphics & Applications*, 18(5):70–78, September - October 1998. ISSN 0272-1716.

[4] P. Ekman and W. V. Friesen. *Facial Action Coding System*. Consulting Psychologists Press, Inc., 1978.

[5] I. Essa, S. Basu, T. Darrell, and A. Pentland. Modeling, tracking and interactive animation of faces and heads using input from video. In *Proceedings of Computer Animation '96 Conference*, June 1996. Geneva, Switzerland.

[6] I. A. Essa and A. P. Pentland. Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, July 1997.

[7] G. Faigin. *The Artist's Complete Guide to Facial Expression*. Watson-Guptill Publications, 1990.

[8] M. K. Greenwald, E. W. C. III, and P. J. Lang. Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli. *Journal of Pyschophysiology*, 3:51–64, 1989.

[9] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. Making faces. In *SIGGRAPH 98 Conference Proceedings*, Annual Conference Series, pages 55–66. ACM SIGGRAPH, Addison Wesley, July 1998.

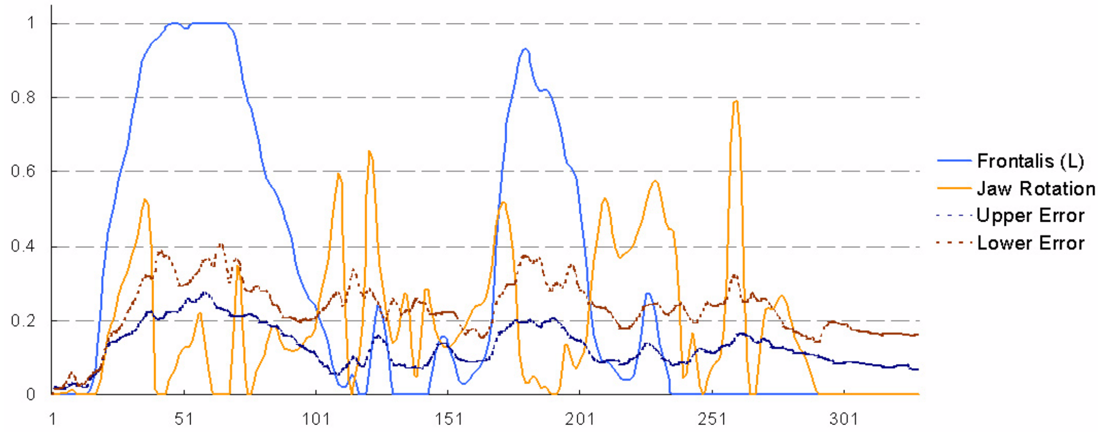

Figure 5. Contractions of 'left frontalis' / 'jaw rotate' and marker errors.

[10] R. M. Koch, M. H. Gross, and A. Bosshard. Emotion editing using finite elements. *Computer Graphics Forum*, 17(3):295–302, 1998. ISSN 1067-7055.

[11] R. M. Koch, M. H. Gross, F. R. Carls, D. F. von Büren, G. Fankhauser, and Y. Parish. Simulating facial surgery using finite element methods. In *SIGGRAPH 96 Conference Proceedings*, Annual Conference Series, pages 421–428. ACM SIGGRAPH, Addison Wesley, Aug. 1996.

[12] C. Kouadio, P. Poulin, and P. Lachapelle. Real-time facial animation based upon a bank of 3D facial expressions. In *Proceedings of Computer Animation '98 Conference*. IEEE Computer Society Press, 1998.

[13] Y. Lee, D. Terzopoulos, and K. Waters. Realistic face modeling for animation. In *SIGGRAPH 95 Conference Proceedings*, Annual Conference Series, pages 55–62. ACM SIGGRAPH, Addison Wesley, Aug. 1995.

[14] J. P. Lewis, M. Cordner, and N. Fong. Pose space deformations: A unified approach to shape interpolation a nd skeleton-driven deformation. *Proceedings of SIGGRAPH 2000*, pages 165–172, July 2000. ISBN 1-58113-208-5.

[15] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, 2nd edition, 1984.

[16] G. M. Nielson. Scattered data modeling. *IEEE Computer Graphics and Applications*, 13(1):60–70, Jan. 1993.

[17] J. Noh and U. Neumann. Expression cloning. *Proceedings of SIGGRAPH 2001*, pages 277–288, August 2001. ISBN 1-58113-292-1.

[18] F. I. Parke and K. Waters. *Computer Facial Animation*. A K Peters, 1996. ISBN 1-56881-014-8.

[19] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin. Synthesizing realistic facial expressions from photographs. In *SIGGRAPH 98 Conference Proceedings*, Annual Conference Series, pages 75–84. ACM SIGGRAPH, Addison Wesley, July 1998.

[20] F. Pighin, R. Szeliski, and D. H. Salesin. Resynthesizing facial animation through 3D model-based tracking. In *Seventh International Conference on Computer Vision (ICCV '99) Conference Proceedings*, pages 143–150, September 1999. Corfu, Greece.

[21] D. Terzopoulos and K. Waters. Physically-based facial modelling, analysis, and animation. *The Journal of Visualization and Computer Animation*, 1:73–80, 1990.

[22] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, June 1993.

[23] K. Waters. A muscle model for animating three-dimensional facial expression. *Computer Graphics (Proceedings of SIGGRAPH 87)*, 21(4):17–24, July 1987. Held in Anaheim, California.

[24] L. Williams. Performance-driven facial animation. *Computer Graphics (Proceedings of SIGGRAPH 90)*, 24(4):235–242, August 1990. ISBN 0-201-50933-4. Held in Dallas, Texas.