Facial Muscle Parameter Decision from 2D Frontal Image

Shigeo MORISHIMA*, Takahiro ISHIKAWA*, Demetri TERZOPOULOS**

Room 1201, Seikei University*
3-3-1 Kichijoji-kitamachi, musashino, Tokyo.
180, Japan
Phone: +81-422-37-3726
E-mail: shigeo@ee.seikei.ac.jp

University of Toronto**
6 King's Colledge Road Toronto
Ontario Canada
Phone: +1-416-978-7777
E-mail: dt@vis.toronto.edu

Abstract

Muscle based face image synthesis is one of the most realistic approach to realize life-like agent in computer. Facial muscle model is composed of facial tissue elements and muscles. In this model, forces are calculated effecting facial tissue element by contraction of each muscle strength, so the combination of each muscle parameter decide a specific facial expression. Now each muscle parameter is decided on trial and error procedure comparing the sample photograph and generated image using our Muscle-Editor to generate a specific face image.

In this paper, we propose the strategy of automatic estimation of facial muscle parameters from 2D marker movements using neural network.

This is also 3D motion estimation from 2D point or flow information in captured image under restriction of physics based face model.

1 Introduction

Recently, research into creating friendly human interfaces has flourished remarkably. In the human-human communication system, facial expression is the essential means of transmitting non-verbal information and promoting friendliness between the participants. We have already developed a facial muscle model as a method of synthesizing realistic facial animation[1][2].

The facial muscle model is composed of facial tissue elements and muscle strings. In this model, forces effecting each facial tissue are calculated by contraction of each muscle. So a combination of muscle strengths decides a specific facial image. Currently, however, we have to manually determine each muscle parameter by trial and error, comparing the synthesized image to a photograph.

This paper proposes a method of automatic estimation of facial muscle parameters from 2D marker movements in a face-frontal image. Our goal is to estimate the muscle parameters exactly and resynthesize a face expression with the original impression.

This paper proposes two methods of automatic estimation of facial muscle parameters from 2D information, i.e., marker movements and optical flow in a frontal face image. In the first, small colored circle markers are attached to a subject's face to measure the quantity of transformation of the face when an expression appears. Then we can find out the difference between any specific expression and a neutral expression.

A neural network which has learned several patterns of facial expressions can convert the marker movements into muscle parameters. This neural network can realized an inverse mapping of the image synthesis process from the 3D muscle contraction to the 2D point movement in the display. So this is also 3D motion estimation from 2D point tracking in captured image under restriction of physics based face model.

2 Muscle Parameters

2-1 Feature Points

A marker is attached on each feature point of the subject's face to measure and model facial expression. A feature point is chosen for each muscle, from the grid point in face model which gives the biggest movement when contracting the muscle. If the feature point has al-

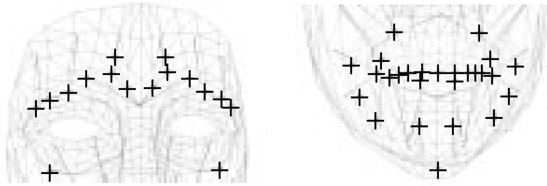


Figure 1. Feature Points

ready been chosen by another muscle, the point which gives second biggest movement is chosen as the feature point. Some of the feature points are appended heuristically whose motion is independent of other points when controlling each muscle. As a result, 16 feature points in the forehead area and 26 feature points in the mouth area are selected as shown in Figure 1.

2-2 Learning Patterns

Learning patterns are composed of the individual contraction of each muscle and their combination. In the case of individual motion, contraction of each muscle between maximum strength and neutral is quantized into 11 steps. In the combination case, we create 6 basic facial expressions consisting of anger, disgust, fear, happiness, sadness and surprise, and quantize the difference between neutral and each of these also into 11 steps.

The total number of learning patterns is 143 in each forehead sub-area. In the mouth area, each muscle contraction does not happen independently. So all learning patterns are composed of combinations, i.e., basic mouth shapes for vowels "a", "i", "u", "e" and "o", and a closed mouth shape. Also 6 basic expressions are appended as in the forehead area, and jaw rotation is specially introduced. They are also quantized into 11 steps. The number of learning patterns is 143 for the mouth sub-area.

Each pattern is composed of a data pair: muscle parameter vector and feature points movement vector. Neural networks were trained by Back Propagation.

3 Evaluation

3-1 Open Test

We attach markers on a real human's face and get movements of the markers from 2D images captured by a camera when any arbitrary expression is appearing.

After normalization, the movement values of the markers are given to the neural network and a facial image is re-generated using the facial muscle model on the parameters from the neural network output. Example images are shown in Figure 2. Muscle parameters are decided only from the 2D image, but the 3D facial image is well regenerated.

Also, some exceptions occur when a facial feature which cannot be generated by any muscle combination in our model is given as the test sample. In the example image in Figure 3 "ANGER", the upper and lower lips are being pulled up at the same time. In our current muscle model, the lower lip does not move up beyond the standard position because the muscle action is only limited to contraction. It is necessary to improve the muscle model by including expansion and relocation of muscles to solve this problem.

3-2 Objective Evaluation

Table 1 shows the error between captured face and synthesized one. 1st column means an averaged error in 42 marker coordinate in the display. Second column is also error in marker location but it cosiders human's sensitivity. At first, 7 people create the face whose impression is very close to the original face using Muscle Editor by trial and error procedure.

And these faces are averaged in muscle parameters and true face is defined. Figure 3 shows an averaged face for Anger. Standard deviation(s) is also calculated for each expression and it means sensitivity of human for expression shift. 2nd column is error in marker position. For example, in case of Anger, error is 5.01σ and σ is 3.41 pixels. 3rd column is error in muscle strength.

a) Original b) Average c) Synthesized

Figure 2. Open Test for Surprise

a) Original b) Average c) Synthesized

Figure 3. Open Test for Anger

	Marker Position Error pixels	Marker Position Error σ	Sigma pixels	Muscle Parameter Error σ	Sigma pixels
Anger	9.79	5.01	3.41	2.37	24.6
Disgust	5.25	1.19	5.33	1.73	23.0
Fear	6.46	1.68	5.42	1.38	17.8
Happiness	3.91	0.91	5.29	1.44	21.1
Sadness	6.91	1.75	4.51	1.47	18.5
Surprise	6.48	1.48	6.38	1.59	15.4

Table 1. Estimated Error for Open Test

These result shows a quantitative standard for evaluation of impression in the synthesized face. Anger is worst one in all cases.

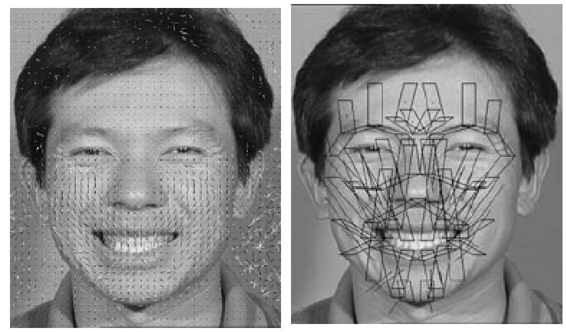
4 To Omit The Marker Location

In the next step, the marker has to be omitted to make the motion capture process easy and solve the problem of strong dependance on initially position of markers. Here new method is introduced. This method uses Optical Flow to get measurement of facial expression. Flow is calculated at each flame, and Finally flow is sumed by each flow. And averaged in the mask on the face. Figure 4(a) is result of optical flow calculation and Figure 4(b) show the masks for averaging movement.

Facial expression is resynthesized using the facial muscle model on the parameter from the neural network output. This neural network had learned couple of optical flow of original facial image and muscle parameter corresponding with original facial image. As a result, we can get images which are almost same as original images. Figure 5 shows an example of synthesized image. More delicate feature of expression can be captured by optical flow than marker tracking method. This evaluation is future subject.

5 Conclusion and Discussion

A method of automatically estimating 3D facial muscle parameters from 2D marker movements is presented in this paper. Parameter conversion from 2D to 3D works well when the model is fitted to the target person's face precisely in the 2D image and the expression variation is within the range of combinations of basic expressions and vowel pronunciations. We currently fit the model to the facial image by manual operation. But it's impossible to decide the location of all grid points precisely in the real facial image. Of course, a marker's movement strongly depends on its initial location and on the target person, and parameter conversion is very sensitive to its



(a) Optical Flow (b) Masks for Averaging

Figure 4. Face Feature from Optical Flow

Captured Synthesized

Figure 5. Synthesis by Optical Flow

effects.

Thus the correspondence between a real face and the model is the next problem to be solved. Now, our facial muscle model requires long computation time. As a result, this method is not real-time processing. Furthermore, from our evaluation test we can see that there is limit to generate any arbitral expression with our model, so relocation of the muscles and a new definition of the physics for the new muscle model are under examination. We had also introduced method with optical flow without markers. More delicate facial expression can be captured by optical flow. This evaluation is next subject.

Reference

[1] Yuencheng Lee, Demetri Terzopoulos and Keith Waters, "Constructing Physics-Based Facial Models of Individuals", Graphics Interface'93, pp.1-8 (1993).

[2] Demetri Terzopoulos and Keith Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, No.6, pp.569-579 (1993).

[3] Paul Ekman and Wallace V. Friesen, "Facial Action Coding System", Consulting Psychologists Press Inc. (1978).