

HUMAN FACE ANIMATION BASED ON VIDEO ANALYSIS, WITH APPLICATIONS TO MOBILE ENTERTAINMENT

JOHN SY-SEN TANG

City University of Hong Kong
itjohn@cityu.edu.hk

ALAN WEE-CHUNG LIEW

Chinese University of Hong Kong
wcliew@cse.cuhk.edu.hk

HONG YAN

City University of Hong Kong
University of Sydney
h.yan@cityu.edu.hk

Received April 15, 2005

Revised June 21, 2005

In this paper, we present a novel non-uniform rational B-spline (NURBS) muscle system for the simulation of human facial expressions and talking animation based on features extracted from video sequences. We construct the facial muscles based on anatomical knowledge and NURBS models. We use 43 feature points to represent facial expressions and apply a lip contour extraction technique to determine lip shapes. Our system is able to generate many different facial expressions and mouth shapes. The system has been used in web and mobile phone based digital entertainment applications.

Key words: Human face animation, facial expressions, lip synchronization, mobile entertainment
Communicated by: D Taniar

1 Introduction

With the fast increase in computer power, 3D human facial animation is gaining many applications in different areas such as entertainment and teleconferencing [1, 2]. Facial modeling and animation by geometric manipulations can be roughly classified into five categories [3], i.e., interpolation-based, parameterization-based, pseudo-muscle modeling, performance-based and physics-based modeling. Interpolation techniques use in-between method to generate the frames between two key-frames [4]. It is fast and relatively easy to generate primitive facial animation but difficult to create a wide range of realistic facial configurations. Parameterization techniques use a set of independent parameter values to specify any possible facial expressions [5, 6]. However, there might be conflict between parameters when they affect the same vertices. The pseudo-muscle modeling based method applies geometric deformation like free form deformation (FFD) animation [2]. As it does not follow the actual facial

anatomy, it is not suitable for facial animation that requires high fidelity. Performance-based method [18] captures real human facial movements and uses the data for animation. With fast advance in motion capture equipment, it can create more realistic expressions for 3D facial animation. However, the capturing equipment is still expensive and is not widely available. The physics-based muscle modeling method tries to simulate real human facial muscles for animation. Since it is based on the human anatomy, it is the closest to realistic human facial animation.

Compared with costly motion capture device, a vision-based approach is inexpensive [20-23]. Model-based optical flow tracking algorithm is a common method [14,15]. It mainly uses optical flow to estimate facial movements. Terzopoulos and Waters [7] track the contour features to drive a muscle base model [11]. They estimate the contraction parameters from the contours. Chai uses 2D video parameters to control motion capture parameters to drive a 3D model [16]. In this paper, we develop a vision-based system to drive a spline muscle model [19,20]. The system tracks the expression and lip shape parameters, which control the contraction of the spline muscle in facial animation.

Our system allows control on linear muscles to simulate a particular expression realistically by modifying the weights of different control points. As it uses NURBS to simulate muscles, the control points can be put on the surface of face mesh based on the facial anatomy. This makes it easier to locate facial muscles. Through the control points, the curve can be formed under the mesh, like the muscle under the face. By changing the weights of the control points, the knots will form a motion vector to control the movement of the mesh vertex within certain region. The number of control points can be determined by the complexity of different parts of the face. To animate the mouth, a lip contour extraction technique [13,20] is employed to extract the lip shape parameters from a video sequence. These parameters will be used to control a virtual model to form different phonemes.

The rest of the paper is organized as follows. Section 2 presents an overview of the NURBS muscle model. Section 3 describes whole system and Section 4 provides simulation results. Section 5 presents applications of our human facial animation system to web and mobile phone based digital entertainment, followed by conclusions in Section 5.

2 NUBRS Muscle Modeling

Our earlier NURBS curves based method separates the human face into five facial units and uses the NURBS curves to control them [24-26]. With this approach, it may be hard to locate a control polygon on the mesh model as the control polygons of the NURBS curves only roughly follow the face mesh. In the system we propose in this paper, each muscle is formed by one NUBRS curve. Each NUBRS curve has three to five control points. It is easy to locate the control points on the surface of the face mesh as it is based on the facial anatomy. Modification of the weights of the control points can force the knots of NURBS curve to move, which will then generate a motion vector to influences the nodes on the 3D mesh. The number of control points of the NUBRS muscle dependeds on the types of muscle and parts of the face.

2.1 NURBS Curves

An NURBS curve of degree n is defined as [8]

$$e(u) = \frac{\sum_{i=0}^n B_{i,n}(u) \omega_i P_i}{\sum_{i=0}^n B_{i,n}(u) \omega_i} \quad 0 \leq u \leq 1 \quad (1)$$

where $e(u)$ is the knot of the curve, which will be used for calculating the motion vector as described in Section 3.1, ω_i are the weights, P_i are the control points and $B_{i,n}(u)$ is the blending function defined as follows,

$$B_{i,n}(u) = \frac{n!}{i!(n-i)!} u^i (1-u)^{n-i} \quad (2)$$

An important property of NURBS curves is the convex hull property, i.e.

$$\sum_{i=0}^n B_{i,n}(u) \omega_i = 1 \quad (3)$$

This property ensures that the polynomial can follow the control points smoothly without erratic oscillations. Another important property is endpoint interpolation, which means that the curve always passes through the first and last control points [8, 9].

2.2 NURBS Linear Muscle and NURBS Sphincter Muscle

The control points of the NURBS model are classified into two groups. They are reference control points and current control points (Fig. 1). A reference control point is used to relate the knot vector and the node of the mesh inside the influence region. A current control point is the control point whose weight is currently being modified.

Figure 2 shows the displacement of the nodes before and after increasing the weight of the current control point. The nodes will contract towards the current control point, which acts like the human muscle contraction.

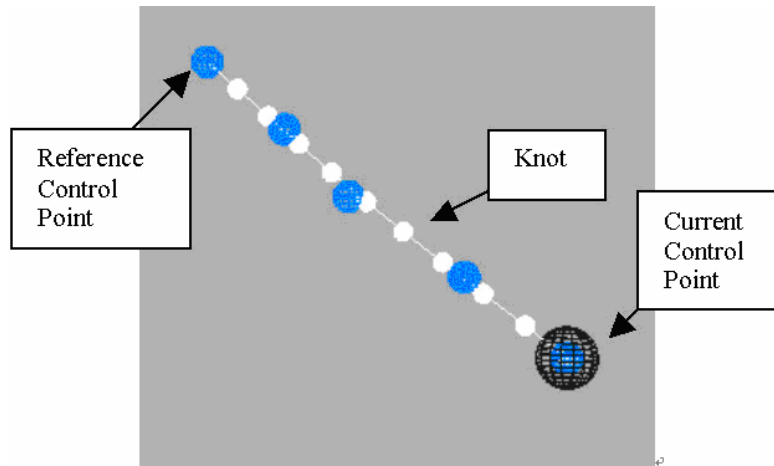


Figure 1: A NURBS curve.

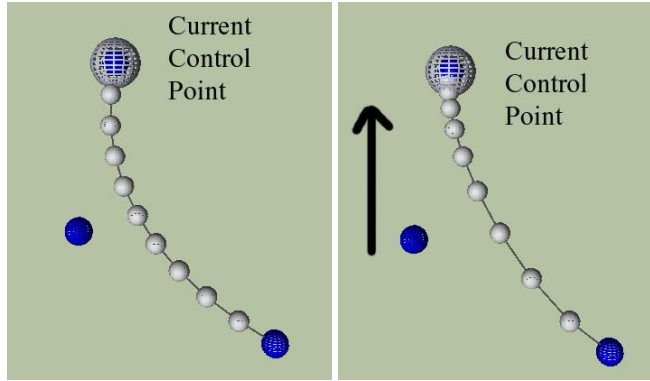


Figure 2: A NURBS muscle before and after contraction.

For NBURS linear muscles, the current and reference control points are illustrated in Fig. 3.

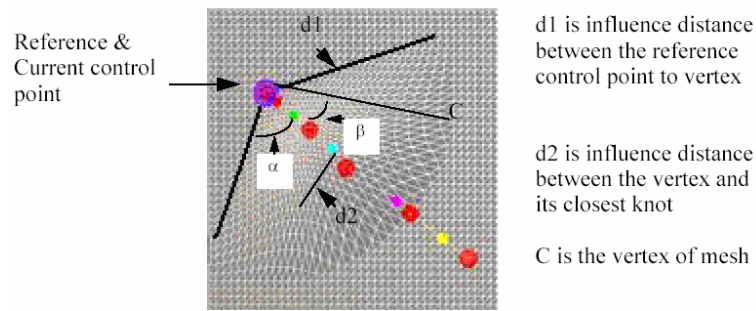


Figure 3: NURBS linear muscle model.

Figure 4 shows the relationship between the reference point, the knot and the vertex. B is the average of the knot positions before movement and B' is average of the knot positions after movement. $\overrightarrow{B'B}$ is a vector formed by the knot movement and can be obtained by

$$\overrightarrow{B'B} = \frac{1}{n+1} \sum_{i=0}^n [e_i(u)' - e_i(u)] \quad (4)$$

where $e_i(u)$ is the node before movement, $e_i(u)'$ is the node after the movement. In Figure 3, C is the vertex of the mesh which is within the influence region. A virtual vector is used to control the vertex direction. If C is repositioned to C' , a vector $\overrightarrow{C'C}$ will be formed. We can use vector $\overrightarrow{B'B}$ to determine vector $\overrightarrow{C'C}$ based on the following geometric relations,

$$\angle BAC = \angle B'AC' \quad (5)$$

$$\angle ABC = \angle AB'C' \quad (6)$$

$\overrightarrow{C'C}$ can be used to calculate the new position of vertex C . Finally, the displacement of the vertex, C'' , can be calculated by adding a displacement vector [8],

$$C'' = C + D \cdot R \cdot \overrightarrow{C'C} \quad (7)$$

where K is contraction factor, R is an angular displacement factor,

$$R = \frac{\cos \beta}{\cos \alpha} - 1 \quad (8)$$

and D is a radial displacement factor given by

$$D = \left(1 - \frac{\|AC\|}{d_1}\right) \cdot \sin \frac{\pi \|AC\|}{2d_1} \quad (9)$$

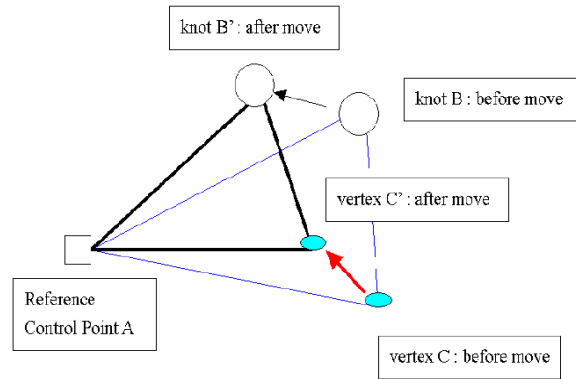


Figure 4: Relationship between the control point, knot and vertex.

As the NUBRS muscle can perform non-linear movement, the motion vectors are calculated independently for each mesh node. This ensures the smooth movement of the mesh. Moreover, it increases the controllability of the muscle system.

Since the sphincter muscle does not have angular displacement, it can be formed by multiplying two radial displacement parameters as shown in Fig. 5:

$$D_2 = \cos \frac{a \pi}{2d_1} \cos \frac{b \pi}{2d_2} \quad (10)$$

where a is the length between the knot and the vertex, b is the length between the control point and the vertex.

2.3 Usages of NURBS Muscles

The linear NURBS muscles are mainly responsible for facial expressions. The sphincter muscle can be used for the deformation of the mouth region. Since human face has fatty tissue under the skin, we

have to consider the effect on fatty tissue when we use muscles to control the skin mesh. To create a realistic expression, our system provides a way to simulate the fatty tissue on a face as shown in Fig. 6. This is achieved by adding additional control points between two end control points. The control points will drag the mesh slightly up, simulating the fatty tissue. As the control point is placed on the surface of a face, the curve will follow the shape of the face.

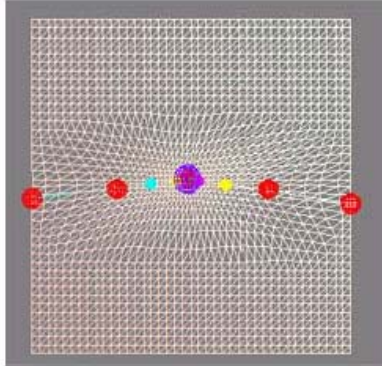


Figure 5: NURBS sphincter muscle model.

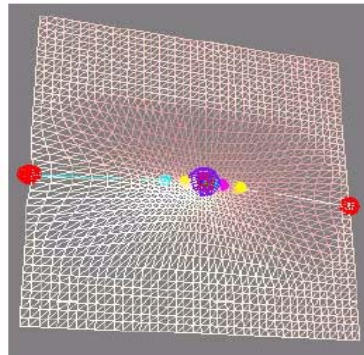


Figure 6: Fatty tissue simulation.

The motion vector is recalculated each time after the NURBS muscle changes its weights (i.e. the nodes move to a new position). The NURBS curve movement is very similar to human muscle contraction. By changing the weights, the nodes contract to the control point. In other words, we can simulate a realistic human muscle movement by following the movement of the NURBS curve. The system uses this dynamic smooth movement vector to guide the vertex of a face model. This is different from the normal static vector muscle, where a vector only points to one static direction but does not follow the shape of the face. As a result, our system can generate smooth and realistic facial animation. Moreover, our system can simulate the fatty tissue. The thickness of the fatty tissue can change the appearance of the expression. This increases the flexibility to generate different effects on the face model. A detailed facial expression can therefore be simulated under our system. Figure 7 shows a smiling expression generated by a normal vector muscle compared with the NURBS vector muscle. We can see that a more realistic smile is simulated by the NURBS vector muscle (see the difference on the cheek).

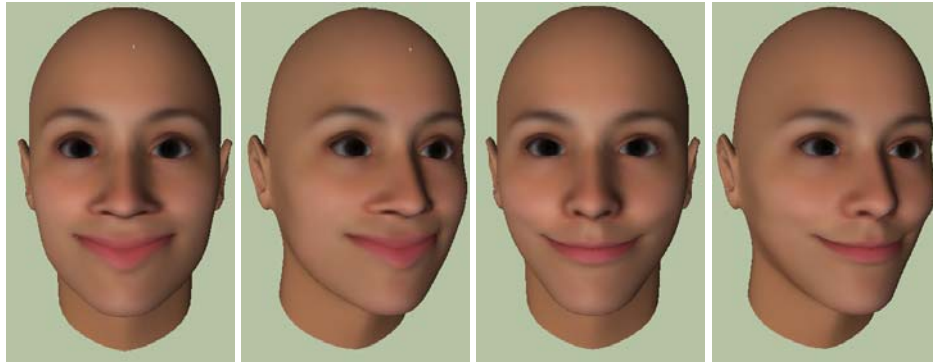


Figure 7: Comparison between smiles generated using normal (left two images) and NURBS (right two images) vector muscles.

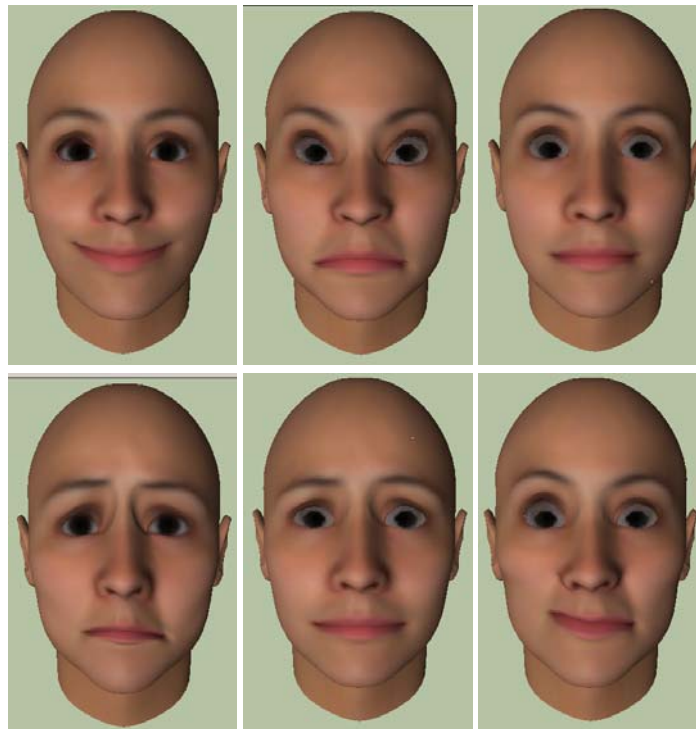


Figure 8: Basic facial expressions. From left to right, they are happiness, anger and surprise in the first row and sadness, fear and disgust in the second row.

Figure 8 shows a set of basic expressions. To make an expression, we just need to simply change the weights of the control points of each muscle. For example, if we want to make the model smile, we can modify the weights of the zygomatic major NURBS muscle. This muscle has three control points. The first control point is mainly responsible for contracting the model to make it look like smile and the second control point is used to simulate the fat tissue. The third control point is used to form this NURBS muscle. Table 1 shows the experiment result of weight distribution of the control points used for each muscle for a smile expression. Different models will have different sizes and structures, and through the experiment we can determine the suitable scale parameter to control the

weighting of the muscles. We can produce the expression weight distribution table for each basic expression in the experiment. This will be very useful for the expression tracking process, which we will describe in Section 3.1.

Table 1: Weight distribution of NURBS muscle for “smile”

Name of Muscle	Number of Control Points	1 st Control Point	2 nd Control Point	3 rd Control Point	4 th Control Point	5 th Control Point
Zygomatic Major	3	1	0.5	0		
Angular Depressor	3	0	0	0		
Frontalis Inner	3	0.2	0.1	0		
Frontalis Major	3	0.3	0.1	0		
Frontalis Outer	3	0.3	0.1	0		
Labi Nasi	3	0	0	0		
Inner Labi Nasi	3	0	0	0		
Lateral Corrigator	3	0	0	0		
Secondary Frontalis	3	0	0	0		
Levator Labii	5	0	0	0	0	0
Orbicularis Oris	5	0	0	0	0	0
Mentalis	5	0	0	0	0	0

Figure 9 illustrates how NURBS curves can control the mouth shape. The reference control point on the curve is fixed to ensure that the mesh would not form an unexpected mouth shape. The current control can be any one of the control points. With such arrangement, the muscle can control the lip to form a desired phoneme shape.

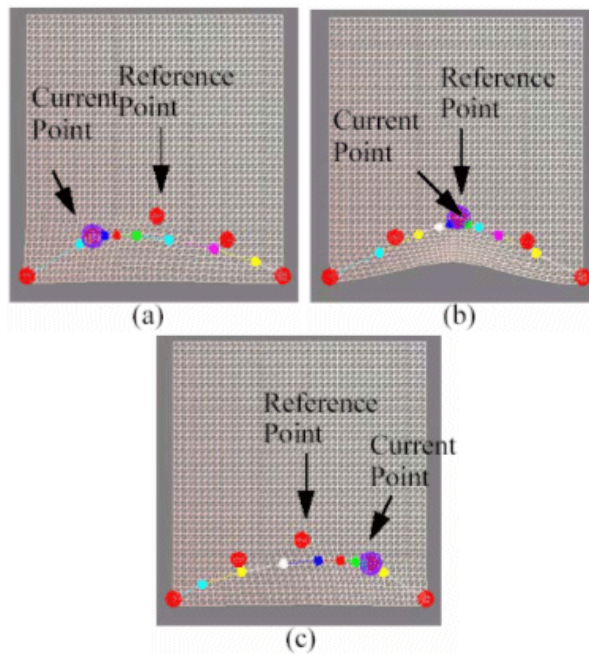


Figure 9: NURBS muscles with same reference point but different control points.

3 The Facial Animation System

In our system, the animation parameters obtained from video images are mapped to the appropriate spline muscles to drive facial animation. Facial expressions are controlled by 43 tracking parameters and the lip shape is tracked by a contour model with 5 parameters. Transforming the 2D data to control the 3D model is the final stage of this animation system. Figure 10 shows the structure of our system.

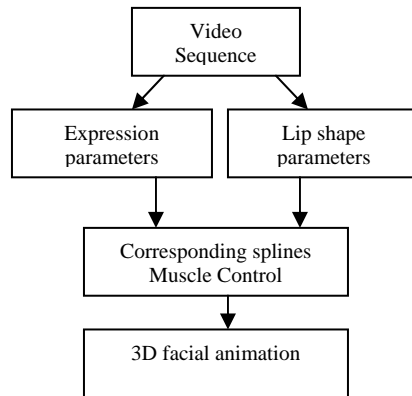


Figure 10: Structure of the facial animation system.

3.1 Facial Expression Extraction

In our system, a set of points are used to track facial expressions (Fig. 11). We follow the distribution of the control points of the NURBS muscle to place the tracking points on the face. There are 43 points assigned for expression tracking. The system extracts the color points from video [10]. The locations of tracking points are manually initialized. We use an 8 by 8 window to search the new locations in the consecutive frame. The magnitude and direction of a tracking point movement will be measured to drive the NURBS muscle model.

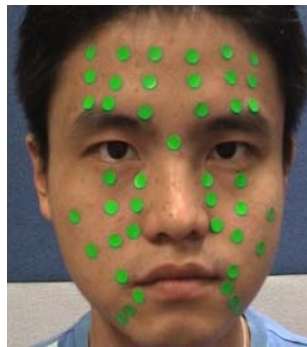


Figure 11: Expression tracking points.

3.2 Expression Control parameters

Our system controls the facial expression using 22 muscles based on the facial action coding system (FACS) [12] to. Figure 12 illustrates the mesh with NURBS muscles. The tracking points follow the locations of the control points. That is, the parameter obtained from a tracking point is used to control

the corresponding control point. The direction of the tracking point will indicate whether this muscle is contracting or releasing. The magnitude of the tracking point movement represents the contraction force. We have

$$\text{Expression Weight} = S_1 E \tag{11}$$

where S_1 is a scaling parameter determined experimentally and may be different for different muscles.

In Section 2.3, we mentioned that each basic expression has a weight table. We can use the weight table to determine the tracked face expression. The probability of each basic expression can be expressed using the following equation:

$$\text{Expression probability} = \frac{1}{100N} \sum M \tag{12}$$

where N is the total number of the muscles, and M is the percentage of the weight difference in each muscle. That is

$$M = \frac{1}{n} \sum |w_c - w_t| \tag{13}$$

where w_c is the weight of the control point in the expression weighting table and w_t is the weight of the correspond tracking point, and n is the number of the control points in the NUBRS muscle.

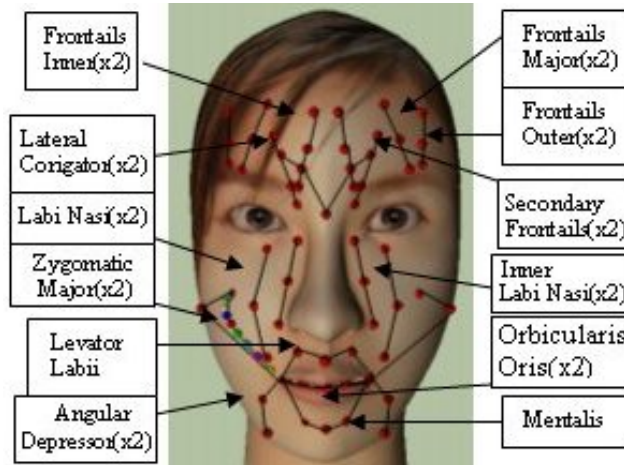


Figure 12: A set of facial muscles.

3.3 Lip Shape Extraction

We apply a technique we developed earlier [13] to extract the mouth boundary. This technique uses color video images to analyze lip movement and a geometric deformable model to describe the lip shape (Fig. 13). This model is formed by two curves. The shape of the model is pre-constrained to the expected lip shape. The lip equations are defined as follows,

$$y_1 = h_1 \left(\left(\frac{x - sy_1}{w} \right)^2 \right)^{1+\delta^2} - h_1 \tag{14}$$

$$y_2 = \frac{-h_2}{(w - x_{off})^2} (x - sy_2 - x_{off})^2 + h_2 \tag{15}$$

for $x \in [-w, w]$ with the origin at (0,0). In this model, parameter s describes the skewness of the lip shape and exponent δ describes the deviation of the curve from a quadratic. We then find an optimal partition of a given lip image into lip and non-lip regions based on pixel intensities and colors. Since this model is defined by a small set of parameters with clear physical interpretation, it is suitable for controlling the NURBS muscles. An NUBRS muscle is formulated based on a few control points, which are determined according to these parameters, as discussed below.

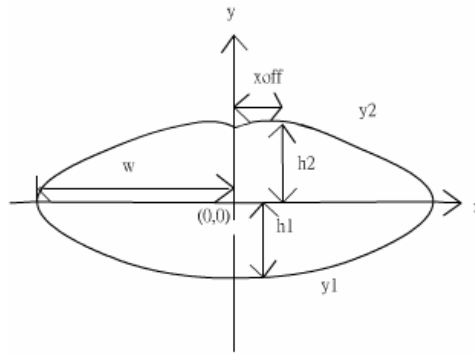


Figure 13: A geometric lip model.

3.4 Lip Contour Extraction Parameters

In Equations (14) and (15), five parameters are used to control the lip contour. They are given by

$$p = \{w, h_1, h_2, x_{off}, \delta\}.$$

The correspondences between muscles in Figure 13 and the lip model parameters in Equations (14) and (15) are as follows:

- Zygomatic Major and Angular Depressor : w
- Mentalis and Orbicularis Oris : h_1
- Levator Labii and Orbicularis Oris : h_2 Orbicularis Oris : x_{off} Mentalis : δ

The parameters determine the weights of different muscle's control points for activating the muscle to form the desire lip shape. The lip shape parameters are linearly proportional to the muscle's weight,

$$\text{Weight of NURBS curves} = Sp \tag{16}$$

where S is a scaling constant, which is determined experimentally and is different in different part of the face model. For example, the width parameter w directly controls the Zygomatic Major and

Angular Depressor muscles since two muscles stretch the lip to make it wider. In practice, these two muscles are controlled by the same weight using equation (16). It ensures that the mouth is only stretched wider but not to other direction.

4 Experimental Results

An NURBS facial modeling system is implemented to test the described method. The modeling system is developed using VC++/OpenGL. It runs on a Pentium 4 1.9GHz PC with 3D graphics interface. Based on the input video sequence, we create a variety of expressions and mouth shapes using NURBS-based muscles. Each expression simulation is done according to the action units of the FACS. Figure 14 shows expressions created according to video analysis.

Figure 15 compares the video lip extraction results with the 3D simulation results of the mouth uttering the digit “four” in Cantonese. The results from lip extraction control the virtual model and lip tracking system gives meaningful parameters to control the NUBRS muscles for creating realistic lip animation.

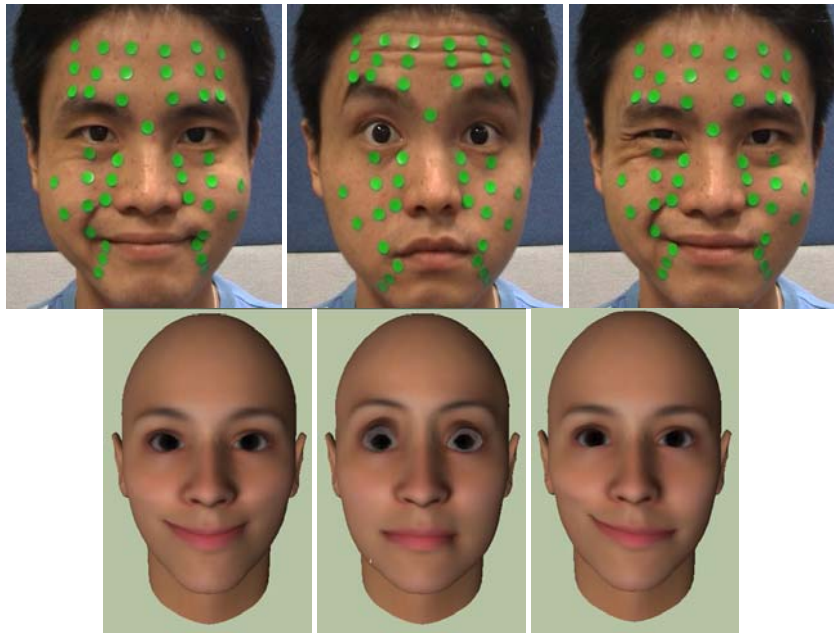
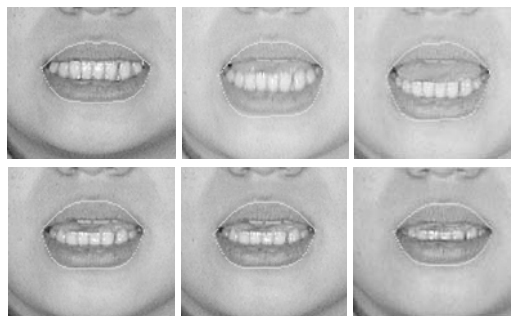


Figure 14: Example of video tracking (first row) and graphics simulation (second row) for happiness, surprise and disgust.



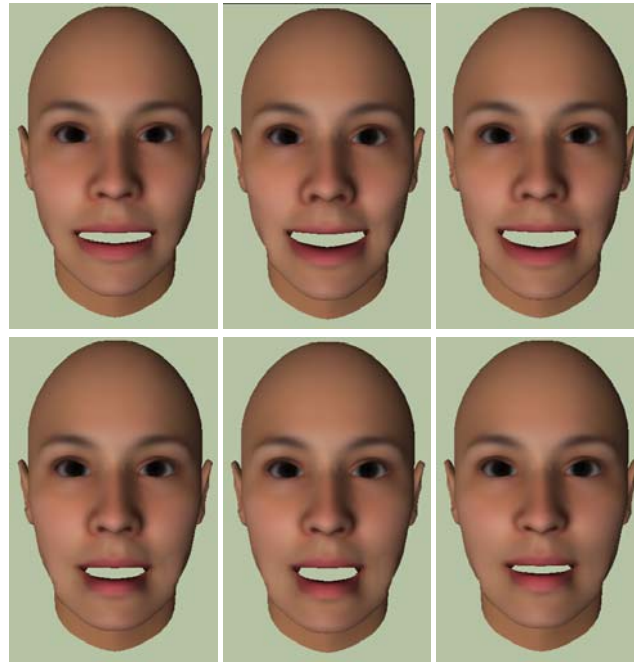


Figure 15: Comparison between real video lip shapes (first two rows) with the animated lip sequence (third and fourth rows) for a mouth uttering the digit “four” in Cantonese.

5 Practical Applications

Our lip synchronization and facial animation system has already been implemented on the web site “<http://www.HyperAcademy.com>”, or simply “<http://www.hy8.com>”. The web site can be accessed from both PCs and mobile phones. It contains many animation movies that can be used for greeting. On the web, one can select a greeting category and a specific message in a desired language, specify whether it is on PC or mobile and choose a favorite character. Then the user can preview the movie, enjoy it or send it out by specifying an email address. An example of the display from the web is shown in Figure 16. Our technologies have been used by internet and communications companies for real-time facial animation on 3G systems and multimedia message services (MMS) on 2G systems.

Our lip synchronization and facial animation system can have many other interesting applications:

- *Greeting characters:* Using our system one can send a voice message with an animated character and music, for example, saying “happy birthday” to another user.
- *Talking celebrities:* We can have animated human characters to read news, tell stories and report stock prices. The system can be used with real human voice or a text to speech engine. We can also show advertisement with virtual characters, such as promoting a company’s products and services.
- *Education and training:* Using our novel lip synchronization and facial animation technologies, educational and training software can be made more user-friendly and more attractive. The student can choose a character he/she likes most and learn from this virtual teacher more effectively.
- *Karaoke:* Currently some Karaoke systems can show images or videos along with a song. It will be more interesting if we can show animated person singing with a user. Using our technology, the user will be able to choose his/her most favorite virtual singer from a large number of characters in a database and change the singers’ appearance (e.g wearing different clothes).

- *Other applications:* The above applications can be implemented on mobile phones as well as desktop and specialized computers. In addition, our system can find a wide range of other applications, such as in talking toys, television, and intelligent tour guide at train stations, airports, large government and business buildings and exhibition centers.

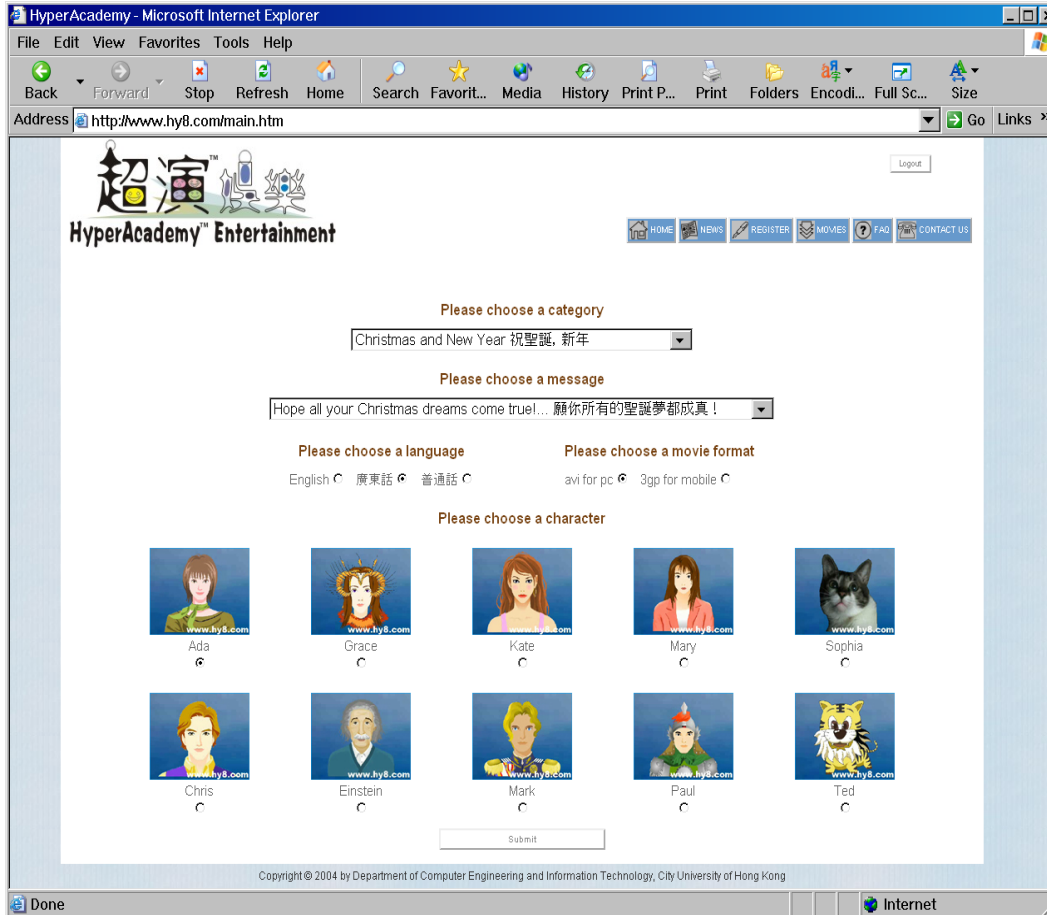


Figure 16: Our web site providing facial animation greeting movies for PCs and mobile phones.

6 Conclusions

In this paper, we have presented a novel method for facial expression animation, which uses NURBS curves to create linear and sphincter muscle and control the mouth shape. Our NURBS-based muscle model can simulate more realistic facial expressions by incorporating fatty tissue simulation. Since the control points are attached to the surface of the mesh, the muscle positions can be easily located. In order to extract the detailed lip shape, we use different approaches to track facial expressions and lip shapes. The flexible facial muscles also allow realistic simulation and animation of talking mouth based on a set of physically meaningful lip shape parameters extracted from video of a talking person. Combing the video input data with this muscle model, we can simulate realistic facial expressions and phonemes. Our lip synchronization and facial animation techniques have many interesting applications in web and mobile phone based digital entertainment systems.

Acknowledgements

This work is supported by a strategic research grant (project 7001556) from City University of Hong Kong.

References

1. G. Breton, C. Bouville and D. Pele, "FaceEngine: A 3D Facial Animation Engine for Real Time Applications", *Proceedings of the 2001 Conference on 3D Technologies for World Wide Web*, February 2001.
2. S. Coquillart, "Extended Free-Form Deformation: A Sculpturing Tool for 3D Geometric Modeling", *Computer Graphics*, 24: 187-193, 1990.
3. J. Noh and U. Neumann, "A Survey of Facial Modeling and Animation Techniques", *USC Technical Report*, 99-105, 1998.
4. F. Pighin, J. Hecker, D. Lishchinski, R. Szeliski and D. H. Salesin, "Synthesizing Realistic Facial Expression from Photographs", *SIGGRAPH Proceedings*, 75-84, 1998.
5. F. I. Parke, "Parameterized Models for Facial Animation", *IEEE Computer Graphics and Applications*, 2(9): 61-68, 1982.
6. J. Noh, D. Fidaleo and U. Neumann, "Animated Deformations with Radial Basis Functions", *ACM Symposium on Virtual Reality Software and Technology (VRST 2000)*, 166-174, October 2000.
7. D. Terzopoulos and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical model", *IEEE Trans. Pattern Analysis and Machine Intelligence*. 15:569-579, 1993.
8. L. Piegl and W. Tiller, *The NURBS Book*, 2nd Ed., Springer, Berlin, 1997.
9. D. Hearn and M. Pauline Baker, *Computer Graphics C version*, 2nd Ed., Prentice-Hall International, Inc., 1998.
10. D. Chai and K. N. Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications", *IEEE Trans on Circuits and Systems for Video Technology*, 9(4), June 1999.
11. K. Waters, "A Muscle Model for Animation Three-dimensional Facial Expression", *ACM SIGGRAPH Computer Graphics, Proceedings of the 14th Annual Conference on Computer Graphics and Interactive techniques*, 21(4), August 1987.
12. P. Ekman and W. V. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA, 1978.
13. A.W.C. Liew, S.H. Leung and W.H. Lau, "Lip Contour Extraction from Color Images Using a Deformable Model", *Pattern Recognition*, 35: 2949-2962, Dec 2002.
14. D. DeCarlo and D. Metaxas. "The Integration of Optical Flow and Deformable Models with Applications to Human Face Shape and Motion Estimation". *Proceedings CVPR '96*, 231-238, 1996.
15. S.B. Gokturk, J.Y. Bouguet, and R. Grzeszczk. "A Data Driven Model for Monocular Face Tracking", *In IEEE International conference on computer vision*, 701-708, 2001.
16. J.X. Chai, J. Xiao and J. Hodgins. "Vision-based Control of 3D Facial Animation", *Eurographics/SIGGRAPH symposium on Computer Animation*, 2003
17. E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice-Hall International, Inc., 1998.
18. I. Buck, A. Finkelstein, C. Jacobs, A. Klein, D. H. Salesin, J. Seims, R. Szeliski and K. Toyama, "Performance-driven hand-drawn animation", *ACM SIGGRAPH, Symposium on Non-photorealistic animation and rendering*, June 2000.
19. S.S. Tang, H. Yan, J. Feng, L.K. Szeto and A.W.C. Liew "A NURBS-based Vector Muscle Model for Generating Human Facial Expressions", *Proceedings of 4th International*

- Conference on Information, Communication & Signal Processing & 4th IEEE Pacific-Rim Conference on Multimedia*, Singapore, 2003
20. S.S. Tang, A.W.C. Liew and H.Yan “Lip-Sync in Human Face Animation Based on Video Analysis and Spline Models”, *10th International Multimedia Modelling Conference*, Jan, 2004
 21. S.K. Karunaratne and H.Yan, “A new efficient expression generation and automation cloning method for multimedia actors”, *Circuits, System and Signal Processing*, 20(3):447-484, 2001
 22. H. Yan, “Image analysis for digital media applications”, *IEEE Computer Graphics and Applications*, (21)(1):18-26, 2001
 23. H. Yan, “Human face image processing techniques”, *Int’l J. Image and Graphics*, 1(2):197-215, 2001.
 24. D. Huang and H. Yan, “Modeling and animation of human expressions using NURBS curves based on facial anatomy”, *Signal Processing: Image Communications*, 17:457-465, 2002
 25. D. Huang and H. Yan, “Modeling of deformation using NURBS curves as controller”, *Signal Processing: Image Communications*, 18:419-425, 2003
 26. D. Huang and H. Yan, “NURBS curve controlled modeling for facial animation”, *Computers & Graphics*, 17:373-385, 2003.