An AAM based Line Drawing Facial Animation Creation Method for Expression Transfer

Qingxiang Wang[1,2], Changhe Tu[1] and Xiaoqiang Ren[2]
[1]*School of Computer Science and Technology, Shandong University*
Jinan, 250100 Shandong, China
[2]*School of Information, QILU University of Technology*
Jinan, 250353 Shandong, China
wangqx@spu.edu.cn, chtu@sdu.edu.cn, renxq@spu.edu.cn

Abstract

A simple method of line drawing animation creation method from a human facial video is presented in this paper. Without much manual intervention, the work could be used in personal cartoon video processing and communication. In this domain, a number of works has been proposed, while most of them need interactive processing or specific data such as 3D model. The method presented in this paper could automatically transfer expressions without much user participation. Given an input video and line drawing model with some basic emotions, active appearance models are used to get the current facial feature points of the video, and then the system can achieve the emotion parameters by solving the emotion function of AAM and transfer them to the same basic emotions model of line drawings. The basic emotions of line drawings are formed by background image and facial feature lines which are drawn with a set of Bézier curves with width. Finally, facial line drawing animation is obtained and shown in the experiment.

Keywords: *Facial Expression Representations, Line Drawings, Active Appearance Models, Animation*

1. Introduction

Facial line drawing is a powerful art form, which takes an important part in the NPR, and is widely used in art creation, games, comics, advertising design and other domains. Previous research has demonstrated the potential of the approach in portrait, e.g. PICASSO System [1] and DCM [2]. Furthermore, facial line drawings are frequently used as a means of providing visual representations of an individual's face in automatic computer vision application alternative solution.

Vision based methods can avoid intrusive markers on the actor's face and usually use a camera. Some work had been published, to get animation, Chai [3] create 3D animation with the help of the motion capture data, Taehyun Rhee [4] use parametric 3D model to track feature points and head movements, Eun-Jung Lee [5] generates Exaggerated line drawings, they rely on 3D models and not suitable for line drawings. Another progress, [6,7] are also use AAM to generate facial sketch, [6] presents a hierarchical animation method for transferring expressions by a neighbour expression transfer model (NET) and [7] proposed a method to get expressive caricatures of basic main expressions, but they need an approximate face existed in the train set.

This paper presents a method for facial line drawing transferring, which uses the captured videos of the user's expressions from a family camera. An overview of the system is shown in

Figure 1. Our main contribution is that we present a method which is easy to use for expression transferring from video to line drawings without manual intervention in the runtime processing.

Our system include two parts: offline pre-processing and runtime processing, in the offline pre-processing processing part, we first use training image set which is consist of 70 pictures of the faces of 10 individuals; to establish Active Appearance model (Active Appearance Models, abbreviation for AAM) [8], image set including neutral and six basic expression, at the same time using Bézier curves to establish neutral and basic expression lines comic character models, for a particular user, with the user's neutral and basic expression images. In run-time, after inputting a video which includes a facial expressions sequence, we first use active appearance model on each frame to get the designated feature point set of the face, then get the expression composition factor with the basis of the positioning of the feature points trained from the basic expression images by linear fitting method, and finally transfer these coefficient to the same basic emotions model of line drawings in the final result.

This paper is organized as follows. Section 2 we will introduce background knowledge of our work, the AAM and Bézier curves. Section 3 shows how to track the facial feature points with AAM. Section 4 describe the facial line drawing models which are defined with Bézier curves and Section 5 illustrate the method of expression transfer from live video to line drawings with the same basic emotions. The followed two sections are the result and conclusion.

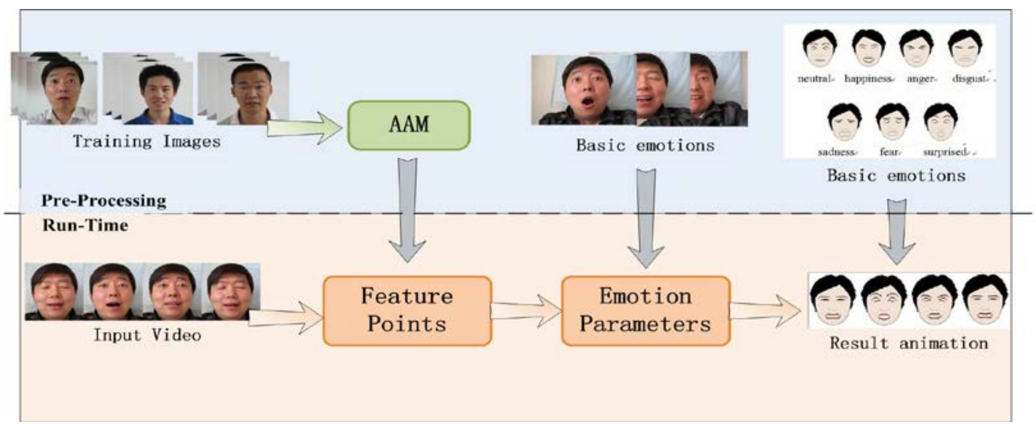

Figure 1. System Overview

2. Background Knowledge

The Active Appearance Models (AAMs) [8] is a successful method for matching the feature points of new facial images and is applied in many applications. The method uses the statistical models and the input image’s appearance to estimate the shape of the face. In the paper we use AAM algorithm to track the positions of the feature points on human face in each frame of the video.

2.1. Active Appearance Models

AAM is a statistical model that is trained by a set of images. The images are manually labeled before training. The labeled points on the images are defined as the shape of the image. The coordinates of n points are defined as a vertices $s = (x_1, y_1, . . . , x_n, y_n)^T$.

AAM assume that shape could be defined as the linear combination of the basic shape and some variation vertices of the points in the basic shape. The basic shape named S_0 which is the mean of shapes in all of the training images after aligning these shapes to the same coordinate with each other using Procrustes and then applies the PCA to the aligned shapes to get the variations of the shape. The model is defined as follow

$$S = S_0 + \sum_{i=1}^{m} p_i S_i \qquad (1)$$

where the coefficients p_i are the shape parameters and The base shape S_0 is the mean shape and the vectors S_i are the eigenvectors corresponding to the m largest eigenvalues.

The appearance of the AAM is also assumed as the linear combination of the mean appearance A_0 and some variation of the appearance. A_0 is defined as the pixels set that lie inside the base mesh S_0. For each training image, the pixels number is different, so we should project them to the same size to mesh S_0 at first with Piece Wise Affine, and then apply PCA on it to get the follow

$$A = A_0(x) + \sum_{i=1}^{l} q_i 2A_i(x) \qquad (2)$$

Where the coefficients q_i are the appearance parameters. As with the shape, the base appearance A_0 and appearance images A_i are usually computed by. The base appearance A_0 is the mean shape normalized image and the vectors A_i are the eigenvectors corresponding to the l largest eigenvalues.

After applying PCA to the shape and appearance of the training images, we can get the S_i and A_i and the Relation matrix R between the variation of coefficient and appearance residual. When an image of face is inputted, AAM minimizes equation (3) to fit a shape to the face by iterative model refinement.

$$\sum_{x \in S_0} \left[A_0(x) + \sum_{i=1}^{m} q_i A_i(x) - I(W(x;p)) \right]^2 \qquad (3)$$

In our experiments, the AAMs are built up with expression face images belonging to 10 individuals and each of them has seven images included neutral, happiness, anger, disgust, fear, sadness and surprised. Each image is manually labeled with 68 points for the training.

2.2. Bézier Curves

Bézier curves are widely used in computer graphics to model smooth curves. As the curve is completely contained in the convex hull of its control points, the points can be graphically displayed and used to manipulate the curve intuitively. Affine transformations such as translation and rotation can be applied on the curve by applying the respective transform on the control points of the curve.

In this paper we choose cubic Bézier curves with width to describe the facial line drawing model and most of the curves are cubic curves with four control points. The control points are saved to file for reused. Some lines share one control point with the line previous or next to them. We will introduce how to use that to form facial line drawing model in Section 4.2.

3. Facial Features Initializing and Tracking

Tracking or identify the facial features and then further extract the expressions information from video to help the users to recognize the expression and provide the input for followed application. Commonly used feature tracking methods, such as optical flow method, need manually specify the initial features used for tracking in advance. While features recognition method, such as ASM, AAM, 3DMM, with a given initial value, can get the location of the feature points by optimization. The number of the features has some different versions like 13 points, 19 points, 58 points, *etc*. Regardless of that, the facial features usually contain the obvious characteristics, such as the corners of the mouth and eyes. ASM is easy to fall into local extremum and the calculation of 3DMM is heavier. The AAM use the global matching method so that avoid the local minima of ASM. In this paper, we choose AAM for facial features tracking in video and use the 68 feature points model.

3.1. Initializing

When an expression video is coming, we first use the Viola-Jones face detect method [9] with the haarcascade_frontalface_alt2 model from OpenCV to find the region of the face, and then Initialize the initial value of AAM iterative coefficients with it. If the current frame has no face detected, the frame will be dropped and deal with the next one until the face appears. Every result of current frame will be the initial value of the next until the next matched feature points are very different with current. The different threshold is measured with mean square error and set to 10.

$$\mathrm{MSE} = \frac{\sum_{i=1}^{N} \| f_i^n - f_i^c \|^2}{N} \tag{4}$$

Where N is the number of the feature points and f_i^n , f_i^c are the feature points of next frame and current frame.

The feature points in the model is composed of a fix number of points and the final fitting result is the same. In this paper, we use the same definition as XM2VTS[10] frontal data, 68 points.

3.2. Tracking

We use the origin AAM method[8] to fit the frames of the video. The tracking result of every frame is a vector of 68 features points and each point is formed by $[x, y]^T$. The locations of the features points are defined as shown in Figure 2. The model is trained with the near frontal parallel views images and the correspondence denote of the facial features points locations. We choose the top 97.5% PCA eigenvalue as the model parameters. The maximum iterations of AAM in our method are set to 30.

We will exchange it to a seven basic emotion space so as to transfer the emotions to line drawing facial animation.

In Figure 2, the red circles on the face are the tracked facial features. We use 68 features points to define the AAM shape. In which 15 points on the contour of the face, 5 points on each of the eyes, 7 points on the eyebrows and 19 points on the mouth. The top left green grid is the mean shape S_0 of the AAM model that is trained in the pre-processing. The red circles on the green grid are the current shape S points without rotation and scale that shows the offset of S from S_0.

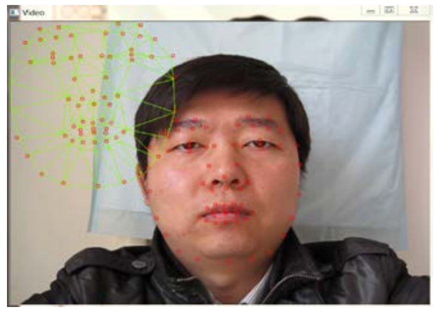

Figure 2. Tracking. The Red Circles on Face is the Features Points. The Top Left Green Grid is the Mean Shape S0 and Red Circles are the Offset of Current Shape S from S0

4. Facial Line Drawing Model

Psychological research has classified six facial expressions which correspond to distinct universal emotions: happiness, anger, disgust, fear, sadness and surprised [11]. The model is based on Scott McCloud's method. Later McCloud defined six basic emotions and considered these basic emotions can be blended to achieve complex emotions [12]. We also defined six basic emotions with line drawings which will be blended to form the final result and using Bézier splines to describe them.

4.1. Face Model

The model includes seven images, a neutral expression and six basic emotions of happiness, anger, disgust, fear, sadness and surprised. Each image consists of a face background (the nose and face contour.) and the facial features(the eyes, eyebrows, lip *etc.*) as shown in Figure 4.

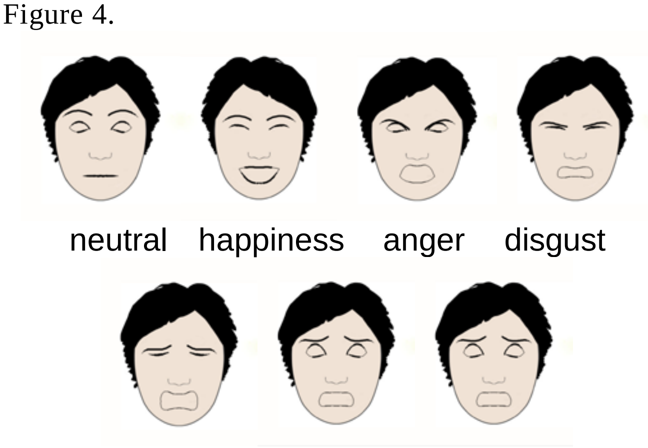

Figure3. Face Model. Each Image Includes Background (Nose, Hair and Contour) and Facial Features (Eyes, Eyebrow, Mouse, etc)

4.2. Facial Lines Definition

The animation figures are shown as the facial lines drawing. On the face, we use splines to describe the facial features such as eyes, eyebrows and mouth lips. Each of the components is consist of several lines.

The splines which we use to describe the facial features are piecewise parametric curves, have two endpoints and some control points in between.

We use the method of Bézier curve to define the strokes and describe the stokes' width at every control point just like Freshman [13], as in (5).

$$\{(x_i, y_i, w_i) \mid 1 \le i \le n\} \quad (5)$$

When drawing, the width of every point is calculated by the segmentation and interpolation of the width recorded in its adjacent control points.

The model is saved into text format and each file contains two parts, the control points' positions with width, and the composition of each curve.

The format of the save model is designed as follow:

```
n_points: 56// number of control points
point 0 365 285 1    //control points position and width
……
point 55 256 480 3
n_lines: 14           //number of lines
line rbow 4 0 1 2 3  // composition of lines
……
```

In the row which starts with point, the number shows the control points, with the order of point identifier, x coordinate, y coordinate, width. In the row which starts with line, the number shows the line name, the count of control points, the control points identifier that comprise it.

This animation figure is defined of 14 lines and all of those lines are quadratic Bézier curves. Each of this figure's eyes is two lines and eyebrows are one. The mouth is more complex which is consists of 8 lines.

Every emotion of the figure is saved as one file and the file contains same points and curves with the same structure. The figure in Figure3 has seven files include of neutral, happiness, anger, disgust, fear, sadness and surprised.

5. Expression Transfer

5.1. Emotion Model

Over the years, many different explanations have been put forward, resulting in various emotion models. Two major approaches are based on several independent dimensions or categorical basic emotions, *e.g.*, [14]. In this paper, we use the basic emotions to express the facial expression and assume every expression could be combined by neutral expression and the six basic emotions of happiness, anger, disgust, fear, sadness and surprised, as in (6),

$$S(x) = S_0 + \sum_{i=1}^{7} \alpha_i E_i(x) \quad (6)$$

Where $S(x)$ is the current AAM shape result and $E_i(x)$ is the shape vector of the neutral and six basic emotions. S_0 is the mean shape of the AAM model. The E_i is defined as the current user's specific emotion shape vector difference to the mean shape vector S_0 in the AAM model. α_i is the ratio of the emotion E_i.

At first, we record seven basic emotion images of the user on near frontal parallel views. Seven basic emotion images that we captured are neutral, happiness, anger, disgust, fear, sadness and surprised. Each of the emotions has one image which is one to one correspondence to the defined animation figures line drawings, see in Figure4. The seven human images in left column near the input frame are the performed emotions. Then fit each of the images with AAM and after that we will get a set of AAM parameters. These parameters are E_i in the equation (6) of current user.

AAM model will introduce rotation and scale after fitting the target image, but what we need is the expression changes, that is only relative to the variation of neutral expression, and facial rotation and scale are not necessary, so after the fitting process by equation (1) of getting the feature point set, rotation and scale matrix do not multiply on the result.

The E_i is the AAM fitting shape result of the basic emotion without rotation and scale which contains 136 values. One of the emotion shapes, Anger, is shown in table 1. So the equation (6) is over determined equation. In Section 2.2, we get the vector of feature points. Because there are seven emotion images and the corresponding points of each individual, we can resolve the approximate solution of the equation (6) by singular value decomposition (SVD) with the known E_n and E_i.

Table 1. The Shape of Anger Emotion

Anger shape
-1.92236 66.8227 -1.7048 104.714 0.34762 146.966
10.2814 195.851 25.8888 224.931 44.9988 258.235
77.837 274.854 132.879 281.434 171.535 274.09 201.131
260.699 222.356 230.321 238.49 196.911 247.645
154.202 245.539 96.8205 242.757 61.1757 215.385
23.9634 195.808 -1.37799 163.823 -1.05573 142.801
14.052 168.909 14.4292 193.024 18.0952 26.0767
35.3795 43.2284 12.9261 74.3477 5.27602 98.542 18.919
70.8567 24.356 47.0508 28.5893 32.4173 71.9019
64.4227 49.5252 94.5389 65.5623 66.1767 73.7709
65.2745 62.7324 212.031 59.6125 180.647 43.4693
151.125 63.305 182.235 67.9855 181.46 55.8205 105.369
62.8164 98.7767 92.6099 86.4826 126.981 104.834
142.905 127.925 143.159 149.319 140.564 164.463
120.015 144.729 90.2076 138.685 59.9698 108.409
132.44 145.206 129.642 78.0704 187.092 93.2955
169.479 108.223 163.207 125.37 165.367 140.807
161.533 158.031 165.776 178.522 182.465 164.574
201.317 148.747 210.188 128.633 211.904 107.068
211.092 92.3631 205.004 105.985 196.223 127.708
197.444 148.23 193.978 145.46 172.873 125.747 172.98
105.76 174.931 126.382 184.496 125.647 114.935

5.2. Transfer

After resolved equation (6) in Section 5.1, with the solved α_i, we used the similar equation (7) of the face line drawing model to get the animation result.

$$A(x) = A_0 + \sum_{i=1}^{7} \alpha_i A_i(x) \quad (7)$$

Figure 4 shows the transfer of one frame, when this frame comes, first our method gets the features points of the face with AAM and then decomposes it into the ratios of seven basic emotions by equation (6). Table 2 shows the ratios of every basic emotion. The number of ratio could be in [-1, 1]. With the same ratios, the basic emotion line drawings are combined to form the final animation figure.

The high and low value of the ratios may not indicate the current emotion composition of the human face, but the closest emotion composition on the basis. The rations only show how to use the feature points of basic emotion to constitute the current track facial feature points. Because of the correspondence of the basic emotion to the basic line drawings, the result could be constituted by the same ratios.

In Figure 4 we could see the ratios of the emotions are normalized and the sum of the ratios is 1.

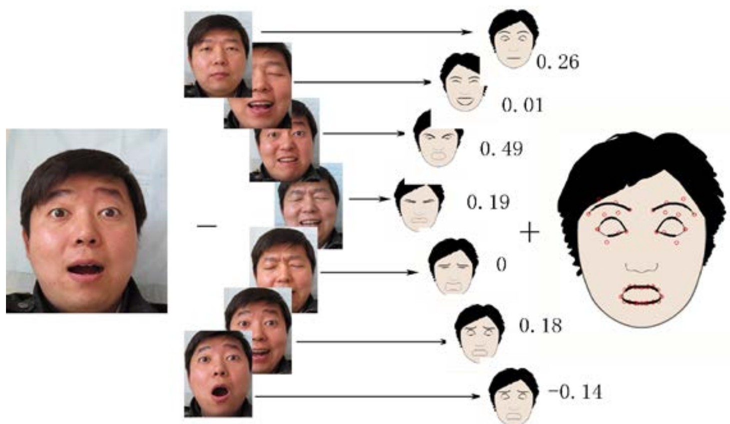

Figure4. The Transfer of One Frame, the Left is Input Frame and the Right is the Output Line Drawing. The Seven Human Images near the Input Frame are Performed Emotions to Get the Ei of Equation (6). The Seven Line Drawings near the Output are the Correspondence Predefined Figure Emotions. The Values are the Ratios of the Emotions

Table 2. The Transfer Ration of the Frame of Figure 4

Emotion	*Ratio*
Neutral	0.26
happiness	0.01
anger	0.49
disgust	0.19
fear	0
sadness	0.18
surprised	-0.14

Some ratios of emotions may be negative that are caused by the solved method of the linear equation and the corresponding line drawing emotion will also be negative. So that does not affect the final result.

6. Experiment Result

We build a tool to define the Bézier curves of facial line drawing model, deal with the video and demonstrate the result. It could also allow the user to edit the coefficients of the basic emotions on the result. The interface is showed in Figure 5.

With the help of the tool, we can adjust the control points of the line drawings. The red dots on the face in Figure 5 are control points of each Bézier curves. Some points will be shared by the adjacent curves.

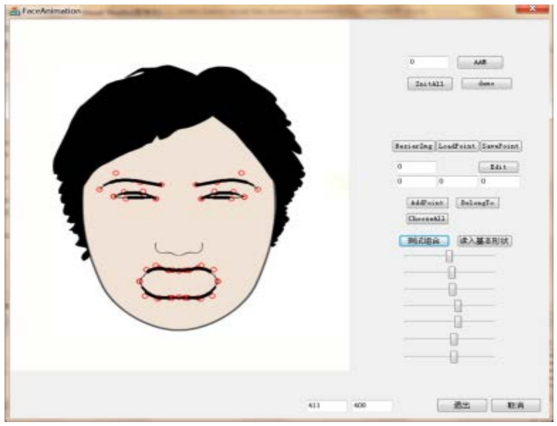

Figure 5. The Interface of System

For feature point training, we build the AAM model by the tools am_tools_win_v3 downloaded from Tim Cootes's web site of Manchester. The line drawings are obtained from the web of grimace project. All the experiments are implemented with C++ and OpenCV2.1 on PC with Intel(R) CPU, 2.33GHz, 2G Memory. We tested our method on several video; each video is 640x480 pixels. Several results are shown in Figure6.

In the experiment, the training set of AAM are the IMM Face Database [14] (http://www2.imm.dtu.dk/~aam/datasets/imm_face_db.zip) and some front facial images of Asian which are manually labeled. We manually label four people, each of them have seven images including neutral expressions and six basic emotions. The facial features points are defined as 68 points like the red circles on the face of Figure3.

The test video sequence has 1269 frames with expression changes, almost frontal parallel views. The individual of the video has been captured seven images of the basic emotions including neutral expression and the six basic emotions of happiness, anger, disgust, fear, sadness and surprised. Each emotion has one image and those images are fitted by AAM to get the AAM model parameters. Total processing time of this sequence is 126.4s which include the I/O operation for writing two images file at each frame to the disk.

Table 3 shows the cost time of this test video sequence, the total processing time is 126.4s and 99.6ms per frame. In which the AAM fitting time is 117.7s and 92.8ms per frame and the transfer and draw time is 8.7s and 6.8ms per frame. Most of the time is costed on AAM fitting because of the faces of the train images and waiting-for-fitted image in the video occupies almost 1/3 of the 640x480 pixels area. The appearance of AAM model is large which causes the fitting process slow on resolving equation (3) because the AAM fitting method is sensitive to the appearance size.

Table 3. The Cost Time of Experiment

	processing time	*AAM fitting time*	*Transfer and draw*
*Total frames:*1269	126.4s	117.7s	8.7s
Per frame	99.6ms	92.8ms	6.8ms

The result shows in the Figure 6. Some result images from the sequence. The singular rows are origin input images from the video sequence and the even rows are output animation figure which has a background face and eyes, eyebrows, lips drawn by the line drawings.

In the result, the fitting coefficient may be a negative value, the value distribution are between [1.1]. From the result we can see that the line drawings of result animation are similar to the input video.

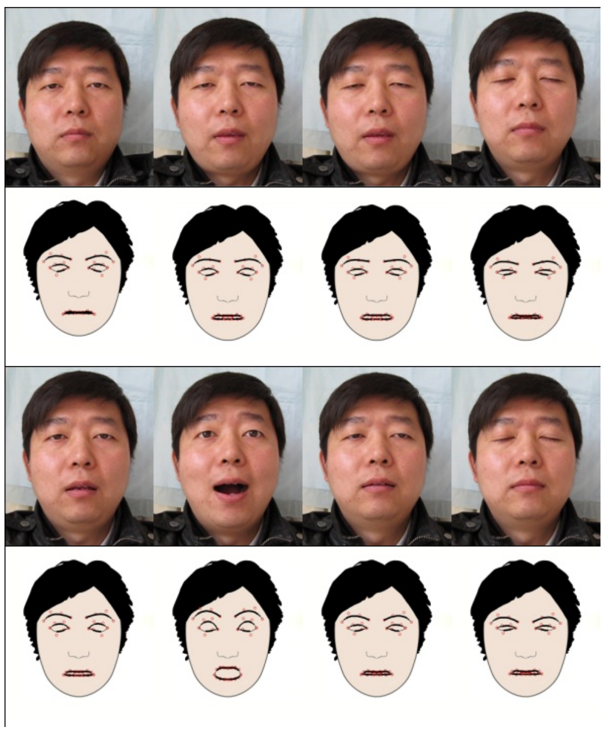

Figure 6. The Animation Result. Images in Upper Row with Human Face are some Frames of the Input Video and that on Lower Row are the Corresponded Animation Result, Red Circles on Faces are Control Points of Bézier Curves

7. Conclusion

We present a method for facial line drawing animation generated from vision by using active appearance models. The line drawing models are defined with Bézier curves and combined to express the expression of live video. The facial line drawing animation only uses seven basic shape parameters that can be used in communication. It achieves a smooth effect of artistic elements in the facial animation, and maintains the expression change of the users. The transfer ratios are only related to the input human face and independence to the output animation figure so that could be transferred to other figures with the same composing emotions, even to the animal face. The number of the control curves could also be increased to get more details. The ratios are only related the input face shape and the basic emotion shape and resolve the approximate solution that is not necessary to find the true expression formation. That make the method is easy to use without careful facial expression analysis.

Since we use AAM algorithm to track the feature points on human face, there are some limitations come from AAM algorithm feature tracking problem, such like when the human face not facing front very well. Another problem is the coherence between two consequent frames. Since the tracked feature points of the two frames may have perturbations on the human face, this will make the control points a little motion on the face between two frames. The third is there should have some wrinkles on specific emotions. All these could be the future work of the method.

Acknowledgements

We would like to thank Tim Cootes for providing calibration tools in his web site and Oliver Spindler and Thomas Fadrus for providing original painting of this grimace in their web site.

This work was supported by NSFC Project (61272242), NSFC-Guangdong Joint Fund Project (U1035004), National High-Tech Program Project (2012AA01A306) and Institute Support Innovation Project of Jinan Administration of Science &Technology (No. 201202012), Jinan star science and technology plan (No. 20120104).

References

[1] Koshimizu, H., Tominaga, M., Fujiwara, T., Murakami, K, IEEE T SYST MAN CY. 6, 294 **(1999)**.
[2] Ching-Ting Tu and Jenn-Jier James Lien., IEEE T SYST MAN CY B, 40, 1158 **(2010)**.
[3] Chai J.-X., Xiao J., Hodgins J., Vision-based control of 3d facial animation, Proceedings of the ACM SIGGRAPH/Eurographics symposium on Computer animation, **(2003)**July. 193–206; San Diego, California.
[4] Taehyun Rhee, Youngkyoo Hwang, James Dokyoon Kim, Changyeong Kim. Real-time Facial Animation from Live Video Tracking, Proceedings of the ACM SIGGRAPH/Eurographics symposium on Computer animation, **(2011)** August 215–224; Sydney, Australia.
[5] Eun-Jung Lee, Ji-yong Kwon, In-Kwon Lee., Comput Animat Virt W, 18, 279 **(2007)**.
[6] Yang Yang, Nanning Zheng, Yuehu Liu, Shaoyi Du, Yuanqi Su, Yoshifumi Nishio, Signal Processing, 91, 2465 **(2011)**.
[7] M. Obaid, D. Lond, R. Mukundan, and M. Billinghurst. Facial Caricature Generation Using a Quadratic Deformation Model, Proceedings of the International Conference on Advances in Computer Enterntainment Technology, **(2009)** June 285-288, Dortmund, Germany.
[8] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor, Ieee T Pattern Anal, 23, 681, **(2001)**.
[9] Paul Viola, Michael J.Jones., Int. J. Comput. Vis. 57, 137 **(2004)**.
[10] K. Messer, J. Matas, J. Kittler, J. Luettin and G. Maitre. XM2VTSbd: The Extended M2VTS Database, Proceedings of 2nd Conference on Audio and Video-base Biometric Personal Verification, **(1999)** March 72-77; Washington, DC, USA
[11] P. Ekman, W. Friesen, and P. Ellsworth. Emotion in the Human Face: Guidelines for Research and an Integration of Findings, Pergamon, **(1972)**.
[12] S. McCloud. Making Comics: Storytelling Secrets of Comics, Manga and Graphic Novels, HarperCollins Publishers, New York**(2006)**, Chapter 2, pp.80-102,.
[13] Freeman W.T., Tenenbaum J.B., Pasztor E.C., An example-based approach to style translation for line drawings", Technical Report TR99-11, MERL, Cambridge **(1999)**.
[14] P. Ekman. Basic emotions. Handbook of cognition and emotion, John Wiley & Sons**(1999)**, pp. 45–60,.
[15] M.M. Nordstrøm, M. Larsen, J. Sierakowski and M.B. Stegmann, The IMM Face Database - An Annotated Dataset of 240 Face Images, technical report, Informatics and Mathematical Modelling, Technical Univ. of Denmark, DTU**(2004)**.

Authors

Qingxiang Wang received M.Sc. degrees in computer software and theory in 2000 and 2006 from Donghua University, Shanghai, China. Currently, he is a lecturer in the School of Information, QILU University of Technology, Jinan, China and a Ph.D. student in Human-computer interaction and virtual reality, Shandong University since September 2009. His research interests include computer vision and character animation.

Changhe Tu is a Professor of School of Computer Science and Technology at Shandong University. He obtained his PhD in 2003 from Shandong University. His research interests include geometric modeling and processing, computer aided geometric design, computer graphics and computer animation. His publications cover such topics as analysis and computation of intersections between two quadric surfaces, distance computation of curved surfaces, mesh processing, and facial expression tracking. He is a member of CCF (China Computer Federation) and ACM.

Xiaoqiang Ren received B.Sc. and M.Sc. degrees in computer science and technology in 2000 and 2008, respectively from Shandong Polytechnic University, Jinan, China and Shandong University of science and technology. Currently, he is a lecturer in the School of Information, QILU University of Technology, Jinan, China. His research interests are data mining and image processing.