Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 706159, 6 pages
http://dx.doi.org/10.1155/2014/706159

Research Article

A New Method of 3D Facial Expression Animation

Shuo Sun[1] and Chunbao Ge[2]

[1] *Department of Mathematics, Tianjin Polytechnic University, Tianjin 300387, China*
[2] *Shengshi Interactive Game, Beijing 10010, China*

Correspondence should be addressed to Shuo Sun; sunshuo@tjpu.edu.cn

Received 4 March 2014; Accepted 6 April 2014; Published 20 May 2014

Academic Editor: Li Wei

Animating expressive facial animation is a very challenging topic within the graphics community. In this paper, we introduce a novel ERI (expression ratio image) driving framework based on SVR and MPEG-4 for automatic 3D facial expression animation. Through using the method of support vector regression (SVR), the framework can learn and forecast the regression relationship between the facial animation parameters (FAPs) and the parameters of expression ratio image. Firstly, we build a 3D face animation system driven by FAP. Secondly, through using the method of principle component analysis (PCA), we generate the parameter sets of eigen-ERI space, which will rebuild reasonable expression ratio image. Then we learn a model with the support vector regression mapping, and facial animation parameters can be synthesized quickly with the parameters of eigen-ERI. Finally, we implement our 3D face animation system driving by the result of FAP and it works effectively.

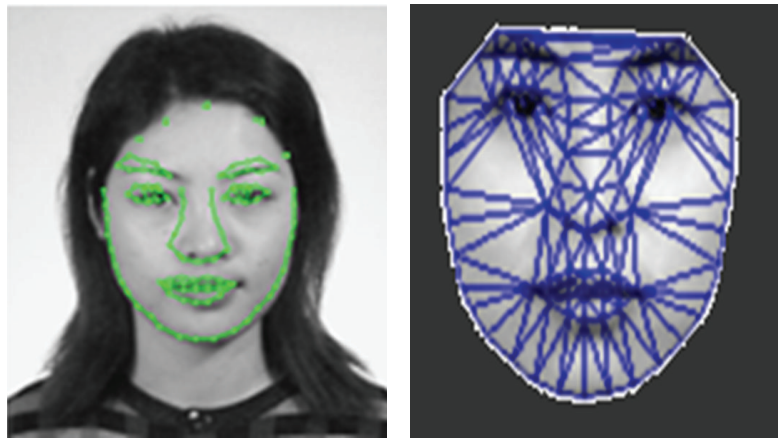
1. Introduction

Facial animation is one alternative for enabling natural human-computer interaction. Computer facial animation has applications in many fields. For example, in the entertainment industry, realistic virtual humans with facial expressions are increasingly used. In communication applications, interactive talking faces not only make the interaction between users and machines more fun, but also provide a friendly interface and help to attract users [1, 2]. Among the issues concerning the realism of synthesized facial animation, humanlike expression is critical. But how to analyze and comprehend humanlike expression is still a very challenging topic for the computer graphics community. Facial expression analysis and synthesis are an active and challenging research topic in computer vision, impacting important applications in areas such as human-computer interaction and data-driven animation. We introduce a novel MPEG-4 based 3D facial animation framework and the animation system driving by FAP that produced from camera videos. The MPEG-4 based framework has the advantages of currency and few data [3–5]. First the system takes 2D video input and recognizes the face area. Then the face image was transformed to ERI [6]. Next, a simple ERI's parameterized method is adopted for generating an FAP driving model by the support vector regression and it is a statistic model based on MPEG-4. The results of FAPs can be used to drive the 3D face model defined by MPEG-4 standard.

The remainder of the paper is organized as follows. We present the related work in Section 2. We then describe how to preprocess video data in Section 3. Section 4 presents how to construct the eigen-ERI. Section 5 describes the extraction of the FAP. In Section 6 we propose a novel SVR-based FAP driving model. Finally, we show the experiments and conclude the paper in Sections 7 and 8.

2. Related Works

Recently, realistic facial animation has become one of the most important research topics of computer graphics. Many researchers focus on the nature of the facial expression. In [7, 8], a sample-based method is used to make photorealistic expression in detail. In particular, Chang et al. [9] implement a novel framework for automatic 3D facial expression analysis in video. Liu et al. proposed an ERI-based method to extract the texture depth information of the photo in [6], and Figure 2 shows the ERI of the frown expression. For reflecting the changes of a face surface, Tu et al. [10] use the gradients of the ratio value at each pixel in ratio images and apply it



(a) AAM (b) Mesh triangle

Figure 1: Mesh model of neutral expression.

Figure 2: ERI of frown expression.

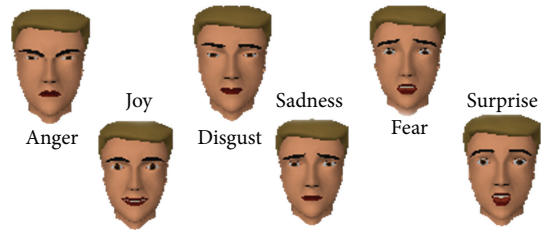


Figure 3: Six basic facial expressions.

to 3D model. In addition, a parameterized ERI method was proposed by Tu et al. [10] and Jiang et al. [11].

Zhu et al. [12] used a method of SVR-based facial texture driving for realistic expression synthesis. In Zhu's work, a regression model was learned between ERI's parameters and FAPs. According to the inputs of FAPs, the model will forecast the parameters of ERIs. Then a reasonable facial expression image was generated with ERIs. On the contrary, our method is to forecast FAPs from the parameters of the eigen-ERIs; furthermore we realize a 3D face animation system driving by FAPs.

In this paper, we realize a 3D face animation system that can generate realistic facial animation with realistic expression details and can apply in different 3D model similar with human. The main problems of our facial animation system are the extraction of the ERI from the camera video and learning the SVR-based model, furthermore building an FAP driving 3D facial expression animation system. Next section will introduce these.

3. Video Data Preprocessing

3.1. Video Data Capture. For the robustness of the algorithm, we get video data in normal light environment, and video equipment is often used to support continuous capture PC digital camera. Then we set the sampling rate of 24 frames per second and the sampling resolution of 320∗240, and the total 1000 sampling frame was captured, of which 200 expressions defined key frame.

In this paper, we adopted the method of the marked points in face to extract facial motion data, and the most difference with other methods is that our method is based on the MPEG-4 standard (Figure 1(a)). The advantage of the standard is that you can share daters, and the data can be used with any standards-based grid. By marked points, we can get more accurate facial motion data, while blue circle chip can be pasted in any place, including eyebrows and eyelids, and have no effects on their movement, which makes it possible to get the whole facial motion data and satisfy with the processing of the next step.

In order to obtain experimental data, first we make a coarse positioning through the blue calibration point using face detection tools and have calibration and reduction transformation operations on the sample data. Then we design a face mesh model in Figure 1 to describe the geometry of the facial animation. Mesh vertices were mapped automatically using active appearance model (AAM) and manual fine tuning to meet subsequent demand for data extraction.

After obtaining the data of the texture and feature points, all texture will be aligned to the average model.

3.2. Features Points. According to MPEG-4, we defined six basic expressions types (see Figure 3) and captured them from video camera. Twenty-four key frames demanded for training and one nonkey frame used for testing per expression type are extracted with resolution of 320∗240 pixels. In Figure 1(a), the

68 feature points are marked automatically with AAM [13]. In our experiment, we adopted the 32 feature points belonging to the FPs in MPEG-4.

4. Computation of Eigen-ERI

4.1. The Computation of the ERI. For building general ERI's model, we used a large of frontal and neutral expression face as sample library. And we extract the outline of the face features by the method of the active appearance model (AAM). Then the following formula was defined to compute the average face's ERI, as the standard shape model:

$$F_m = \frac{1}{N}\sum_{i=1}^{N} F_i. \tag{1}$$

Here F_m means average face, and F_i means any one face $i = 1, 2, \ldots, N$, and N means the number of the frontal and neutral expression faces.

For getting facial expression details, we compute the ERI from all key frames sequences continuously. The first frame is denoted by F and regarded as a neutral face image. F_i $(1 \le i < n)$ denoted the rest of expressive face images samples, where n is the total number of key frames. Each expressive face sample F_i will be aligned to F. Then, the ERI of the sample sequence can be computed as follows:

$$R_i(u, v) = \frac{F_i(u, v)}{F(u, v)}. \tag{2}$$

Here, the (u, v) denote the coordinates of a pixel in the image, and F_i and F denote the color values of the pixels, respectively.

Because the computation of ERI is each point one by one according to [14], the results will be unpredictable if the face features cannot be aligned exactly. So we revise the definition of ERI's computation

$$R(u, v) = \frac{F'(u, v)}{Ave(F(u, v))}. \tag{3}$$

4.2. Eigen-ERI. The matrix of ERI is the 2D gray image $I(x, y)$. We represent it with the $W \times H$ dimension vector Γ. The training sets are $\{\Phi_i \mid i = 1, \ldots, M\}$; the M means the sum of the images. The average vector of the all images is

$$\psi = \frac{1}{M}\sum_{i=1}^{M} \Gamma_i. \tag{4}$$

The difference value of the ERI's Γ_i and the average of the ERI's ψ were defined as Φ_i:

$$\Phi_i = \Gamma_i - \psi; \quad i = 1, \ldots, M. \tag{5}$$

The covariance matrix is represented as follows:

$$C = AA^T, \tag{6}$$

where $A = [\Phi_1, \ldots, \Phi_M]$.

Feature face is composed of the orthogonal eigenvector of the covariance matrix.

ERI calculation only takes the M eigenvector corresponding to the largest eigenvalue. The M is decided by the threshold of the θ_λ:

$$J = \min_r \left\{ r \mid \frac{\sum_{i=1}^{r} \lambda_i}{\sum_{j=1}^{M} \lambda_j} > \theta_\lambda \right\}. \tag{7}$$

In this paper, we select the maximal 21 variables in the eigen-ERI space to represent 96% variation information in the sample sets.

5. Extraction of FAP

5.1. Definition of the FAP (Facial Animation Parameter). The FAP is defined by a set of facial animation parameters in MPEG-4 standard. FAP based on the face of small actions is very close to the facial muscle movement. In fact, FAP parameter represents a set of basic facial movements, including the head movement control, tongue, eyes and lips, facial expression and lip can reproduce the most natural move. In addition, like those of humans did not exaggerated facial expressions cartoon, FAP parameter can be traced.

There are six basic expressions types, defined in MPEG-4 (see Figure 3). The MPEG-4 has a total of 68 FAP. The value of FAP is based on FAPU (facial animation parameter unit) as the unit, so that FAP has the versatility. The calculation formulation is

$$\mathrm{FAP}_i = \frac{\mathrm{FP}' - \mathrm{FP}}{\mathrm{FAPU}}. \tag{8}$$

Among them, $i = 3, \ldots, 68$ FP and FP$'$ are neutral, facial feature points on the corresponding parts.

5.2. The Implement of Face Animation Based on FAP. Facial animation definitions table defines three parts. First, the FAP range is divided into several segments. Second, we want to know which grid points are controlled by the FAP in the mesh. Third, we must know the motion factor of the control points in each segment. Each FAP needs to find out which parts of the three parts in facial animation definition table, then according to the MPEG-4 algorithm, calculated by the displacement of the FAP control all grid points. For a set of FAPs, each FAP calculated the effective grid point rod size; by the shift-and-add up you will get a vivid facial expression (Figure 3). The concrete implement may refer to [15].

6. SVR-Based FAP Driving Model

6.1. Support Vector Regression (SVR). In a typical regression problem, we are given a training set of independent and identically distributed (*i.i.d.*) examples in the form of n ordered pairs, $\{(x_i, y_i)\}_{i=1}^{n} \subset R^d \times R$, where x_i and y_i denote the input and output, respectively, of the ith training example. Linear regression is the simplest method to solve the regression problem where the regression function is

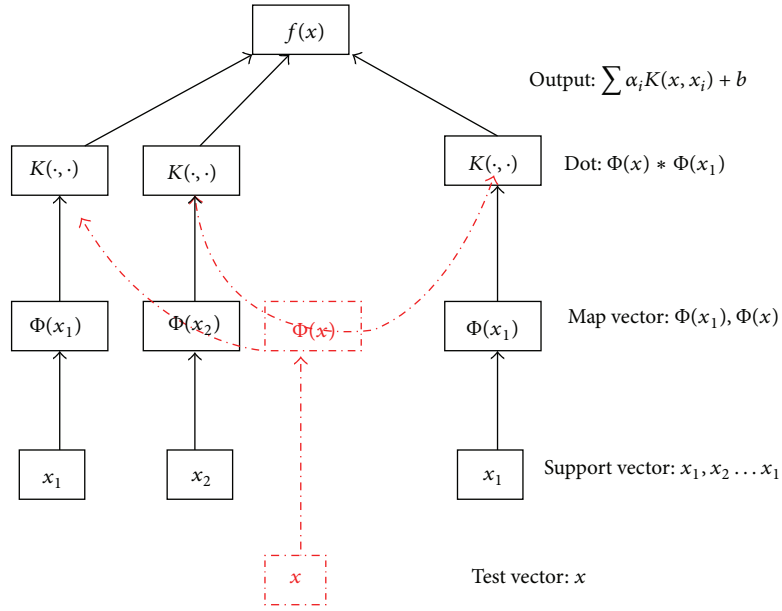


Figure 4: The steps of SVR processing.

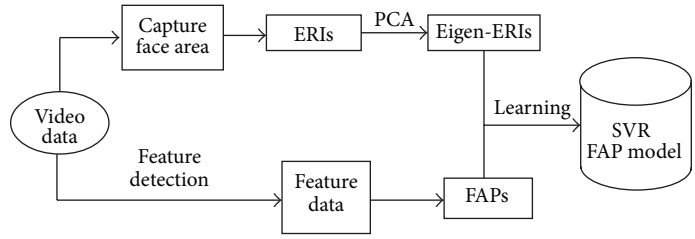


Figure 5: Offline learning progress.

a linear function of the input. As a nonlinear extension, support vector regression is a kernel method that extends linear regression to nonlinear regression by exploiting the kernel trick [16, 17]. Essentially, each input $x_i \in R^d$ is mapped implicitly via a nonlinear regression map $\phi(\cdot)$ to some kernel-induced feature space F where linear regression is performed. Specifically, SVR learns the following regression function by estimating $w \in F$ and $w_0 \in R$ from the training data:

$$f(x) = \langle w, \phi(x) \rangle + w_0, \tag{9}$$

where $\langle *, * \rangle$ denotes the inner product in F. The problem is solved by minimizing some empirical risk measure that is regularized appropriately to control the model capacity.

One commonly used SVR model is called ε-SVR model, and the ε-insensitive loss function

$$|y - f(x)|_\varepsilon = \max\{0, |y - f(x)| - \varepsilon\} \tag{10}$$

is used to define an empirical risk functional, which exhibits the same sparseness property as that for support vector classifiers (SVC) using the hinge loss function via the so-called support vectors. If a data point x lies inside the insensitive zone called the ε-tube, that is, $|y - f(x)| \le \varepsilon$, then it will not incur any loss. However, the error parameter $\varepsilon \mid 0$ has to be specified a priori by the user. The primal optimization problem for ε-SVR can be stated as follows:

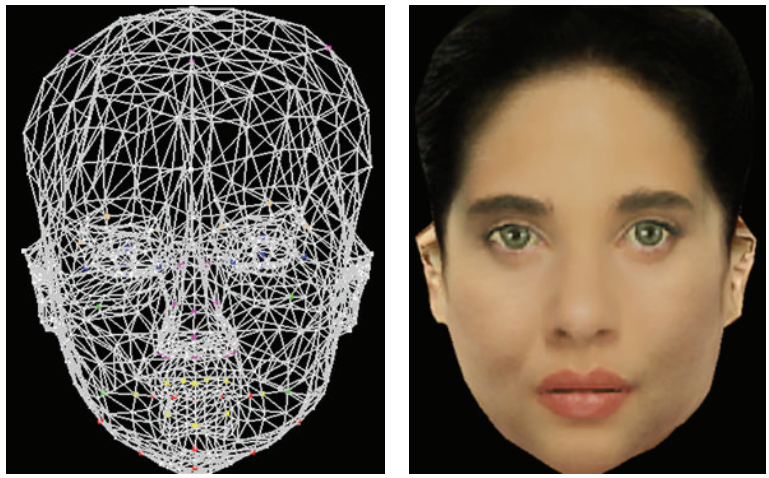
$$\begin{aligned} \min_{W,\xi(*)} \quad & \frac{\lambda}{2}\|W\|^2 + \sum_{i=1}^{n} (\xi_i + \xi_i^*) \\ & y_i - (\langle W, \phi(x_i) \rangle + w_0) \le \varepsilon + \xi_i \\ & (\langle W, \phi(x_i) \rangle + w_0) - y_i \le \varepsilon + \xi_i^* \\ & \xi_i^* \ge 0. \end{aligned} \tag{11}$$

Relatively, (3) transforms into

$$f(x) = \sum_{i=1}^{l} (-\alpha_i + \alpha) K(x_i, x) + b. \tag{12}$$

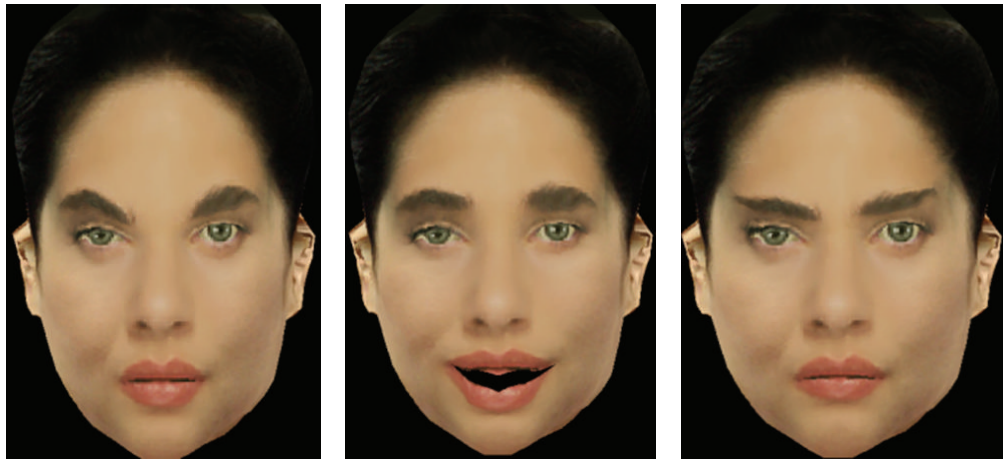
Here, α_i denotes Lagrange multiplier, $K(x_i, x)$ is kernel function, and $f(x)$ is decision function. For predicting, it is only a dot product operation and costs very low in the real time. Figure 4 shows the steps of the SVRM mapping ERI to FAP parameters.

6.2. SVR-Based Algorithm. According to the above theory, an FAP driving model for its every parameter is built. Given



(a) 3D mesh (b) Synthesis face

Figure 6: 3D facial expression animation system.



(a) Anger (b) Joy (c) Disgust

(d) Sadness (e) Surprise

Figure 7: Five basic 3D expressions' animation.

a set of training data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\} \subset X \times Y$, where X denotes the space of the ERI and $x_1 \in R$ is the ERI's parameter and Y denotes the space of the feature of FAP and $y_1 \in R$ is an FAP. We regard it as a regressive problem and represent it by the support vector regression method, including ε-SVR [16] and v-SVR [17]. The offline learning progress is as Figure 5.

An offline learning process was showed in Figure 5. In the above section, we have got the learning parameters of the FAPs and ERIs. Then we can get a regressive model from the ERI's parameters to the FAPs vectors through the SVR model. In a statistical sense, the results of FAPs can reflect the expression relative with the ERI.

There are several steps used to explain how the facial expression animation is driven by video camera.

Step 1. Establish the MPEG-4 based 3D facial expression animation system driving by FAP; see Figure 6.

Step 2. Capture and detect face image from video camera, and compute its ERI.

Step 3. Forecast the FAPs of the current frame according to its ERI.

Step 4. Compute the motion of the feature points in the face mesh based on the new FAP and animate the 3D human face mesh.

7. Experiment Results

We have implemented all the techniques described above and built an automatic 3D facial expression animation system on Windows environment. The result was showed in Figure 7. We represent six basic expressions driving by FAP which come from the forecast through ERI.

8. Conclusion

In this paper, we realize a 3D face animation system that can generate realistic facial animation with realistic expression details and can apply in different 3D model similar to human. It is capable of generating the statistical realistic facial expression animation while only requiring simply camera device as the input data and it works better in any desired 3D face model based on MPEG-4 standard.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported partly by NSFC Grant no. 11071279.

References

[1] E. Cosatto, *Sample-based talking-head synthesis [Ph.D. thesis]*, Swiss Federal Institute of Technology, 2002.

[2] I. S. Pandzic, "Facial Animation Framework for the web and mobile platforms," in *Proceedings of the 7th International Conference on 3D Web Technology (Web3D '02)*, pp. 27–34, February 2002.

[3] S. Kshirsagar, S. Garchery, and N. Magnenat-Thalmann, "Feature point based mesh deformation applied to MPEG-4 facial animation," in *Proceedings of the IFIP TC5/WG5.10 DEFORM'2000 Workshop and AVATARS'2000 Workshop on Deformable Avatars (DEFORM '00/AVATARS '00)*, pp. 24–34, Kluwer Academic Press, 2001.

[4] F. Parke and K. Waters, *Computer Facial Animation*, A. K. Peters, Wellesley, Mass, USA, 1996.

[5] "MPEG-4 Overview, ISO/IEC JTC1/SC29N2995," 1999, http://web.itu.edu.tr/~pazarci/mpeg4/MPEG_Overview1_w2196.htm.

[6] Z. Liu, Y. Shan, and Z. Zhang, "Expressive expression mapping with ratio images," in *Proceedings of the Computer Graphics Annual Conference (SIGGRAPH '01)*, pp. 271–276, August 2001.

[7] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing realistic facial expressions from photographs," in *Proceedings of the Annual Conference on Computer Graphics (SIGGRAPH '98)*, pp. 75–84, July 1998.

[8] T. Ezzat and T. Poggio, "Facial analysis and synthesis using image-based models," in *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition*, pp. 116–120, October 1996.

[9] Y. Chang, M. Vieira, M. Turk, and L. Velho, "Automatic 3D facial expression analysis in videos," in *Proceedings of the 2nd International Conference on Analysis and Modelling of Faces and Gestures (AMFG '05)*, 2005.

[10] P.-H. Tu, I.-C. Lin, J.-S. Yeh, R.-H. Liang, and M. Ouhyung, "Expression detail for realistic facial animation," in *Proceeding of the Computer-Aided Design and Graphics (CAD '03)*, pp. 20–25, Macau, China, October 2003.

[11] D.-L. Jiang, W. Gao, Z.-Q. Wang, and Y.-Q. Chen, "Realistic 3D facial animations with partial expression ratio image," *Chinese Journal of Computers*, vol. 27, no. 6, pp. 750–757, 2004.

[12] W. Zhu, Y. Chen, Y. Sun, B. Yin, and D. Jiang, "SVR-based facial texture driving for realistic expression synthesis," in *Proceedingsof the 3rd International Conference on Image and Graphics (ICIG '04)*, pp. 456–459, December 2004.

[13] Y. Du and X. Lin, "Emotional facial expression model building," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 2923–2934, 2003.

[14] W. U. Yuan, "An algorithm for parameterized expressionmapping," *Application of Computer Research, Journal*. In press.

[15] D. Jiang, Z. Li, Z. Wang, and W. Gao, "Animating 3D facial models with MPEG-4 facedeftables," in *Proceedings of the 35th Annual Simulation Symposium*, pp. 395–400, 2002.

[16] V. N. Vapnik, *Statistical Learning Theory*, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, New York, NY, USA, 1998.

[17] B. Scholkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, Mass, USA, 2002.