Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Data-driven facial animation based on manifold Bayesian regression*

WANG Yu-shun, ZHUANG Yue-ting, WU Fei
(*School of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*)
(*The Microsoft State Education Ministry Key Laboratory, Zhejiang University, Hangzhou 310027, China*)
E-mail: yswang@cs.zju.edu.cn; yzhuang@cs.zju.edu.cn; wufei@cs.zju.edu.cn
Received June 5, 2005; revision accepted July 22, 2005

Abstract: Driving facial animation based on tens of tracked markers is a challenging task due to the complex topology and to the non-rigid nature of human faces. We propose a solution named manifold Bayesian regression. First a novel distance metric, the geodesic manifold distance, is introduced to replace the Euclidean distance. The problem of facial animation can be formulated as a sparse warping kernels regression problem, in which the geodesic manifold distance is used for modelling the topology and discontinuities of the face models. The geodesic manifold distance can be adopted in traditional regression methods, e.g. radial basis functions without much tuning. We put facial animation into the framework of Bayesian regression. Bayesian approaches provide an elegant way of dealing with noise and uncertainty. After the covariance matrix is properly modulated, Hybrid Monte Carlo is used to approximate the integration of probabilities and get deformation results. The experimental results showed that our algorithm can robustly produce facial animation with large motions and complex face models.

Key words: Facial animation, Manifold, Geodesic distance, Bayesian regression
doi:10.1631/jzus.2006.A0556 **Document code:** A **CLC number:** TP391

INTRODUCTION

Realistic facial animation is highly important in the computer graphics field as it is an essential facility for human-computer interface and virtual reality and is also a difficult task because there are so many non-rigid motions besides rigid motion of heads as expression changes. Instead of modelling all the complicated facial motions, data-driven facial animation just exploits facial motion data captured in real scenes. Most motion capture systems rely upon the placement of markers on the surface of the face, and the use of vision algorithms to track the movement of those points over time. The tracked feature points are then used to "interpolate" the whole facial mesh. Traditional scattered data interpolation techniques such as Radial Basis Functions (RBF) or B-splines are effective for mesh deformation and reconstruction problems with a small part of data loss (Lee *et al*., 1997; Noh *et al*., 2000). However, when deforming a face mesh using these methods, significant artifacts may occur because these methods take advantage of only the 3D positions of the mesh vertices but throwing away the information on edges. For example, when the markers on the outer contour of lips move outwards to indicate opening mouth, the lips of the deformable model may be stretched as if they were becoming thicker instead of mouth being opened. On the other hand, the noises in the data, which are inevitable in motion capture and 3D scans, are also not properly considered in such techniques.

This paper proposes a solution to data-driven facial animation named manifold Bayesian regression. Our goal is to couple temporally dense facial motion data and a spatially dense static model to provide high resolution in both temporal and spatial domain. Facial motion capture data indicates the movement of

* Project supported by the National Natural Science Foundation of China (No. 60272031), the National Basic Research Program (973) of China (No. 2002CB312101) and the Technology Plan Program of Zhejiang Province (No. 2003C21010), China

marker points on the face mesh. Bayesian regression is like an interpolation method which is trying to find the movement of other points and additionally to smooth out noise. The "interpolated" movement is affected by both the movement of markers and the distance between the point itself and the markers. Euclidean distance is mostly used to measure the nearness between two certain vertices. However, it provides little knowledge about the connectivity of mesh vertices, which is essential in the deformation of meshes. In this paper, the geodesic manifold distance (Tenenbaum *et al*., 2000) is used to replace the Euclidean distance as a novel distance metric in facial deformation algorithms. A static face model can be regarded as a manifold, which is a topological space that is locally Euclidean. The geodesic distance, or the shortest path, is a widely accepted concept in manifold learning to provide spatial understanding of nonlinear topology. We put facial animation into the framework of Bayesian regression. Bayesian approaches provide an elegant way of dealing with noise and uncertainty.

After briefly reviewing related work in Section 2, we introduce the details of our algorithm in Sections 3 and 4. Convincing experimental results are shown in Section 5. Conclusions are given in Section 6.

RELATED WORK

After the pioneering work of Williams (1990) who creatively tracked facial motion from video streams for animating faces, the idea of video based animation became a stream of animation techniques (Pan *et al*., 2000; Zhuang *et al*., 1999; Liu *et al*., 1999) and various excellent researches were carried out to apply facial motion data to realistic facial animation.

Example based approaches (Na and Jung, 2004; Pyun *et al*., 2003; Joshi and Tien, 2003; Blanz *et al*., 2003) use motion vectors to estimate the deformation parameters or blending shape coefficients. Predefined morph targets are then blended with respect to the estimated parameters or coefficients. These approaches are attractive for their stable and accurate results. Nevertheless, the processes of handcrafting of morph targets themselves are expensive and time-consuming.

Noh and Neumann (2001) proposed a retargeting method to map the displacements of vertices on source model to target meshes. The motion displacements are scaled and rotated with respect to the local detail geometry of source and target meshes for pre-processing. This method works well with the mapping between dense source models and similar target meshes. However, animation by mapping sparse facial motion data to static face models remains unsettled.

Guenter *et al*.(1998) captured both 3D geometry and texture information for facial animations. Photorealistic facial expressions can be reconstructed by deforming the underlying face model with respect to the captured data. The generated head highly resembles the performer. Therefore it is not suitable for animating a given static face model.

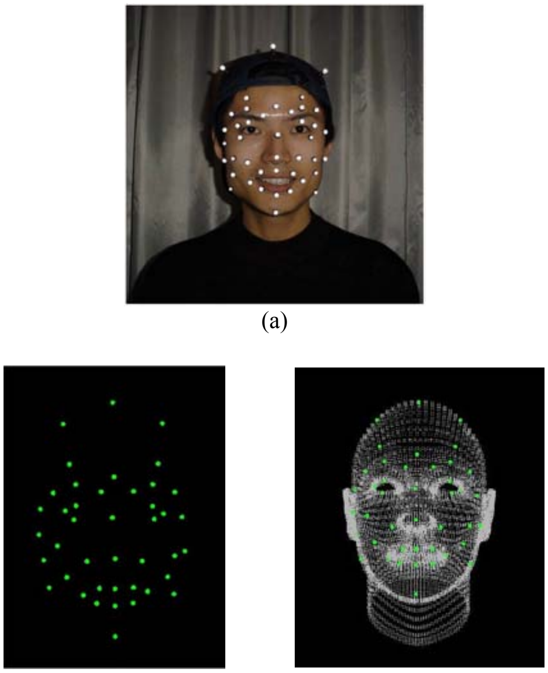
Sánchez Lorenzo *et al*.(2003) used facial motion capture data as control points to deform a face mesh. Reformulation of free-form deformation is applied in which the precise control points are difficult to decide for complex face meshes. They also need manual region segmentation to indicate mesh discontinuities which may be difficult and tedious for a new face model.

DATA-DRIVEN FACIAL ANIMATION

Motion data pre-processing

Three dimensional facial motion data may be generated by motion capture devices or reconstructed from two dimensional motion data tracked in video streams. The face models can be obtained via 3D scanner or exported from modelling software. Highly realistic faces can be modelled from video or images (Blanz *et al*., 2003). The face models and facial motion data may be from notably different scenarios and must be aligned in the first step.

Approximate affine transformations are computed and applied to facial motion data at each frame with respect to the correspondences between markers and face mesh vertices so as to align the three dimensional proportion and alleviate structural disparity between the motion data and the face model, as shown in Fig.1. In motion capture context, several markers will usually be added on the top of head for the purpose of global motion estimation, so that separation of the rigid movement and non-rigid expressions can be faster.



(a) (b) (c)

Fig.1 Correspondences between markers and head model vertices are made manually or with the aid of some heuristic search. (a) Facial motion performer with 42 markers (Including 3 markers for global motion estimation); (b) Facial motion data captured by MoCap device; (c) Deformable head model (in vertex rendering context) with markers mapped onto it

The concept of warping kernels was first proposed by Williams (1990) as deformation initiators or feature points to drive facial animations. In this paper, we refer to the facial motion markers which have been already mapped onto the neutral mesh as warping kernels to provide pivots of regressions. The problem of facial deformation can be formulated as estimating data approximate functions $\boldsymbol{t}=F(\boldsymbol{s}, \theta)$ in nonlinear regression with sparse warping kernels, where $\boldsymbol{s}$ is an input vector indicating the original 3D coordinates of a vertex, $\boldsymbol{t}$ is a vector of the deformed results, and θ is the hyper-parameter which must be learned from the sparse warping kernels. The approximate functions are estimated at each frame of facial motion data to generate sequential facial animation.

Geodesic manifold distances

Traditional regression techniques such as neural networks and radial basis functions favour the Euclidean distances for measuring the similarity between vertices in input space. However, the Euclidean distances are not suitable for modelling data distributed on complex geometry and topology such as human faces. As shown in Fig.2, points on different sides of lips are close in the Euclidean metric (straight line segment on the right figure), but far away from each other in its actual topology and kinematics properties.

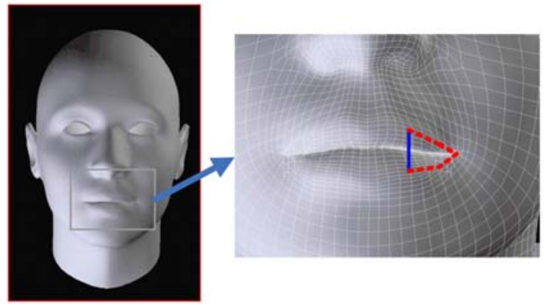


Fig.2 The Euclidean distance (straight line segment on the right figure) is not suitable for modelling distance between points on different sides of discontinuities. The geodesic distance (dot folding line segments on the right figure) is proved to be a better approach to measure distances in globally non-Euclidean spaces

Nonlinear manifold modelling techniques were developed recently during the research on subspace learning and face recognition (Zhang *et al*., 2004; Huang *et al*., 2004), including ISOMAP, locally linear embedding (LLE) and Laplacian eigenmap (LEM). ISOMAP (Tenenbaum *et al*., 2000) performs nonlinear dimensionality reduction by applying multi-dimensional scaling (MDS) on the geodesic distance matrix. LLE (Roweis and Saul, 2000) and LEM (Niyogi and Belkin, 2002) are local algorithms that represent nonlinear manifold by focusing on the preservation of local neighbour structure. Geodesic manifold distance (Tenenbaum *et al*., 2000), which is computed as the shortest path, is a widely accepted concept in the field of manifold learning. A face mesh can be regarded as a manifold, which is a topological space that is locally Euclidean. Points may appear deceptively close by measuring their straight-line Euclidean distance, however far apart on the underlying manifold, as measured by their geodesic manifold distances.

We exploit the geodesic manifold distances to explore the complex geometry and topology of human face models. We take the given face mesh as an undirected graph, where its nodes and arcs are represented by mesh vertices and edges respectively. The sparse matrix of the constructed graph has 22984 non-zero entries in the 5832 by 5832 matrix in the neutral quadrangle mesh, i.e., only 0.068% non-zero entries. The geodesic distances are then computed as the shortest path δ_{ij} between nodes $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ in the graph

$$\delta_{ij}=dijk(\boldsymbol{v}_i, \boldsymbol{v}_j)$$

using Dijkstra algorithm (*dijk*).

Furthermore, we construct a distance map from feature points to all other points on the neutral mesh and stored for later expression generations for acceleration. The geodesic distance can be intuitively adopted to extend traditional regression processes such as radial basis functions (RBF) by replacing Euclidean distance γ with geodesic distance δ in the approximate function, e.g., RBF approximate function with multi-quadrics:

$$F(\boldsymbol{x}_j, \boldsymbol{W}) = \sum_{i=1}^{n}\left[w_i \sqrt{\gamma_{ji}^2 + \sigma_i^2} \right],$$

where $\boldsymbol{x}_j$ is a vector representing the 3D coordinates of the input point, $\boldsymbol{W}=[w_1,w_2,\ldots,w_n]^{\mathrm{T}}$ are parameters, γ_{ji} is Euclidean distance between the feature point and the input point, and σ_i is the stiffness constant that regulates the local or global effects of the feature points.

The geodesic distances can accelerate the generation of facial animation by automatic modelling of the discontinuities, which is one of the key problems of data driven facial animation. Traditional facial animation methods facilitate affection volume masks (Sánchez Lorenzo *et al*., 2003), which use virtual masks to represent different face regions, or other region segmentation techniques to model the discontinuities which may be difficult and tedious for users. Our method simplified the pre-processing of deformable face models and can produce plausible results in both continuous and discontinuous regions on face models, as shown in Fig.3.

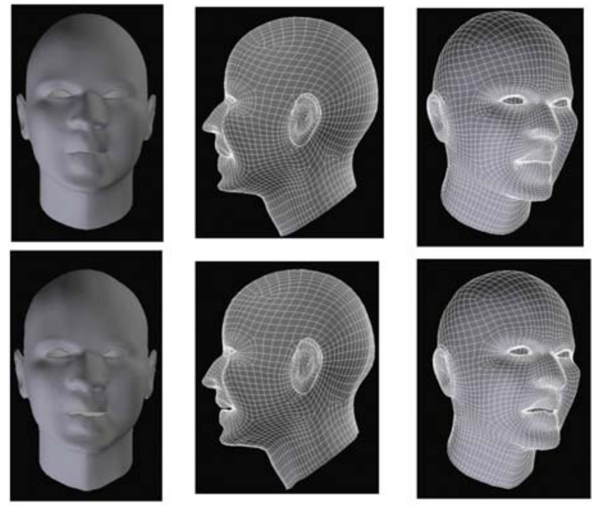


Fig.3 Generated anger expressions from facial motion data. Expression in the first row was generated via regression in Euclidean distance metric, where the lips (one of the most important discontinuities on human face) are stretched instead of opening mouth. The second row expression is from the same process except blending some Geodesic distance metric. The motions of lips are correctly modelled with mouth open

Blending of Euclidean and geodesic distances

A potential error may occur when the meshes contain only sparse triangles or quadrangles, in which cases the geodesic distances may be discrete and result in undesirable artifacts and distortions. We solve this problem by blending the geodesic and the Euclidean distances formulated as

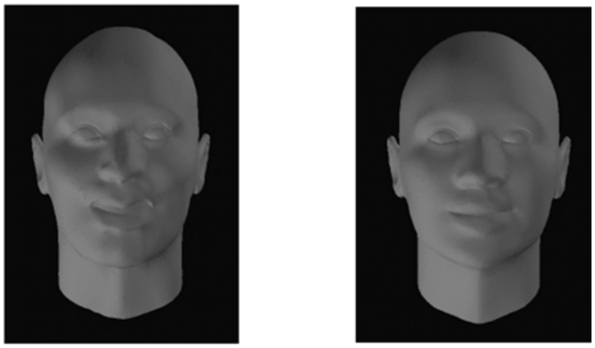
$$\delta_{ij} = w \times \delta_{ij} + (1-w) \times \gamma_{ij},$$

where $w\in[0,1]$ is the blending coefficient. The Euclidean distance γ works as a smoother part to eliminate the artifacts. The value of w can be decided empirically.

BAYESIAN REGRESSION

While the geodesic manifold distance provides a tool for modelling the spatial structure of face models, the Bayesian regression can find the temporal features of facial animation. Although regression techniques like radial basis functions are enough for providing facial deformation results under noise-free circumstances, the processes of facial motion capture and 3D face scanning can hardly be noise-free.

Bayesian approaches to regression (Hertzmann, 2003; MacKay, 1998; Williams and Rasmussen, 1996) provide an elegant way of dealing with noise and uncertainty, as well as a framework of learning the statistical features of data. We use Bayesian approaches for exploring facial deformation driven by warping kernels, as shown in Fig.4.



(a) (b)

Fig.4 Bayesian marginalization is applied to eliminate noise in motion or model data. The artifacts resulted from noisy data in (a) are smoothed in (b) in the generated expression

Framework of Bayesian regression

In Bayesian interpretation of the regression problem in Section 3.1, a nonlinear function $y(\boldsymbol{s})$ parameterized by the hyper-parameter θ is assumed to underlie motion data $\{\boldsymbol{s}^{(n)}, t_n\}_{n=1}^{N}$ where N is the number of observations, or feature points. We denote coordinates of feature points at the neutral frame by $\boldsymbol{S}_N=\{\boldsymbol{s}^{(n)}\}_{n=1}^{N}$, and the correspondences at a target frame by three vectors $\boldsymbol{t}_N=\{t_n\}_{n=1}^{N}$ for respective warping channels of x, y or z.

The inference of $y(\boldsymbol{s})$ by inferring the hyper-parameter θ is depicted by the posterior probability distribution:

$$P(\theta \mid \boldsymbol{t}_N, \boldsymbol{S}_N)=\frac{P(\boldsymbol{t}_N \mid \theta, \boldsymbol{S}_N)P(\theta)}{P(\boldsymbol{t}_N \mid \boldsymbol{S}_N)}. \quad (1)$$

Minimizing the minus log of posterior probability distribution as an object function:

$$\theta=\arg\min_{\theta}(-\log(P(\boldsymbol{t}_N \mid \theta, \boldsymbol{S}_N)P(\theta)). \quad (2)$$

Maximize a posterior (MAP) and maximum likelihood (ML) are commonly used approaches for predicting outputs. However, they sometimes converge to local minima (Williams and Rasmussen, 1996) and are not convenient for generation of facial animations. Bayesian predictions can also be made without estimated value of θ by marginalizing over the hyper-parameter. We perform integrations in Eq.(3) by sampling the possible values of θ from $P(\theta|\boldsymbol{t}_N, \boldsymbol{S}_N)$ using Markov chain Monte Carlo method of Hybrid Monte Carlo (Duane *et al*., 1987). The sampling process saves 200 sampled values of the hyper-parameter.

$$P(\boldsymbol{t}_{N+1} \mid \boldsymbol{t}_N, \boldsymbol{S}_N)=\int P(\boldsymbol{t}_{N+1} \mid \theta, \boldsymbol{S}_N)P(\theta \mid \boldsymbol{t}_N, \boldsymbol{S}_N)\mathrm{d}\theta. \quad (3)$$

Modulation of covariance matrix with the geodesic distances

There are many choices of covariance matrix under the condition of non-negative definite property for any set of points $\{\boldsymbol{s}^{(1)},\dots, \boldsymbol{s}^{(n)}\}$. We empirically choose the following covariance matrix used in (Williams and Rasmussen, 1996) to adopt the geodesic distances.

$$\boldsymbol{C}(\boldsymbol{s}^{(i)}, \boldsymbol{s}^{(j)})=v_0\exp\left\{-\frac{1}{2}\sum_{l=1}^{d} w_l(\boldsymbol{s}_l^{(i)}-\boldsymbol{s}_l^{(j)})^2\right\} + a_0 + a_1\sum_{l=1}^{d}\boldsymbol{s}_l^{(i)}\boldsymbol{s}_l^{(j)} + v_1\rho(i,j). \quad (4)$$

where θ=log(v_0,v_1,w_1,…,w_n,a_0,a_1) represents the hyperparameter since the variables in Eq.(4) are positive scale parameters. θ corresponds closely to hyper-parameters in neural networks to be adapted through learning.

The covariance function in Eq.(4) consists of three parts: the weighted sum term, the linear regression term (including a_0 and a_1) and the noise term. The first term expresses that the nearby inputs produce similar outputs and that the weights of w_l are usually set to ones. This leads to the equation below:

$$\boldsymbol{C}(\boldsymbol{s}^{(i)}, \boldsymbol{s}^{(j)})=v_0\exp\left\{-\frac{1}{2}\sum_{l=1}^{d}(\boldsymbol{s}_l^{(i)}-\boldsymbol{s}_l^{(j)})^2\right\} + a_0 + a_1\sum_{l=1}^{d}\boldsymbol{s}_l^{(i)}\boldsymbol{s}_l^{(j)} + v_1\rho(i,j), \quad (5)$$

where

$$\sum_{l=1}^{d}(\boldsymbol{s}_l^{(i)}-\boldsymbol{s}_l^{(j)})^2 \quad (6)$$

is the Euclidean distance.

We substitute the geodesic distance $\delta_{ij}=dijk(\boldsymbol{s}^{(i)}-\boldsymbol{s}^{(j)})$ for Eq.(6) and get:

$$\boldsymbol{C}(\boldsymbol{s}^{(i)},\boldsymbol{s}^{(j)})=v_0\exp\left\{-\frac{1}{2}dijk(\boldsymbol{s}^{(i)}-\boldsymbol{s}^{(j)})^2\right\}+a_0+a_1\sum_{l=1}^{d}\boldsymbol{s}_l^{(i)}\boldsymbol{s}_l^{(j)}+v_1\rho(i,j). \tag{7}$$

With the covariance matrix $\boldsymbol{C}$, Eq.(2) can be deduced for N training pairs as:

$$\theta=\arg\min_{\theta}\left(\frac{1}{2}\log\det(\boldsymbol{C})+\frac{1}{2}\boldsymbol{t}_N{}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{t}_N+\frac{N}{2}\log 2\pi\right). \tag{8}$$

EXPERIMENTAL RESULTS

The data used in this paper are captured with Motion Analysis motion capture system. Facial motion data is captured with a set of 42 markers (including 3 markers on a head mounted jig for global motion estimation) at a frame rate of 60 fps. Eight representative expressions are extracted, as shown in Fig.5, and used to drive an ordinary quadrangle face mesh via manifold Bayesian regression, as in Fig.6. The results showed that Bayesian method for regression and the geodesic distances are suitable for representing the underlying deformation paradigm and the complex topology of human faces. As shown in Fig.6, the generated expressions are smooth and the correct movement of discontinuities in the mouth region shows the power of manifold. The situation in the eyes' regions may be confusing as they show hardly any difference between the left and the right expressions. The reason is that they differ from the mouth region in the neutral expression as they have proper spatial gaps between the upper and the lower eyelids by which the Euclidean distances can equally well express the discontinuities.

We have built a prototype system with Visual C++ and Matlab. We used a 1.5 GHz Pentium-IV PC in our experiments. The process of constructing undirected graph from face mesh with 5832 vertices and 5746 faces takes about 1 s and finding the shortest path from feature points to all other vertices takes about 0.5 s. The Bayesian regression process takes about 2 s.

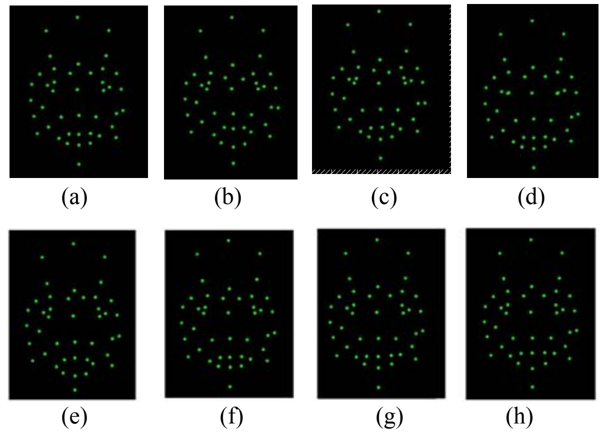


Fig.5 Eight key expressions extracted from facial motion data. (a) Neutral; (b) Anger; (c) Disgust; (d) Eye-close; (e) Fear; (f) Sad; (g) Smile-close; (h) Smile-open

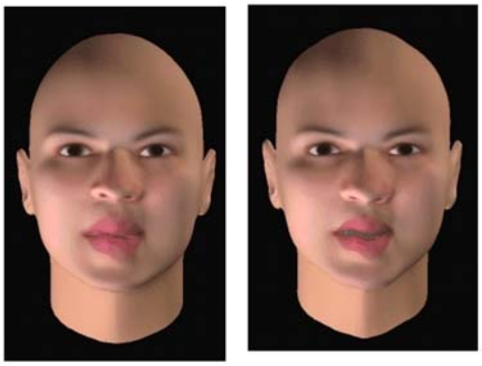
Besides the whole process of facial expression generation requires little user intervention except some markers' manual correspondences.

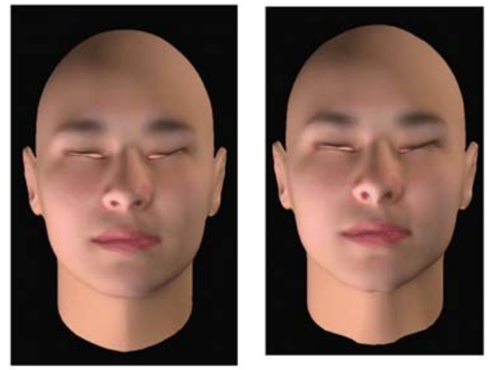
CONCLUSION AND FUTURE WORK

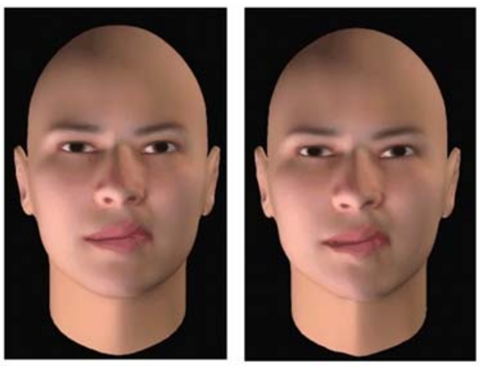
In this paper, we present a novel and effective method for driving facial animations based on motion data. Geodesic manifold distances are adopted in the framework of Bayesian learning to automatically model the discontinuities and complex topology of human faces. After the covariance matrix is properly modulated with the geodesic distances, Hybrid Monte Carlo method is then used to compute the integral of probabilities to predict results under noisy circumstances. The techniques in this paper allow facial motion data to be applied to any facial mesh. Previously, adapting facial motion data to an individual mesh required much more artistic intervention.

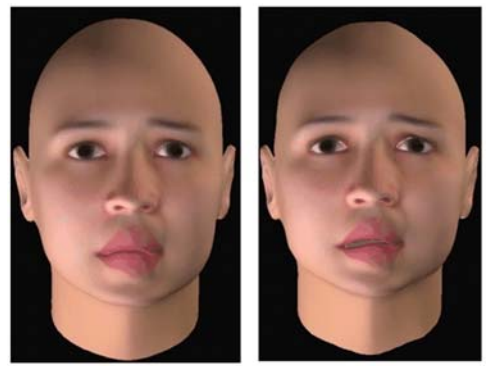
Techniques in this paper can be used in many scenarios, such as human computer interaction and digital entertainment. They also have prospective applications in data compression for 3D streaming media, or distant meeting (Cutler *et al.*, 2002) for facial communication between participants as the computing power grows.

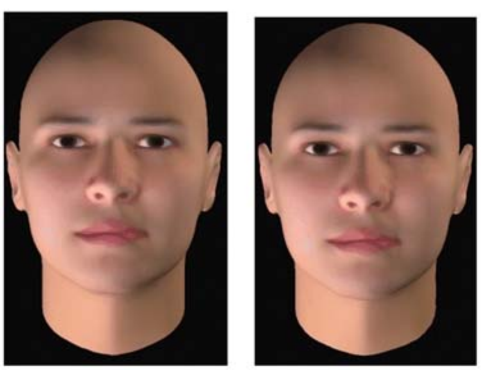
Although the generated facial expressions are plausible in the big picture, there are several possible improvements. The selection of neighbours in manifold learning may be adjusted to provide better modelling of face topology. The wrinkles and detailed

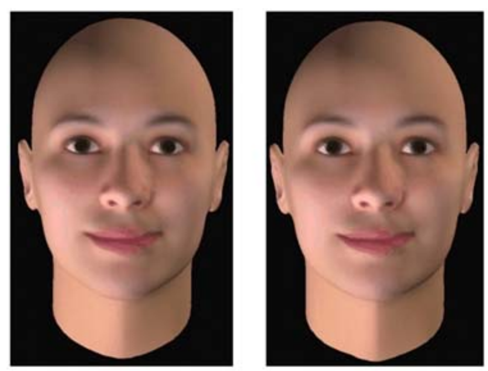

(a)


(b)


(c)


(d)


(e)


(f)

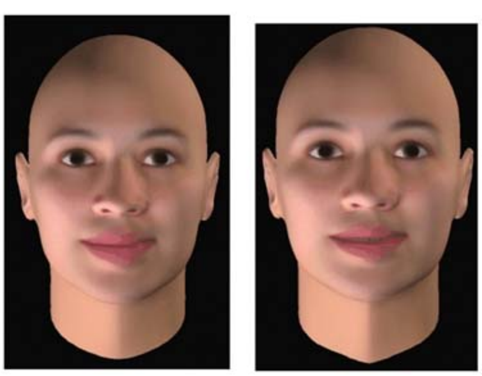

(g)

Fig.6 More results by Bayesian regression with texture mapping on different expressions. The left ones of pairs are generated in the Euclidean distance metric only and the right ones are in the mixture metric of geodesic distances and Euclidean distances with a blending weight of 0.39, as described in Section 3.3. The generated expressions are smooth and the movement of lips demonstrates the power of manifold. The eyes' regions show hardly any difference with the aid of manifold since they differ from the mouth region in the neutral expression as they have proper gaps between the upper and the lower eyelids by which the Euclidean distances can equally well express the discontinuities. (a) Anger expression; (b) Eye-close expression; (c) Disgust expression; (d) Fear expression; (e) Sad expression; (f) Mouth-closed smile expression; (g) Mouth-open smile expression

textures have also not been properly tackled in the existing techniques. These problems ought to be considered in future work.

References

Blanz, V., Basso, C., Poggio, T., Vetter, T., 2003. Reanimating Faces in Images and Video. Computer Graphics Forum, EUROGRAPHICS 2003, Granada, Spain, **22**(3):641-650.

Cutler, R., Rui, Y., Gupta, A., Cadiz, J., Tashev, I., He, L.W., Colburn, A., Zhang, Z.Y., Liu, Z.C., Silverberg, S., 2002. Distributed Meetings: A Meeting Capture and Broadcasting System. Proc. of ACM Mulitmedia 2002, Juan-les-Pins, France.

Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D., 1987. Hybrid Monte Carlo. *Physics Letters B*, **195**(2):216-222. [doi:10.1016/0370-2693(87)91197-X]

Guenter, B., Grimm, C., Wood, D., Malvar, H., Pighin, F., 1998. Making Faces. SIGGRAPH 98 Proceedings, p.55-66.

Hertzmann, A., 2003. Machine Learning for Computer Graphics: A Manifesto and Tutorial. Proceeding of Pacific Graphics 2003, Banff, Alberta, p.22-36.

Huang, X.S., Li, S.Z., Wang, Y.S., 2004. Learning with Cascade for Classification of Non-Convex Manifolds. First IEEE Workshop on Face Processing in Video, Washington, D.C., USA.

Joshi, P., Tien, W.C., 2003. Learning Controls for Blend Shape Based Realistic Facial Animation. Proceedings of EUROGRAPHICS03, p.187-192.

Lee, S., Wolberg, G., Shin, S.Y., 1997. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*. **3**(3):228-244. [doi:10.1109/2945.620490]

Liu, X., Zhuang, Y.T., Pan, Y.H., 1999. Video Based Human Animation Technique. ACM Multimedia99, Orlando, Florida, USA, p.353-362.

MacKay, D.J.C., 1998. Introduction to Gaussian Processes. *In*: Bishop, C.M.(Ed.), Neural Networks and Machine Learning. Kluwer Academic Press, p.133-166.

Na, K., Jung, M., 2004. Hierarchical Retargeting of Fine Facial Motions. Proceedings of EUROGRAPHICS04, 687-695.

Niyogi, P., Belkin, M., 2002. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Technical Report, University of Chicago.

Noh, J., Neumann, U., 2001. Expression Cloning. Proceeding of SIGGRAPH01, p.271-288.

Noh, J., Fidaleo, D., Neumann, U., 2000. Animated Deformations with Radial Basis Functions. Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST), p.166-174.

Pan, Y.H., Zhuang, Y.T., Liu, X.M., 2000. Video motion capture in VBA—video-based animation. *Journal of Zhejiang University SCIENCE*, **1**(1):1-7.

Pyun, H., Kim, Y., Chae, W., Kang, H.Y., Shin, S.Y., 2003. An Example-Based Approach for Facial Expression Cloning. ACM SIGGRAPH/Eurographics Symposium on Computer Animation, p.167-176.

Roweis, S., Saul, L., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**(5500):2323-2326. [doi:10.1126/science.290.5500.2323]

Sánchez Lorenzo, M.A., Edge, J.D., King, S.A., Maddock, S.C., 2003. Use and Re-use of Facial Motion Capture Data. Vision, Video, and Graphics, VVG 2003. University of Bath, UK, p.135-142.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**(5500):2319-2323. [doi:10.1126/science.290.5500.2319]

Williams, L., 1990. Performance-Driven Facial Animation. SIGGRAPH 90 Proceedings, p.235-242.

Williams, C.K.I., Rasmussen, C.E., 1996. Gaussian Processes for Regression. Proceeding of NIPS 8, MIT Press.

Zhang, J.P., Li, S.Z., Wang, J., 2004. Manifold Learning and Applications in Recognition. *In*: Tan, Y.P., Yap, K.H., Wang, L.P.(Eds.), Intelligent Multimedia Processing with Soft Computing. Springer-Verlag, Heidelberg.

Zhuang, Y.T., Liu, X.M., Pan, Y.H., 1999. Video Motion Capture Using Feature Tracking and Skeleton Reconstruction. IEEE International Conference on Image Processing (ICIP'99), Kobe, Japan, p.232-236.