Research on Local Face Motion in 3D Facial Expression Generation without Mesh

Mei XUE † Shogo TOKAI ‡ and Hiroyuki HASE ‡

† Faculty of Engineering, University of Fukui 5-32-27 Bunkyou, Fukui, 910-8507 Japan

E-mail: † xmwindxmwind@hotmail.com, ‡ {tokai, haseh}@u-fukui.ac.jp

Abstract Our research in this paper is a kind of method directly processing points based upon intrinsic geometry and color information with the advantages of no necessity to set up face model or expression library in advance and direct processing of points in PC (Point Cloud) without any mesh or markers. The animation of facial expression is a kind of combination of global face motion and local face motion which is a set of partial movements on the face. Correct corresponding points finding process improves the result of the interpolated PC for facial expressions between two different facial point clouds. This paper introduces the related works in PCs processing and facial expression modeling. Then the pipeline of facial expression generation has been introduced in detail with every step. Finally the experimental results and future works have been discussed.

Keyword 3D Graphics, Realism Animation, Point Clouds, Facial Expression

1. Introduction

Animation has gained increasing importance in many fields of computer graphics, including 3D game engines, computer animation for feature films, surgery simulation, and virtual reality. Among these applications, convincing facial expressions are essential to capture the audience in computer-animated films. In existing research of facial expression generation, except of popular modeling methods, Point clouds (PCs) which are 3D data acquired from special measurement equipments are attracting more and more interests. The traditional process of 3D objects reconstruction based upon PCs has mainly five steps: 3D Acquisition, Surface Reconstruction, Processing & Modeling, Rendering and Animation [1] in which reconstruction and deformation has been under research greatly. Among those, Surface Reconstruction of the model is complex and time consuming [2] [3]. Mesh construction based upon PCs is another option for obtaining 3D shape. Repeated Delaunay operations can adversely affect the stability and accuracy of the calculations, imposing undesirable restrictions on the time step. Therefore, meshless methods are well suited for handling large deformations due to their flexibility when locally refining the sampling resolution.

In the animation of facial expression, the application based upon PCs from range images has been seldom used directly for simulation attracting more and more attention. Our research focus on how to generate facial expressions in computer with several PCs derived from measuring equipment such as range image camera. We observed that facial expressions include global face motion and local face motion. The face motion is the combination of these two kinds of motions. Global face motion is a kind of rigid motion including translation and rotation of a head. On the other hand, local motion is a set of partial movements on the face by facial expression. We make use of these face motions to decide the interpolated PCs from several PCs therefore to generate the facial expression in computer.

This paper introduces the related works in PCs processing and facial expression modeling. Then the pipeline of facial expression generation has been introduced in detail with every step. Finally the experimental results and future works have been discussed.

2. Facial Expression Generation Pipeline

In this paper, the PCs are derived from the range image camera MINOLTA VIVID 700 (Figure 1). The size of range images is 200 pixels × 200 pixels, the distance between the face and the camera is 1.5m, Reduction Rate 1/1, Fill Holes Off, Remove Boundary, Filter None and Auto-focus.

The PCs derived have the following features: only the coordinates and color information of points available from range images; unorganized PC which means points in one PC locate randomly without any discipline; uncertainty of PCs which means points in different PCs have no relationship between each other; noise and error exist.

The pipeline of PC-based facial expression generation has 6 steps. By use of 3D acquisition measurements such as range image camera, the PCs of the same face with different expressions have been obtained with information of coordinates and color. To simplify the afterwards processing, the PCs need to be preprocessed for refinement and uniformities according to local density since the PCs under research are unorganized with uncertainty, noise and error. To interpolate new ones for PCs acquired for different facial expressions, the relationship between two PCs need to be found which decide the motions from one PC to another. The motions involved are the combination of Global Motion and Local Motion. With the interpolated new PCs, facial expressions can be generated for computer animation. With the interpolated new PCs, facial expressions can be generated for computer animation.

Obtaining facial expressions between two PCs at time t0=0 and t1=1, is a kind of deformation time-varying. The research of reconstruction of deforming geometry from time-varying point clouds discussed in [4] is based upon a Bayesian statistical model. On the basic idea of moving points (surfel) over time, our research is only involving simple explicit information and implicit information of points and avoid complex physically considerations. In our research, Facial expressions include global face motion and local face motion. Global face motion is a kind of rigid motion including translation and rotation while local face motion is a set of partial movements of the face. The advantages of our method are: without mesh construction, there is no need to set up any face model or expression library in advance. Deformation decision is simplified to decide the Global Motion and Local Motion of PCs and there is no need to set up markers on real face or consult to any complex physical theories. The interpolated points are derived from the PCs directly. There is no need to set up any mesh model and avoid time-consuming Delauny reconstruction.

Analysis of the global face motion is a similar process as initial alignment of two PCs. To align of them, the ICP (iterative closest point) algorithm [5], [6], [7] which is a popular method, has been used to obtain the global face motion.

ICP starts with two PCs and an initial guess for their relative rigid-body transform, and iteratively refines the transform by repeatedly generating pairs of corresponding points on the PCs and finding the transformation that minimizes an error metric, such as the sum of squared point-to-point distances.

The method in ICP algorithm has been used to decide the translation and rotation in global face motion.

$$\epsilon = \sum_i [(Rp_i + t - q_i) \times n_i]^2 \quad , \ i=1 \cdots n \qquad (1)$$

where, ε is the alignment error between the two PCs, R is the rotation matrix, t is the translation vector, pi and qi are PCs for different facial expressions, ni is the normal vectors of points in pi which can be calculated from coordinates with the method of local plane fitting [8]. This equation can be changed into a linear system in the form of Cx=B which can be solved by Cholesky decomposition since C is symmetric matrix.

3. Local Face Motion

Local face motion is a set of partial movements of the face therefore points in one PC should be selected and matched to the closest points on the other PCs. One of the PCs is moved such as to minimize the distances between pairs of points and the process is iterated until convergence.

Local face motion involves deformation rules decision which is very difficult to be decided. According to local curvature discontinuity, the optimized eye corners and mouth corners can be decided for face partition. The corresponding partitions of faces in different PCs are the constraints at time t0=0 and t1=1. Every point Pi can find its several corresponding points Qij (j=1⋯k, k is number of Kd-nearest neighbors) in another PC. Then the points at time t=T for intermediate frames can be decided to simulate the expression animation as shown in Figure1.

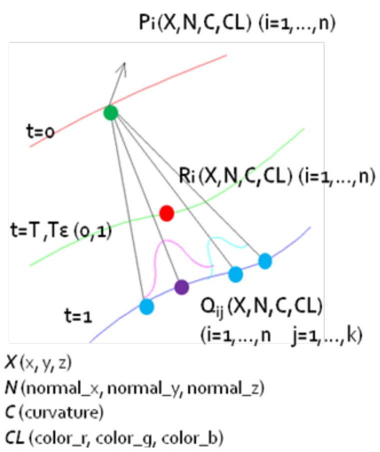
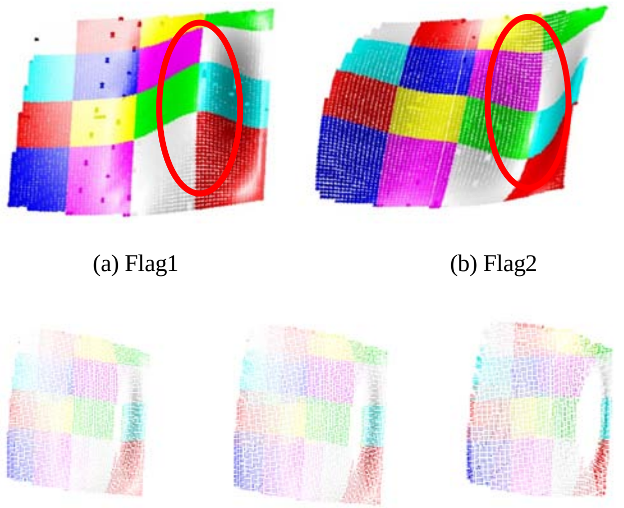


Figure 1: *Interpolated Point Cloud from 2 Point Clouds with Different Facial Expressions*

In the process of looking for corresponding point, 2 PCs are pressed into octree grids individually firstly. Since human's facial expression won't change points' positions greatly, the process of looking for the corresponding points Qi for every Pi can be shorten by 2 steps: first is to find the grid geometrically nearest to the one where Pi belongs to; second is to find the corresponding points Qij in the grid found in the first step. The width of octree grids is decided by the density of the PC adaptively.

But if only coordinates considered in corresponding points, there will be incorrect corresponding points found resulting blanks in interpolated Flags at different time t1, t2 and t3 since Flag and Flag2 have large geometry difference in the area around red circles as shown in Figure 2.



(a) Flag1 (b) Flag2

(c) Interpolated PC at time t=t1 (d) Interpolated PC at time t=t2 (e) Interpolated PC at time t=t3

Figure 2: *Incorrect Corresponding Points Found Resulting Blank In Interpolated PCs.*

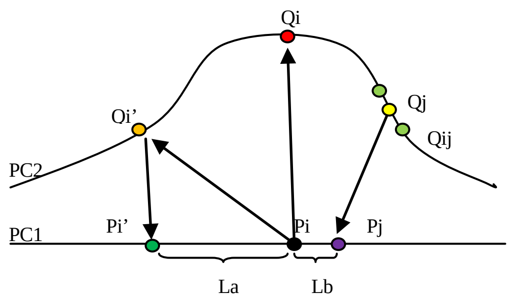


Figure 3: *Process of Correct Corresponding Points Finding*

In Figure 3, it shows the process of looking for correct corresponding points. The correct corresponding point of Pi in PC1 should be Qi in PC2. When the 2 PCs have large geometry differences, Qi' will be found instead resulting blanks for interpolated PC. Therefore the steps of correct corresponding points finding are as following:

1. For every point Pi in PC1, Looking for Qi' in PC2;

2. Looking for Pi' in PC1 for Qi';

3. If La, the distance between Pi and Pi' is larger than some threshold ε1, go to step 4, else Qi' is the final result;

4. For every point Qj in PC2, looking for Pj in PC1, if Lb, the distance between Pi and Pj is smaller than some threshold ε2, Qj is the final result.

5. Since there is difference between Pi and Pj, Kd-nearest neighbors of Qj will be used as final corresponding points for Pi.

After every point Pi find its corresponding points Qij in another PC, the points at t=T for intermediate frames can be decided to simulate the expression animation as shown in Figure 1.

RBF (Radial Basis Function) has been used for the different influence weights to Pi of every corresponding point as shown in the equation (2). Normal vectors of points can be calculated by the method mentioned in Global Face Motion. Color information comes from the texture of range image.

$$R_i(X,N,C,CL) = \frac{1}{k}\sum_{j=1}^{k}\left(Q_{ij}(X,N,C,CL) - P_i(X,N,C,CL)\right)\varphi(x)t + P_i(X,N,C,CL)$$

$$i=1\cdots n,\ t=(0,1) \qquad (2)$$

In equation (2), Rij is the vector consisting of coordinate vector X, normal vector N, and color vector C of the interpolated points. t is the time ranging for 0 to 1. φ(x)is the Radial Basis Function in the form of Gaussian function $\varphi(x) = e^{-\frac{2x^2}{h^2}}$ where x indicates the difference of the normal vector and color information between Qij and Pi as shown in equation (3).

$$x = \sum_{j=1}^{k}\left\|Q_{ij}(N,C) - P_i(N,C)\right\|, \quad (i = 1\cdots n) \qquad (3)$$

4. Implementation and Results

The system has been programmed with VS2008 and C++. The PCs derived from the range image camera MINOLTA VIVID 700 as shown in Figure 4, Figure 5 and Figure 6. From range images of different facial expressions, intrinsic information of PCs are derived or calculated. Then the PCs are preprocessed with operations of refinement and uniformity. The ICP algorithm has been employed for rotation matrix and translation vector in Global Motion. Local motion has been processed for interpolated new PCs. Figure 7 displays the two original PCs for t0=0 and t1=1 in color red and blue and one interpolated PC at $tm=t0+m\times TS$, (m= 1…T/TS) in black.

In Figure 8, (a) and (d) are the PCs rendering with surfel rendering mode for two different facial expressions. (b) is the interpolated PC at time t=0.5 for the pair of PCs (a) and (d) rendered with surfel. From the experimental result, it is evident that the final visible result depends heavily upon the rendering process which will be our next work to be improved.

(c) is the computer generated facial expression which has narrow difference with the real one (d).

Figure 4: *PCs Derived From the Range Image Camera*

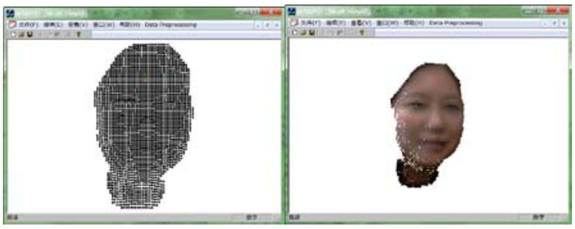
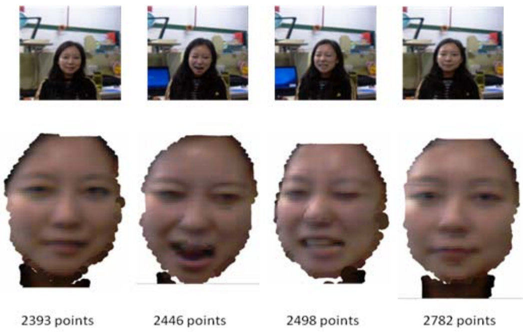


Figure 5: *Displaying of PC Before and After Preprocessing.*



(a) (b) (c) (d)

Figure 6: *PCs of Different Facial Expressions*

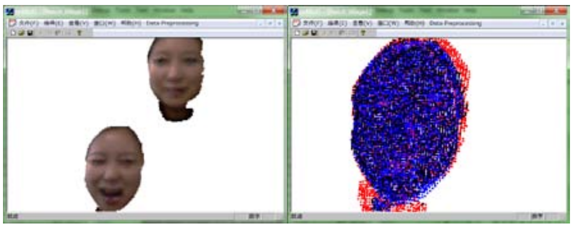
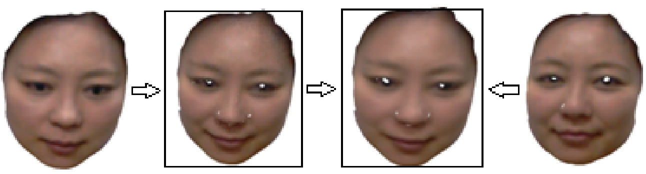


Figure 7: *The Two Original PCs For t0 and t1 In Color Red and Blue And One Interpolated PC At tm*



(a) (b) (c) (d)

Figure 8: *PCs of 2 Different Facial Expressions (a) and (d)*

Table 1 shows more experimental results of PCs of different facial expressions. From the results, it could be see that if the two facial expressions differ greatly, it will bring unsatisfactory results. And the final result depends on the result of Global face motion heavily. If the result is not satisfactory, another iteration of Global face motion should be made until the final result could be obtained.

Comparing with the methods existing, as for 3 main considerations for face and facial expression simulation, deformation model, deformation decision and deformation process, the method in this paper has the following advantages:

- **Deformation Model:** There is no need to set up face model or expression library in advance.
- **Deformation decision:** Deformation decision is simplified to decide the Global Motion and Local Motion of PCs and there is no need to set up markers on real face or consult to any complex physical theories.
- **Deformation process:** The interpolated points are derived from the PCs directly. There is no need to set up any mesh model and avoid time-consuming Delauny reconstruction.

5. Conclusions

Animation is a sequence of images. Indicate enough interpolated PCs at different time may generate fluent animation. The method in this paper deals with several PCs taken from one face at different time and interpolates new PCs for smooth changing from one PC to another. It is a kind of method directly processing from points in PC with no mesh and markers and avoids the time consuming calculation involved in the surface reconstruction and meshes maintaining and has the superiority both off line and on line.

Reference

[1] M. Gross, H. Pfister, Point-Based Graphics, Publisher: Morgan Kaufmann, June 29, 2007.

[2] GREINER, G., LOOS, J., AND WESSELINK, W. 1996. Data dependent thin plate energy and its use in interactive surface modeling. Computer Graphics Forum 15, 3 (August), 175–186. ISSN 1067-7055.

[3] WITKIN A. P., AND HECKBERT P. S. 1994. Using particles to ssample and control implicit surfaces. In Proceedings of the 21st annual conference on Computer graphics and interactive techniques, ACM Press,269–277.

[4] Michael Wand, Philipp Jenke, Qixing Huang, Martin Bokeloh, Leonidas Guibas, Andreas Schilling, Reconstruction of Deforming Geometry from Time-Varying Point Clouds, SGP

'07Proceedings of the fifth Eurographics symposium on Geometry processing, 2007.

[5] Besl, P., McKay, N., A method for registration of 3-D shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, p.239-256, 1992.

[6] D.W. Eggert, A. Lorusso, R.B. Fisher, Estimating 3-D rigid body transformations: a comparison of four major algorithms, International Journal of Machine Vision and Applications, Vol. 9, p.272–290, 1997.

[7] Szymon Rusinkiewicz, Marc Levoy, Efficient variants of the ICP algorithm, Proceedings of the International Conference on 3D Digital Imaging and Modeling (3DIM), 2001.

[8] H. Hoppe, et.al, Surface reconstruction from unorganized points, SIGGRAPH 1992 Proceedings, p.71-78, July 1992.

PCs		Sample points		Time costs	Interpolated PC
Face 1	Face 2	Face 1	Face 2		
		27503	27414	59 sec	
Face 1	Face 3	Face 1	Face 3		
		27503	27444	75 sec	
Face 2	Face 3	Face 2	Face 3		
		27414	27444	18 sec	

Table 1: *Experimental results for PCs of Different Facial Expressions*