Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/100440

3D Mesh Sequence Compression Using Thin-plate Spline based Prediction

Yuan Gao[1,2,*]

[1] Beijing Key Laboratory of Multimedia Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, 100124 Beijing , China
[2] Department of Communication Engineering, Beijing Electronic Science and Technology Institute, 100070 Beijing , China

Received: 17 Apr. 2016, Revised: 3 Jun. 2016, Accepted: 4 Jun. 2016
Published online: 1 Jul. 2016

Abstract: We propose an efficient scheme for compressing mesh sequence with different connectivities using a thin-plate spline (TPS) based multi-reference prediction coding. The scheme combines the single-reference prediction with the P-structure prediction, and predicts each frame of the sequence using the TPS transformation. Compared with traditional compression scheme, our scheme imposes the TPS transformation in the prediction coding which utilizes both spatial and temporal correlations of the mesh sequence with different connectivities. We test our scheme for 3D facial meshes and experimental results show that our scheme produces both accurate reconstruction meshes and high compression efficiency.

Keywords: mesh compression, Thin-plate Spline (TPS), mesh sequence, matching closest point

1 Introduction

With the rapid development of computer graphics technologies, 3D meshes are widely used in computer graphical applications. The representation and compression of 3D meshes are of significant importance for the aim to storing and transmitting 3D meshes data.

Compression of mesh sequence is a popular and important topic in data compression, which is useful for compressing and transporting animation meshes in a band-limited network. Previous researches focus on static individual 3D mesh, where the spatial correlations of the mesh facilitate the compression precess. Utilizing temporal correlations of the mesh sequence shall improve the compression efficiency. See Karni and Gotsman [1] who combine principle component analysis (PCA) with linear prediction coding (LPC) in dynamic 3D mesh sequence compression.

However, due to the procedure of mesh generation using professional devices such as Cyberware 3D Scanner [7], multi-camera [6], the connectivity of meshes of different frames is not uniform. On one hand, imposing traditional compression schemes (e.g. [1]) of nonuniform connectivity mesh sequence do not work; on the other hand, the frame-by-frame compression schemes of static meshes have a low efficiency. Previous work of mesh sequence compression, such as Hou et al. [8], compress geometry video using a model-based joint bit allocation scheme, which allocates reasonable bitrate to each x, y, z-dimension of rate-distortion model.

In this paper, we propose a mesh sequence compression scheme using TPS transformation as the prediction coding. The prediction coding imposes TPS transformation on the input mesh sequence, so that all meshes are remeshed with a uniform connectivity with respect to a reference mesh, and inter-frame prediction is simultaneously completed by encoding the prediction residuals and the coefficients of the transformation matrix between adjacent frames. Compared with traditional approaches for compressing mesh sequence, our scheme encodes a mesh sequence with transformation matrices, residuals and the reference mesh, which utilizes both spatial and temporal correlations of the mesh sequence. Experimental results show that, our scheme achieves good compression bit rate and promising qualitative results, compared with frame-by-frame compression schemes.

The rest of this paper is organized as follows: Section 2 introduces the method of applying the thin-plate spline on 3D facial meshes registration, and how to match the closest point. Section 3 describes the framework and

* Corresponding author e-mail: gaoyuan_edu@126.com

procedure of our whole compressing scheme in details. Experiment results and discussion are shown in Section 4. Finally, Section 5 concludes this paper with future probable research work.

2 Thin-plate spline transformation and Matching closest point algorithm

Before proposing the compression scheme of mesh sequence, we introduce the thin-plate spline transformation [9] and matching closest point algorithm in this section.

2.1 Thin-plate spline transformation

Thin Plate Spline (TPS) was firstly introduced by R. L. Harder and R. N. Desmarais [9]. TPS has a unique function to decompose a space into an affine transformation and a non-affine transformation [10], which makes TPS a useful tool in shape analysis and able to approximate almost all biological deformation.

Hu et al. [3] present a non-rigid registration method for 3D human faces. The registration is based on a tool for surface deformation: the TPS transformation. Essentially, TPS is a kind of radial basis function [2] , which is given by:

$$\phi = r^2 \log r \tag{1}$$

where r denotes the Euclidean distance between two points in Cartesian coordinates. When the TPS role curved surface deformation from space $\mathbb{R}^3$ to $\mathbb{R}^3$, we denote TPS as f: $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ [3]. There are two sets of points F_1, F_2,

$$\begin{cases} F_1 = \{P_{1i} | P_{1i} = (x_{1i}, y_{1i}, z_{1i}),\ i = 1, \ldots, N_1\} \\ F_2 = \{P_{2j} | P_{2j} = (x_{2j}, y_{2j}, z_{2j}),\ j = 1, \ldots, N_2\} \end{cases} \tag{2}$$

$$N_1 \neq N_2$$

According to the front, the TPS transformation has the following style:

$$f(P) = Pd + Kw \tag{3}$$

where P is the matrix of homogeneous coordinates of a vertices set, d, w are the affine transformational matrix and non-affine transformational matrix respectively, and K is the matrix of values of TPS basic function. To obtain the transformational matrix, we need to select out a group of corresponding control point sets respectively in F_1 and F_2 for TPS transformation and denote them as M_1 and M_2:

$$\begin{cases} M_1 = \{J_{1i} | J_{1i} = (x_{1i}, y_{1i}, z_{1i}),\ i = 1, \ldots, M\} \\ M_2 = \{J_{2j} | J_{2j} = (x_{2j}, y_{2j}, z_{2j}),\ j = 1, \ldots, M\} \end{cases} \tag{4}$$

According to equation (3), we can determine the matrix d and w, and then F_2 can be obtained from F_1.

The registration of the paper [3] takes TPS transformation on the reference 3D face mesh, after obtaining the transformational matrices d and w , take global TPS transformation on reference face mesh F_1 through the equation (3) to get a registration result approximate face mesh F_1'. If the control points are appropriate, F_1' will be approximate to target mesh F_2. Then take the F_1' in target mesh F_2 to find the vertices of the nearest points, and the reconstructed target mesh F_2' can be obtained, which will be approximate to target mesh.

A weakness of the method in the paper [3] is that they only pick out one mesh as the reference mesh when aligning a mesh sequence and the registration efficiency of the meshes far away from the reference mesh is always worse than the ones near to the reference mesh.

2.2 Matching closest point algorithm

In the process of the above, an important step is matching the closest points. By using a classic K-dimensional binary search tree (KD-tree) proposed by Bentley [4], M_2 can be obtained by M_1, and F_2' can be obtained by F_1'. M_1 is uniformly selected from the reference mesh F_1, F_1' is obtained through the TPS transformation.

KD-tree is an algorithm which is widely used in searching nearest neighbor points [11] and it is a rigid matching closest point process. Each node of KD-Tree represents a point of n-dimensional space $\mathbb{R}^n$, and each layer of the Tree makes decisions according to the layer of discriminator. After many divisions, leaf node can be obtained. Afterwards, the neighboring points are determined using the backtracking algorithm. The closest point search is completed by searching the root node space along a certain path.

3 Compress 3D mesh sequence

This section proposes the main result of this paper: a mesh sequence compression scheme using TPS transformation as the prediction coding. The main idea integrates the TPS transformation with the inter-frame prediction of video coding, in order to improve the compression efficiency using the temporal correlations.

Our scheme consists of a mesh grouping phase (Section 3.1), and the TPS-based prediction phase (Sections 3.2-3.4). The mesh grouping phase divides all the mesh frames of the input sequence into several groups according to the order of the sequence, on each of which shall be imposed prediction coding. Typically the length of each group is set by an odd number and the middle frame shall be selected as the reference frame. The TPS-based prediction phase proposes three prediction codings: TPS-based single-reference prediction, TPS-based P-structure prediction and TPS-based multi-reference prediction. TPS-based single-reference prediction considers the middle frame as the single reference frame based on the assumption that the middle frame have a relative good correlation with both the former and latter frames. TPS-based P-structure prediction starts from the middle frame and predicts the adjacent frames in a frame-to-frame way. During the prediction coding, the last decoding frame is considered as the reference frame of the current encoding. TPS-based multi-reference prediction integrates two preceding prediction codings together, by selecting the reference frame using a composition of TPS transformation.

3.1 A grouping of mesh sequence based on video compression

In a 3D mesh sequence, as each frame of a 2D video is a static picture, we consider that each frame of a 3D video is a static 3D mesh. Take Fig. 1 as an example, where the video starts from the *smile* frame to the *close the mouth* frame. We decompose the video into a number of static 3D meshes. In video coding techniques, frames are segmented into many groups called group of pictures (GOP) and every GOP contains the key frame I and other prediction frames B and P. For 3D mesh sequences likewise, frames are also segmented into many groups known as group of meshes (GOM). In every GOM, there is one frame I and several frames P, in which the frame I is compressed as a static 3D mesh and the reference mesh for TPS transformation of all the frames P, and the frames P are as the target meshes in TPS transformation. Moreover, in video coding standards like H.264, the frames in a GOP are usually ranged as the following style: IBBPBBPBBP.... Differently, we organize the 3D mesh frames of a GOM in such style: IPPP.... What is the same in both styles is that the length of GOP/GOM is variable. What is different is that in our framework of 3D mesh sequences, the position of the frame I is variable; moreover, the corresponding frame I of every frame P of the sequence is variable in the multi-reference prediction coding. In following three subsections, we will describe all these methods.

Fig. 1: A sample of 3D mesh sequence.

3.2 TPS-based single-reference prediction

According to TPS transformation and the high correlations between the reference mesh and the target mesh, the higher quality of the approximate mesh is reconstructed. In a GOM, the middle frame shares the best correlations with its previous and following frames, so the middle frame selected as the reference frame I, and the others are frame P.

There is a 3D sequence $\{F_1, F_2, \ldots, F_m\}$ whose length is m, it is divided into a number of GOM. Set the length of each GOM is $2n-1$, the meshes of GOM are written as $\{F_1, F_2, \ldots, F_{2n-1}\}$. The mesh locating in the middle is F_n and the frame of F_n is the reference frame I. The others are frames P. Coding order is as shown in Fig. 2(a). In this kind of GOM structure, each frame P is transformed from frame I.

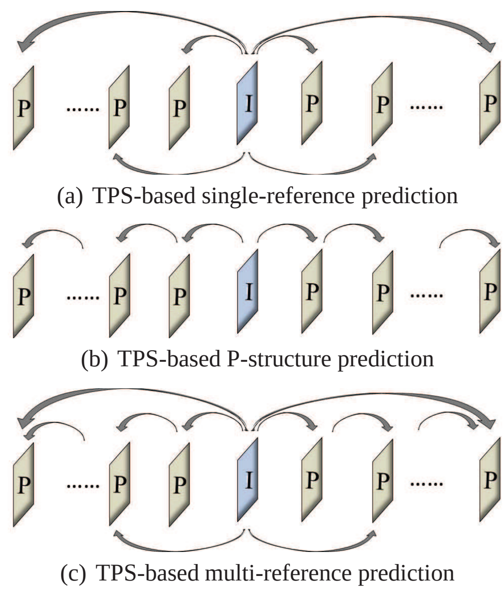


Fig. 2: Three prediction codings.

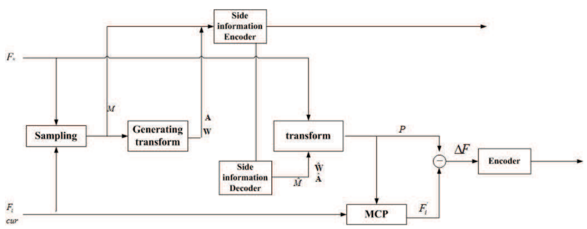
The coding framework is as shown in Fig. 3(a), the meshs of a GOM are

$$\{F_1, F_2, \ldots, F_{n-1}, \boldsymbol{F_n}, F_{n+1}, \ldots, F_{2n-1}\}.$$

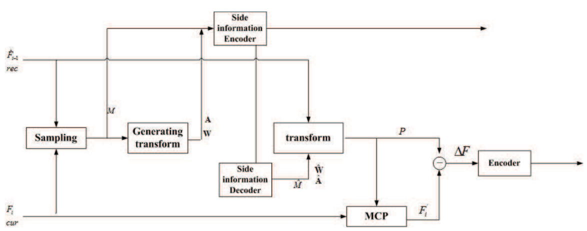
Set the mesh of current frame is F_i, and the mesh of the reference frame I is F_n. First sample the control points M. We uniform sample the control points in F_n set M_1, and using a classic K-dimensional binary search tree (KD-tree) [4] to search out corresponding closest points set M_2 in F_i. Then use the equation (3) and $M_1'^{\top} w = 0$, we get the TPS transformation matrices A and W. Continuing transformation, we can get the P, which has the same topological structure with F_n. We can find the nearest points in F_i of P by using a classic K-dimensional binary search tree (KD-tree) [4] which process we called matching closest point (MCP), and we get the F_i' which is the approximate mesh of F_i. Considering the correlations with P and F_i', we obtain the residuals ΔF.

Using the TPS transformations we get the transformational matrices A and W, control points M_1 and residuals ΔF. We use 7-zip to compress lossless F_n and the side information consisting of A, W and M_1, and use one-dimensional discrete wavelet transformation (1D DWT) to compress the residuals ΔF.

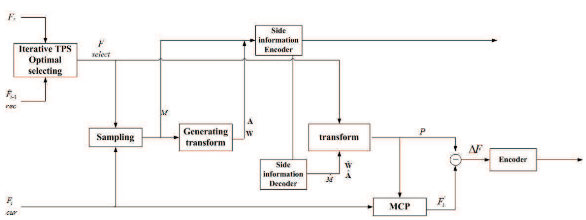
The decoding steps are given as follows, unzip the side information and $\widetilde{\Delta F}$, F_n. Using the TPS transformation with A, W, M_1 and F_n, we get P. Add the residuals $\widetilde{\Delta F}$ and P, the result is the reconstruction of the mesh $\widetilde{F_i}$.



(a) TPS-based single-reference prediction



(b) TPS-based P-structure prediction



(c) TPS-based multi-reference prediction

Fig. 3: The framework of three prediction coding.

3.3 TPS-based P-structure prediction

As described in Section 3.2, according to the theory of TPS transformation, the high correlations the reference frame and the target frame have, the higher quality of the approximate mesh can be reconstructed. We try to use the previous frame of current frame as the reference mesh of TPS transformation. Coding order is as shown in Fig. 2(b).

For a GOM, we still select the middle mesh F_n as the reference frame I. First, get corresponding target meshes F_{n-1}' and F_{n+1}' by using TPS transformation on the meshes F_{n-1} and F_{n+1} which are the previous and following frames of F_n. To obtain the biggest correlation, then we take decoded $\widetilde{F_{n-1}}$ and $\widetilde{F_{n+1}}$ as reference meshes for their adjacent frames in the GOM until every frame in the GOM is transformed. After the whole GOM is transformed by TPS, we will get the transformational matrices A and W, control points M_1 and residuals ΔF. We still use 7-zip to compress lossless F_n and the side information consisting of A, W and M_1, and use one-dimensional discrete wavelet transformation (1D DWT) to compress the residuals ΔF. In decoding, we also use the above mentioned parameters with TPS transformation, and add the middle mesh P and the residuals $\widetilde{\Delta F}$ to get the reconstruction of the mesh $\widetilde{F_i}$. Coding framework as shown in Fig. 3(b).

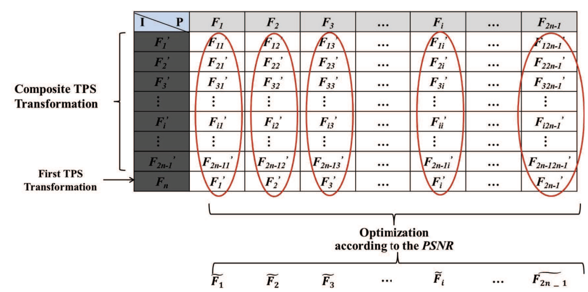


Fig. 4: The optimization process of I-frame.

3.4 TPS-based multi-reference prediction

Given the fact that the quality of the reconstruction mesh acquired from TPS transformation directly decides the quality of the decoded mesh, we consider to combine the methods of Sections 3.2 and 3.3, coding order is as shown in Fig. 2(c). Through iterative TPS transformations, we expect to improve the quality and compression ratio of the mesh reconstructed.

Obtain the reconstruction of the meshes of a GOM by using the transformation method of Section 3.2, and we get a set $F_1', F_2', \ldots, F_{2n-1}'$. Now we put these meshes F_j' in iterative TPS optimizer, and use every F_j' as the reference frame I of the mesh to TPS transformation with the every target mesh F_i, finally get mesh F_i' with contrast F_i and F_{ji}' in accordance the standard of optimization reconstruction mesh quality PSNR (later mention), that is shown in Fig. 4.

Until the whole GOM is transformed by TPS, we will get the transformational multiple matrices A and W, control points M_1 and residuals ΔF. We still use 7-zip to compress lossless F_n and the side information consisting of A, W and M_1, and use one-dimensional discrete wavelet transformation (1D DWT) to compress the residuals ΔF. In decoding, we also use the above mentioned parameters with composite TPS transformation, and add the middle mesh P and the residuals $\widetilde{\Delta F}$ to get the reconstruction of the mesh $\widetilde{F_i}$, Coding framework as shown in Fig. 3(c).

4 Experimental results

To verify the effectiveness of our proposed framework, this scheme has been implemented in a 3D mesh sequence of smiling facial expression from BU-4DFE: 3D dynamic facial expression database [12]. We select 17 frames of the more obvious expression changes, the frames are as a GOM which length is n (n=17), and the frame I is the middle ninth frame (m=9) and 200 control points are selected uniformly. The one-dimensional discrete wavelet transformation (1D DWT) is used to deal with the residuals in coding.

The objective quality of the 3D mesh sequence reconstructed is measured by Peak Signal to Noise Ratio (PSNR)

$$PSNR = 20\log_{10}\frac{peak}{d}$$

and the unit is dB, where the peak is the length of bounding box diagonal, d is the Hausdorff distance [5] between the original mesh and the reconstructed mesh. We measure PSNR for 3D mesh with 1024 bits per vertex (kbpv).

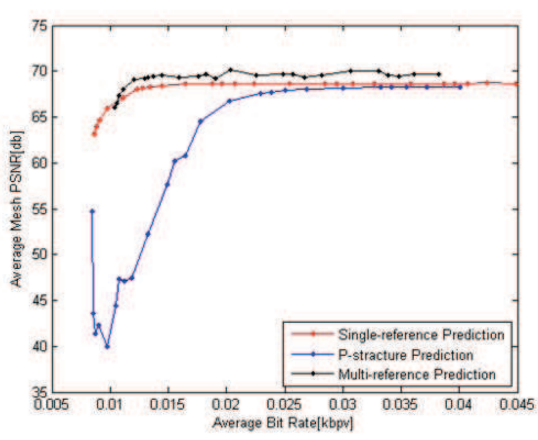


Fig. 5: The compression result of three frameworks.

We selected $50,40,30,20,10$ these five quantization steps to test our three coding frameworks proposed in this paper. And the values of quantitative accuracy are selected $1,2,3,4,5$. Take this to adjust the average bit rate and obtain the average PSNR in the different bit rate. In addition we use 1D DWT methods to compress each frame mesh individually, and we compare the subjective quality and compression efficiency for these methods. Then we compare the average PSNR of the three frameworks when the rate is between 0.005 to 0.045(kbpv), the experiment results show that, the average PSNR of TPS-based multi-reference prediction is better than other two frameworks. The result is as shown in Fig. 5. We can also find to obtain the higher compression performance by using our proposed frameworks and compressing the residuals using 1D DWT than compressing frame-by-frame compressing meshes using 1D DWT. Finally, we show that the compression quality of the frame of contrast value $\{F_1, F_6, F_{11}\}$, through the original mesh, the reconstructive mesh compressed by using our proposed framework of TPS-based multi-reference prediction and frame-by-frame compression by 1D DWT. This is shown in Fig. 6 and Fig. 7. Through comparing the subjective quality under average bit rate, we find that the average bit rate of our proposed framework is less than half of the method of frame-by-frame compression, when the subjective quality of our proposed framework is equal to or better than the frame-by-frame compression coding method.

Experimental results show that, compared with TPS-based single-reference prediction and P-structure prediction , the TPS-based multi-reference prediction coding has the higher average PSNR to the reconstruction of 3D mesh sequence under different bit rate. Comparing with frame-by-frame compression method, our proposed TPS-based reference prediction compression methods can obtain the lower compression average bit rate.

5 Conclusions

This paper proposes to compress a mesh sequence with different connectivities using a TPS-based multi-reference prediction coding. The prediction coding first selects the middle frame as the reference frame, and imposes the TPS transform for predicting all frames. The predicted frame is then used for the reference frame and re-imposed the TPS transform for predicting each frame. All the prediction results of a single frame are optimally selected using the PSNR results. Then we encode each mesh frame with the reference frame, a TPS transform matrix, the control points, and the residual term. We implement our scheme for compressing 3D facial mesh sequence. Experimental results show that the compression scheme produces promising results in both the subjective quality and the objective quality of reconstruction of the mesh sequence.

In future work, we shall generalize our scheme for handling more different kinds of models. We shall use

Fig. 6: The quality comparison of several 3D meshes frame in the higher average bit rate.

Fig. 7: The quality comparison of several 3D meshes frame in the lower average bit rate.

more effective control method to minimize the residuals, which is reasonably expected to achieve high reconstruction quality and lower compression bit rate.

References

[1] Z. Karni and C. Gotsman, Computers & Graphics **28**, 25-34 (2004).
[2] J. C. Carr, W. Fright, R. K. Beatson, IEEE Transactions on Medical Imaging **16**, 96-107 (1997).
[3] Y. L. Hu, M. Q. Zhou, Z. K. Wu, Proceedings of the 2nd International Congress on Image and Signal Processing 1-6 (2009).
[4] J. L. Bentley, Communications of the ACM **18**, 509–517 (1975).
[5] N. Aspert, D. Santa-Cruz, T. Ebrahimi, IEEE International Conference on Multimedia & Expo **1**, 705-708 (2002).
[6] M. Tejera and A. Hilton, Articulated Motion and Deformable Objects 12-25 (2012).
[7] Y. L. Hu, B. C. Yin, S. Q. Cheng, C. L. Gu, Proceedings of 2004 International Conference on Machine Learning and Cybernetics **7**, 4362-4367 (2004).
[8] J. H. Hou, L. P. Chau, Y. He, M. Q. Zhang, IEEE Transactions on Circuits and Systems for Video Technology **23**, 1537-1541 (2013).
[9] R. L. Harder, R. N. Desmarais, Journal of Aircraft **9**, 189-191 (1972).
[10] F. L. Bookstein, IEEE Transactions on Pattern Analysis and Machine Intelligence **11**, 567-585 (1989).
[11] M. Greenspan, M. Yurick, Proceedings of the 4th International Conference on 3D Digital Imaging and Modeling 442-448 (2003).
[12] L. J. Yin, X. C. Chen, Y. Sun, T. Worm, M. Reale, The 8th International Conference on Automatic Face & Gesture Recognition 1-6 (2008).

Yuan Gao is Ph.D. candidate at College of Metropolitan Transportation, Beijing University of Technology and Lecturer at Department of Communication Engineering, Beijing Electronic Science and Technology Institute. Her research interest covers multimedia, 3D mesh sequence compression, video image processing and codec.