3D SPATIAL RECONSTRUCTION AND COMMUNICATION FROM VISION FIELD

Xun Cao, Qifei Wang, Xiangyang Ji, Qionghai Dai

TNlist (Tsinghua National Laboratory for Information Science and Technology)
Department of Automation, Tsinghua University, Beijing, China, 100084

ABSTRACT

Vision field describes the real world visual information by summarizing the seven-dimensional plenoptic function into three domains: view, light, time, from which a better understanding of previous 3D capture and reconstruction systems can be provided. In this paper, we first show how to reconstruct 3D spatial information from all the three attributes of the vision field, namely full-space vision field reconstruction. Then, based on Laplacian iterative geometry prediction, a 3D mesh coding algorithm with cascaded quantization is presented to facilitate the communication of the reconstructed 3D models from vision field. At last, experimental results of both the 3D spatial reconstruction and the 3D mesh coding are demonstrated.

Index Terms**—**Three dimensional TV, Animation, Virtual Reality, Predictive Coding

1. INTRODUCTION

The real world we see is in three dimensions (3D) and the reproduction of the 3D spatial information has drawn more and more attention in recent years. To encode the real-world visual information, plenoptic function [1] gives us a ray-based explanation with seven variables to describe light rays received by human eyes. According to the plenoptic function, however, a very important dimension-*depth* is lost during the prevalent projective imaging procedure by common cameras. The absence of depth information presents much difficulty when we are trying to recover the 3D real world from the daily pictures. Consequently, ever since plenoptic function was proposed, a series of concepts have been further developed in the past two decades, aiming at re-build the 3D spatial information, among which some sparkling ideas like ray-space [3], light field [2], photometric stereo [10] have inspired more and more researches in the realm. An efficient and accurate 3D spatial reconstruction is still a major goal of many communities such as signal processing, computer vision and computer graphics.

In this paper, based on the concept of vision field [11][12], we show how to combine both the multiple camera views and multiple lights to achieve accurate 3D spatial reconstruction. We further extend the 3D reconstruction at single time instant over time axis to fulfill a dynamic modeling (or animation). The 3D reconstruction with all the three – view, light, time information of vision field is called full-space vision field reconstruction. On the other hand, storing and transmitting the 3D models are also crucial to the 3D applications such as 3D-TV, so we also provide the audience with a progressive 3D mesh coding algorithm with cascaded quantization. Experimental results show that our approach has better R-D (rate-distortion) performance than the state-of-the-arts. In the following, though the concept of vision field has been interpreted in our previous papers [11][12], we would like to start with a brief review of the vision field and its relationship with other methods.

2. VISION FIELD AND RELATED WORK

The plenoptic function is composed of seven variables, which includes the position of viewer in 3D space (x, y, z), the angle of shooting light (θ, ϕ), the wavelength of the light (λ), and the time instant (t). Vision field summarizes this representation by the observation that all the seven variables can be categorized into three domains: view, light, time. From this perspective, we compare the vision field and related concepts in Table 1, the DoF is the degree of freedom for variables, note the 3D reconstruction system column means that which attribute of the vision field can be manipulated by that method. A detailed description about all these plenoptic function's derivatives and related 3D reconstruction systems can be found in [12].

Table 1. the comparison between vision field and related concepts and systems.

Related Concepts	DoF	3D reconstruction system
plenoptic function [1]	7	no system presented
ray-space [3]	4	View
light filed [2]	4	View
FVV/FTV [4]	5	View, Time
reflectance field [5]	5	Light
opacity hulls [6]	5	View, Light
vision field [11]	7	View, Light, Time

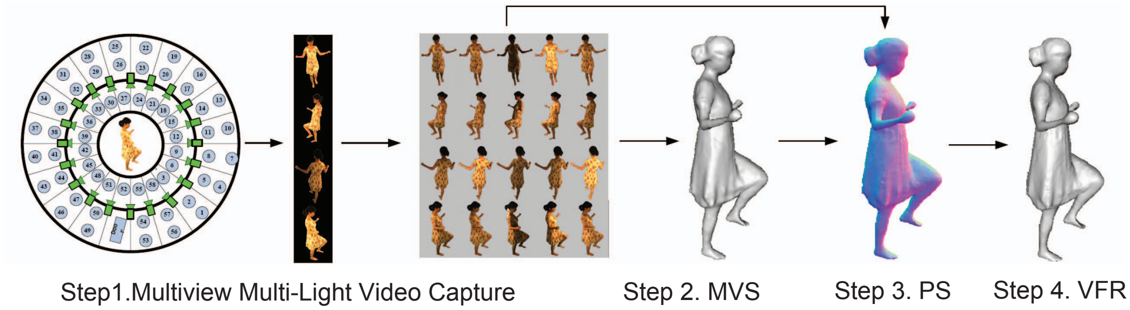

Figure 1. The flow of vision field reconstruction: (1). Multiview and Multi-light dynamic scene capture using our MVML capture system [9]; (2). Multiview View Stereo (MVS, in the *view* domain of the vision field); (3). Photometric Stereo (PS, in the *light* domain of the vision field); (4). Vision Field Reconstruction (VFR, combine MVS and PS over *time* domain).

3. VISION FIELD RECONSTRUCTION

In our previous work [12], we have demonstrated how to implement 3D reconstruction by each subspace of the vision field, for example, if only the view subspace is applied, it is the classical multiview stereo (MVS) problem [8]. Similarly, if only the light subspace is utilized, it is an emerging technique called photometric stereo (PS) [10], which can recover the 3D surface by integrating the normal map obtained by lights from different directions. It's intuitive to think: can more accurate 3D reconstruction be achieved if we combine both the view and light information? The answer is positive based on our experiments of using both MVS and PS techniques [7]. Here, we further extend the 3D reconstruction over time axis by tracking the temporal correspondences.

As illustrated in Figure 1, the multiview video sequence is first captured by our vision field capture system [9] with varying illumination to control the light from different directions. Then, the MVS and PS are performed in the *view* domain and *light* domain respectively. Optical flow is employed to build the correspondences over *time* axis under varying illumination; this not only enables us to re-build 3D models for dynamic scene, but also provides us with more accurate 3D reconstruction results than only in one or two subspaces of the vision field, as verified in [13]. Recovering 3D information in all the *view-light-time* domains is called full-space vision field reconstruction.

4. PROGRESSIVE 3D MESH CODING WITH CASCADED QUANTIZATION

The 3D models generated by the vision field are represented by meshes. However, 3D meshes often cause a huge amount of data, which presents great difficulty in storage and transmission. To efficiently compress and transmit the 3D meshes, many 3D mesh coding algorithms were proposed to reduce the redundancy in the raw data format [14]. Multi-Resolution 3D Mesh Coding (MR3DMC) [15] was proposed to provide high performance as well as scalabilities. In this coding algorithm (Triangle FAN-based compression-TFAN), 3D meshes were transmitted and rendered with different levels of details. One of the most significant MR3DMC schemes is PTFAN [16]. PTFAN is a multi-resolution extension of the TFAN technique, which supports both quality and spatial scalability and outperforms most of the state-of-the-art coding algorithms in rate-distortion (R-D) performance. However, since PTFAN does not fully exploit the Laplacian iterative geometry prediction, the quality scalability in PTFAN is not sufficiently flexible to meet the transmission bandwidth variation. Besides, the maximum decoding error does not decline significantly until receive all the bit-stream of the current resolution layer. To achieve flexible quality scalability, we propose a 3D mesh coding algorithm with cascaded quantization based on PTFAN. In the proposed coding algorithm, the mesh geometry prediction residuals are partitioned into a number of iterative layers. Each iterative layer is split into several quality layers to provide flexible quality scalability. To improve the R-D performance, the quantization parameter of the first quality layer is determined by the importance of the corresponding iterative layer. Our 3D mesh coding algorithm with cascaded quantization can also be applied in other progressive 3D mesh coding framework.

4.1. Laplacian Iterative Geometry Prediction

In [18], the Laplacian matrix is proposed to derive the least-squares mesh. Denote $G = (V, E)$ as the given mesh graph G, V represent the vertices and E denotes the edges. The Laplacian matrix L of the 3D mesh can be defined as:

$$L = \begin{cases} 1 & i = j \\ -1/d_i & (i,j) \in E \\ 0 & other \end{cases} \quad (1)$$

where d_i denote the valence of the vertex i. In [19], based on the assumption that each vertex lies in the center of gravity of its immediate neighbors, a fairness and smoothness condition for a vertex is defined as:

$$v_i - \frac{1}{d_i} \sum_{j:(i,j)\in E} v_j = 0 \tag{2}$$

where v_i and v_j denote the geometry location of vertex i and j. The least-squares mesh is derived by solving the linear equations [6].

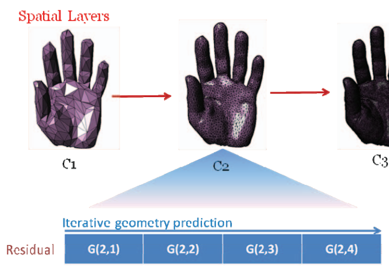

Figure 2. Iterative geometry prediction in PTFAN.

As shown in Figure 2, PTFAN simplifies the 3D mesh into three resolution layers. In the first layer, it uses all the vertices of the closest lower resolution layer as an initial set of control vertices to generate a smooth least-square mesh. In the second layer, the prediction residue is partitioned into four clusters to perform Laplacian iterative geometry prediction. During the encoding process, the control vertices set is iteratively updated with each refined clusters.

4.2. Cascade Quantization

In our proposed cascade quantization, each iterative layer is split into several quality layers. The quantization step-size of each layer is determined by the importance of the iterative layer. Since the vertices in each iterative layer are predicted from the vertices in its neighboring lower iterative layer, the distortions of vertices in the lower layer will propagate to the current decoded vertices. Therefore, the vertices at the lower quality layer are more important than those in the higher quality layer. Figure 3 illustrates such quantization framework. In Figure 3, the $QP(n,l)$ denote the corresponding quantization parameters QP of the n^{th} quality layer and l^{th} iterative layer, the details about how to compute the cascaded quantization parameters are included in [17].

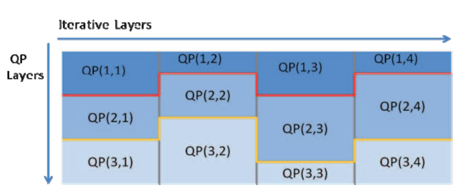

Figure 3. Cascaded quantization framework.

The proposed progressive 3D mesh coding algorithm with cascade quantization is generally implemented as follows:

1) Suppose the total vertices number is m, the number of iterative layer is L. Thus, the refined vertices number of each iterative layer is defined by the following function as:

$$c_i = \begin{cases} \frac{1}{3} \times \frac{m}{2^{L-2}} & if\ i = 1 \\ \frac{2}{3} \times \frac{m}{2^{L-2}} & if\ i = 2 \\ 2^{i-1} \times \frac{m}{2^{L-2}} & if\ i > 2 \end{cases} \tag{3}$$

2) Assign the quantization step-size of each quality layer with the pre-defined QP of the lowest quality layer. Generally, the QP of the lowest quality layer is set according to importance of the current iterative layer which is studied in [17].
3) Perform cascaded quantization on the geometry prediction residue. Subsequently, encode and transmit the quantized coefficient from high quality layer to low quality layer. For each quality layer, the quantized coefficients are encoded and transmitted in the increasing order of the iterative layer index.

5. EXPERIMENTAL RESULTS

The vision field construction is compared with typical 3D reconstruction methods which belong to only one or two subspaces of vision field. We found that the VFR has better performance on recovered 3D model details by different datasets. Due to the page limitation, interested readers may watch the video demo of VFR results on various video sequences on [21].

As for the proposed 3D mesh coding algorithm with cascaded quantization, we evaluate the R-D performance on the two static meshes: *Fish* and *Feline*. The features of these models are listed in Table 2. The meshes are split into three resolution layers. In cascaded quantization, the numbers of the iteration layer and quality layer are set to 4 and 2, respectively. The QPs of the two quality layers are set to (4, 16, 16, 16) and (4, 4, 4, 4), respectively. The MAX error metric defined in [20] is adopted to measure the distortion of the reconstructed 3D meshes. The R-D curves of the 3D meshes are plotted in Figure 4.

Table 2. features of the test 3D meshes.

Model	Number of vertices	Number of triangles
Fish	24980	49956
Feline	49864	99732

Fig. 4 shows the MAX error/bit-rate performance of the proposed coding algorithm, PTFAN, and the scalable 3D mesh compression standard (SC3DMC). In the simulation, the MAX error/bit-rate performance is obtained at different bit-rate levels. As the curves shown in Fig. 4, the proposed coding algorithm achieves significant bit-rate savings than SC3DMC, especially at low bit-rate case. Compared with PTFAN, the proposed coding algorithm can improve the MAX error/bit-rate performance within a certain bit-rate ranges. Moreover, our coding algorithm suffers lower distortion before receiving all the bit streams in any resolution layer.

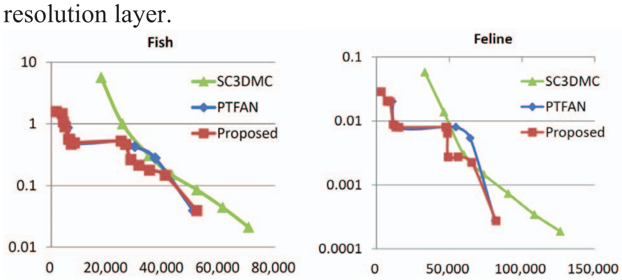

Figure 4. MAX error/bitrate for *Fish* and *Feline* of different coding algorithms.

6. CONCLUSIONS

Following our previous work [11][12], we continue to introduce the concept of vision field and first demonstrate how to recover 3D scenes using full-space vision field. The results shows that full-space vision field reconstruction can achieve better performance than using just one or two subspaces of the vision field. A progressive 3D meshing coding algorithm with cascaded quantization is also presented in this paper as an important module for the future applications requiring 3D communication.

ACKNOWLEDGMENT

The authors would like to thank the support of the National Basic Research Project (No.2010CB731800), the Project of NSF China (No.61035002 & 61170194).

REFERENCES

[1] E H. Adelson, J R. Bergen. “The Plenoptic Function and of Early Vision”, Computational Models of Visual Processing, MIT Press, Cambrige, MA, 1991:3-20

[2] M.Levoy, P.Hanrahan. “Light field rendering”, International Conference on Computer Graphics and Interactive Techniques (Proc.SIGGRAPH), 1996.

[3] T. Fujii, T.Kimoto, and M.Tanimoto, “Ray Space Coding for 3D Visual Communication,” Picture Coding Symposium (PCS), 447-451, 1996

[4] M. Tanimoto, “Overview of FTV (Free-Viewpoint Television)”, Proceedings of the IEEE international conference on Multimedia and Expo, 2009: 1552-1553

[5] P. Debevec, T.Hawkins, C.Tchou, H-P. Duiker, W.Sarokin, M.Sagar. “Acquiring the Reflectance Field of a Human Face”, Proceedings of the 27th annual conference on Computer Graphics, 145-156. 2000

[6] W.Matusik, H.Pfister, A.Ngan, "Image-based 3D Photography using Opacity Hulls". ACM Transactions on Graphics. 21(3) 2002

[7] C.Wu, Y.Liu and Q.Dai, "Fusing Multiview and Photometric Stereo for 3D Reconstruction under Uncalibrated Illumination", IEEE Trans. Visualization and Computer Graphics.(In Press)

[8] Y.Liu, X.Cao, Q.Dai and W.Xu, "Continuous Depth Estimation for Multi-view Stereo", IEEE Int'l Conf. Computer Vision and Pattern Recognition(CVPR), 2009

[9] http://media.au.tsinghua.edu.cn/dome.jsp

[10] L. Zhang, B. Curless, A. Hertzmann, S. M. Seitz, ”Shape and Motion under varying illumination: Unifying Stucture from Motion, Photometric Stereo, and Multi-view Stereo”, Proc. IEEE int’l conf. Computer Vision, 2003

[11] Q.Dai, X.Ji, X.Cao, "Vision Field Capturing and its applications in 3DTV" , in Picture Coding Symposium, 2010,

[12] X.Cao, Y.Liu, Q.Dai, "Vision Field Capture for advanced 3DTV applications ", Visual Communication & Image Processing (VCIP), 2011

[13] Y.Fu, Y.Liu, and Q.Dai, "Dynamic Shape Capture via Periodical-illumination Optical Flow Estimation and Multi-view Photometric Stereo", International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT, , 2011

[14] J. Peng, C.-S. Kim, and C.-C. Jay Kuo, “Technologies for 3D mesh compression: A survey,” J. Vis.Commun. Image Rep., vol. 16, no. 6, pp. 688-733, Apr. 2005.

[15] K.Mamou, T.Zaharia, and F.Prêteux, “TFAN: A low complexity 3D mesh compression algorithm,” Journal of Visualization and Computer Animation, vol. 20, no. 2-3, pp. 343-354, Jun. 2009.

[16] K. Mammou and C. Dehais, “PTFAN: a multi-resolution extension for the TFAN codec,” in ISO/IEC JTC1/SC29/WG11, MPEG2010/17178, Kyoto, Japan, Jan. 2010.

[17] L.Zhang, X.Ji, Q.Dai, and K.Mamou, “Results of core experiment CE3: SNR scalable coding based on PTFAN,” in ISO/IEC JTC1/SC29/WG11, MPEG2010/M17943, Geneva, Switzerland, Jul. 2010.

[18] Z. Karni and C. Gotsman, “Spectral compression of mesh geometry”, in: ACM SIGGRAPH, pp. 279-286, Jul. 2000.

[19] D. Chen, D. Cohen-Or, O. Sorkine, and S. Toledo, “Algebraic analysis of high-pass quantization,” ACM Transactions on Graphics, vol. 24, no. 4, pp. 1259-1282, Oct. 2005.

[20] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, “MESH: Measuring Error between Surfaces using the Hausdorff distance,” in Proc. of IEEE Int. Conf. Multimedia and Expo (ICME), Jun. 2002, pp. 705-708.

[21] http://media.au.tsinghua.edu.cn/vfr.jsp