

Image-Based Compression, Prioritized Transmission and Progressive Rendering of Circular Light Fields (CLFS) For Ancient Chinese Artifacts

X. Z. Yao, S. C. Chan, Z. Y. Zhu, K. T. Ng

Dept. of Electronic and Electrical Engineering, The University of Hong Kong

{xyzao, scchan, zyzhu, ktng}@eee.hku.hk

H. Y. Shum

Microsoft Corporation

hshum@microsoft.com

Abstract—This paper proposes an efficient algorithm for the compression, prioritized transmission and progressive rendering of circular light field (CLF) for ancient Chinese artifacts. It employs wavelet coder to achieve spatial scalability and divide the compressed data into a lower resolution base layer and an additional enhancement layer. The enhancement layer is coded as in JPEG2000 into packets where the base-layer is coded using disparity compensation prediction (DCP). The frame structure is designed to provide efficient access to the compressed data in order to support selective transmission and decoding. The depth and alpha maps are coded analogously. A prioritized transmission scheme which support interactive progressive rendering is also proposed to further reduce the latency and response time of rendering.

Keywords—Image-based rendering, circular light field, compression, prioritized transmission/rendering.

I. INTRODUCTION

Image-based rendering/representation (IBR) [1-7, 11] is an emerging and promising technology for rendering new views of scenes from a collection of densely sampled images or videos. It has potential applications in virtual reality, immersive 3D/multiview television and advanced visualization systems. Important IBR representations include the 2D panorama, plenoptic modeling, light fields [2], lumigraphs [3], 3D concentric mosaics, ray-space representation, plenoptic videos (PVs) [4, 7], etc.

With the advance of technology, 3D and multi-view displays are becoming available commercially. This breakthrough motivates us to study in this paper on the important problem of cultural heritage preservation and dissemination of ancient Chinese artifacts using the image-based approach. To avoid possible damages to the artifacts and to speed up the capturing process, we propose in [17] to employ the image-based approach instead of using 3D laser scanners. A circular array consisting of multiple digital still cameras (DSCs) was constructed to capture the circular light fields (CLFs) of the objects. We also developed novel techniques for rendering new views of the artifacts from the images captured using the object-based approach [5]. The multiple views so synthesized enable the ancient artifacts to be displayed in modern multi-view displays. A number of ancient Chinese artifacts from the University Museum and Art

Gallery at the University of Hong Kong were captured and excellent rendering results are obtained.

Since IBR representations usually have huge data sizes, efficient compression techniques [4, 6-8, 10-16] are essential to their storage and transmission. While classic compression techniques, such as the MPEG standard and other extensions could achieve good compression efficiency; careful modifications are usually required to address the random access and selective decoding requirements in IBR. In this paper, we study the compression and prioritized rendering of the CLFs captured in [17]. In particular, we propose a coder which provides efficient access to compressed data to support selective decoding, prioritized transmission, and progressive rendering.

The proposed coder employs wavelet coder to achieve spatial scalability and divide the compressed data into a lower resolution base layer and an additional enhancement layer, which contains the remaining high frequency wavelet coefficients. The enhancement layer is coded as in JPEG2000 [9] into packets. This enables the enhancement layer to be progressively transmitted and decoded. The base-layer is arranged into group of pictures (GOPs) with pictures which are either intra (I-) coded, predicted- (P-) or bi-directionally (B-) predicted. To provide efficient access to the compressed data, the frame structure in [4, 5, 11] is utilized where I-pictures in each GOP are used as references for predicting P- or B-pictures. B-picture has higher compression ratio due to bi-directional prediction but their decoding complexity is also slightly higher than the P-pictures, which only require one reference picture for decoding. Here, we only employ P-pictures to simplify the selective decoding/transmission. Instead of using MPEG-2 as in [4, 5, 11] for coding the P- and B-pictures in the base layer, the H.264 standard is employed because of its better coding efficiency. The depth and alpha maps are coded analogously. A prioritized transmission scheme which supports interactive progressive rendering is also proposed. The rest of the paper is organized as follows: in section II, the proposed system and associated techniques concerning the capturing and rendering of the CLFs for the ancient Chinese artifacts are reviewed. Section III is devoted to the proposed compression algorithm of the circular light

field (CLF). Experimental results are presented in Section IV and final conclusions are drawn in Section V.

II. THE PROPOSED CLF SYSTEM

Central to IBR is the plenoptic function [1], which describes all the radiant energy that can be perceived by the observer at any point in space and time. The plenoptic function is a 7-dimensional function of the viewing position, the azimuth and elevation angle, time, and wavelengths. Traditional images and videos are 2D and 3D special cases of the plenoptic function. Depending on the functionalities required, there is a spectrum of image-based representations [6]. They differ from each other in the amount of geometry information being used.

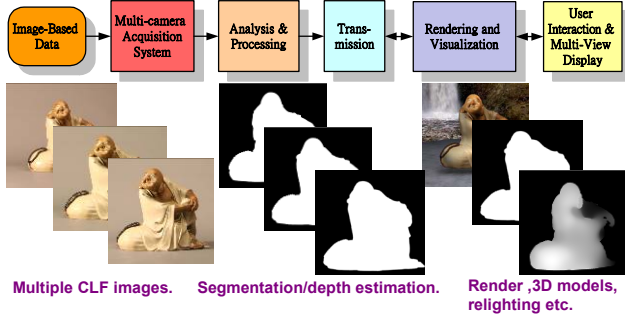


Fig.1 Image-based rendering of Ancient Chinese Artifacts

In [17], we proposed a multiple camera approach for image-based rendering of ancient Chinese artifacts. It employs the concept of pop-up light field and object-based plenoptic videos [5] to synthesize new views of the artifacts from images captured by a circular array of cameras. To reduce rendering artifacts, the objects in [5] are extracted using semi-automatic segmentation and tracking techniques. From the segmented objects, approximate depth information for each IBR object can be estimated to render new views at different viewpoints. The operations are illustrated in Fig. 1.



Fig.2 Circular camera array constructed

Fig. 2 shows the circular camera array that we have constructed, which consists of 13 canon 450D digital still cameras with an angular spacing of 3 degrees and a radius of 3 meters. Before the cameras can be used for depth estimation and image-based rendering, they are calibrated to determine the intrinsic parameters as well as their extrinsic parameters. In [17], the photometric invariant features were first employed to extract the foreground from the monochromatic screen background. Then, matting was applied to obtain soft segmentation information of the object in forms of the alpha maps. Example CLF image, alpha map and depth map

extracted are shown in fig. 3. Once the alpha and depth maps have been estimated, virtual views can be synthesized. Possible holes are filled by in-painting.



Fig.3 Example (a) CLF image and corresponding (b) alpha map and (c) depth map

III. COMPRESSION AND PRIORITIZED TRANSMISSION/RENDERING OF CLFS

Since the CLFs have high resolution and involve large amount of data, they need to be compressed before transmission or storage. In the receiver, the images are decompressed and rendered online according to the user's interaction. In our case, we need to transmit the texture and associated alpha and depth maps to the receiver for rendering. Considerable effort has been devoted to the efficient compression methods of various image-based representations such as the light fields, lumigraphs and concentric mosaics [4-7, 12, 13]. Apart from good coding efficiency, it is important for the compression algorithm to provide efficient access to the compressed light field data so that they can be selectively decoded to reduce the storage requirement and decoding complexity [4, 6, 12]. Moreover, a prioritized transmission scheme using scalable coding technique is highly desirable in order to reduce the latency in launching the application and improve the response time. We now propose a new CLF compression algorithm, which offers the two desirable properties above.

A. Spatial scalability

The Scalable coding, in forms of spatial scalability, SNR, and temporal scalabilities, is an effective method to prioritize the transmission of compressed information to support users with different bandwidth/decoding complexities and to reduce the latency in progressive viewing. This is particular desirable in IBR compression because the data size is large and the data is highly correlated. The proposed coder incorporates spatial scalability in our previous proposed PV coding scheme [4, 5, 11] to support users with different bandwidth/decoding complexity and to reduce the latency in interactive rendering through prioritized transmission of the compressed CLF data. In the proposed coder, we first perform wavelet transform on the light field images so as to produce a multi-resolution representation of the CLF images as shown in Fig. 4. The low resolution images form the base layer while the higher frequency wavelet coefficients form the enhancement layer. The base layer is intended to provide a reasonable rendering quality of the artifacts to the users. Together with the enhancement layer, high quality rendering at full resolution

can be achieved. The base layer can serve to reduce the transmission bandwidth during initial launching of the application or heavy network traffic due to its reduced transmission bandwidth. We now describe the compression of the base and enhancement layers to support selective decoding/transmission. Using these functionalities, we can then perform prioritized transmission to reduce the loading latency and improve response time.

B. Coding of the Layers to support selective transmission/decoding

As can be seen from Figure 4, the base layer consists of a sequence of light field images with reduced resolution. To facilitate efficient access and selective decoding of the image data, we choose to avoid the excessive inter-dependency in predicting the light field images as in other approaches [10]. Therefore, we adopt a similar frame structure in [4,5,12] where images are either coded as intra (I-) pictures, predicted (P-) or bi-directionally predicted (B-) pictures which are predicted from the reference I-pictures. In our case, we adopt the JPEG2000 standard to compress the I-pictures and then divide the compressed bit stream into the base and enhancement layers. This approach has the advantages that the I-pictures can be decoded by conventional JPEG2000 standard for preview etc. The enhancement layer, which consists of the high frequency wavelet coefficients and the residual in the base layer, is coded as in JPEG2000 standard.

To explore the correlation between consecutive light field images, the encoded I-pictures are then used as reference for predicting adjacent P-pictures. To avoid too much inter-dependency between the images which will greatly increase the complexity in selective decoding of desired images for rendering, we do not employ P-pictures as reference. A group of consecutive light field images are grouped together to form a group of pictures (GOP). Inside each GOP, I-pictures are used as references for predicting P- or B-pictures. B-pictures have higher compression ratio due to bi-directional prediction but their decoding complexity is also slightly higher than the P-pictures, which only require one reference picture for decoding. In Figure 4, the light field images are divided into GOPs, each consists of 3 consecutive images. Here, we only employ P-pictures to simplify the selective decoding/transmission. Other configurations can be used at the expense of slightly higher decoding complexity, but possibly with a better coding efficiency. Instead of using MPEG-2 as in [4] for coding the P- and B-pictures, we have adopted the H.264 standard because of its better coding efficiency. The depth and alpha maps are coded analogously.

C. Prioritized transmission for interactive progressive rendering

Since the light field data is huge, it may take long time to transmit even for the base layer before the rendering application can be launched. Usually, the users may wish to start rendering or viewing once the application is invoked. Under these circumstances, it is important to selectively

transmit the compressed data according to some priority schemes in order to satisfy the rendering requirement of the users as much as possible. As most users would generally start viewing from the center and gradually move to the two ends, the center views of the light field images should be given a relatively higher priority for transmission. For instance, suppose that the first light field image to be displayed is the right center one, i.e. V7. It is natural to transmit the base layer of V7 first, followed by those of V6 and V8 in the same GOP. To improve rendering speed while avoiding excessive storage for the light field images, the decoded intra-pictures are stored in the memory while the P-pictures are stored in their compressed form. After complete transmission of this GOP, the base layer of the next I-picture such as V4 can be transmitted, followed by their adjacent P-pictures in the same GOP and vice versa. This is the normal transmission order.

During the transmission, if the user wishes to render a new view at a certain position, then higher priority will be given to those data required to render that view, with the intra-pictures given the higher priority. For instance, if a view between V1 and V2 is required, then V1 will be given the highest priority for transmission followed by V2. In order to reduce the overheads to specify the location of the compressed data due to switching from one view to the other, the compressed data in the I- and P-pictures are divided into packets. Transmission of the highest priority view will begin only after the completion of the current packet. It can be seen that the proposed coding scheme greatly improves the effectiveness of this prioritized transmission scheme due to the independency of the GOPs and the reduced dependency of the inter-coded pictures inside each GOP. Otherwise, if the pictures are predicted one by one, then the entire compressed light fields have to be transmitted before rendering can proceed.

IV. EXPERIMENTAL RESULTS

We now evaluate the performance of the proposed system. The resolution of the light field image is 1088×1472 ; 2-level of wavelet transform is performed. The enhancement layer consists of the compressed high frequency wavelet coefficients and they are organized as packets to facilitate interactive prioritized transmission. The coding results are shown in Figure 5, where the PSNR versus BPP (bits per pixel per frame) is plotted. For clarity of presentation, the averaged coding performance for a) base layer for I-pictures, b) base layer for P-pictures, c) base plus enhancement layers for I-pictures, and d) base plus enhancement layers for P-pictures are separately shown. For the base layer, it can be seen that the P-pictures have better coding efficiency than the I-pictures due to the effectiveness of disparity compensation/prediction (DCP).

Results for the coding of alpha and depth maps are also presented in Figure 5. They have considerably better coding efficiency compared to the texture image, as more redundancies exist in the alpha and depth maps.

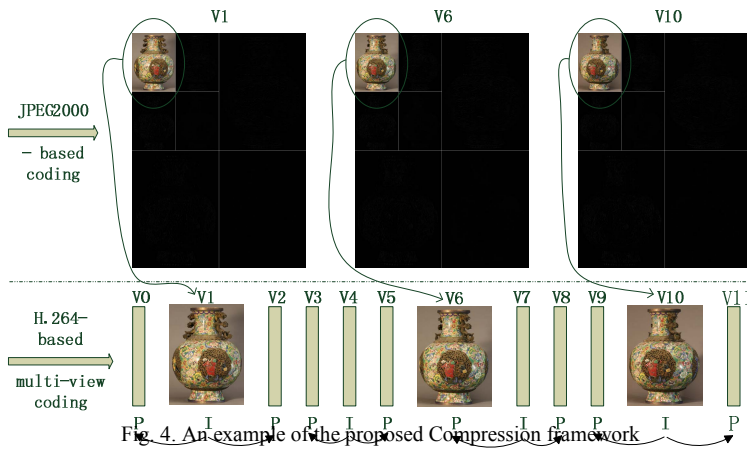


Fig. 4. An example of the proposed Compression framework

V. CONCLUSION

In this paper, an efficient CLF compression algorithm for ancient Chinese artifacts has been presented. Apart from good coding efficiency, the proposed coder offers the desired properties of selective decoding and progressive rendering to reduce the storage and decoding complexity. Based on the proposed GOP structure and spatial scalability, a prioritized transmission scheme is also proposed to further reduce the latency and response time of rendering by taking into account the user interaction.

REFERENCES

- [1] E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, pp. 3-20, MIT Press, Cambridge, MA, 1991.
- [2] M. Levoy and P. Hanrahan, "Light field rendering," *SIGGRAPH'96*, pp. 31-42, Aug. 1996.
- [3] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," *SIGGRAPH'96*, pp. 43-54, Aug. 1996.
- [4] S. C. Chan, K. T. Ng, Z. F. Gan, K. L. Chan and H. Y. Shum, "The plenoptic videos," *IEEE Trans. CSVT*, vol. 15, no. 12, pp. 1650-1659, Dec., 2005.
- [5] S. C. Chan, Z. F. Gan, K. T. Ng, H. Y. Shum and K. L. Ho, "An Object-Based Approach to Image/Video-Based Synthesis and Processing for 3-D and Multiview Televisions," *IEEE Trans. on CSVT*, vol. 19, no. 6, pp. 821-831, June 2009.
- [6] H. Y. Shum, S. C. Chan and S. B. Kang. *Image-based rendering*. Springer, 2007.
- [7] S. C. Chan, H. Y. Shum, and K. T. Ng, "Image-based rendering and synthesis: technological advances and challenges," *IEEE Signal Processing Magazine: Special Issue on MVI and 3DTV*, Nov, 2007.
- [8] A. Smolic, et al "Coding algorithms for 3DTV—A survey," *IEEE Trans. CSVT*, vol. 17, no. 11, pp. 1606-1621, Oct. 2007.
- [9] D. Taubman and M. W. Marcellin, *JPEG2000 Image Compression: Fundamentals, Standards and Practice*, Kluwer Academic Publishers, 2002.
- [10] T. Wiegand et al, "Overview of the H.264/AVC video coding standard," *IEEE Trans. CSVT*, vol. 13, pp. 560-576, July 2003.
- [11] K. T. Ng, S. C. Chan, H. Y. Shum, and S. B. Kang, "On the data compression and transmission aspects of panoramic video," in *Proc. of IEEE ICIP*, vol. 2, pp. 105-108, Oct. 2001.
- [12] J. Li, H. Y. Shum, and Y. Q. Zhang, "On the compression of image based rendering scene: a comparison among block, reference and wavelet coders," *Int. J. Image Graph.*, vol. 1, no.1, pp. 45-61, 2001.
- [13] M. Magnor and B. Girod, "Data compression for light-field rendering," *IEEE Trans. CSVT*, vol. 10, no. 3, pp. 338-343, Apr. 2000.
- [14] ISO/IEC JTC1/SC29/WG11, "Survey of Algorithms Used for Multi-view Video Coding (MVC)," Doc. N6909, Hong Kong, January 2005.
- [15] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *IEEE Trans. CVST (Special Issue 3DTV MVC)*, vol. 17, no. 11, pp. 1461-1473, Oct. 2007.
- [16] M. Flierl and B. Girod, "Multiview video compression," *IEEE Signal Processing Mag.*, vol. 24, no. 7, pp. 66-76, Nov. 2007.
- [17] Z. Y. Zhu, K. T. Ng, S. C. Chan and H. Y. Shum, "Image-Based Rendering of Ancient Chinese Artifacts for Multi-view Displays - a Multi-Camera Approach," to appear in *Proc. IEEE ISCAS 2010*.

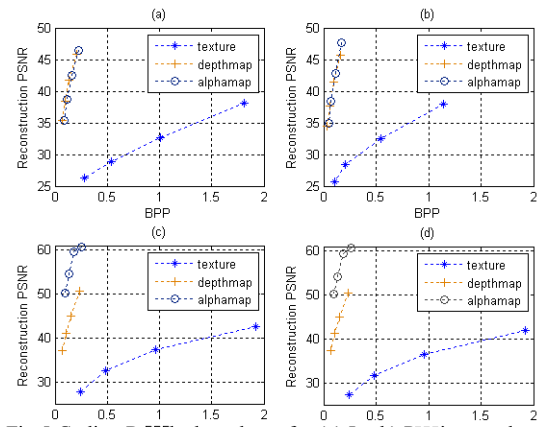


Fig.5 Coding Results base layer for (a) I- , b) P-pictures, base plus enhancement layers for (c) I- pictures, and d) P-pictures.