THE COMPRESSION ISSUSES OF PANORAMIC VIDEO

King-To Ng, Shing-Chow Chan

Department of Electrical and Electronic Engineering,
The University of Hong Kong.
{ktng, scchan}@eee.hku.hk

Heung-Yeung Shum

Microsoft Research, China.

hshum@microsoft.com

ABSTRACT

This paper proposes efficient data compression techniques for panoramic video. Panoramic videos have been used as a means for representing dynamic scenes or paths along a static environment. They allow the user to change viewpoints interactively at a point in time or space. High-resolution panoramic videos, while desirable, consume a significant amount of storage and bandwidth for transmission, and make real-time decoding very computational intensive. A high performance MPEG-like compression algorithm, which takes into account the random access requirements and the redundancies of the panoramic video, is presented. The transmission aspects of panoramic video over cable network, LAN and Internet are also briefly discussed.

1. INTRODUCTION

Image and videos are effective means of conveying information of objects and scenes. With increasing demand for better user experience in interactive applications such as virtual walkthrough, computer games, and medical simulation, virtual reality techniques are becoming increasingly more important. Image based rendering (IBR) using the plenoptic function [1] has recently emerged as a simple yet powerful photo-realistic representation of real world scenes [2-6]. Its basic principle is to render new views of a scene using rays that were previously captured in densely sampled pictures taken from the scene. A simple example of such IBR representation is the panorama. During rendering, part of the panoramic image is re-projected onto the screen to emulate the effect of panning and zooming. Other more sophisticated representations include the Light Field [4], Lumigraph [5], and Concentric Mosaics [6].

Most IBR representations are of static scenes. This is largely attributed to the logistical difficulties in capturing and transmitting dynamic representations, which involve huge amounts of data. The compression of the Light Field, Lumigraph, and Concentric Mosaics were addressed in [4,7-9]. It is envisioned that data compression will continue to be an important issue in IBR. More recently, panoramic videos are used to capture dynamic scenes for applications such as tele-presence and autonomous vehicles [11-13]. Much emphasis was placed on the construction of the panoramic video and how they can be constructed and rendered. Behere (http://www.behere.com) is one of the first companies offering streaming panoramic video. Nevertheless, the amount of data associated with panoramic videos can be very high, which poses a problem when good resolution and interactive speeds are required.

In this paper, we propose efficient means for compression and transmission of high-resolution panoramic videos. To see the severity of the problem of transmission, consider a 2048 × 768 panoramic image, which occupies about 4.5 MB. A 25 frames/sec video at this resolution would require 112.5 MB/s of transmission bandwidth. The transmission bandwidth and digital storage can be significantly reduced if data compression techniques are used. Another problem associated with high-resolution panoramic videos is its high computational complexity in software-only real-time decoding. To remedy these problems, a high performance MPEG-like compression algorithm, which takes into account these requirements and the redundancies of the panoramic video, is presented. The transmission aspects of panoramic video over cable network, LAN and Internet are also briefly discussed.

The paper is organized as follows: the principle of panoramic video, its construction and rendering are discussed in Section 2. Section 3 is devoted to the proposed compression algorithm. The transmission issues of panoramic video will then be discussed in Section 4. We provide concluding remarks in Section 5.

2. PANORAMIC VIDEO

2.1. Panoramic Mosaic

A panoramic mosaic is a high-resolution image obtained by projecting a series of images (registering and stitching), taken on a plane when a camera is moving along a given axis, on a cylindrical or spherical surface. Figure 1 shows the construction of a panoramic mosaic. Because it is obtained by stitching several images together, its resolution is usually very large (e.g. 2048×768). Several algorithms for constructing such mosaic or panoramas were previously reported in [2,14,15]. Using the panorama, it is possible to emulate "virtual camera panning and zooming" by projecting the appropriate portions of the panorama onto the user's screen [2].

2.2. Capturing of Panoramic Video

As mentioned earlier, a time-varying environmental map can be obtained by taking panoramas at regular time interval either at a given location or along a trajectory. Such time-varying environmental map or panoramic video closely resembles a video sequence with very high resolution. There are different methods for capturing a panorama video [11,12]. For example, in the FlyCam system [11], multiple cameras are mounted on the faces of an octagon 10cm wide. While in [12], the camera is fitted with a mirror to provide the panoramic video.

In this paper, we employ the panoramic video reported in [16]. It was captured using an omni-directional setup [16]

comprising a catadioptric omni-directional imaging system [17] with a 1300×1100 pixel camera, all placed on a movable cart. To capture the panoramic video, four video streams of the omni-directional video at different camera orientations (front, left, back, right) along the same path were taken. This was done because each omni-directional image has blind spots in the middle, and has only about 200 degree field of view from side to side. The resulting panoramic video (with frame resolution of 2048×768) was created by stitching these four video streams frame by frame.

2.3 Rendering of a Novel Video View

Figure 2 is a flow chart showing the decoding of the panoramic video. In the viewer, the compressed videos are decoded and rendered to create a scene at a given viewing angle. The resolution of the panoramic video is usually very large. The decoding or transmission of the whole panoramic video is very often time-consuming. This problem can be remedied by reducing the resolution of the decoded video and/or decoding only a given portion in the whole video frame. Actually, in virtual walk through applications, it is unnecessary to decode the entire video frame because only a fraction of the panorama will be used for rendering the novel view. Because of this reason, the panorama is usually divided into tiles to simplify decoding and the data transfer from slower devices such as CD ROM [2].

In our system, each panoramic video frame is divided into 6 vertical tiles as shown in Figure 2. If the whole panorama has a view of 360 degrees, then the maximum viewing angle of each tile will be 360/6 = 60 degrees, which is sufficient for most applications. It is therefore only necessary to decode simultaneously, at most two tiles at a time. According to the current view angle, the tiles involved (the shaped ones) will be decoded and placed in the decoding buffer. Appropriate portion of the panorama inside the buffer will be used to render the novel view. Tiles switching might happen when the user changes his/her viewpoint during the playback of the panoramic video. Therefore, additional mechanism must be provided in the compressed data stream to provide fast tile seeking. This will be discussed in the following section on the compression of panoramic video.

3. COMPRESSION OF PANORAMIC VIDEO

3.1. MPEG2 Video Coding of Sub-tiles

As mentioned earlier, panoramic video sequences have large spatial resolution. Therefore, they have to be compressed to reduce the amount of digital storage and bandwidth for transmission. For instance, the entire panoramic video sequences "cafeteria" consists of 25 (2048×768) video frames per second which will require 112.5 Mbytes/sec of storage, if no compression is employed. Like video, successive panoramic images have significant amount of temporal and spatial redundancies, which can be exploited by prediction techniques similar to motion estimation in video coding.

Also from Section 2, we know that it is better to divide each mosaic image into smaller tiles to avoid decoding the whole panoramic video and to reduce the data transfer from slower secondary devices. It is therefore natural to treat each of these tiles as a video sequence and compressed them individually. If a panoramic video with a resolution of (2048×768) is divided into 6 non-overlapping tiles, we end up with 6 video sequences each has a resolution of (352×768). To provide functionalities such as fast forward/backward and to make the panoramic video compatible to most decoders, it is natural to employ the commonly used MPEG2 video coding standard [10] to compress each of these video streams. Another advantage of MPEG2, as we shall see in Section 4, is that it is very efficient in compressing the high resolution panoramic video with a compression ratio of more than 100 times, yet with reasonably good reconstruction quality. Next, we shall consider the organization of the compressed video streams to provide effective access to individual tile during decoding.

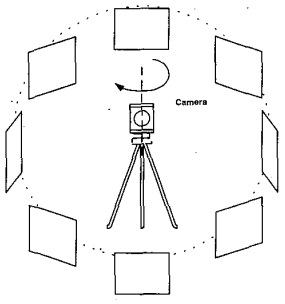

Figure 1. Construction of a panoramic mosaic.

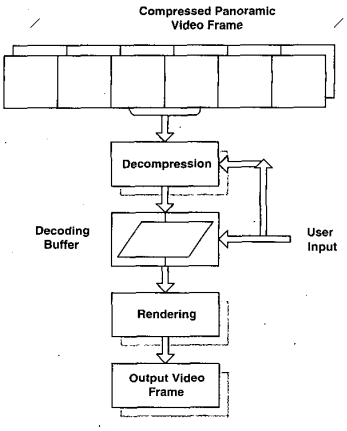

Figure 2. Rendering of panoramic video.

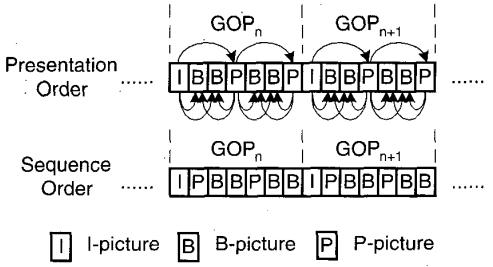

Figure 3. GOP setting in MPEG2 bitstream

3.2. The Selective Decoding Problem (Tile Seeking)

For transmission and digital storage of the panoramic video, the individual tiles must be organized in an efficient manner so that fast switching between tiles can be performed during decoding. Figure 3 shows the format of a tile or video stream encoded using the MPEG2 standard. Consecutive image frames of a given tile are arranged in groups called Group of Pictures (GOP). In each GOP, the image frames are encoded as I-, P-, or B- pictures. I-pictures are intra-coded and are used as references for predicting the next P- and other B-pictures in between using motion estimation. P-pictures are predicted using motion estimation from the previous I- or P-pictures. B-pictures are bi-directionally predicted from the nearest reference pictures. The arrows in Figure 3 show the inter-dependency of this prediction between various pictures in a GOP. In the proposed coder, there are 7 pictures in each GOP consisting of one I-picture, two P-pictures and 4 B-pictures in between. Also shown in Figure 3 is the sequence order of the compressed image frames to be transmitted. Note that the reference pictures are transmitted before the B-pictures because they must be decoded before the B-pictures, as they serve as references for reconstructing the B-pictures in between.

Figure 4 shows how the 6 tiles (video streams) in our panoramic video are multiplexed. Each tile is encoded by the MPEG2 standard with the same GOP structure shown in Figure 3. The compressed data of the tiles in the same panoramic video frame are packed together. This allows the decoder to locate very quickly the corresponding I-pictures for decoding the required tiles. Individual picture in each tile can be accessed randomly by searching for the appropriate picture headers. During decoding, the viewer can selectively decode the tiles required by the user, for example streams 1 and 2 in Figure 4. The novel view can then be generated by re-mapping appropriate pixels in the tiles onto the user's screen.

When the viewing angle is changed in such a way that some of the required pixels are no longer in the tiles currently being decoded, switching to the new tile(s) has to be performed. If this happens during the decoding of the P- and B-pictures in a GOP, tile switching can only begin in the next GOP, because the I-pictures of the new tiles in the current GOP might not be available (in practice, previously decoded data is usually not buffered). Because of this reason, the separation of the I-pictures in the panoramic video streams should not be very large. Otherwise, it will introduce unacceptable delay in switching from one stream to another. As mentioned earlier, there are 7 images in each Group of picture (GOP). Therefore, at a frame rate of 25f/s, the maximum delay during tile switching is 0.28 second, which is quite acceptable. Other values can be chosen according to one's tradeoff between the compression performance and the response time delay. The synchronized I-pictures also allows us to preserve the fast forward and fast backward capability in the MPEG2 standard. Notice that the number of P- and B-pictures in GOPs from different tiles can be different (as well as GOP from the same tile), as long as their I-pictures are synchronized. This helps to improve the compression performance, at the expense of more complicated encoding and decoding processes.

3.3. Experimental Results

The "cafeteria" panoramic video sequence described in Section 2 is compressed using the proposed coding algorithms. The six tiles of the panoramic video are encoded independently using the MPEG2 video-coding standard. Each stream has a Group of Picture (GOP) consisting of seven image frames with two B-frames between successive I- or P- pictures, Figure 3. Table 1 shows the compression performance of the panoramic video sequence using the proposed algorithm at different bit rates (target bit rate of 1 and 1.5Mb/s per tile). Figures 5 and 6 show a typical panorama and the decompressed tiles of the panorama, respectively. It shows good quality reconstruction with a compression ratio of 162. Next, we briefly outline the transmission aspect of panoramic video over cable networks, LAN and Internet.

Bitrate (Mb/s)	Compression Ratio	Mean PSNR (dB)		
		Y	U	V
5.727	162	42.16	45.69	44.74
8.596	108	45.40	47.10	46.60

Table 1. Compression performance of the panoramic video sequence. (25 frames per second, resolution: 352×768×6).

4. TRANSMISSION OF PANORAMIC VIDEO

In order to deliver the interactive virtual walkthrough experience offered by panoramic video, the compressed data stream can be broadcasted or transmitted through Video On Demand (VOD) systems over for example the Internet, LAN or cable networks. For broadcasting applications say over cable network, the whole panoramic video can be transmitted through a few cable TV channels with each channel carrying one or more tiles of video streams. According to the user input, the set top box can be designed to decode the appropriate tiles in the panoramic video. Due to the division of the panoramic video into tiles, only a limited number of tiles, 2 in the proposed system, have to be decoded. Additional hardware is required for rendering the novel view from the decoded video streams. For broadcasting over LAN, the decoding and rendering will most likely be performed by a workstation or PC. With nowadays technology, the real-time rendering and decoding of the panoramic video will not present significantly problems. In applications where the channel has limited and dynamic bandwidth such as internet, the tiles can be transmitted on a "on-demand" basis, where only the required video streams will be transmitted. Further reduction of bandwidth for transmission can be achieved by creating a scalable bit stream using for example multiresolution techniques.

5. CONCLUSION

We have described new compression and transmission techniques for panoramic video. A panoramic video allows the user to change viewpoint interactively in a dynamic scene or along a trajectory. High-resolution panoramic videos, while desirable, consume significant amounts of storage and bandwidth for transmission, and make real-time decoding compute-intensive. In this paper, we propose a high performance MPEG-like compression algorithm which takes into account the random access requirements and the redundancies of the panoramic video. The transmission aspects of panoramic video over cable network, LAN and Internet have also been briefly discussed.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Sing-Bing Kang at Microsoft Research for providing us the panoramic video data used in this paper. This work is supported by the Hong Kong Research Grants Council, Area of Excellent in Information Technology (AOE IT) project.

REFERENCES

[1] E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, pages 3-20. MIT Press, Cambridge, MA, 1991.
[2] S. E. Chen, "QuickTime VR – an image-based approach to virtual environment navigation," in *Computer Graphics (SIGGRAPH'95)*, pp. 29-38, August 1995.
[3] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Computer Graphics (SIGGRAPH'95)*, pp. 39-46.
[4] M. Levoy and P. Hanrahan, "Light field rendering," in *Computer Graphics Proceedings (SIGGRAPH'96)*, pp. 31-42, August 1996.
[5] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Computer Graphics (SIGGRAPH'96)*, pp. 43-54.
[6] H. Y. Shum and L. W. He, "Rendering with Concentric Mosaics" in *Computer Graphics (SIGGRAPH'97)*, pp. 299-306, August 1999.
[7] T. Chen, "From image and video compression to computer graphics", in *Proc. ICIP 2000*, Vol. 2, pp. 9-12, 2000.
[8] J. Li, H. Y. Shum and Y. Q. Zhang, "On the compression of image based rendering scene," in *Proc. ICIP 2000*, Vol. 2, pp. 21-24, 2000.
[9] H. Y. Shum, K. T. Ng and S. C. Chan, "Virtual reality using the concentric mosaic: construction, rendering and data compression," in *Proc. ICIP 2000*, Vol. 3, pp. 644-647, 2000.
[10] ITU-T Rec. H.262|ISO/IEC 13818-2, "Generic Coding of Moving Pictures and Associated Audio Information: Video," Nov. 1994.
[11] J. Foote and D. Kimber, "FlyCam: Practical Panoramic Video and Automatic Camera Control", in *Proc. IEEE International Conference on Multimedia and Expo*, 2000, vol. 3., pp. 1419-1422.
[12] J. Baldwin, A. Basu and H. Zhang, "Panoramic video with predictive windows for telepresence applications", in *IEEE International Conference on Robotics and Automation*, 1999, vol.3 pp. 1922-1927.
[13] T. Boult, "Remote reality demonstration," in *Conference on Computer Vision and Pattern Recognition*, 1998, pp. 966-967.
[14] R. Szeliski, "Video mosaics for virtual environments," in *IEEE Computer Graphics and Applications*, vol.16 (2), pp22–30, March 1996.
[15] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and texture-mapped models," in *Computer Graphics (SIGGRAPH'97)*, pp. 251–258, August 1997.
[16] S. B. Kang, "Catadioptric self-calibration," in *Conference on Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 201-207.
[17] S. Nayar, "Catadioptric omnidirectional camera," in *Conference on Computer Vision and Pattern Recognition*, 1997, pp. 482-488.

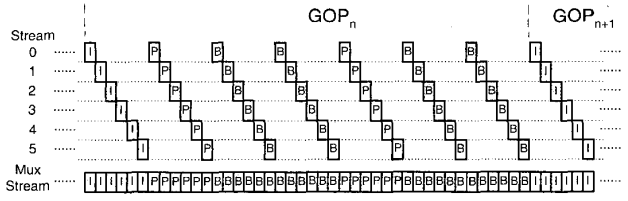

Figure 4. Multiplexing of the tiles (streams) in the MPEG2 compressed panoramic video.

(a) (b)

Figure 6. Frame 8 of the decompressed streams 3 (a) and 4 (b) with bit rate 1Mb/s per tile. (Luminance component)

Figure 5. Frame 8 of the "cafeteria" panoramic video sequence. (Luminance component)