ICSP2006 Proceedings

Transcoing-based Data Transmission of Sphere Panoramic Videos

Feng Dai(1)(2) Yong-dong Zhang(1) Yan-fei Shen(1) Shou-xun Lin(1)

(1) Institute of Computing Technology, Chinese Academy of Sciences
(2) Graduate University of Chinese Academy of Sciences
Beijing 100080, P. R. China
E-mail: fdai@ict.ac.cn

Abstract: Omnidirectional video is a kind of new emerging media. The filed of view of an omnidirectional video can be 360 degrees in vertical and horizontal. The users can navigate interactively through the scene and change their view angel. The scene of an omnidirectional video is often projected on the surface of a sphere or a cylinder to store, which is a panoramic video. Panoramic videos are often high-resolution and consume a significant amount of bandwidth for transmission. To resolve the problem, tiles-based data transmission is applied in some systems but it is not efficient for sphere panoramic video and transmits a mass of redundant bits to the users. In this paper, we proposed an efficient transcoding-based data transmission technique for panoramic videos, which reduced the amount of data transmission at most.

Key words: panoramic video tile-based data transmission transcoding-based data transmission

I. INTRODUCTIONS

Today, with the increasing processing power of CPUs and graphics adapters, more complex data can be processed by computers. On the other hand, Internet has helped us to connect and communicate with others anywhere in the world. The increasing band width of Internet provided us with the possibility to transmit larger amount bits than before. These technological advances make it is possible for people to demand for better user experience in interactive application such as virtual walkthrough and computer game.

In computer graphics literature, the methods for scene representation are often classified into two categories[1]. The first one represents the scene by classical 3-D computer graphics, which is called geometry-based modeling. The other one represents the scene by real images or videos, which is called image-based modeling. The geometry-based modeling provides the high interactivity in general than image-based modeling, but it suffers from expensive cost in both modeling and rendering. The image-based approach can avoid these drawbacks and brings the users the natural feeling in the real scene.

The image-based methods can be derived from the theory of the plenoptic function. The 7-dimensional function has initially been postulated by Adelson and Bergen [2]:

$$p = P(V_x, V_y, V_z, \theta, \phi, \lambda, t) \qquad (1)$$

By dropping the time variable t and the wavelength of light λ, fixing the camera location (V_x, V_y, V_z), the plenoptic modeling is degraded to 2-dimensional function:

$$p = P(\theta, \phi) \qquad (2)$$

Omnidirectional video is the dataset of 2-dimensional plenoptic function in space. It is a kind of new emerging media. The field of view of an omnidirectional video can be 360 degrees in vertical and horizontal while that of conventional videos are usually 60~70 degrees. Users could change their view angle interactively to the direction they desired at any time and navigate interactively through the scene.

Omnidirectional video capturing systems are capable of capturing a 360-degree horizontal field of view at one time. The omnidirectional video capturing systems are usually classified into two categories. One is multi-camera system, such as FlyCam[3],Ladybug2[4], Telemmersion System[5]. Each camera captures a portion of scene. The other one is single camera system, which uses hyperbolic mirrors to project an omnidirectional view of a smaller portion of a sphere onto camera sensors[6][7]. Usually the multi-camera system captures the panoramic video with higher quality than single camera system.

Sphere or cylinder panoramic video can be obtained when the scene of omnidirectional video is projected on a sphere or a cylinder surface. This wide view characteristic of panoramic video makes them suitable for applications like interactive walkthroughs and surveillance. More recently, panoramic videos have been proposed to capture dynamic environment maps for applications such as tele-presence and autonomous vehicles[6].

Because of the large amount dataset, the resolution of panoramic videos is usually very large. For example, the resolution of sphere panoramic video acquired by Telemmersion System is up to 2400x1200[5]. Compared with conventional videos, large amount of transmission bandwidth is needed. To resolve the problem, the panoramic video is divided into tiles before compression in many systems[8][9][10]. But the tiles-based data transmission does not suit for sphere panoramic video and produces a large amount of redundant data.

Our goal is to overcome the transmission problem for sphere panoramic video, and propose a transcoding-based data transmission scheme for sphere panoramic video.

The rest of paper is organized as follows. Section II briefly reviews and analyzes the tiled-based data transmission for panoramic videos. Data transmission based on transcoding for sphere panoramic video is presented in section III. In Section IV, experimental results are reported. We conclude and outlook the paper in Section V.

II. TILES-BASED DATA TRANSMISSION

As the resolution of a panoramic video is very large, transmitting the entire panoramic video is very often time-consuming. Certainly we can remedy the problem by reducing the resolution of the panoramic video, but the very coarse video will be provided to the users because the panoramic video is much larger than the conventional video.

Fortunately, building the perspective view of a given view angle does not need all of the data in the panoramic video. In order to avoid transmitting the entire high resolution image to the user, in [8] the tiles-based compression and transmission for panoramic videos is initially proposed. Many researchers also adopt the tiles-based transmission (TLDT) in their systems.

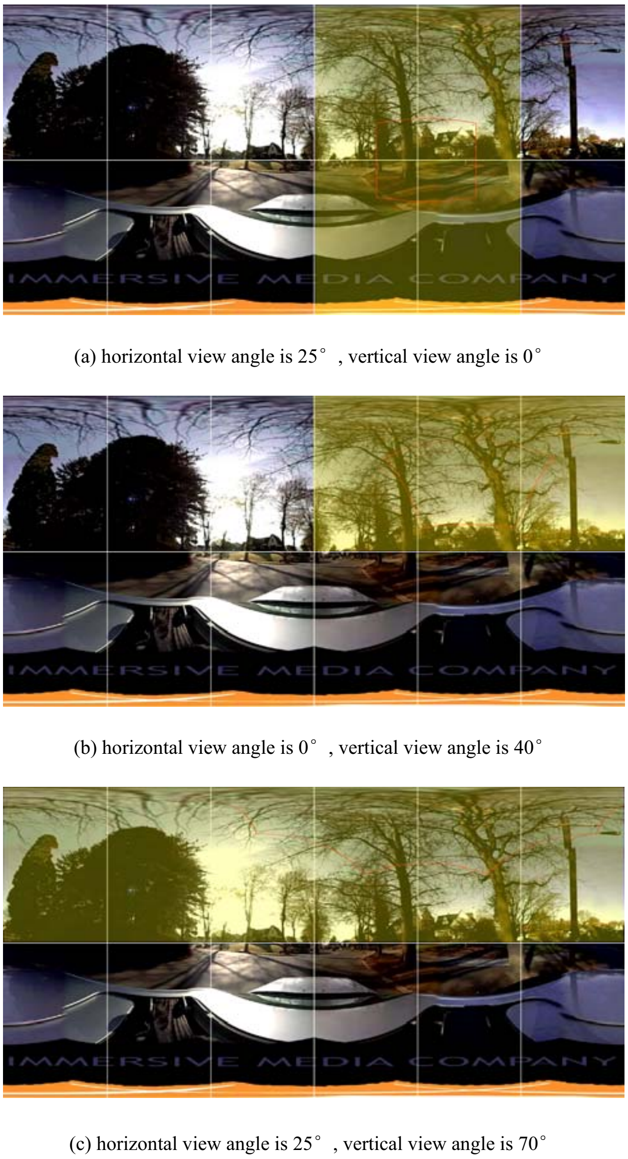
Tiles-based transmission divided the high-resolution panoramic videos into tiles. Each tile is compressed and decompressed individually. So the server only needs to transmit the tiles overlapping the current view orientation. Appropriate

0-7803-9737-1/06/$20.00 ©2006 IEEE

portion of the panorama inside the tiles is used to render the novel perspective view.

When the panoramic video is cylinder panoramic video, because of the limited field of view in vertical, the vertical view angle is usually fixed. When the panoramic video is a sphere panoramic video, according to the projection from a plane to a sphere, the position of corresponding area is changed with the horizontal view angle and the shape of corresponding area is changed with the vertical view angle. Fig.1 shows the *village* sequence divided into 12 tiles, and the corresponding area for different view angle is enclosed by the red edge. The resolution of the panoramic video is 2000x1000 and the resolution of perspective view is 352x288.

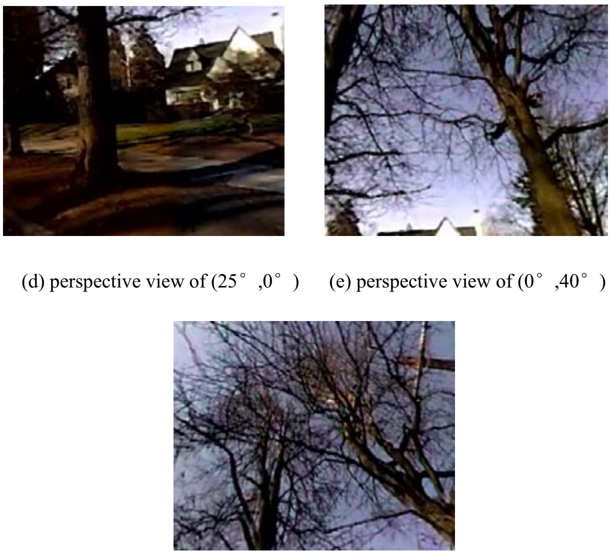
From Fig.1 we can see that although only the tiles overlapping the corresponding area needed to transmit to the users, a mass of redundance of data must be transmitted. Larger the vertical view angle is, larger the corresponding area is. When the vertical view angle is near 90 degrees, half of the tiles are involved. At this time, half of the frame must be transmitted. Of course we can divide the panoramic video into more tiles, which will reduce the amount of redundant data, but the small size of tiles will bring the very inefficient performance of compression.



(a) horizontal view angle is 25° , vertical view angle is 0°

(b) horizontal view angle is 0° , vertical view angle is 40°

(c) horizontal view angle is 25° , vertical view angle is 70°



(d) perspective view of (25° ,0°)   (e) perspective view of (0° ,40°)

(f) perspective view of (25° ,70°)

Fig.1. Perspective views of three given different view angle, and the corresponding area and involved tiles in panoramic view employing tiles-based data transmission. (The corresponding area in the red edge contributes to the perspective view. The tiles covered by yellow shadow is involved tiles that needs to be transmitted to users)

III. TRANSCODING-BASED DATA TRANSMISSION

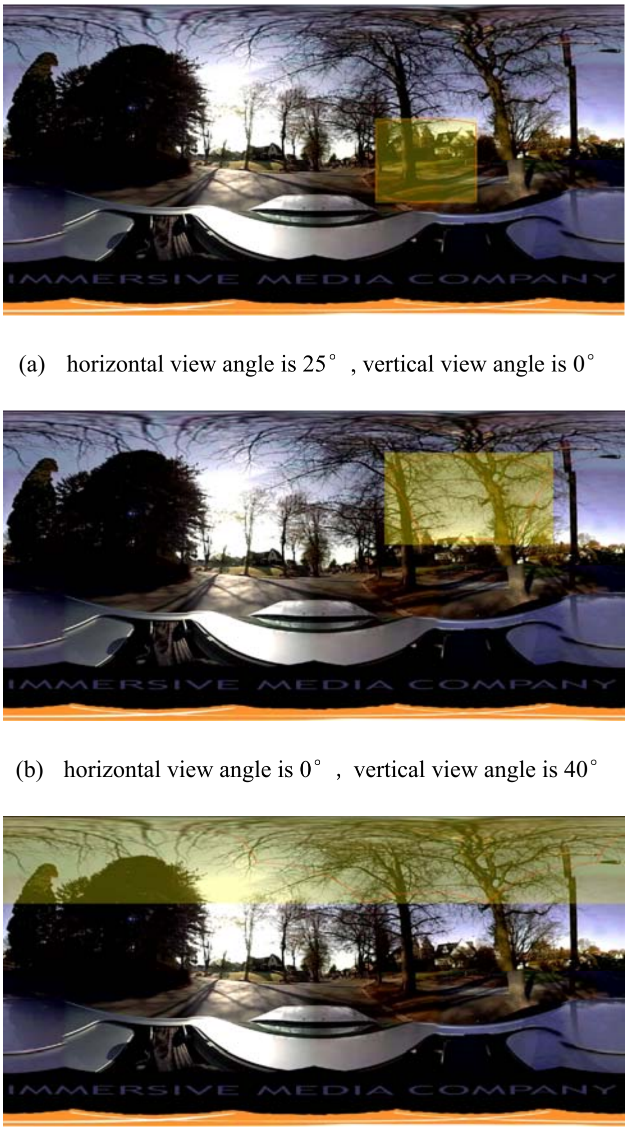
Because of the drawback of tiles-based data transmission (TLDT), we propose an efficient transcoding-based data transmission (TSDT) for sphere panoramic videos, which reduced the amount of data transmission at most.

Because users only need the data in corresponding area to build the perspective view, we can only transmit the data in the corresponding area. As we know that the corresponding area is not a rectangle, so we transmit the rectangle which right covers the corresponding area in the panoramic video, which is covered by yellow shadow in the Fig.2. Table 1 shows the amount of bits transmitted to users in tiles-based data transmission (TLDT) and transcoding-based data transmission (TSDT). We can see that the TSDT saves large numbers of bits than TLDT.

Table 1. Amount of data transmitted using TLDT and TSDT

view angle	Amount of bits transmitted		bits saved
	TLDT	TSDT	
(25° , 0°)	667x1000	336x272	86.3%
(0° , 40°)	1000x500	560x288	77.7%
(25° ,70°)	2000x500	2000x288	42.4%

Now the problem changes to how to obtain the compressed data in the rectangle and transmit the bits to the users. The most straightforward way to achieve this is to decode the panoramic video, and fully encode the reconstructed signal in the rectangle then transmits the bits stream of extracted video to the users. This mean is much time-consuming because the video encoding is significant complexity. A fast motion estimation is used to produce the extracted video bits stream.



(a) horizontal view angle is 25° , vertical view angle is 0°

(b) horizontal view angle is 0° , vertical view angle is 40°

(c) horizontal view angle is 25° , vertical view angle is 70°

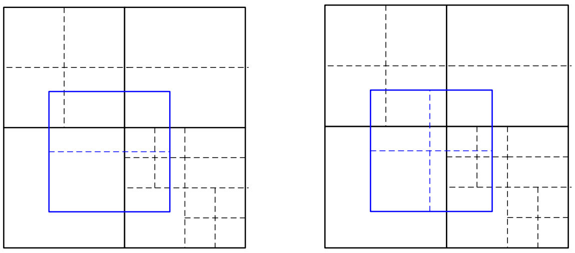
Fig.2. Corresponding area in panoramic view and the rectangle covering the corresponding area employing transcoding-based data transmission. (The corresponding area in the red edge contributes to the perspective view. The rectangle covered by yellow shadow is the extracted video frame, which needs to be transmitted to users)

Fast Motion Re-estiamtion

If a macroblock in the extracted video is to be intercoded, a new motion vector needs to be estimated for performing motion-compensated prediction. Although the full search motion estimation of a typical video encoder can be used to estimate the new motion vector, this approach is not preferred for real-time applications because motion estimation is arguably the most computationally intensive function in many video coding standards; for example, it comprises more than 60% of the total computational cost in H.264/AVC.

Fortunately the development of transcoding technology provides an efficient way to reenode the reconstructed signal we need. We reuse the motion vector information in precoded the panoramic video to reduce the complexity. Yap-peng Tan compares the performance of five fast motion estimation methods in [11], and proposes a weighted vector median method (AWVM) to re-estimate the new motion vectors for H.264/AVC[12].

As different macroblock partitions of an extracted video may correspond to different proportions of macroblocks in the precoded video. We use weighted vector median method to re-estimate the new motion vectors required.



(a) two 16x8 partitions

(b)four 8x8 partitions

Fig.3. The partitions of a macroblock in mode 16x8 and 8x8 and their corresponding macroblock partitions from the precoded panoramic video.

Let mv_i be the motion vector and A_i the area of the composing region from the ith corresponding macroblock partition. V is the set of K corresponding motion vecters. $V = \{mv_1, mv_2, ..., mv_k\}$ .The area-weighted vector median mv is obtained as

$$mv = \arg\min_{mv_j \in V} \sum_{i=1}^{K} \left\| mv_j - mv_i \right\|_2 \qquad (3)$$

In the motion estimation of area-weighted vector median method, if a corresponding partition is intracoded, it will not be involved in estimating the new motion vector. For this reason, if all the corresponding partitions are intra-coded, the new macroblock partition will also be transcoded with intra mode. Each mode gets its best motion vector and the best mode is that minimize the cost.

Process of transcoding-based data transmission

The steps of transcoding-based data transmission for sphere panoramic video are followed.

1. The client sends the desired view angle to the server.
2. According to the formula that transforms perspective view to sphere panoramic view, the server determines the corresponding area and the rectangle to be transmitted in the panoramic view.
3. Employing fast motion re-estiamtion of area-weighted vector median method, the server encodes the extracted video then transmits the bits stream to the client.
4. After receiving the extracted video bits, the client transforms sphere panoramic view to the perspective view.
5. When the user changes view orientation, the client sends the new view angle to the server.

IV. EXPERIMENTS RESULTS

Three sequences are tested in our experiments. Each sequence is precoded in the H.264/AVC at 2000x1000 resolutions. We encode the extracted video at three given vertical view angle and four horizontal view angle. When vertical view angle is 0° , 40° and 70° , 310K/bps ,400K/bps and 600Kbps bits stream is produced. We compare the performance of TSDT with the full search motion estimation at each view angle. From the table we can see that the full search motion estimation performs an average of 0.57 dB in PSNR better than the weighted vector median method。When the vertical view-angle is 70° , because the corresponding area crosses the entire width of the sphere panoramic video as shown in Fig2, the rectangle need to be transmitted is same in different horizontal view angles. Considering the much time saved, the weighted vector median method is acceptable.

Table 2. PSNR obtained by encoding extracted video from *village* sequence using TSDT and full search motion estimation

View angle		PSNR(dB)	
vertical	horizontal	TSDT	Full Search
0°	0°	24.53	25.14
0°	90°	22.96	23.67
0°	180°	26.24	27.06
0°	270°	29.64	30.24
40°	0°	23.11	23.76
40°	90°	20.33	20.99
40°	180°	19.78	20.32
40°	270°	22.65	23.14
70 °	0°	20.18	21.34
70 °	90°	20.18	21.34
70 °	180°	20.18	21.34
70 °	270°	20.18	21.34

Table 3. PSNR obtained by encoding extracted video from *city* sequence using TSDT and full search motion estimation

View angle		PSNR(dB)	
vertical	horizontal	TSDT	Full Search
0°	0°	37.97	38.48
0°	90°	35.28	35.57
0°	180°	35.53	36.08
0°	270°	34.54	34.87
40°	0°	34.48	34.83
40°	90°	32.71	33.08
40°	180°	36.71	36.97
40°	270°	33.20	33.56
70 °	0°	36.33	36.69
70 °	90°	36.33	36.69
70 °	180°	36.33	36.69
70 °	270°	36.33	36.69

Table 4. PSNR obtained by encoding extracted video from *hall* sequence using TSDT and full search motion estimation

View angle		PSNR(dB)	
vertical	horizontal	TSDT	Full Search
0°	0°	28.73	29.42
0°	90°	22.97	23.76
0°	180°	26.45	27.10
0°	270°	24.14	24.61
40°	0°	24.32	25.19
40°	90°	27.65	28.72
40°	180°	25.74	26.41
40°	270°	26.17	27.01
70 °	0°	25.18	25.45
70 °	90°	25.18	25.45
70 °	180°	25.18	25.45
70 °	270°	25.18	25.45

V. CONCLUSION AND OUTLOOK

We have proposed an effective transcoding-based data transmission method for sphere panoramic video. It much reduces the amount of redundant data comparison with the tiles-based data transmission. Experiments results shows that the transcoding-based data transmission can obtain slightly inferior extracted video quality than full search motion estimation, but it is computationally more efficient.

VI. ACKNOWLEDGMENT

This work is supported by Beijing Science and Technology Planning Program of China (D0106008040291) and National Nature Science Foundation of China under grant number 60302028.

REFERENCES

[1] Smolic. A and, Kauff. P "Interactive 3-D video representation and coding technologies". Proceedings of the IEEE. Vol. 93, No.1, Jan 2005, pp:98 – 110

[2] E. H. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in Computational Models of Visual Processing. Cambridge, MA: MIT Press, 1991, pp. 3–20.

[3] J. Foote and D. Kimber, "FlyCam: Practical panoramic video and automatic camera control," in Proc. IEEE Int. Conf. Multimedia and Expo, vol. 3, 2000, pp. 1419–1422.

[4] http://www.ptgrey.com

[5] http://www.immersivemedia.com/

[6] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams." Computer Vision and Image Understanding, 1998, Vol.71, No.2, 154--165.

[7] S. Nayar, "Omnidirectional video camera," in Proc. DARPA Image Understanding Workshop, 1997, pp. 235—241.

[8] S. E. Chen, "Quicktime VR - an image-based approach to virtual environment navigation." SIGGRAPH 95 Conference Proceedings, pages 29--38, Aug. 1995

[9] King-To Ng, Shing-Chow Chan, and Heung-Yeung Shum, "Data compression and transmission aspects of panoramic videos", Circuits and Systems for Video Technology, IEEE Transactions on, Jan. 2005, Vol.15, No.1, pp. 82- 95

[10] Grunheit. C, Smolic. A, and Wiegand, T. "Efficient representation and interactive streaming of high-resolution panoramic views", Image Processing, International Conference on. Vo. 3, pp. 24-28 , June 2002.

[11] Tan Y.-P, Liang Y, and Sun H "On the methods and performances of rational downsizing video transcoding", Signal Processing: Image Communication, Volume 19, Number 1, January 2004, pp. 47-65(19)

[12] Yap-Peng Tan, and Haiwei Sun, "Fast motion re-estimation for arbitrary downsizing video transcoding using H.264/AVC standard", Consumer Electronics, IEEE Transactions on, Vol. 50, No. 3, pp: 887- 894, Aug. 2004 .