THE MOST EFFICIENT TILE SIZE IN TILE-BASED CYLINDER PANORAMIC VIDEO CODING AND ITS SELECTION UNDER RESTRICTION OF BANDWIDTH

Feng Dai[(1)(2)] *Yanfei Shen*[(1)] *Yongdong Zhang*[(1)] *Shouxun 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

Cylinder panoramic videos are a 360° representation in horizontal of a certain scene. The users can navigate interactively through the scene and change their view angels. Because of large field of view, panoramic videos are often high-resolution and consume a significant amount of bandwidth for transmission. To resolve the problem, tile-based panoramic video coding and transmission is applied in some systems. With tile-based panoramic video coding and transmission, the panorama is divided into tiles and only the tiles involved with perspective view are transmitted and decoded. In the tile-based video coding, the selection of tile size is a problem. Different tile sizes will bring different quality of perspective view with the same bandwidth. In this paper, we proposed a method to select the most efficient tile size. With this method, the most efficient tile size could be fast selected and users can build the best quality perspective view from cylinder panoramic videos with the given bandwidth.

1. INTRODUCTION

Panoramic videos are a 360° representation of a certain scene. The field of view of a panoramic video can be 360° in vertical and horizontal direction while that of conventional videos are usually 60~70°. The users can navigate interactively through the scene and change their view angel with a special panoramic video player. Now panoramic video are emerging to widely be used in telepresent, surveillance, entertainment and other applications.

Cylinder projection is one of the most popular projections in applications. Cylinder panoramic videos are 360° in horizontal direction and limited in vertical view direction. And users can change their view direction in horizontal direction.

Because of the large field of view and amount dataset, the resolution of panoramic videos is usually very large. For example, the resolution of panoramic video acquired by Telemmersion System is up to 2400x1200[1]. Compared with conventional videos, large amount of transmission bandwidth is needed.

Tile-based panoramic video coding divides the panorama into tiles. Each tile is compressed and decompressed individually. Only the tiles involved with current perspective view need to be transmitted.

In tile-based panoramic video coding and transmission, the selection of tile size is a problem. In the same condition, Different tile sizes will bring different compression efficiency and transmitted macroblocks. An appropriate tile size could bring best visual quality when the band width is same. Until now, some researchers have worked on the problem but did not obtain satisfying result [2]. In this paper, we propose a method to find the most efficient tile size for cylinder panoramic video coding, which brings the best quality perspective view with the same bandwidth.

The rest of paper is organized as follows. Section 2 briefly introduces the tile-based coding and transmission for cylinder panoramic videos. An efficient method is present to select the most efficient tile size under the restriction of bandwidth in section 3. In Section 4, experiment result is reported. We conclude the paper in Section 5.

2. TILE-BASED CYLINDER PANORAMIC VIDEO CODING AND TRANSMISSION

As the resolution of a panoramic video is very large, transmitting the entire panoramic video is very often time-consuming. Fortunately, building the perspective view of a given view direction does not need all of the data in the panorama. In order to avoid transmitting the entire high resolution image to users, in [3] the tile-based coding and transmission for panoramic videos is firstly proposed. Many researchers also adopt the tiles-based transmission in their systems [4][5][6].

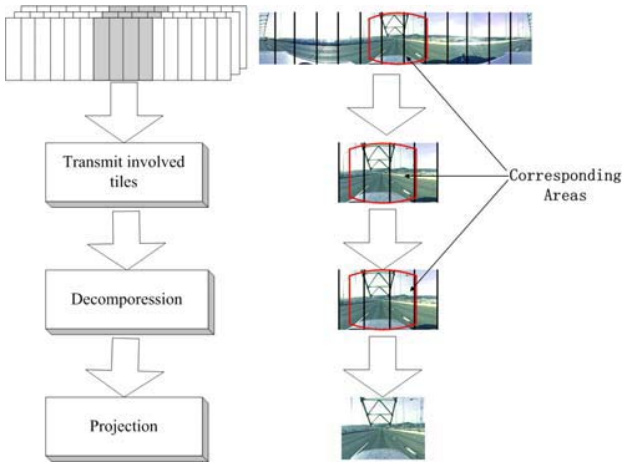

Fig.1 Process of building a perspective view from a cylinder panoramic video

Tiles-based panoramic video coding and transmission divides the high-resolution panorama into tiles. Each tile is compressed and decompressed individually. The server transmits the tiles involved with the perspective view. Appropriate portion of the panorama inside the tiles is used to render the perspective view, which we call *corresponding areas*.

Because of the limited field of view in vertical, the vertical view angle is usually fixed. Only the horizontal view angle is changed. So we only need to divide the panorama into tiles in horizontal direction and not in vertical direction as shown in fig.1.

3. THE SELECTION OF THE MOST EFFICIENT TILE SIZE

Tiles-based panoramic video coding and transmission reduce the data to be transmitted. But how to choose an appropriate tile size to encoding a panoramic video is still a problem. Smaller tile size will bring less transmitted macroblocks but lower compression efficiency. Larger tile size will bring higher compression efficiency but more transmitted macroblocks.

3.1. Tile Size

3.1.1 Relation between tile size and compression efficiency

Firstly we discuss the relation between tile size and compression efficiency for panoramic video. The division of the panorama will bring more tile boundaries and more macroblocks located on the boundaries. Because each tile must be encoded and transmitted individually, they could not refer to each other. During intra prediction, the valid intra prediction modes for macroblocks located on the boundaries are less than those inside the tiles. During inter prediction, the motion vectors for macroblocks located on the boundaries have more possibility to point to the outside of the slice than the ones inside the tile. So for the same panoramic video, smaller tile size is, more the number of macroblocks on the tile boundary is, and lower the compression efficiency is.

3.1.2 Relation between tile size and involved macroblocks

When the tile size a and width of corresponding areas M in macroblock units is given, the width of involved tiles W_a may be $\lceil \frac{M}{a} \rceil$ or $\lceil \frac{M}{a} \rceil + 1$ depending on the position of the *corresponding areas* (fig.2). When we encode a panoramic video, we do not know which tiles would be involved and transmitted in the future. To ensure the decoder buffer not to overflow, we must consider the worst condition. The number of involved tiles N_a could be calculated as following equation (1). $\lceil \ \rceil$ is ceil function.

$$N_a = \left\lceil \frac{M}{a} \right\rceil + 1 \tag{1}$$

So the width of involved tiles in macroblock units can be calculated using equation (2)

$$W_a = N_a \cdot a \tag{2}$$

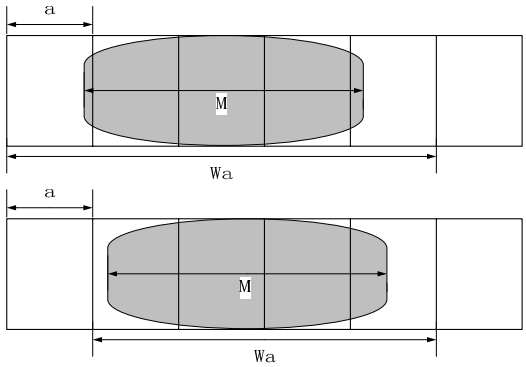

Fig.2 Different width of involved tiles for same filed of view

3.2 Target bit rate

In the application of transmission of panoramic video on Internet, the bandwidth is only used to transmit the data of involved tiles. Different number of involoved tiles will bring the different target bit rate of panoramic videos.

The target bit rate of panoramic video R_a is calculate as following:

$$R_a = \frac{W}{W_a} \cdot R \tag{3}$$

W is the width of panorama in macroblock units, R is the given bandwidth. For example, when the resolution of panoramic video to be compressed is 1920x352, the bandwidth R is 300kps, and the width of corresponding areas M is 20 macroblock, if we divide the panorama into 60 tiles (the tile size a is 2), 22 involved macroblocks are transmitted and the target bit rate is 1636kps; if the panorama is divided into 30 tiles (that is the tile size is 4),

24 involved macroblocks is transmitted and the target bit rate is 1500kps.

The most directly way to find the most efficient tile size is to list the result of PSNR coding employing different tile size *a*, and the tile size which brings the maximum PSNR is the most efficient tile size. Table.1 is a result of PSNR for *hall* sequence with *M* is 20 and the bandwidth is 300kps. R_a is the target bit rate of the panoramic video. From the table we can clearly see that the when the tile size is 4, we can get the best video quality.

Table.1 Result of PSNR of different tile sizes

Tile Size	W_a	N_a	R_a***(kps)***	***PSNR***
1	21	21	1714	24.92
2	22	11	1636	25.55
3**	24	8	1500	25.40
4	24	6	1500	25.57
5	25	5	1440	25.47
6*	30	5	1200	24.54
7	28	4	1286	24.99
8*	32	4	1125	24.28
9*	36	4	1000	23.65
10	30	3	1200	24.64
11*	33	3	1091	24.15
12*	36	3	1000	23.75
13*	39	3	923	23.36
14*	42	3	857	22.99
15*	45	3	800	22.76
…	…	…		…

3.3 Candidates of most effcient tile size

Enocoding the panoramic video with every possible tile size then list and compare the obtained result of PSNR is complicated. We use S represents the set of candidates of most efficient tile size. Firstly, S contains all the possible tile sizes.

$$S = \{1, 2, 3, 4, 5 \ldots W\}$$

We reduce number of elements in *S* by following two rules:

Rule 1: If $N_{a_1} = N_{a_2}$ & $a_1 > a_2$, a_1 is remove from *S*.

Because the number of involoved tiles and the bandwidth are both same, obviously the smaller tile size will bring better visual quality.

Rule 2: If $W_{a_1} \geq W_{a_2}$ & $a_1 \leq a_2$, a_1 is remove from *S*.

Because the bandwidth is same, smaller tile size and more involved macroblocks both bring lower compression efficiency. In the condition, the larger tile size brings better visual quality.

In table 1, the tile sizes with asterisk are removed from *S* according rule 1 and the ones with double asterisk are removed from *S* according rule 2. The number of elements in *S* reduces to 6 from 120.

$$S = \{1, 2, 4, 5, 10, 20\}$$

List and compare the tile size in *S* is a way to find the most efficient tile size. But it is also much complicated because we need encode the panoramic video using all of the tile sizes in *S*.

The most efficient tile size must be in *S*, we only need to select the most efficient tile size from these candidates. We look for the most efficient tile size by look for the index of the most efficient tile size.

Experiments results show that the bandwidth *R*, width of panorama *W* and width of coressponding areas *M* are the factors to decide the index of most efficient tile size in *S*. Different details in the videos affect the result slightly. Experiments results show that when the ratio *W* to *R* increases 3 times or *M* increases 10, the index of most efficient tile size will increase 1. Based on experiments, we propose a model to calculate the index of most efficient tile size as following:

$$I = \log_3\left(\frac{W}{R}\right) + \frac{M}{10} - 1.5 \qquad (3)$$

Following is the process to find the most efficient tile size:

1) Put 1 to *W* to *S*; calculate the corresponding W_a and N_a by the given *M* using equation (1) and (2).
2) Remove elements from *S* according rule 1 and rule 2
3) Sort and index the tile size in *S* ascending, the index of smallest tile size is index as 0.
4) According equation (3), calculate *I*.
5) Encode the panoramic video using the two tile sizes with the indexes are *I* and *I-1*.
6) Compare the two result of PSNR; choose the tile size which brings better PSNR as the most efficient tile size.

4. EXPERIMENTAL RESULTS

Five cylinder panoramic video sequences as shown in fig.3 are tested in the experiment. All the sequences are from Immersive Media Corp [1]. Videos with three different resolutions are tested in the experiments. For each resolution, three different *M* is tested, and for each *M*, four different bandwidths are tested. In our experiment, the value of *M* ensures that the field of perspective view is between 60 degrees and 90 degrees as in most applications.

(a) *forest*

(b) *bridge*

(c) *village*

(d) *hall*

(e) *lake*

Fig.3. Five cylinder panoramic video sequences

Table.2 Result of most efficient tile size with 960x176

M	*R(kps)*	*I*	most efficient tile size									
			bridge		*hall*		*village*		*forest*		*lake*	
			pro	*full*	*pro*	*full*	*pro*	*full*	*pro*	*full*	*pro*	*full*
10	100	2	2	2	2	2	2	2	2	2	2	2
	150	1	2	2	2	2	2	2	2	2	2	2
	200	1	2	2	2	2	2	2	2	2	2	2
	250	1	2	2	1	1	1	1	1	1	1	1
12	100	2	3	3	3	3	3	3	3	3	3	3
	150	1	2	2	2	2	2	2	2	2	2	2
	200	1	2	2	2	2	2	2	2	2	2	2
	250	1	2	2	2	2	2	2	2	2	2	2
15	150	2	3	3	3	3	3	3	3	3	3	3
	200	1	3	3	3	3	3	3	3	3	3	3
	300	1	3	3	3	3	3	3	1	1	3	3
	400	1	3	3	1	1	1	1	1	1	1	1

Table.3 Result of most efficient tile size with 1920x352

M	*R(kps)*	*I*	most efficient tile size									
			bridge		*hall*		*village*		*forest*		*lake*	
			pro	*full*	*pro*	*full*	*pro*	*full*	*pro*	*full*	*pro*	*full*
20	200	3	4	4	4	4	4	4	4	4	4	4
	300	2	2	2	4	4	2	2	2	2	4	4
	400	2	2	2	2	2	2	2	2	2	2	2
	500	2	2	2	2	2	2	2	2	2	2	2
25	300	3	5	5	5	5	5	5	5	5	5	5
	400	2	5	5	5	5	5	5	5	5	5	5
	500	2	5	5	5	5	5	5	5	5	5	5
	600	2	5	5	5	5	5	5	5	5	5	5
30	300	3	5	5	5	5	5	5	5	5	5	5
	400	3	3	3	3	3	3	3	3	3	3	3
	500	3	3	3	3	3	3	3	3	3	3	3
	600	3	3	3	3	3	3	3	3	3	3	3

Table.4 Result of most efficient tile size with 2400x448

M	*R(kps)*	*I*	most efficient tile size									
			bridge		*hall*		*village*		*forest*		*lake*	
			pro	*full*	*pro*	*full*	*pro*	*full*	*pro*	*full*	*pro*	*full*
25	300	3	5	5	5	5	5	5	5	5	5	5
	500	2	2	2	2	2	2	2	2	2	2	2
	600	2	2	2	2	2	2	2	2	2	2	2
	800	2	2	2	2	2	2	2	2	2	2	2
30	300	3	5	5	5	5	5	5	5	5	5	5
	500	3	3	3	3	3	3	3	3	3	3	3
	600	3	3	3	3	3	3	3	3	3	3	3
	800	3	3	3	3	3	3	3	3	3	3	3
40	400	4	5	5	5	5	5	5	5	5	5	5
	600	4	4	4	5	5	4	4	5	5	4	4
	800	4	4	4	4	4	4	4	4	4	4	4
	1000	3	4	4	4	4	4	4	4	4	4	4

A feature called *Flexible Macroblock Ordering* (FMO) is supported by H.264/AVC. FMO specifies a pattern that assigns the macroblocks in a picture to one or several slice groups. Each slice group is transmitted separately. It is very suit the tiles-based video coding. In the experiments, H.264/AVC is adopted and we choose the slice group type *foreground with left-over* to encode each tiles as a slice group and transmit the tiles as slices.

In our experiments, the modified reference software JM10.2 is used. We modified JM10.2 because the reference encoder only supports 8 slices at most, while in our experiments, much more slices are used.

Table 2, Table 3, and Table 4 list the results of our experiment. *pro* is the result got by our method, and *full* is the result by encoding with every possible tile size. In our experiment, the right ratio is 100%.

5. CONCLUSION

The selection of most efficient tile size is an important issue for panoramic video coding, which could transmit the best quality video with the limited bandwidth. Our proposed method could select the most efficient tile size for cylinder panoramic videos efficiently avoiding encoding with every possible tile size. We also can see that the most efficient tile size will increase when the *M* increases or *R* decreases. We can further decrease the complexity by encoding only a segment of the panoramic video when perform step (6).

ACKNOWLEDGEMENT

This work is supported by Beijing Science and Technology Planning Program of China (D0106008040291).

REFERENCES

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

[2] R. Kitaura , H. Habe, H Kimata and T. Nomura "Evaluation result of divided omni-directional video using AVC (EE1)" ISO/IEC JTC 1/SC 29/WG 11 MPEG2003/M10416. December 2003, Waikaloa.

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

[4] K. T. Ng, S. C. Chan and H. Y. Shum, "Data compression and transmission aspects of panoramic videos", Circuits and Systems for Video Technology, IEEE Transactions on, Vol.15, No.1, Jan. 2005, pp. 82-95.

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

[6] S. Heymann, A. Smolic, K. Müller, Y. Guo, J. Rurainski, P. Eisert and T. Wiegand, "Representation, Coding, and Interactive Rendering of High-Resolution Panoramic Images and Video Using MPEG-4", 2. Panoramic Photogrammetry Workshop, Berlin, Germany, February 24-25, 2005.