

A Multi-reference Viewpoint Selection Algorithm Based on Viewpoint Similarity Judgment

Jian-ping Wu^{1,a}, Jian Zhang^{2,b,*}

1School of Business, Lingnan Normal University, Zhanjiang, 524048, China

2 Southern Institute of Data Science and Technology, foshan, 528000, China

aguanwudan@126.com, bzj3319896@163.com

Abstract

This paper proposes a type of reference viewpoint selection method based on the viewpoint similarity standard that reduces computation complexity and increases coding performance. Experimental results show that the proposed method can reduce the computational complexity in the multiview video coding with the same quality.

Keywords: characteristic vector of image block, Video compression, similarity determination, multiview video

With the continuous development of multimedia technologies, the free viewpoint video technology will become an important integrated part of the video technologies in the future, for it can provide interactive operation functions and the depth perception of images, able for users to choose and operate on audio-video objects^[1]. The multiview video technology, being one of the core segments of the free viewpoint video technology, wherein the coding compression, transmission and storage of data exist mostly as multiview video data in a viewpoint video system. The multiview video data are obtained from shooting by use of a number of camera arrays, which data volume will multiply with the increase of the number of viewpoints; thus the compression of multiview video signals has become a puzzle to be solved in multiview video storage and transmission^[2]. Spatial redundancy exists to certain extent among various viewpoints in multiview video signals, which, however, can be effectively eliminated by use of reasonable disparity estimation methods to achieve the purpose of compressing multiview video data. Disparity estimation is one of the key technologies in multiview video coding^[3]. Kimata^[4] have proposed the method of multiple reference viewpoints to increase the precision of disparity estimation and improve the coding effect, however, the computation volume under such method is so large that real-time treatment is unfavorable; Clewer, et^[5] al. have recommended the variable block disparity estimation method by use of the quadtree splitting method that can improve the precision of

disparity estimation to some extent while its ensued computation volume does not increase that much, however, its precision of disparity estimation is still insufficient compared with the method of multiple reference viewpoints; Wang^[6], . have used the method of extracting video object from natural scenes in multiview video treatment, which, although has solved the problem of compressing multiview videos and establishing virtual viewpoints concurrently, has a similarly large computation volume, and meanwhile, certain problems still exist in extracting video objects, making its application in actual video coding systems yet immature.

Reference viewpoints of disparity estimation in multiview coding

In the disparity estimation of cubic video coding, it's just sufficient to take one of its viewpoints as a reference viewpoint and another viewpoint as a target one for disparity estimation. From the perspective of the information theory, whichever viewpoint taken as the reference viewpoint is not distinguishable for the coding efficiency. However, multiple viewpoints exist in a multiview coding system, placing crucial importance to the selection of appropriate reference viewpoints.

A perceptive method is selecting proper viewpoints as reference viewpoints as same as the cubic video coding system, while other viewpoints are taken as target viewpoints for disparity estimation^[7]. Such disparity estimation method based on single reference viewpoint is a direct application of the available disparity estimation algorithm in cubic videos to multiview coding systems. Its advantages include the simple system structure, the possibility to use the available cubic video coding system for direct expansion and fairly uncomplicated computation. However, as only the information of one of its viewpoints is used as reference, information entropy cannot be provided sufficiently for the disparity estimation of some target viewpoints, and under the circumstance that there are quite a number of viewpoints, certain bit numbers will be wasted during coding, causing sufficient compression ratios unable to be acquired.

To increase the estimation precision of multiview disparities and improve the coding performance, multiple reference viewpoints can be used to estimate the disparities. This method uses the information from two viewpoints as reference, allowing the prediction efficiency of disparity estimation compensation to increase significantly, as indicated in Fig.2, wherein using multiple reference viewpoints to estimate the coding performance is far better than the single reference viewpoint disparity estimation. However, this coding system has a complicated structure and its computation complexity multiplies with the increase of the number of reference viewpoints, causing overly high computation cost consumed in disparity estimation across the entire system, unfavorable for real-time treatment and application.

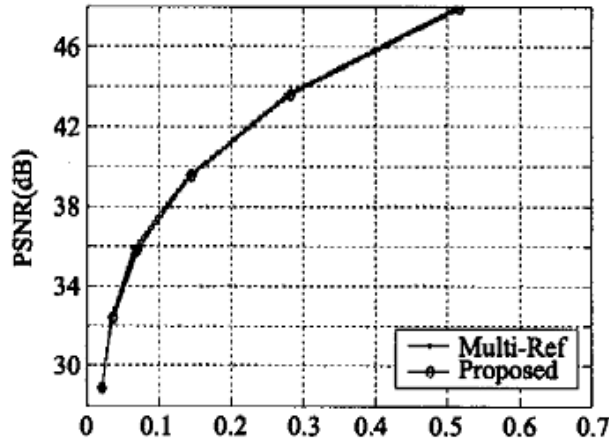


Fig. 1 Impact of Different Quantities of Reference Viewpoints to Multiview Signal Coding Performance

Fig.1 gives the result of predicting the multiple reference viewpoints disparity estimation of the “Cup” test sequence (Fig.2(b)), which shows that with the increasing number of reference viewpoints, the computation complexity will also multiply, but the coding performance of the system will not increase further after the coding performance of the system reaches certain extent. As during the disparity estimation of the “Cup” test sequence, the information entropy obtained from two reference viewpoints can sufficiently reach the system coding requirement, several proper reference viewpoint are sufficient to be selected in actual application to encode multiview videos appropriately.

In multiview video treatment, the use of multiple reference viewpoints for disparity estimation is the main method in estimating multiview disparity, however, its facing problem is the overly complicated computation, and due to the use of multiple reference viewpoints, multiple viewpoints need to be referred during disparity estimation, which further increase the load of the entire coding system.

A multiple reference viewpoint selection algorithm based on viewpoint similarity judgment

The selection of proper reference viewpoints in a multiview coding system will directly affect the coding efficiency within the entire system, for which a type of reference viewpoint selection program is proposed based on the judgment of multiview image similarity in this paper, in addition to an analysis of its comparison with other programs.

Program I: Prediction of Single Reference Viewpoint Disparity Estimation Compensation It applies the method directly using the single reference viewpoints in Fig.1(a) for encoding multiview video signals. This program has relatively low complexity, simple in system structure, and is able to utilize the available cubic video coding system for direct expansion. However, its disparity

estimation precision is not sufficient, unable to reconstruct high quality target viewpoints at the decoding end, resulting in unsatisfactory coding efficiency.

Program II: Perdition of Multiple Reference Viewpoint Disparity Estimation Compensation, Multiple viewpoints are selected in multiview videos as reference viewpoints for disparity estimation. This program uses the method illustrated in Fig.1(b), able to increase largely the precision of disparity estimation, however, the computation complexity also multiplies accordingly, unsuitable for use in sites where the instantaneity requirement is high. In this paper, taking into account the placement of cameras, generally two viewpoints as the most left and the most right are used as the reference viewpoints to realize the maximum information entropy. Excessive reference viewpoints have no significant impact to the system coding performance.

In a multiview coding system, the use of multiple reference viewpoints for disparity estimation can increase the disparity estimation/compensation efficiency, bringing about apparent improvement effect to the coding quality of the entire coding system. However, this is at the cost of increasing the computation complexity, for if more than two viewpoints are used as reference viewpoints, its computation complexity will multiply, and the computation cost the entire system consumes on disparity estimation will be too high be favorable to real-time treatment. Meanwhile, due to the use of multiple reference viewpoints, the disparity field will become complicated and the bit rate required by the coding disparity field will increase accordingly.

Program III: Judge and select proper viewpoints by use of viewpoint similarity in favor of predicting the disparity estimation compensation. The viewpoint similarity detection algorithm proposed in this paper detects scales based on the histogram characteristics of images among different viewpoints (on the mean value and variance basis). The scaling function of the similarity between Viewpoint Numbered (n+p) and Viewpoint Numbered n is

$$MSE_{n+p,n} = \sum_{i=1}^I \sum_{j=1}^J (\bar{x}_{n+p,i,j} - \bar{x}_{n,i,j})^2 + |\sigma_{n+p,i,j}^2 - \sigma_{n,i,j}^2| \quad (1)$$

In the above term, (i,j) is the position of the macroblock in the image. As for the CIF format multiview video sequence (the image has 352×288 pixels), if the macroblock has 16×16 pixels, i=1,2,...,22; j=1,2,...,18; $\bar{x}_{n+p,i,j}$ is the mean value of the brightness or tone component of Viewpoint Numbered (n+p) in the (i,j) position, i.e.

$$\bar{x}_{n+p,i,j} = \frac{1}{N \times M} \times \sum_{u=1}^N \sum_{v=1}^M x_{n+p,i,j}(u,v) \quad (2)$$

$\sigma_{n+p,i,j}^2$ is the variance of the brightness or tone component of Viewpoint Numbered (n+p) in the (i,j) position, i.e.

$$\sigma_{n+p,i,j}^2 = \frac{1}{N \times M} \times \sum_{u=1}^N \sum_{v=1}^M (x_{n+p,i,j}(u,v) - \bar{x}_{n+p,i,j})^2$$

(3)

In the above term, M and N are the width and height of the macroblock, and if the macroblock has 16×16 pixels, M=16, N=16.

When the similarity between a target viewpoint and a certain reference viewpoint approximates to a certain extent, the main reference when making multiple reference viewpoint disparity estimation is this reference viewpoint, and other reference viewpoints has no significant impact on disparity estimation.

$$I_A = \begin{cases} True & MSE_{A,L} < MSE_T \\ False & MSE_{A,L} \geq MSE_T \end{cases}$$

(4)

In Term (4), IA represents the disparity estimation method used by Target Viewpoint A; MSE_{A,L} represents the viewpoint similarity between Target Viewpoint A and Reference Viewpoint L; MSET represents the similarity judgment threshold. In this paper, 20% of the maximum MSE of the target viewpoint and the reference viewpoint is taken as the similarity judgment threshold MSET. When IA is True, Reference Viewpoint L is used as the single reference viewpoint for disparity estimation, otherwise multiple reference viewpoints are used for disparity estimation. Thus, under the premise that the disparity estimation precision is ensured, the computation complexity is reduced and the practicability of the multiview coding system is enhanced.

Experimental result and analysis

In order to verify the effectiveness of the algorithm (Program III) proposed in this paper, three multiview image test sets as Xmas, Cup and Note indicated in Fig.2 have been taken for experimental analysis. 10 viewpoints in each test set are aligned in one horizontal line, wherein the viewpoint image size is 640×480, in the YUV(4:2:0) format, and the spacing between two viewpoints of Xmas and Note is 30mm, the spacing between two viewpoints of Cup is 15mm.



(a) Three Viewpoints among 10 Viewpoints in Xmas Multiview Test

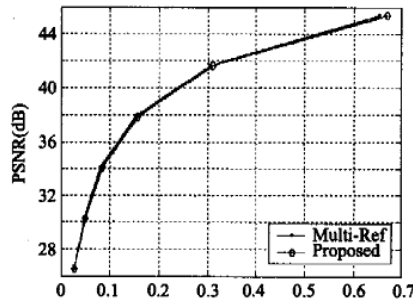


(b) Three Viewpoints among 10 Viewpoints in Cup Multiview Test Set

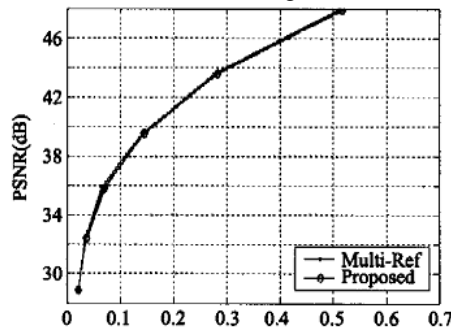


(c) Three Viewpoints among 10 Viewpoints in Note Multiview Test Set
 Fig.2 three multiview image test sets as Xmas

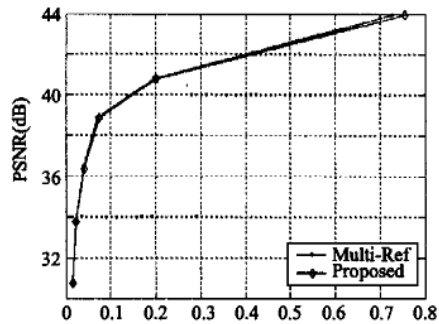
Fig.2 compares the disparity estimation and coding peak signal-to-noise ratio for three multiview images of Xmas, Cup and Note in three programs. In this figure, Scheme1, Scheme2 and Scheme3 respectively represent Program I, Program II and Program III. Program I uses the disparity estimation method of single reference viewpoint; Program II uses the disparity estimation method of two reference viewpoints, whereas if it is more than three reference viewpoints, it will not be favorable to the enhancement of coding performance, on the contrary, the computation volume will be increased, thus only two reference viewpoints as the most left and the most right ones are used; Program III is the reference viewpoint selection program based on the judgment of viewpoint similarity proposed in this paper.



(a) Comparison Result of Three Programs in Xmas Sequence



(b) Comparison Result of Three Programs in Cup Sequence



(c) Comparison Result of Three Programs in Note Sequence

Fig.3 Comparison Result

The experimental result indicates that the application of the algorithm proposed in this paper is better than the single reference viewpoint regarding the coding performance, for the peak signal-to-noise ratio (PSNR) has a significant increase, which coding quality is basically the same as the multiple reference viewpoints. The entire computation complexity (i.e., including the entire process of disparity estimation, coding, etc.) performance comparison is indicated in Table I. Apparently, under the circumstance that its signal-to-noise ratio is identical to that in Program II, the program proposed in this paper reduces the computation volume of disparity estimation, thus decreases the computation complexity of the entire multiview image coding process and the bit rate has also been lowered slightly.

Table 1 Performance Comparison of the Computation Complexity in Three Programs

Scheme	Xmas			Cup			Note		
	PSNR	Bits	Time	PSNR	Bits	Time	PSNR	Bits	Time
Scheme1	41.40	1 105 000	56.1%	43.60	1 006 170	60.1%	43.90	2 899 050	60.2%
Scheme2	41.70	937 903	100%	43.60	856 360	100%	43.90	2 233 480	100%
Scheme3	41.60	926 431	74.3%	43.60	844 760	74.6%	43.90	2 246 280	74.5%

Conclusions

The multiview video treatment technology is the core technology in the free viewpoint system. It solves the problems in programs such as coding, transmission and storage, etc. of free viewpoint data. The disparity estimation based on multiple reference viewpoints is a unique method in multiview video treatment and one of its key technologies, able to eliminate greatly the spatial redundancy among various viewpoints and realize the purpose of the effective

coding compression of multiview videos, to solve the problems of multiview videos in transmission and storage. Further studies will be conducted on the fast algorithm of multiple reference viewpoint disparity estimation, to reduce the computation volume occupied by disparity estimation in a multiview video treatment system and make it more practical.

Fund project

Doctoral Fund of Lingnan Normal University (No. 201106), the national Spark Program (No. 2013GA780086).

References

- [1] Li C W, Zhang J, Xie Q, Triple-network Convergence of Power Lines as a Transmission. *Applied Mechanics and Materials*. April, 2013, 321, 1264-1269.
- [2] Goryn Daniel, Hein Soren. On the Estimation of Rigid Body Rotation from Noisy Data, *IEEE Transactions on pattern Analysis and Machine Intelligence*, April 2013, 17(12), 219-122.
- [3] A view-dress based algorithm for rebuilding 3D models[J], *The 13th IEEE Joint International Computer Science and Information Technology Conference*, 2011, (01), 344-347.
- [4] S. Aljoscha, M.C. Chen. 3DAV exploration of video-based rendering technology in MPEG[J], *IEEE Transaction on Circuit and Systems for Video Technology*, 2010, 14(03), 348-356.
- [5] Eggert DW, Lorusso A, Fisher R B . Estimating 3-D Rigid Body Transformations: A Comparison of four Major Algorithms, *Machine Vision and Application*, 2013, 9, 272-290.
- [6] Zhang Jian, Yu sheng-sheng. Research on the Bit Computation Model for Time Variable Channel Bandwidth, *Wuhan University Journal of Natural Sciences*[J], 2007, 12(3), 442-446.
- [7] Yu Q S, Zhang J, Peng J X. A Whole Rate-Distortion Optimize Pronounce Based on Wireless Sensor Network. *Advanced Materials Research*, April. 2013, 756: 890-894.