Research on Transmission Performance of MVC Schemes Based on Improved Free-Viewpoint Video System Framework

QIUMIN XU[1], MEI YU [1, 2], GANGYI JIANG [1, 2], RANDI FU [1]

[1] Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[2] National Key Laboratory of Machine Perception, Peking University, 100871, China

Abstract: Based on the analysis of the classical free viewpoint video (FVV) system, an improved framework of FVV system for high interaction is proposed, which focuses on the requirements and capabilities of receiver side, and aims to achieve low cost of bandwidth, decoder and view generation. Under the framework, disparity search should be completed in server side so as to make view generation to be simple in receiver side. The disparity information is obtained by multi-view video encoder, and transmitted to the receiver side. On the other hand, in order to evaluate various schemes of multi-view video coding (MVC) more comprehensively, a new parameter called as Transmission Compression Efficiency is defined. The performance of transmission is analyzed among several MVC schemes, and Simulcast scheme is outstanding.

Key-Words: multi-view video coding, transmission performance, free viewpoint video system framework, compression efficiency

1 Introduction

Free viewpoint video [1] is one of the most attracting three-dimensional video in the future, for which scenes are captured from different angle of view with several cameras simultaneously, and the user can choose and watch the video with respect to arbitrary viewpoint interactively. The processes of a general FVV system [2] include acquisition, correction, multi-view video coding, transmission /storage, decoding, view generation and display. The system has to take into account the interrelations between all of these components. At the receiver side, multi-view images are reconstructed and used for FVV view generation. The view generation procedure should interpolate virtual viewpoint video from some existed views. Free-viewpoint images are displayed on a 2D/3D display.

The key aspects of multi-view video coding are that it provides high coding efficiency, operates with reasonable complexity and memory requirements, and facilitates random access. Several different prediction structures have been presented for efficient MVC. Different from conventional video system, high interaction between the server side and the receiver side (such as free-viewpoint television set) is the most important requirement in FVV system. Therefore, bandwidth cost and complexity of receiver are crucial factors to evaluate FVV system.

This paper focuses on high interactive FVV system, aims to achieve low bandwidth cost and view generation efficiently. It is organized as follows: In section 2, based on three different FTV frameworks, an improved FVV framework is proposed and its advantages have been described. In Section 3, a new parameter to evaluate the performance of transmission rate-distortion is defined, which different from compression efficiency for storage. Section 4 gives the comparison results of several MVC schemes in compression efficiency and transmission compression efficiency, and finally the conclusion is given.

2 An Improved Framework of FVV System

Nagoya University, NTT, and KDDI have proposed three different orders to construct FTV system [3], as shown in Fig.1. In their proposed FTV system, view generation is performed in the ray-space, so the function of view generation is divided into depth search and interpolation. In case A and case C, both depth search and interpolation are entirely performed at receiver side or at server side. In case B, depth search is performed at server side while interpolation is at the receiver side.

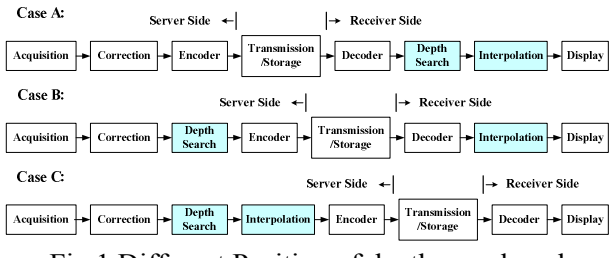


Fig.1 Different Position of depth search and interpolation in FTV system

Because of high complicated process of view generation, it is realized that in case A, processing at server side is heavy while processing at receiver side is light, the situation in case C is the opposite. In case B, depth search and interpolation are put in the server side and the receiver side respectively. Considering FVV application and services, the three frameworks can be applied in different conditions, such as download/package services, broadcast services and communication services.

In classical FVV system, the receiver side needs to decode multi-view video streams, and realizes view generation. However, the capability of receiver side is usually poor. It is assumed that the receiver side is a resource constrained terminal such as TV set, so its capability of memory and processing is not powerful enough to process the complicated operations. In this paper, we propose an improved user oriented framework of FVV system, as shown in Fig.2.

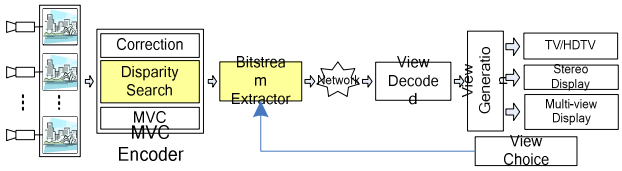


Fig.2 An improved user oriented framework of FVV system

Our framework is rooted on the framework in Case B, however, to achieve high efficient viewpoint access and generate virtual view in real time, it posses several different points as follows.

1) In our framework, view generation is performed in pixel-based form. Therefore obtaining accurate disparity information becomes an important procedure in view generation [4]. The disparity information for view generation is computed and encoded into bit streams by MVC encoder and transmitted to the receiver side. However, the increasing bit rate should be controlled within 10% to keep high compression efficiency.

2) Compared with Case B, our framework adopts the technique of Bitstream Extraction [5] after MVC encoder. In this case, only partial of bitstream should be transmitted to receiver side according to the feedback of users, thus the bandwidth to transmit the interactive views can be reduced effectively.

3 The Evaluation of the Performance of Transmission

Highly interaction is a primary feature of FVV system. The requirements of FVV system have been put forward in JVT and MPEG documents [2,3], which include data format, compression, rendering, and so on. Compression efficiency, low complexity, low delay, and random access are the main factors to evaluate the performance of FVV system, which have been discussed mostly. It is assumed that higher compression efficiency means less bandwidth cost. However, our FVV system, using the technique of Bitstream Extraction, is not the case.

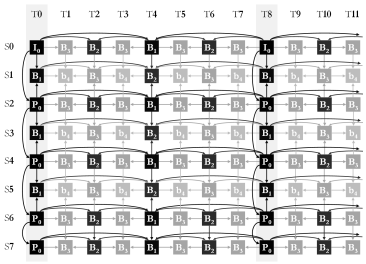
As we known, to generate the video with respect to the selected viewpoint, the server side is asked to deliver bit-streams of the demanded viewpoints or its neighboring viewpoints so as to generate the selected virtual views. Thus bandwidth has to be considered in FVV system design. However, the demand on transmission bandwidth is a choke point in communication of FVV system. Although, to reduce the bandwidth cost, high compression efficiency is no doubtfully to be necessary for MVC, just like conventional video coding, it is not satisfied if only the compression efficiency is taken into account because of huge quantity of multi-view data. By contrast, Bitstream Extraction is a simple and useful scheme to decrease transmitted bitstreams and reduce the cost of bandwidth, thus it has been proposed as a core technique in multi-view video coding. In this case, the server side extracts relevant parts from MVC bitstreams according to the feedback of user's choice and transmits the partial bitstreams to the receiver side.

However, the compression efficiency is unable to exhibit the performance of transmission rate-distortions. Thus, it is necessary to define a new parameter to exhibit the demand on transmission bandwidth. In this paper, we use "Transmission Compression Efficiency (TCE)" to denote the transmission rate-distortion performance. Different from compression efficiency, transmission compression efficiency only contains partial frames depending on the user's choice.

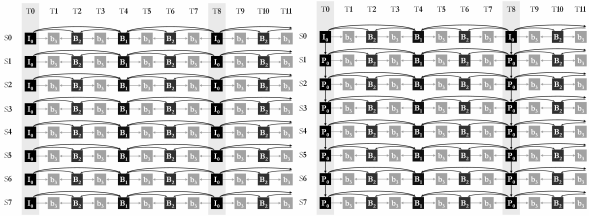
In many cases, the user chooses certain one viewpoint for a relatively long time. If the accessing position exists a viewpoint video which is captured by real camera, the server side only needs to transmit the chosen single viewpoint. Analogously, if the accessing position has no real viewpoint video or display terminal is a 3D display, it's needed to generate a virtual viewpoint to display. In this paper, we use the performance of single view transmission and double view transmission to represent Transmission Compression Efficiency.

4 Experimental Results

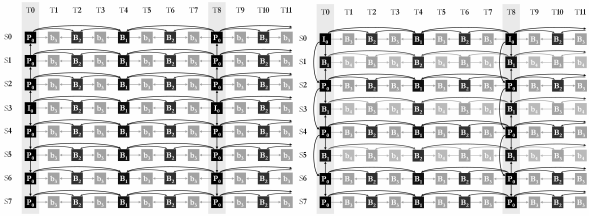
Up to now, various prediction structures for MVC have been proposed. In this paper, we compare two typical MVC schemes for analyzing the requirements in FVV system. HBP structure is one of the most efficient MVC schemes, which benefits from the carefully designed prediction structure in temporal domain as well as spatial domain. Inter-view-temporal prediction structure based on AVC using hierarchical B pictures is proposed by HHI [6], as shown in Fig.3(a). Here, "HBP-MVC" is short for the prediction structure. Because of its superior characteristics such as compression efficiency and the performance of temporal scalability, HBP-MVC has become an efficient scheme for MVC. However, if HBP-MVC structure is used in FVV system, in which high interaction becomes to be an important function to be realized, the structure seems to be un-competent, especially considering the performance of transmission rate-distortions. Fig.3(b) shows several MVC schemes, which have different inter-view prediction in the first time-stamp, and in temporal domain still based on HBP structure [7]. In this paper, we compare these MVC schemes in compression efficiency and transmission compression efficiency.



(a). Hierarchical B pictures structure.



Simulcast KS_IPP



KS_PIP KS_IBP

(b). Four MVC schemes from JVT-V132.

Fig.3 Experimental MVC schemes

Experiments have been made based on JMVM 2.1 [8] to compare the performance of "HBP-MVC", "Simulcast", "KS_IPP", "KS_PIP", "KS_IBP" schemes in compression efficiency and transmission compression efficiency, where multi-view video "ballroom" is used as the test sequence. "Ballroom" is a multi-view video with 8 viewpoints, the size of which is 640 x 480, which is showed in Fig.6. Table 1 lists the MVC parameters in the experiments.

Table 1. The parameters of MVC configure

Frame Rate	25
Search Range	± 96
Entropy Coding	CABAC
Basis QP	22, 27, 32, 37
QP Delta Quant	Default

Fig.4 gives the comparison between the five MVC schemes in the rate-distortion performance for compression efficiency. The result shows that for the same bit-rate PSNR of HBP-MVC schemes is highest, and that of Simulcast scheme is lowest, the difference is about 1.0~1.2 dB. For test sequence "ballroom", other schemes are mostly close to HBP-MVC. By contrast, Fig.5 shows the transmission rate-distortion performance about transmitting single viewpoint and double viewpoints. The results are obviously different from rate-distortion performance for compression efficiency as shown in Fig.4. From Fig.5, it is seen that Simulcast scheme surpasses other schemes significantly, in terms of the transmission compression efficiency, especially transmitting single viewpoint. For the same bit-rate, the PSNR of Simulcast scheme is higher than HBP-MVC scheme at about 3 dB when transmitting single viewpoint, and the gain is about 1.5 dB when transmitting double viewpoints.

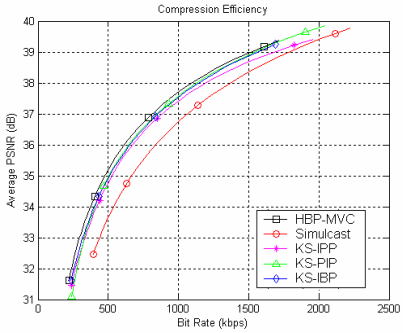
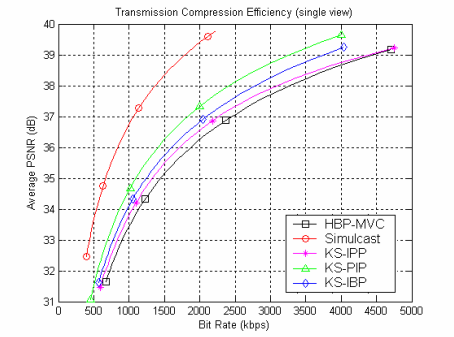
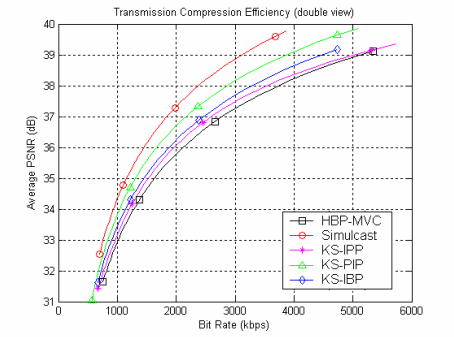


Fig.4 The rate-distortion performance for compression efficiency



(a) The transmission rate-distortion performance about transmitting single viewpoint



(b) The transmission rate-distortion performance about transmitting double viewpoints

Fig.5. Rate-distortion performance for transmission

5 Conclusion

We have proposed an improved framework of FVV system based on the requirements of FVV system. In the proposed framework, the information of disparity information is obtained by MVC encoder and transmitted so as to make view generation to be simple in the receiver side. Considering Bitstream Extraction used in FVV system for reducing the cost of bandwidth, we define a new parameter to evaluate the transmission performance of the prediction structure of MVC. Several MVC schemes are analyzed with compression efficiency and transmission compression efficiency. From the experimental results, it is realized that the rate-distortion performance for compression efficiency reflects the cost of memory needed to store the total bitstream in server side, while the transmission rate-distortion performance shows the cost of bandwidth for transmission. Therefore, the defined transmission compression efficiency can represent the performance of prediction structures in FVV system with heave interactivity.

Acknowledgement

This work was supported by the Natural Science Foundation of China (grant 60472100, 60672073), NSFC/KOSEF Joint Research Project, the Program for New Century Excellent Talents in University (NCET-06-0537), Natural Science Foundation of Ningbo (grant 2007A610037), the Key Project of Chinese Ministry of Education (grant 206059) , and K.C. Wong Education Foundation in Hong Kong.

References:

[1]. Masayuki Tanimoto, Overview of Free Viewpoint Television, Signal Processing: Image Communication, Vol. 21, No. 6, July, 2006, pp. 454-461.

[2]. ISO/IEC JTC1/SC29/WG11. Requirements on Multi-view Video Coding v.6, N8064, Montreux, Switzerland, April, 2006.

[3]. ISO/IEC JTC1/SC29/WG11. Proposal on Requirements for FTV, W127, San Jose, California, USA, April, 2007.

[4]. ISO/IEC JTC1/SC29/WG11, View synthesis for multiview video, M13123, Montreux, Switzerland, April, 2006.

[5]. ISO/IEC JTC1/SC29/WG11, Joint Scalable Video Model JSVM-7, T202, Klagenfurt, Austria, July, 2006.

[6]. K. Mueller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Oelbaum, T. Wiegand, Multi-View Video Coding Based on H.264/MPEG-AVC Using Hierarchical B Pictures, Picture Coding Symposuim'2006, Beijing, April, 2006.

[7]. ISO/IEC JTC1/SC29/WG11. Comparative Study of MVC Prediction Structures, V132, Marrakech, Morocco, January, 2007.

[8]. ISO/IEC JTC1/SC29/WG11. Joint Multiview Video Model (JMVM) 2.0, U207, Hangzhou, China, October, 2006.

Fig.6 Multi-view video "ballroom" with 8 viewpoints