International Journal of Hybrid Information Technology
Vol. 10, No. 8 (2017), pp.91-100
http//dx.doi.org/10.14257/ijhit.2017.10.8.07

Content-Based Scalable Multi-View Video Coding Using 4D Wavelet

Yi Lai[1,2,3*] , Qian Wang[1,2,3] and Yin Gao[1,2,3]

[1]School of Telecommunication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, PR China
[2]Key Laboratory of Electronic Information Application Technology for Scene Investigation, Ministry of Public Security, Xi'an 710121, PR China
[3]International Joint Research Center of Shaanxi Province, Xi'an 710121, PR China
laiyi0614@163.com

Abstract

Most existing scalable multi-view video coding (SMVC) can provide temporal, spatial, SNR as well as view scalabilities, but not support content scalability. In this paper, a region of interest (ROI) coding is developed for scalable multi-view video codec to show the high scaling capability in terms of content scalability. The key component of this algorithm is that the bit-planes of the ROI can be lifted according to users' requirements. And in this way the ROI can be allocated more bits even at low bit-rate and can secure more bits to get a better visual quality. As a result of our experiments, the peak signal to noise ratio (PSNR) within ROI can be improved by 2~4dB and the subjective visual quality of ROI in the proposed scheme can be improved compared to the conventional SMVC algorithm.

Keywords*: Multi-view video coding; ROI; lifting*

1. Introduction

Multi-view video (MVV) can provide consumers with depth effect to the observed scene as if it really existed in front of consumers, allow consumers to freely change views, and interactively modify the properties of a scene. Multi-view video coding (MVC) has recently gained significant attention because of its importance for future multimedia applications including 3DTV (3D Television), FTV (Free Viewpoint Television) and immersive teleconference etc [1-4]. For example, in immersive teleconference, there is an interaction between consumers. Participants at different geographical sites meet virtually and see each other in either free viewpoint or 3DTV style [5-10]. The immersive teleconference provides a more natural way of communications [11-12]. They expand the user experience beyond what is offered by traditional media. Multi-view video shows the same 3D scene from different viewpoint at the same time interval, resulting in a tremendous amount of raw video data. Therefore highly efficient compression is necessary in storage and delivery of multi-view video than other single view data [13-15]. The straight-forward solution for this would be to encode all video signals independently using a state-of-the-art video codec such as H.264/AVC [16]. In the mean while, the different camera signals contain a large amount of statistical dependencies. The point for better coding efficiency lies in efficient exploitation of these inter-view correlations in addition to temporal and spatial correlations within a single view that have been studied well in the conventional video coding [17]. In [18], pioneering work on multi-view video coding was reported.

ISSN: 1738-9968 IJHIT

At present, most MVC technologies can be grouped into two categories: the traditional hybrid DCT-based video coding schemes and the wavelet-based ones. The former has utilized the temporal and view correlations alternatively in the prediction. The JMVM (Joint Multiview Video Model) multi-view video coding method developed within the JVT (Joint Video Team) is an extension of a conventional block-based predictive video codec. But they neither entirely exploit the redundancy among different views nor provide an easy way of implementation for scalabilities, which have become a more and more important feature for video coding and communications [19]. On the other hand, the wavelet-based coding has been proved to be a good way to get both coding performance and full scalabilities. Combination of scalability with MVC was investigated, e.g. in [20-23]. They all have in common that a high-dimensional wavelet transform for the decorrelation of the multi-view video is applied. It is generally composed of lifting-based motion compensated temporal filtering (MCTF) in the temporal dimension, disparity compensated view filtering (DCVF) in the view dimension, and 2D discrete wavelet transform (DWT) in the spatial dimensions. These codecs can be regarded as extensions to exiting single-view wavelet video codecs to fully explore the correlation across views.

However, on most occasions, users pay much more attention to the interested area, while less attention to the other area, named the background. Hence, the scalable enhancement of visual quality (VQ) of ROI according to its importance defined by the user has become a major concern, especially in low bit-rate condition. This is reported less about the exiting MVC technologies. In such low bit rate communication channel as the wireless mobile network, owing to the insufficient bit budget at the compression process, MVV compression will cause the heavy loss of visual detail information, including the region which people are particularly interested in. Therefore we propose a novel content-based scalable multi-view video coding (CSMVC) method to solve this problem. In this scheme, the bit-planes of the ROI is lifted according to users' demands and transmission conditions and a higher bit-rate is allocated to the ROI for scalability in content, thus maintaining a better visual quality of the salient information even at low bit-rate. Experimental results confirm the efficiency of the proposed approach.

The rest of this paper is organized as follows. Section 2 introduces the ROI technology for MVC. In section 3, the content-based SMVC method is presented in detail. Experimental results are given in section 4. And finally we conclude the whole paper in section 5.

2. ROI Technology for SMVC

In visual content there are typically objects or regions that particularly draw the viewer's attention, usually referred to as ROI. Therefore ROI is coded earlier in the codestream than the background in order to achieve the ROI with better quality than the background while maintaining a fair amount of compression.

After accomplishing segmenting and tracking of the original video, the region of interest for each frame can be obtained. The information for rectangle shape or arbitrary shape of the ROI is given. Then we can get the two-value template (MASK), represented by $M(x,y)$ [24]. We have:

$$M(x,y)=\begin{cases}1 & \text{wavelet coefficient } (x,y) \text{ is need}\\ 0 & \text{others}\end{cases} . \tag{1}$$

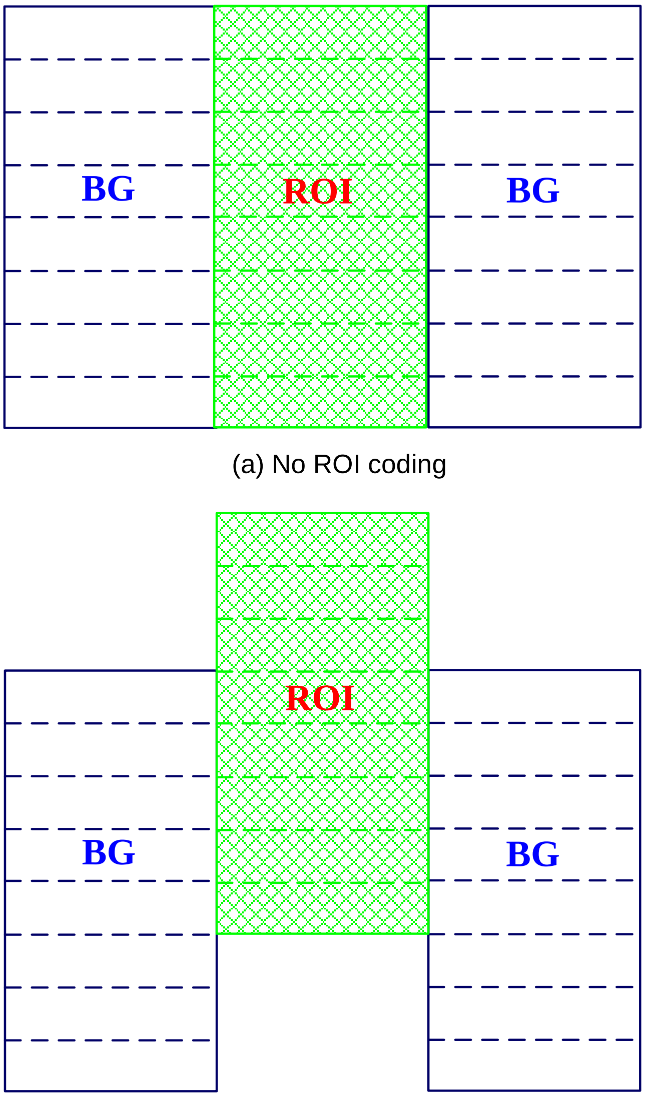
The mask shows which quantized transform coefficients must be encoded with better quality so that the ROI can be encoded with better quality.

The lifting of bit-planes is usually carried out when quantization has been completed. The quantized transform coefficients within the ROI are scaled up in order to place the bits associated with the ROI in higher bit-planes than the background. This means that when the entropy coder encodes the quantized transform coefficients, the bit-planes associated with the ROI are coded before the information associated with the background. The lifting of bit-planes can be formulated by:

$$\left|q_b(u,v)\right| = \left|q_b(u,v)\right| \cdot 2^s \quad . \qquad (2)$$

where $q_b(u,v)$ is the quantized transform coefficients of a given sub-band b and s is the number of lifting.

The procedure of the lifting of bit-planes within ROI is illustrated in Figure 1. BG refers to a background. Note that, the quantized transform coefficients within the ROI are placed in higher bit-planes when the coder encodes and as a result, ROI is coded earlier than the background in the bitstream. Therefore, the decoder can construct the desired region with better quality than the background.



(a) No ROI coding

(b) General scaling ROI-based coding

Figure 1. The Procedure of Lifting of Bit-Planes Within ROI

3. Proposed System Architecture

Low bit-rate video compression will cause loss of some detailed information, even the region which people are particularly interested in. In this paper, a 4D wavelet transform with the lifting of bit-planes within ROI is proposed for multi-view video coding and then we can easily realize scalable multi-view video coding system which shows the high scaling capability in terms of temporal, view, spatial, SNR, and especially content scalability.

A block diagram of the ROI coding scheme can be seen in Figure 2. In order to decorrelate the multi-view video data temporally, MCTF is applied to each video sequence of each camera. The motions of each view video are estimated by using hierarchical variable size block matching (HVSBM). According to the motion vectors, the pixels in current macro-blocks can be mapped to the reference samples. MCTF can be performed along the motion trajectory using LG5/3 or other lifting filters. At the end of the first stage, the low-pass frames are input into the second-level filtering, while high-pass frames are directly sent to the 2D spatial DWT module. At last, the t-LLL band of every view can be input into DCVF module. The use of a wavelet lifting structure guarantees perfect invertibility of this step, and as a consequence of its open-loop architecture, temporal scalability are attained.

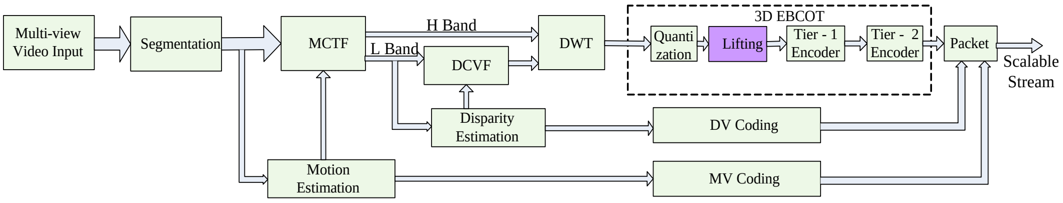


Figure 2. The Block Diagram of the Proposed Content-Based Scalable Multi-View Video Coding

Disparity can also be regarded as the motion along the view axis. Therefore disparity vectors can be estimated by using the same block matching as the temporal transform. Disparity vectors are usually larger than motion vectors in most natural video sequences. Hence a bigger search range should be taken when disparity estimation is carried out. Meanwhile, temporal low-pass subbands can be treated as a frame rate reduced version of the original video. If the structure of the MCTF is the same for all views, then the inter-view correlation of the L-frames must be nearly equal to the subsampled original frames [21]. Hence, in order to remove redundancy between different views, a hierarchical decomposition in a similar way to the temporal direction is applied to L-frames which is constituted by the t-LLL band of every view. This process is called DCVF. This view directional filtering reduces the correlation across the different views and gets the scalability in view direction at the same time.

Next, the high-pass frames of MCTF and the output of DCVF are transformed by wavelet, which can easily support spatial scalability. As stated earlier, the bit-planes of the quantized transform coefficients within ROI can be lift according to the user’s demand and transmission conditions, so that the ROI can be allocated more bits even at low bit-rate, and can secure more bits to get a better visual effect. The MASKs are delivered in the way of odd frames in accordance with low-frequency ones, even frames corresponding to high-frequency ones. Then content scalability can be elegantly provided.

The embedded block coding with optimized truncation is employed to encode the 4-D wavelet coefficients to bitstream. The motion vectors and disparity vectors are encoded by context-adaptive binary arithmetic coding, and the ROI shape parameters are encoded into bitstream.

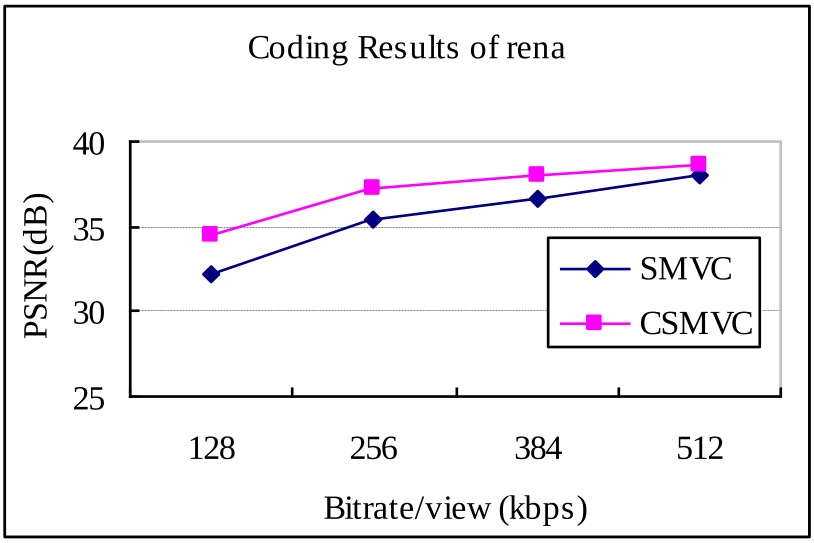
4. Experimental Results

To verify the performance of the proposed MVC scheme, we carry out experiments for two multi-view sequences, called ‘rena’ and ‘flamenco’ [25]. Only a subset of cameras are used from two sequences since processing all views would require too much effort. The spatial resolution of both sequences used in experiments is 640×480 at 30 fps. The number of levels for the view decomposition has been chosen in all experiments to be 2 so that the number of view is set to be 4. The number of decomposition is 3 both in temporal and spatial domain for the experiments. LG5/3 lifting filter is used for MCTF and DCVF in proper order, and CDF9/7 lifting filter is then utilized for DWT. The search range is set to be 16 and 64 pixels for block matching in MCTF and DCVF, respectively.

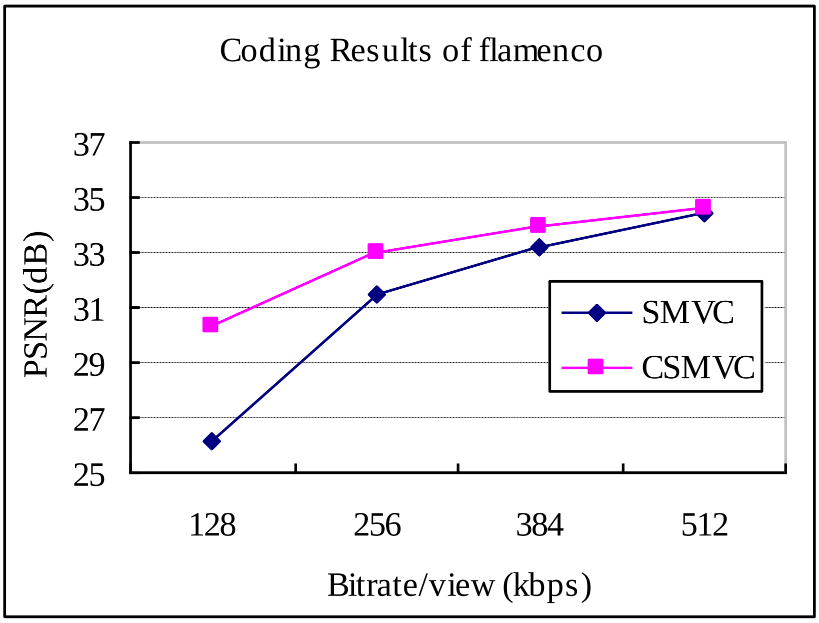
The objects or regions that significantly draw the viewer’s attention are usually selected as ROI. Hence, we give more concern about PSNR within ROI in our experiments. For simplicity, the rectangle shape ROI is used. The only one girl is chosen as ROI in sequence ‘rena’ and the girl wearing yellow skirt is chosen as ROI in sequence ‘flamenco’. The bit-planes are lifted by 3 in all experiments.

Above all, the proposed CSMVC is compared with SMVC. The PSNR curves of the two methods are given in Figure 3. The PSNR is the average PSNR over 4 views within ROI of sequences. Some gain in quality is achieved for both sequences, especially more at low bit rate. The average PSNR of our proposed method is 2.26 dB for ‘rena’ and 4.15 dB for ‘flamenco’ higher than SMVC at the bit rate of 128 kbps per view. This is because the proposed schemes allocate more bit rates to the ROI than the background by lifting bit-planes of ROI while SMVC averagely allocate bit rates to the whole image plane. In addition, ‘flamenco’ contains motion of other dancers so that the PSNR gain is lower than ‘rena’. As the bit rate increases, the background can be allocated enough bit rate at the same time that the ROI is allocated more bit rate. Therefore the PSNR gains become smaller while the bit rate increases. But our proposed is still more efficient than SMVC.

Figure 4 and 5 show some decoding results of ‘rena’ and ‘flamenco’ with SMVC and CSMVC, respectively. In Figure 4 (a) and (b), the face and arms of the girl is not clear, but a better quality of the ROI can be achieved in Figure 4(c) and (d). Certainly, the quality of the background in Figure 4(c) and (d) is worse than Figure 4 (a) and (b), which is at the cost of lifting the bit-planes of the ROI. The similar improvements have been depicted in Figure 5. As seen from these results, the reconstructed subjective quality of the ROI with the proposed coding can be saliently improved compared to SMVC. This proves the content-based scalable multi-view video coding is efficient.

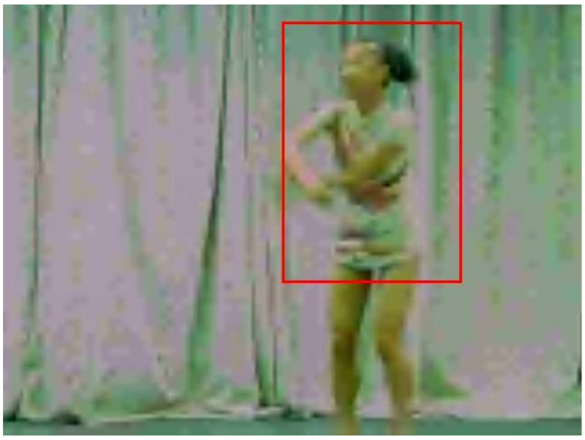


(a) Rena

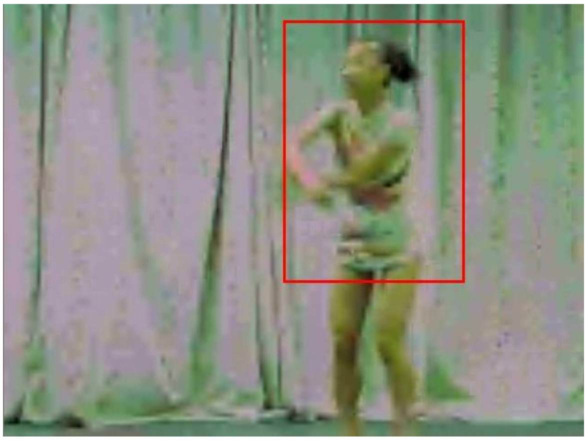


(b) Flamenco

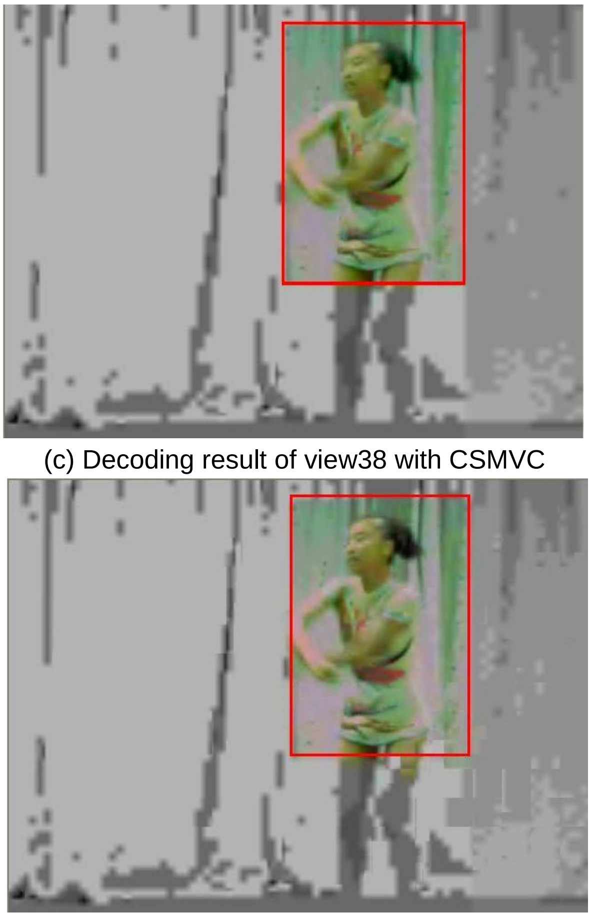
Figure 3. Performance Comparison of SMVC with CSMVC for Tested Multi-View Sequences



(a) Decoding result of view38 with SMVC



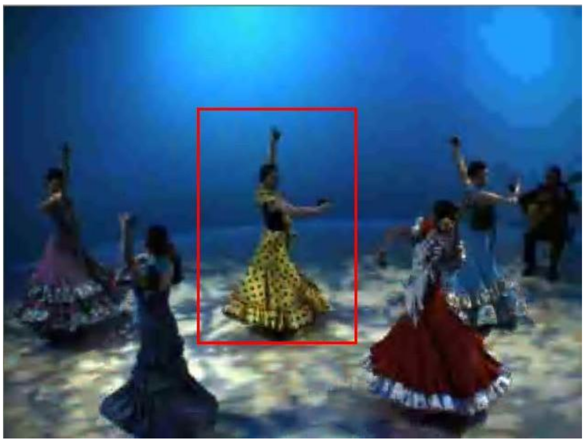
(b) Decoding result of view39 with SMVC



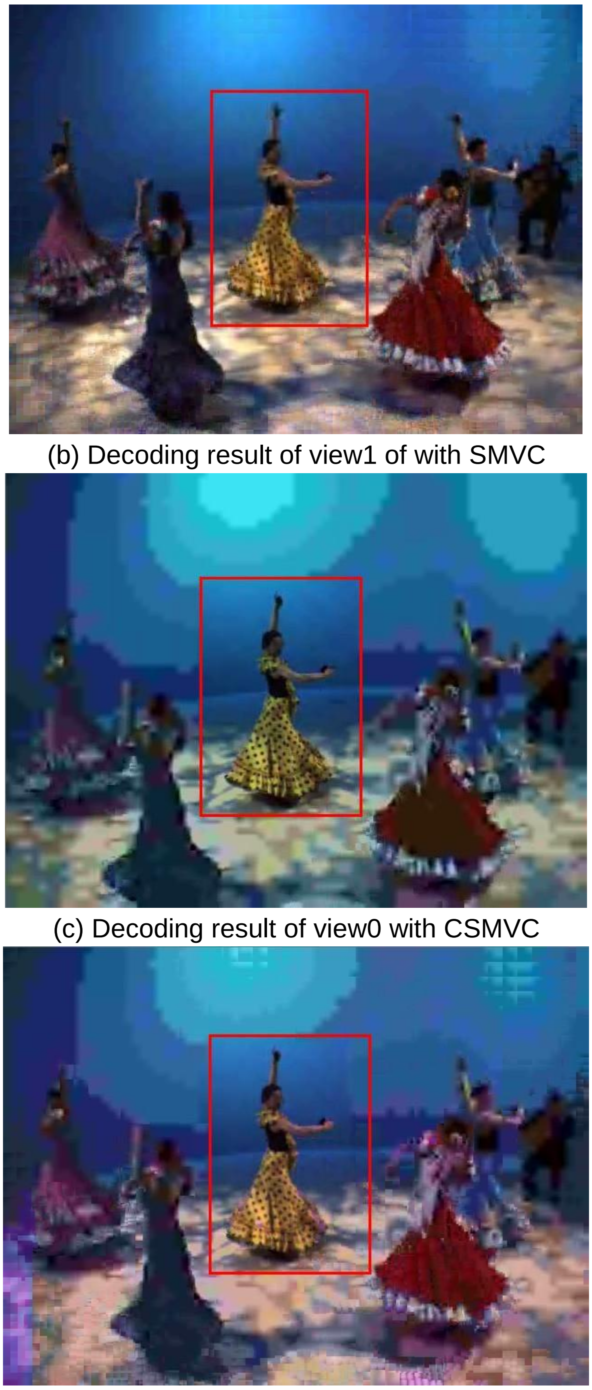
(c) Decoding result of view38 with CSMVC

(d) Decoding result of view39 with CSMVC

Figure 4. Decoding Results of Frame7 of Rena at 128kbps



(a) Decoding result of view0 with SMVC



(b) Decoding result of view1 of with SMVC

(c) Decoding result of view0 with CSMVC

(d) Decoding result of view1 with CSMVC

Figure 5. Decoding Results of Frame21 of Flamenco at 256kbps

5. Conclusions

In this paper, a novel scalable multi-view video coding scheme based on lifting the bit-planes of ROI is proposed. It can support temporal, view, spatial, SNR as well as content scalabilities. Experimental results demonstrate the efficiency of the proposed method. In the future work, we will further study the adaptive ROI quality adjustable rate control approach for CSMVC. More efficient disparity estimation is another very important future research topic.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (41504115), Shaanxi Natural Science Foundation Research Project (2015JQ6223), and Strengthening Police science and technology Foundation form Ministry of Public Security (2015GABJC50).

References

[1] A. D. Abreua, L. Tonia, N. Thomosc, T. Maugeyd, F. Pereirab, P. Frossarda. “Optimal layered representation for adaptive interactive multiview video streaming”, Journal of Visual Communication and Image Representation, vol. 33, no. 11, **(2015)**, pp. 255-264.

[2] Y. C. Fan, Y. C. Chen, and S. Y. Chou, “Vivid-DIBR based 2D–3D image conversion system for 3D display”, Journal of Display Technology, vol. 10, no. 10, **(2014)**, pp. 887-898.

[3] J. Xiao, M. Hannuksela, T. Tillo, et al. “Scalable bit allocation between texture and depth views for 3-D video streaming over heterogeneous networks”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 1, **(2015)**, pp. 139 - 152.

[4] C. Jiang, S. Nooshabadi, “A scalable massively parallel motion and disparity estimation scheme for multiview video coding” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 2, **(2016)**, pp. 346 - 359.

[5] Wang RS, Wang Y, “Multiview video sequence analysis, compression, and virtual viewpoint synthesis” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 3, **(2000)**, pp.397-410.

[6] Droe M, Clemens C, Sikora T, “ Extending single-view scalable video coding to multi-view based on H.264/AVC” , In Proceedings of 2006 IEEE International Conference on Image Processing, Los Atlanta, USA, **(2006)** 2006 May 2977-2980.

[7] Yang WX, Lu Y, Wu F, “ 4-D wavelet-based multiview video coding” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 11, **(2006)**, pp. 1385-1396.

[8] Anantrasirichai N, Canagarajah CN, Redmill DW, “ In-band disparity compensation for multiview image compression and view synthesis” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 4, **(2010)**, pp. 473-484.

[9] Garbas JU, Pesquet-Popescu B, Kaup A, “Methods and tools for wavelet-based scalable multiview video coding” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 2, **(2011)**, pp.113-126.

[10] Smolic A, Mueller K, Stefanoski N, “ Coding algorithms for 3DTV - A survey” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, **(2007)**, pp.1606-1621.

[11] Jarusirisawad S, Saito H, “3DTV view generation using uncalibrated pure rotating and zooming cameras” , Signal Processing-Image Communication, vol. 24, no. 1, **(2009)**, pp.17-30.

[12] Shum HY, Kang SB, Chan SC, “ Survey of image-based representations and compression techniques” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 11, **(2003)**, pp.1020-1037.

[13] Takahashi K, “ Theoretical Analysis of View Interpolation With Inaccurate Depth Information” , IEEE Transactions on Image Processing, vol. 21, no. 2, **(2012)**, pp.718-732.

[14] Yamamoto K, Kitahara M, Kimata H, “ Multiview video coding using view interpolation and color correction” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, **(2007)**, pp.1436-1449.

[15] Asai T, Kanbara M, Yokoya N, “ 3D Modeling of outdoor environments by integrating omnidirectional range and color images” , In Proceedings of Fifth International Conference on 3-D Digital Imaging and Modeling, Los Alamitos, USA, **(2005)** May 447-454.

[16] ITU-T Rec. & ISO/IEC 14496-10 AVC, “Advanced Video Coding for Generic Audiovisual Services, Version3”, **(2005)**.

[17] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, “Efficient Prediction Structures for Multiview Video Coding” , IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, **(2007)**, pp.1461-1473.

[18] M. Magnor, P. Ramanathan, and B. Girod, “Multi-view coding for image-based rendering using 3-D scene geometry”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 11, **(2003)**, pp. 1092–1106.

[19] W. Yang, F. Wu, Y. Lu, J. Cai, K. N. Ngan, and S. Li, “Scalable Multiview Video Coding Using Wavelet”, IEEE International Symposium on Circuits and Systems , Kobe, Japan, **(2005)** May 6078-6081.

[20] W. Yang, Y. Lu, F. Wu, J. Cai, K. N. Ngan, and S. Li, “4D Wavelet-Based Multi-view Video Coding”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 11, **(2006)**, pp. 1385–1396.

[21] J. Garbas, U. Fecker, T. Tr¨oger and A. Kaup. “4D Scalable Multi-View Video Coding Using Disparity Compensated View Filtering and Motion Compensated Temporal Filtering”, International Workshop on Multimdedia Signal Processing, Victoria, Canada, (**2006)** October 54-58.

[22] M. Dr¨ose, C. Clemens, and T. Sikora, “Extending single-view scalable video coding to multi-view based on H.264/AVC”, 2006 IEEE International Conference on Image Processing, Atlanta, USA, (**2006)** October 2977 - 2980.

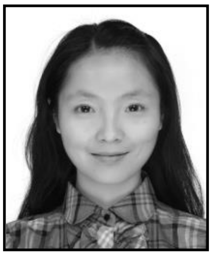
[23] N. Ozbek and A. M. Tekalp, “Scalable multi-view video coding for interactive 3DTV”, 2006 IEEE International Conference on Multimedia and Expo, Toronto, Canada, (**2006)** July 213 - 216.

[24] JTC1/SC29, “Information technology-JPEG 2000 image coding system: Core coding system”, ISO/IEC15444-1:**(2004)**(E),

[25] ISO/IEC JTC1/SC29/WG11 Doc.N7567, “Updated Call for Proposals on Multi-view Video Coding”, Nice , France, **(2005)** October.

Authors

Yi Lai, he received the Ph.D. degree in control science and engineering at Xi'an Jiaotong University in 2013. He is currently a lecture of school of telecommunication and information engineering at Xi’an University of Posts and Telecommunications. His research interests include image processing and analysis, computer vision and multi-view video coding.



Qian Wang, she was born in 1983. She received the Ph.D. degree in electronic engineering from Xidian University in 2011. Since 2011, she has been with Xi`an University of Posts and Telecommunications, where she is currently an associate professor in the school of telecommunication and information engineering. Her research interests include image processing and analysis, computer vision and video compression.

Yin Gao, she received her bachelor’s degree from Xi’an University of Posts and Telecommunication in 2015. She is currently pursuing her master’s degree at the school of telecommunication and information engineering, Xi’an University of Posts and Telecommunication. Her research interests include video retrieval and shot detection .