CODING OF 3D HOLOSCOPIC IMAGE BY USING SPATIAL CORRELATION OF RENDERED VIEW IMAGES

Deyang Liu, Ping An, Chao Yang, Ran Ma, Liquan Shen

School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China

ABSTRACT

Holoscopic imaging is a prospective acquisition and display solution for providing natural and fatigue-free 3D visualization. However, large amount of data is required to represent the 3D holoscopic content. Therefore, efficient coding schemes for this particular type of image are needed. In this paper, an effective coding scheme is proposed by exploring the spatial correlation among the view images with different perspectives rendered from 3D holoscopic image. We utilize the interlaced view image to descript such spatial correlation. A linear prediction method is used on the interlaced view image instead of the original holoscopic image directly. Experimental results show that the proposed coding scheme performs better than HEVC intra standard and screen content coding extension of HEVC with around 2.41dB and 0.42 dB average quality improvement respectively.

Index Terms— 3D holoscopic image, image coding, spatial correlation, interlaced view image, HEVC

1. INTRODUCTION

With the 3D autostereoscopic display technologies becoming more and more mature, users can enjoy a more immersive glassless 3D experience. However, a fundamental limitation of the most existing autostereoscopic display systems is that they tend to cause some uncomfortable feelings for users, such as eye strain or headache after prolonged viewing.

Holoscopic imaging, also referred to as integral, light field, or plenoptic imaging, captures both spatial and angular information of a 3D scene and is considered to be a prospective acquisition and display solution to supply a more nature and fatigue-free 3D visualization [1]. So far, many research groups have considered the standardization of 3D holoscopic application. The JPEG working group starts a new study, known as JPEG Pleno [2], aiming at richer image capturing, visualization, and manipulation. The MPEG group has also started the third phase of Free-viewpoint Television (FTV) since 2013, targeting super multiview, free navigation and full parallax imaging applications [3].

In the simplest form, the holoscopic imaging system with full parallax consists of a lens array mated to a digital sensor. In the acquisition side, the light rays emanated from the 3D scene are captured by the lens array and then recorded by a 2D image sensor. Each lenslet can capture one perspective view of the 3D scene at a slightly different angle to its neighbors. The recorded light rays through different lenses are called elemental images (EIs), shown in Fig.1. In the visualization side, the 3D object is constructed by the intersection of the ray bundles emitted from the recorded holoscopic image behind the same lens array.

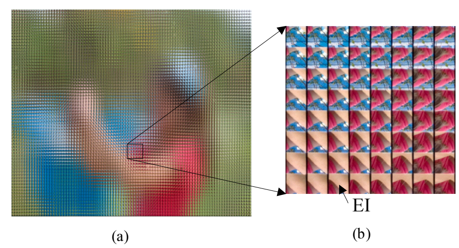

Fig. 1. 3D holoscopic image “Laura”: (a) full image (b) partial enlargement

In order to represent the captured 3D holoscopic image with adequate resolution, large amount of data is required and, consequently, effective compression schemes become of paramount importance for such particular type of content. Due to the fact that the EIs exhibit repetitive patterns and a large amount of redundancy exists between the neighboring EIs, several 3D holoscopic image coding frameworks have been proposed to remove such redundancy. In [4], a prediction coding work is proposed, in which the Self-Similarity (SS) mode [5] is introduced into HEVC to improve the coding efficiency of 3D holoscopic image. The SS mode is similar to the Intra Block Copying (IntraBC) mode [6], which is used to code the screen contents. The IntraBC mode has then been considered in HEVC range extension developments [7-8]. In [9], a disparity compensation based 3D holoscopic image coding algorithm is put forward to derive a better prediction. A coding method combining the locally linear embedding (LLE) is proposed in [10] to predict the coding block.

We have mentioned above that the 3D holoscopic imaging captures both spatial and angular information of a 3D scene. Therefore, views with different perspectives can be rendered from the 3D holoscopic image. High spatial correlation also exists among the rendered view images and such correlation can also be used to compress the 3D holoscopic image. However, such high spatial correlation is not fully explored by the above mentioned coding schemes. Therefore, in this paper, a 3D holoscopic image coding scheme fully exploring such spatial correlation is proposed. In order to descript the spatial correlation among the rendered view images clearly, we propose to use the interlaced view image to represent the 3D holoscopic image. A linear prediction method is used on the interlaced view image instead of the original holoscopic image directly. In the linear prediction method, we first search for the *K*-NN patches of the coding block within the reconstructed neighboring view images to constitute the prediction supports. Then a linear combination of the prediction supports is utilized to estimate the coding block. Since HEVC standard can significantly improve the compression performance of high definition videos with half of the bit rate saved compared to the H.264/AVC for the same perceptual video quality, HEVC standard promises to improve the coding efficiency for holoscopic imaging coding. Therefore, we choose HEVC to carry out the proposed coding method.

The remainder of the paper is organized as follows. Section 2 presents the proposed 3D holoscopic image coding scheme. The experimental results are presented and analyzed in section 3, while the concluding remarks are given in section 4.

2. PROPOSED CODING SCHEME

The proposed coding method aims to improve the coding efficiency by exploring the spatial correlation among the rendered view images from 3D holoscopic image. Fig.2 and Fig.3 present the overview diagram of the proposed encoding and decoding scheme. The details of each block in the diagrams will be explained in the following subsections.

2.1. Encoding

2.1.1. Interlaced view image obtainment

3D holoscopic image contains both the spatial and angular information of a 3D scene. Viewpoint images can be rendered from 3D holoscopic image, where the viewpoint images represent the orthographic projections of the captured 3D scene in different directions. The simplest way to construct a single viewpoint image is to extract one pixel with the same relative position from each EI of a given 3D holoscopic image. However, extracting only one pixel from each EI results in disappointingly low resolution and the rendered view image suffers from severely blocky artifacts. This may weaken the spatial correlation of the neighboring viewpoint images. Therefore, in the proposed method, we construct a single viewpoint image by extracting a patch with the same relative position from each EI [11]. The process to construct viewpoint images from 3D holoscopic image is shown in Fig.4. Suppose that a $P \times P$ patch is extracted with the same relative position from each EI with size $n_x \times n_y$. With $N_x \times N_y$ EIs in 3D holoscopic image, the final rendered view image is of size $P \cdot N_x \times P \cdot N_y$.

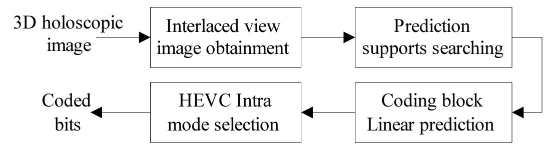

Fig. 2. The proposed 3D holoscopic image encoding system

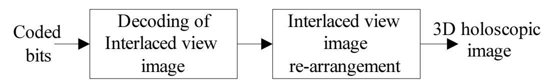

Fig. 3. The proposed 3D holoscopic image decoding system

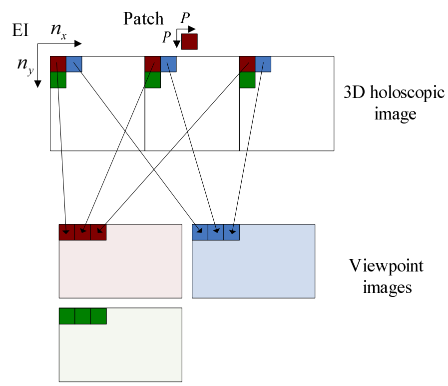

Fig. 4. Process to construct viewpoint images from 3D holoscopic image

By varying the relative position of the patch in each EI, viewpoint images with different horizontal and vertical viewing angles can be derived. After all the patches in each EI are extracted, an array of viewpoint images can be acquired. We then stitch them together to construct the interlaced view image. Fig.5 gives an example of interlaced view image.

2.1.2. Prediction supports searching

Let the pixel values in the current coding block be stacked in a column vector $\mathbf{x}_0$ and the pixel values in its neighbour templates with template thickness being *T* are compacted in a column vector $\mathbf{y}_0$. The principle of acquiring the prediction supports of the coding block is to first search for *K*-NN patches of $\mathbf{y}_0$ within the reconstructed neighboring viewpoint images under Euclidean distance, and then obtain the prediction supports of coding block according to *K*-NN patches of $\mathbf{y}_0$. Since the searching is very time-consuming, specified searching windows are used to reduce the searching complexity.

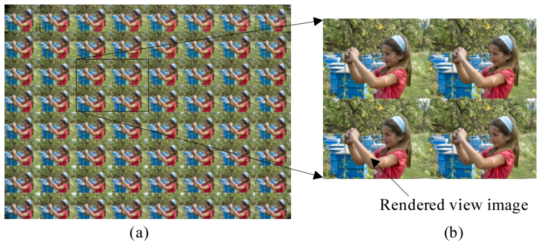

Fig. 5. Interlaced view image: (a) full image (b) partial enlargement

Suppose the coordinate of current coding block is (c_x, c_y), then the coordinates of the corresponding blocks with the same relative position in left, top and top-left neighbouring viewpoint images are $(c_x - W_{view}, c_y)$, $(c_x, c_y - H_{view})$ and $(c_x - W_{view}, c_y - H_{view})$, shown in Fig.6. The W_{view} and H_{view} represent the width and height of viewpoint image, where $W_{view} = P \cdot N_x$ and $H_{view} = P \cdot N_y$. The searching range in left neighbouring viewpoint image is set to $2H$ in horizontal direction. In top neighbouring viewpoint image, the searching range is appointed to $2V$ in vertical direction. In top-left neighbouring viewpoint image the searching range is specified to $V \times H$. The searching ranges in neighbouring viewpoint images are illustrated by the embedded red rectangles in Fig.6.

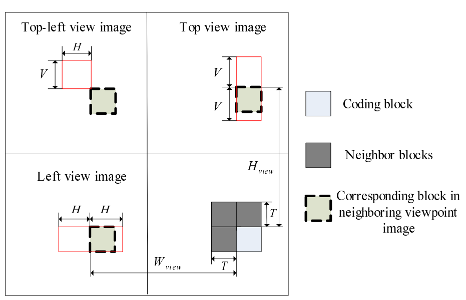

Fig. 6. Searching windows in reconstructed neighbouring viewpoint images

2.1.3. *Coding block linear prediction*

Suppose vectors $\{\mathbf{y}_k | k = 1, 2, ..., K\}$ represent the K-NN patches of $\mathbf{y}_0$, and vectors $\{\mathbf{x}_k | k = 1, 2, ..., K\}$ represent the prediction supports of vector $\mathbf{x}_0$. The main goal of linear prediction of coding block is to estimate vector $\mathbf{x}_0$ by using a linear combination of the prediction supports, which can be given by

$$\widetilde{\mathbf{x}}_0 = \sum_{k=1}^{K} \mathbf{w}_k \mathbf{x}_k \quad (1)$$

where $\widetilde{\mathbf{x}}_0$ is the estimator of vector $\mathbf{x}_0$, $\mathbf{w}_k$ is the weight vector.

In order to obtain the weight vector, we adopt a more direct way. That is to minimize the residual energy $\varepsilon(\mathbf{w})$ by solving a squared error function, where $\varepsilon(\mathbf{w})$ is defined by

$$\varepsilon(\mathbf{w}) = \| \mathbf{y}_0 - \sum_{k=1}^{K} \mathbf{w}_k \mathbf{y}_k \| \quad (2)$$

2.1.4. *HEVC intra mode selection*

The linear prediction method is implemented in HEVC by replacing the least statistically used mode of the 35 intra prediction modes [17]. In other words, the prediction samples that are generated by the substituted directional mode are replaced by the outputs produced by the linear prediction. It is worth noting that in the linear prediction method, all the coding blocks are predicted under the same replaced intra mode. The replaced intra mode is called linear prediction mode in this paper. HEVC rate-distortion optimization (RDO) procedure is adopted to choose the optimal block partition and the prediction mode. In addition, the proposed framework can also be used to predict chrominance blocks.

2.2. Decoding

2.2.1. *Decoding of the interlaced view image*

The decoding procedures of the interlaced view image are uniform to the ones in the encoding. It means that, in the decoder side, if linear prediction mode is selected as the optimal prediction mode, the decoder has to do the same prediction supports searching procedure as the encoder. After the prediction supports are derived, a linear combination of the prediction supports is used to derive the estimator of the coding block. Otherwise, HEVC angular intra prediction is utilized to acquire the prediction of the coding block.

2.2.2. *Interlaced view image re-arrangement*

After the interlaced view image is decoded correctly, the 3D holoscopic image can be obtained by re-arranging the interlaced view image. The interlaced view image re-arrangement is an inverse process of the interlaced view image obtainment, where each viewpoint image is subdivided into patches and each patch is mapped to its original position in the 3D holoscopic image.

3. EXPERIMENTAL RESULTS

3D holoscopic images Bike, Fountain, Laura and Seagull [12] are used in the test. These images are densely sampled with a different depth distribution and scene. The original images have a resolution of 7240×5432 and each EI is of size 75×75 with a rectangular shape. Since vignetting appears at the edge of each EI, we cut the EIs into size of 64×64 from approximately the center position of each EI and attached them together to form a processed 3D holoscopic image. The processed images are of size 6080×4544. All images are transformed into YUV 4:2:0 formats. HEVC Test Model (HM) reference software version 13.0 is modified for the proposed coding scheme. The configuration parameter is set as the "All Intra-Main", which is defined in [13]. Four tested Quantization Parameters 22, 27, 32 and 37 are used. The proposed coding method is evaluated against four prediction schemes utilized to compress the processed holoscopic image directly: the original HEVC intra standard (referred to as "HEVC"), the Range Extension HEVC reference software version 6.0 [7], where IntraBC prediction is used (referred to as "HEVC RExt"), the disparity compensation based 3D holoscopic image coding method proposed in [9] (referred to as "DCCM"), and the screen content coding extension of HEVC [8] (referred to as "HEVC SCC"). The configuration parameter of HEVC RExt is set to "All Intra" [14]. And configuration parameter of HEVC SCC is set to "All Intra" defined in [15].

In the interlaced view image obtainment, the patch size is set to 8, which allows an artifact free rendering for the presented parts [16]. In prediction supports searching, the template thickness *T* is set to the size of coding block. In the proposed method, we search for 3 nearest patches in each neighbouring view images to construct the prediction supports. The search range *H* and *V* are all set to 16.

Table 1. Rate Distortion Gains Over HEVC Intra Standard

Images	Coding Methods	BD-PSNR (dB)	BD-Rate (%)
Bike	HEVC RExt	1.35	-18.51
	DCCM	1.63	-23.21
	HEVC SCC	1.69	-24.57
	Proposed	**2.00**	**-30.66**
Fountain	HEVC RExt	1.52	-22.11
	DCCM	1.77	-26.25
	HEVC SCC	1.84	-27.33
	Proposed	**2.25**	**-34.85**
Laura	HEVC RExt	1.53	-20.40
	DCCM	1.69	-22.80
	HEVC SCC	1.88	-25.36
	Proposed	**2.49**	**-34.45**
Seagull	HEVC RExt	2.07	-31.77
	DCCM	2.24	-35.32
	HEVC SCC	2.56	-39.23
	Proposed	**2.90**	**-46.74**

Table 1 shows the rate distortion gains of four prediction methods over HEVC intra standard. From Table 1, we can see that the performance of the proposed method is clearly superior to the other methods. An average gain of up to 2.41dB has achieved by the proposed coding scheme when compared with HEVC. Compared to HEVC RExt, 0.79dB average RD performance gain is obtained. The average gains of the proposed method over DCCM and HEVC SCC are 0.58dB and 0.42dB. The reason that the proposed method can achieve a high coding efficiency may lie in three aspects. First, interlaced view image is used to represent the 3D holoscopic image, where the interlaced view image is only a data rearrangement of the 3D holoscopic image. Second, spatial correlation of the interlaced view image is fully explored to predict the coding block. Third, a linear prediction method is utilized to further improve the coding efficiency.

Table 2. Coding Time Ratios to HEVC

Image	HEVC RExt	DCCM	HEVC SCC	Proposed
Bike	2.81	3.31	10.35	5.84
Fountain	3.21	4.10	11.35	6.99
Laura	3.97	4.15	29.76	6.05
Seagull	3.04	4.57	26.01	6.38
Average	3.26	4.03	19.37	6.32

Table 2 gives the coding time ratios of HEVC RExt, DCCM, HEVC SCC, and the proposed method to HEVC. From Table 2, we can find that around 6 times coding complexity is needed for the proposed method compared to HEVC. The reason is that the proposed method has to search for the prediction supports and derive the weight vectors. However, less then one third of the coding time is needed for the proposed method compared to HEVC SCC. But the coding complexity of the proposed method is a little higher than HEVC RExt and DCCM. The order of the encoding complexity can be generalized as: HEVC SCC>the proposed coding method>DCCM>HEVC RExt>HEVC.

The detail execution time of the proposed prediction method both on encoder and decoder sides for Laura is illustrated in Table 3. Since decoding of the interlaced view image has to perform the same process as the encoder side, the time ratio to HEVC in decoder side is higher than that in encoder side.

Table 3. Run-Time at the Encoder Side and Decoder Side for Laura

prediction Method	Encoding Time (sec)	Decoding Time (sec)
HEVC	245	3.89
Proposed	1481	52.89

4. CONCLUSION

This paper discusses a coding scheme for 3D holoscopic image by exploring the spatial correlation among the rendered view images. Interlaced view image is used to describe such high spatial correlation. A linear combination of the prediction supports is utilized to obtain a better prediction of the coding block. The experimental results show that the proposed coding scheme can achieve a higher coding efficiency, compared to the HEVC intra standard and several other coding methods in this field.

5. ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China, under Grants 61571285, U1301257, 61422111, and 61301112.

6. REFERENCES

[1] A. Aggoun, E. Tsekleves, M. R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, L. D. Soares, “Immersive 3D Holoscopic Video System,” *IEEE MultiMedia.*, vol. 20, no. 1, pp. 28-37, Jan. 2013.

[2] T. Ebrahimi, JPEG PLENO Abstract and Executive Summary, ISO/IEC JTC 1/SC 29/WG1 N6922, Sydney, Australia, 2015.

[3] M. P. Tehrani, S. Shimizu, G. Lafruit, T. Senoh, T. Fujii, A. Vetro, et al., Use Cases and Requirements on Free-viewpoint Television (FTV), ISO/IEC JTC1/SC29/WG11 MPEG N14104, Geneva, Switzerland, 2013.

[4] C. Conti, P. Nunes, and L. D. Soares, “New HEVC prediction modes for 3D holoscopic video coding,” in *2012 19th IEEE International Conference on Image Processing.*, pp. 1325-1328, Sep. 2012.

[5] C. Conti, L. D. Soare, and P. Nunes, “Influence of self-similarity on 3D holoscopic video coding performance,” in *Proceedings of the 18th Brazilian symposium on Multimedia and the web.*, pp. 131-134, Oct. 2012.

[6] D. Flynn, J. Sole, and T. Suzuki,High Efficiency Video Coding (HEVC) Range Extensions Text Specification: Draft 4, Joint Collaborative Team Video Coding (JCT-VC), document JCTVC-N1005, Apr. 2013.

[7] C. Rosewarne, K. Sharman, M. Naccari, G. Sullivan, HEVC Range Extensions Test Model 6 Encoder Description, JCTVC-P1013, San Jos, USA, 2014.

[8] R. Joshi, J. Xu, R. Cohen, S. Liu, Z. Ma, Y. Ye, Screen Content Coding Test Model 1 (SCM 1), JCTVC-Q1014, Valencia, Spain, 2014.

[9] D. Liu, P. An, R. Ma, L. Shen, “Disparity compensation based 3D holoscopic image coding using HEVC,” in *2015 IEEE China Summit & Int. Conf. Signal and Information Processing (ChinaSIP).*, pp. 201-205, Jul. 2015.

[10] L. F. R. Lucas, C. Conti, P. Nunes, L. D. Soares, N. M. M. Rodrigues, C. L. Pagliari, E. A. B. da Silva, S. M. M. de Faria, “Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC,” in *2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO).*, pp. 11,15,1-5, Sep.2014.

[11] T. Georgiev and A. Lumsdaine, “Focused plenoptic camera and rendering,” *J. Electron. Imag.*, vol. 19, no. 2, p. 021106, Apr. 2010.

[12] T. Georgiev, Jan.2013 [Online]. Available: http://www.tgeorgiev.net/, Website [Online].

[13] F. Bossen, Common HM test conditions and soft-ware reference configurations, Document JCTVC-L1100, 2013.

[14] D. Flynn, K. Sharman, C. Rosewarne, Common test conditions and software reference configurations for HEVC range extensions, Document JCTVC-O1006, 2013.

[15] H. Yu, R. Cohen, K. Rapaka, J. Xu, Common conditions for screen content coding tests, Document JCTVC-Q1015, 2014

[16] Y. Li, M. Sjostrom, R. Olsson and U. Jennehag, “Scalable Coding of Plenoptic Images by Using a Sparse Set and Disparities,” *IEEE Transactions on Image Processing.*, vol. 25, no. 1, pp. 80-91, Jan. 2016.

[17] S. Cherigui, C. Guillemot, D. Thoreau, P. Guillotel and P. Perez, “Hybrid template and block matching algorithm for image intra prediction,” *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*,pp. 781-784, 2012.