EFFICIENT INTRA PREDICTION SCHEME FOR LIGHT FIELD IMAGE COMPRESSION

Yun Li, Mårten Sjöström, Roger Olsson and Ulf Jennehag

Dept. of Information and Communication Systems
Mid Sweden University
SE-851 70 Sundsvall Sweden

ABSTRACT

Interactive photo-realistic graphics can be rendered by using light field datasets. One way of capturing the dataset is by using light field cameras with microlens arrays. The captured images contain repetitive patterns resulted from adjacent microlenses. These images don't resemble the appearance of a natural scene. This dissimilarity leads to problems in light field image compression by using traditional image and video encoders, which are optimized for natural images and video sequences. In this paper, we introduce the full inter-prediction scheme in HEVC into intra-prediction for the compression of light field images. The proposed scheme is capable of performing both uni-directional and bi-directional prediction within an image. The evaluation results show that above 3 dB quality improvements or above 50 percent bit-rate saving can be achieved in terms of BD-PSNR for the proposed scheme compared to the original HEVC intra-prediction for light field images.

Index Terms— Light Field, Compression, Microlens array, HEVC

1. INTRODUCTION

The way of rendering interactive photorealistic graphics by using light field datasets is fairly simple compared to those techniques in demand of scene geometry and surface shading model [1]. The light field images captured by focused plenoptic cameras [2] with microlens arrays exhibit repetitive patterns with spatial and angular information of a scene as shown in Figure 1. Each microlens image resembles a small size frame in a video sequence. The inter-prediction scheme of HEVC [3] is suitable for the prediction of such image patterns. Therefore, the question is how the coding efficiency is by introducing the inter-prediction concept with bi-directional prediction for the intra-prediction of these images.

Light field cameras enable enhanced post-processing capabilities for their captured images, e.g., image refocusing,
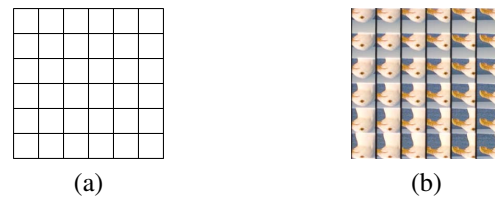
Fig. 1. Light field image captured by using cameras with microlens arrays:(a) microlens array;(b) captured image.

multi-view viewing. The range of applications for light field techniques is enormous, from consumer cameras, computer gaming, 3D video to medical appliances etc. A high quality light field image containing repetitive Elementary Images (EI) captured from microlenses implies a large resolution, which poses the necessity of an efficient coding of the image.

In previous works, vector quantization approach [4] utilized a subset of vectors to represent an entire vector space. The vectors are arranged from the 4D representation of light field images. Discrete Wavelet Transform (DWT) is commonly used in progressive coding [1] [5], which provides the decoder with progressive scalability. Predictive hybrid coding schemes have been applied to light field data set in [6], in which the EI is transformed into Sub Images (SI) and the SIs are encoded by using the multi-view extension of H.264. In addition, Self Similarity (SS) modes were proposed in [7] and [8] for H.264 and HEVC to exploit the repetitive patterns resulted from microlens arrays. The SS modes allow the encoder to encode the light field images efficiently without the requirements of any camera geometrical information, which can be used to perform the EI to SI transformation. However, the SS modes have been proposed in an uni-directional prediction manner, i.e., by using only one reference block.

In this paper, we introduce the entire inter-prediction scheme in HEVC into intra-prediction for the coding of light field images captured by cameras with microlens arrays. This is intended to encode the unprocessed raw images from the camera directly. The proposed scheme can perform inter-microlenses prediction efficiently without knowing camera geometries. The novelties of the papers are: (1) full HEVC inter-prediction schemes are utilized to exploit the inter-microlens redundancy. More specifically, compared to the

prior arts, the proposed scheme is capable of performing bi-directional prediction to further minimize prediction errors; (2) the rate-distortion is investigated and compared for the uni-directional prediction and the bi-directional prediction cases in our proposed scheme. In the context of this paper, we name the proposed scheme with bi-directional prediction as B-coder and with uni-directional prediction as P-coder. The P-coder is similar to the SS modes proposed in [8].

The overall aim of the work is to improve the compression efficiency for the coding of light field images. The work is limited to evaluate the compression on the light field images captured with microlens arrays, more specifically, with focused plenoptic cameras. The goal is to investigate the rate-distortion ratio for the proposed B-coder and the P-coder.

The remainder of the paper is organized as follows. The prediction scheme in HEVC is illustrated in Section 2 and the proposed scheme in Section 3. Test arrangements and evaluation criteria are presented in Section 4. Section 5 shows the results and analysis. We conclude the paper in Section 6.

2. PREDICTION IN HEVC

In HEVC, the Coding Unit (CU) is formed by one luma Coding Block (CB), two chroma CBs and its associated syntax [3]. The largest block is called the Coding Tree Unit (CTU), which is a basic processing unit. CU can be split recursively. In each depth of splitting, Prediction Unit (PU) of different sizes that further separates the CU is the smallest fragmented element, upon which a prediction is performed. The best CU splitting depth, PU fragmentation, and PU prediction modes are selected based on a Rate-Distortion Optimization (RDO) criterion, by using which, the encoder minimizes the distortion of coded images at a bit-rate constraint. In relation to our proposed scheme, the prediction in HEVC is briefly described in the following subsections. The details are referred to the document in [3].

2.1. Intra prediction

A PU can be predicted with 33 directional modes plus a planner and a DC mode. The prediction is based on the up and left boundary samples of the spatially neighboring reconstructed PUs as illustrated in Figure 2. The intra modes are encoded by deriving the most probable mode from the previously coded neighboring PUs.
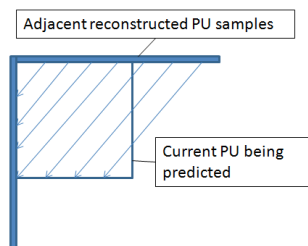

Fig. 2. Intra-prediction: the current PU is predicted from the boundary of its neighboring reconstructed PUs.

2.2. Inter-prediction

In this section, video frames are also referred to as pictures. Pictures are divided into one or more slices. We assume that one frame is one slice for the following description.

Besides using the intra-prediction, a frame can be either uni-directional predicted (P frame) or bi-directional predicted (B frame) as illustrated in Figure 3 (a) and (b). Two reference picture lists, list 0 (L_0) and list 1 (L_1), are utilized to accommodate reference pictures. For the prediction of P frames, all reference pictures are placed in the reference list L_0. For the B frames, past and future reference pictures are placed in L_0 and L_1, respectively. A reference picture is a reconstructed video frame used to predict a current frame. The prediction is performed in a block-wise manner. The best matching block from a reference picture for a current PU is determined by a searching process from the reference pictures within a preset search range. The best is measured in terms of the minimum rate-distortion. For predicting the current PU in P frames, the best matching block from the reference pictures in the list L_0 is chosen as a predictor. For B frames, there are three predictor candidates: the best matching block produced by a weighted sum of two blocks from the reference pictures in each of the lists (L_0 and L_1), the best matching block in L_0, and the best matching block in L_1. The best candidate of the three is chosen as the reference for the current PU.
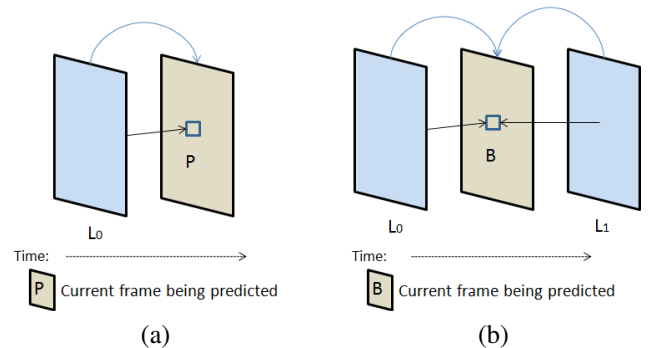

Fig. 3. Inter-prediction: (a) P frame;(b) B frame.

In the Advanced Motion Vector Prediction (AMVP) [3] of HEVC, motion vectors are not encoded directly. In order to reduce the encoding bit-rate, an index for a motion vector candidate and the motion vector difference between the current PU and the candidate are encoded. The motion vector candidates are found for the current PU from its neighboring and temporal collocated PUs. In the HEVC merge mode, motion vectors can be derived from the neighboring and temporal collocated PUs for the current PU.

3. PROPOSED SCHEME

The intra-prediction in HEVC only explores the adjacent encoded PUs for prediction (i.e., up and left boundary parts).

The intra-prediction method is, however, unable to exploit the repetitive patterns characterized by the light field images from plenoptic cameras.

The concept of the proposed coding scheme is presented in Figure 4. Instead of taking the references from adjacent frames for prediction, the proposed scheme takes adjacent microlens images as references, which are the adjacent reconstructed regions in the same image. The adjacent reconstructed regions are determined by the search range parameter in HEVC within the reconstructed CU blocks. An example of the adjacent reconstructed regions are marked with color blue and red in Figure 5 and these regions are neighboring the current PU, block B. For the inter-prediction, HEVC employs two reference lists L_0 and L_1 that store the past and the future reference pictures, as mentioned in the previous section. Therefore, to simulate the prediction of a B frame for our proposed inter-microlenses prediction, the adjacent reconstructed regions are separated into two parts shown in Figure 5 with color blue and red. These two parts are then available for bi-directional prediction and similar to the corresponding reference pictures in the list L_0 and L_1, respectively. We then assume that these two parts are two reference pictures for our proposed scheme, which addresses the inter-microlenses prediction described in detail below.
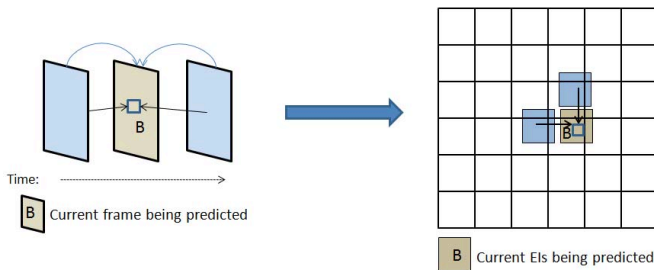

Fig. 4. Introduce inter-prediction into intra-prediction for light field image compression.

B-coder: for the proposed B-coder, there are three predictor candidates for the prediction of a current PU, block B in Figure 5. The first two candidates are the best matching blocks from the two reference pictures in the list L_0 and L_1 (blocks not shown in the figure). The third candidate is produced by $\frac{P_0+P_1}{2}$, where P_0 and P_1 are two blocks from the reference pictures in their respective lists, L_0 and L_1. More specifically, block P_0 is obtained by a refinement search in the neighborhood of the best matching block from L_0, and block P_1 is obtained from L_1 in the same way as block P_0. This refinement search aims at finding the best matching signal $\frac{P_0+P_1}{2}$ to the current PU within the refinement search range (i.e. bi-prediction refinement search range in HEVC). Consequently, the best candidate of the three is selected to predict the PU. Such a prediction scheme forms the proposed B-coder.

P-coder: for the proposed P-coder, a best matching block is searched on the entire adjacent reconstructed region (e.g.

with red and blue in Figure 5). As a result, the best matching block is selected as the reference block to predict the current PU block.

The AMVP is modified accordingly. The motion vector predictor candidates and the merging vector candidates are confined to the spatially neighboring PUs for the current PU.

The B-coder is expected to improve the compression efficiency over the P-coder. The rationale for such an improvement is demonstrated in the multi-hypothesis analysis found in [9]. The analysis shows that the average of multiple reference signals reduces prediction noises, and better prediction is achieved when the number of hypotheses increases. Limited by the number of reference picture lists in the HEVC architecture, we only consider the scenario of using maximum of two hypotheses in this work.

The proposed P-coder and the B-coder use the same syntax as the P slice and the B slice in HEVC. Both coders have inherited the full capability of the inter-prediction scheme in HEVC. However, in the context of light field image coding, we switch the concept of inter-prediction from inter-frame prediction to inter-microlenses prediction, and the PU can cross the boundaries between microlens images.
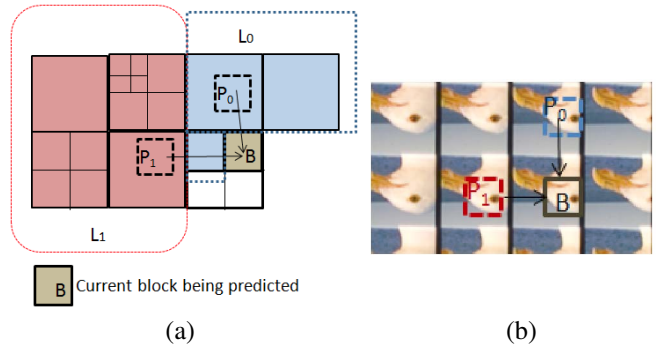

(a)   (b)

Fig. 5. Bi-prediction within an image. (a) Two parts in color blue and red are assumed as two reference pictures and available in the reference list L_0 and L_1; (b) an illustration of the prediction on a light field image.

4. TEST ARRANGEMENT AND EVALUATION CRITERIA

Light field images Seagull and Sergio [10], which are captured by focused plenoptic cameras, are used in the test. These two light field images have a resolution of 7240 by 5236. The resolution of an EI for these images is 75 by 75, and similarities exist between EIs. The Seagull is depicted in Figure 6.

The images were transformed into YUV 4:2:0 format. HEVC Test Model (HM) reference software version 11 [11] was modified for the proposed scheme. The configuration parameters for HEVC were defined as the low delay setting in JCTVC-L1100 [12] with fast search and deblocking filter
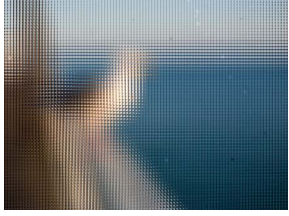
Fig. 6. Light field image Seagull.

enabled. The search range was set to 192, and the search range refinement for bi-prediction set to 4. The tested Quantization Parameters (QP) were selected to be 20, 30, 40, and 50.

The bit-rate, bits per pixel (bpp), was acquired from the bit-stream for all YUV components, whereas the quality is measured on the luminance component by using PSNR. The rate-distortion curves of PSNR versus bpp are plotted for the original HEVC, the P-coder, and the B-coder. The BD-PSNR [13] is also calculated for measuring bit-rate changes for the P-coder and the B-coder over the original HEVC.

5. RESULTS AND ANALYSIS

The results for the image Seagull are plotted in Figure 7. The quality of the decoded image is improved by around 2 dB for the P-coder compared to the original HEVC at the tested bit-rates. Furthermore, above 3 dB improvement is registered for the B-coder over the original HEVC intra for all the tested bit-rates and above 1 dB for the B-coder over the P-coder at the higher bit-rates. Results for the Sergio image in Figure 8 illustrate a similar rate-distortion pattern, in which the decoded light field image quality is significantly higher (above 2dB) for the B-coder than the original HEVC.
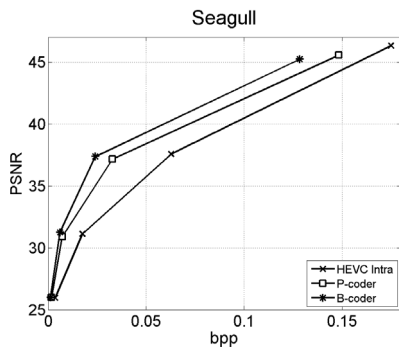

Fig. 7. PSNR over bpp for Seagull.

The bit-rate reduction for the proposed scheme is not convincing at the lowest bit-rate point. This is because of the scaling problem in which the reduction of QP increases the bit-rate nonlinearly. Therefore, the bit-rate changes have been scaled down at the lower bit-rates in the figure due to the larger bit-rate changes at the higher bit-rates. However, for the four bit-rate points, the BD-PSNR shown in Table 1 illustrates
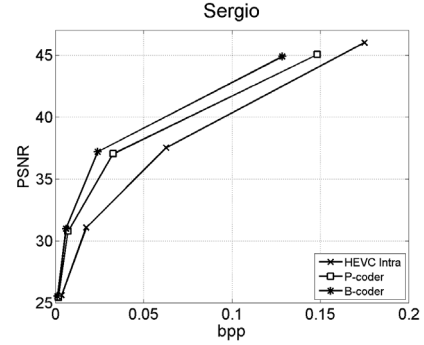

Fig. 8. PSNR over bpp for Sergio.

that the proposed B-coder produces 3.80 dB higher quality than the original HEVC for the decoded image of Seagull, and in terms of bit-rate changes, 56.71 % bit-rate reduction is achieved. Similar decoding quality is acquired for the Sergio image where 3.47 dB increase or 50.60 % bit-rate reduction is obtained.

The improvements for the B-coder over the P-coder can also be seen in Table 1, in which, around 1.16 dB and 0.68 dB quality improvements are achieved for Seagull and Sergio, respectively.

Table 1. BD-PSNR compared to the original HEVC

Images	P-coder		B-coder	
	BD-PSNR (dB)	Bit-rate (%)	BD-PSNR (dB)	Bit-rate (%)
Seagull	+2.67	-42.89	+3.80	-56.71
Sergio	+2.90	-42.95	+3.47	-50.60

5.1. Conclusion

In this paper, we have introduced the full inter-prediction scheme in HEVC into the intra-prediction for coding the light field images captured by cameras with microlens arrays. The proposed scheme is able to exploit the inter-microlenses redundancy without the requirement of the geometrical information from light field images. For the proposed B-coder, the adjacent reconstructed region of a light field image is separated into two parts, which are available for a bi-directional prediction performed within the same image.

The results reveal that light field images are compressed efficiently by using the proposed scheme. The improvement is significant compared to the original HEVC intra. Bit-rate reduction of above 50 percent or PSNR increase of above 3.4 dB is achieved for the tested light field images. The improvement is significant throughout the tested bit-rates.

Future works include the coding efficiency analysis between the proposed scheme and the coding methods by using SI-based pseudo video sequence, comparing the results from our encoders and from the methods with SS modes, and the coding of light field video sequences.

6. REFERENCES

[1] Chuo-Ling Chang, Xiaoqing Zhu, Prashant Ramanathan, and Bernd Girod, "Light field compression using disparity-compensated lifting and shape adaptation.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 15, no. 4, pp. 793–806, Apr. 2006.

[2] Todor Georgiev and Andrew Lumsdaine, "Focused plenoptic camera and rendering," *Journal of Electronic Imaging*, vol. 19, no. 2, pp. 021106, Apr. 2010.

[3] T Wiegand, WJ Han, JR Ohm, and GJ Sullivan, "High efficiency video coding (HEVC) text specification working draft 10," *JCT-VC Document, JCTVC-L1003*, 2013.

[4] M Levoy and Pat Hanrahan, "Light field rendering," *Proceedings of the 23rd annual conference on computer graphics and interactive techniques*, pp. 31–42, 1996.

[5] X Dong, D Qionghan, and X Wenli, "Data compression of light field using wavelet packet," *ICME '04. 2004 IEEE International Conference*, pp. 1071–1074, 2004.

[6] J. Dick, H. Almeida, L. D. Soares, and P. Nunes, "3D Holoscopic video coding using MVC," *2011 IEEE EUROCON - International Conference on Computer as a Tool*, pp. 1–4, Apr. 2011.

[7] Caroline Conti, João Lino, and Paulo Nunes, "Improved Spatial Prediction for 3D Holoscopic Image and Video Coding," *Proc. of the European Signal Processing Conference (EUSIPCO)*, pp. 378–382, 2011.

[8] Caroline Conti, Paulo Nunes, and Luis Ducla Soares, "New HEVC prediction modes for 3D holoscopic video coding," *2012 19th IEEE International Conference on Image Processing*, pp. 1325–1328, Sept. 2012.

[9] Markus Flierl, Thomas Wiegand, and Bernd Girod, "Video codec incorporating block-based multihypothesis motion-compensated prediction," *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, vol. 4067, pp. 238–249, 2000.

[10] "Todor Georgiev website," http://tgeorgiev.net/, retrieved: 08, 2013.

[11] "HM reference software version 11 ," https://hevc.hhi.fraunhofer.de/svn/, retrieved: 08, 2013.

[12] F Bossen, "Common test conditions and software reference configurations," *Joint Collaborative Team on Video Coding (JCT-VC), JCTVC-L1100*, 2011.

[13] Gisle Bjontegard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG-M33*, 2001.