Pseudo sequence based 2-D hierarchical coding structure for light-field image compression

Li Li, *Member, IEEE*, Zhu Li, *Senior member, IEEE*, Bin Li, *Member, IEEE*, Dong Liu, *Member, IEEE*, Houqiang Li, *Senior member, IEEE*

***Abstract*—In this paper, we propose a pseudo sequence based 2-D hierarchical coding structure for light-field image compression. In the proposed scheme, we first decompose the light-field image into multiple views and organize them into a 2-D coding structure according to the spatial coordinates of the corresponding micro-lens. Then we mainly develop three algorithms to optimize the 2-D coding structure. First, we propose a 2-D hierarchical coding structure with a limited number of reference frames to exploit the inter correlations among various views. To be more specific, we divide all the views into four quadrants, and all the views are encoded one quadrant after another to reduce the reference buffer size as much as possible. Inside each quadrant, all the views are encoded hierarchically in both horizontal and vertical directions to fully exploit the correlations among different views. Second, we propose to use the distance between the current view and its reference views instead of the picture order count difference as the criterion for selecting better reference frames for each inter view. The distance based criterion is also applied to the motion vector scaling process to obtain more accurate motion vector predictors. Third, an optimal bit allocation algorithm taking the influence of the various views on the following encoding views into account is proposed to further exploit the inter correlations among various views and improve coding efficiency. The entire scheme is implemented in the reference software of High Efficiency Video Coding. The experimental results demonstrate that the proposed novel pseudo-sequence based 2-D hierarchical structure can achieve maximum 28.4% bit-rate savings compared with the previous pseudo sequence based light-field image compression method.**

***Index Terms*—light-field image, pseudo sequence, bit allocation, hierarchical coding structure, high efficiency video coding**

I. Introduction

THE light-field (LF) image [1], also known as the plenoptic image, contains the information about not only the intensity of light in a scene but also the direction of the light rays in space. The capture of the light or motion is becoming more and more advanced along with the development of the capturing methods [2]. Typically, the LF image is acquired with a plenoptic camera, which places an array of micro-lenses in front of a conventional image sensor. The light beams coming from the object with various angles are firstly refracted through the microlens array. Then they will be captured by the traditional 2-D image sensor to generate the raw sensor data of the LF image. In this way, each micro-lens works as an individual small low resolution image camera conveying a particular perspective of the object in slightly different angles. The raw sensor data will then be converted into the LF image after demosaicing, devignetting, and data structure conversion. As shown in Fig. 1, the LF Toolbox v0.4 [3] can convert the raw sensor data generated by the commercial LF camera Lytro Illum in Fig. 1 (a) into the LF structure as shown in Fig. 1 (b). In Fig. 1 (b), each rectangle represents a 2-D view obtained from the contributions of all micro-lenses of the camera. From Fig. 1 (b), we can see that the LF image is actually with 4-D information including not only the spatial information (similar to 2-D images) but also the angular information (different views).

Since the LF image records the light rays of a scene of interest, it can naturally provide the benefits of rendering new views not only for the changed viewpoint but also for the changed focal point. Recently, especially due to the emergence of the commercial LF cameras, the LF image is becoming a more and more attractive solution to 3-D imaging and sensing. However, the widely use of the LF image is still restricted by its massive size. Since the LF image is with 4-D information, even if the spatial resolution of one view is quite small, the raw data of a LF image with hundreds of views is still very large. For example, the resolution of a raw LF image generated by Lytro Illum [4] containing the captured field of light information is 7728×5368 pixels. Besides the huge image size, since the LF image is generated from a micro-lens array, its characteristic is entirely different from the general 2-D image as shown in Fig. 1 (a), which makes it even harder to compress.

The recent researches for LF image compression can be mainly divided into two kinds both making full use of the classic image/video coding standards, such as JPEG [5], H.264/Advanced Video Coding [6], and H.265/High Efficiency Video Coding (HEVC) [7]. One kind of methods, called as the self-similarity based LF image compression [8], tries to compress the LF image using the commonly used image

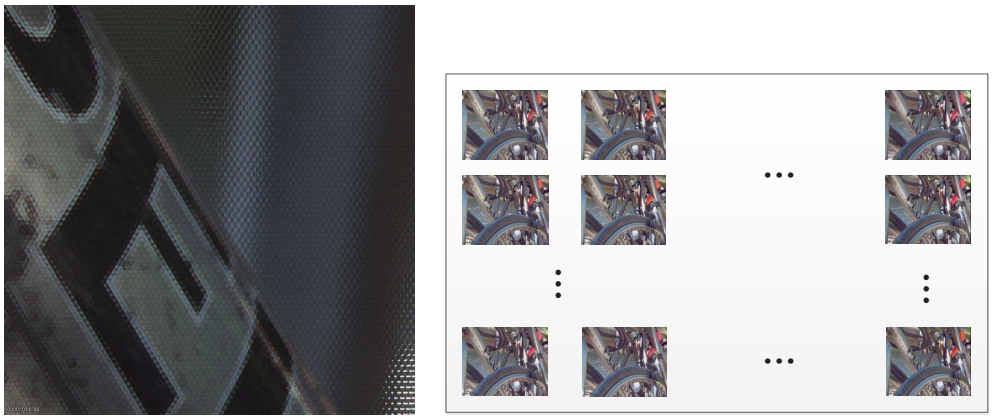
Manuscript received December 28, 2016; revised April 16. 2017 and May 21, 2017; accepted June 26, 2017. This work was recommended by Guest Editor Dr. Yebin Liu. This work was partially supported by UMKC start up grant. This work was also partially supported by Natural Science Foundation of China (NSFC) under Contract 61325009, 973 Program under Contract 2015CB351803.

L. Li and Z. Li are with the University of Missouri-Kansas City, 5100 Rockhill Road, Kansas City, MO 64110, USA. L. Li is also with the CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China. The email addresses are (lil1, lizhu)@umkc.edu.

B. Li is with the Microsoft Asia, No. 5 Danlin Street, Beijing 100010, China. The email address is libin@microsoft.com.

D. Liu and H. Li are with the CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, No. 443 Huangshan Road, Hefei 230027, China. The email addresses are (dongeliu, lihq)@ustc.edu.cn.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2017.2725198, IEEE Journal of Selected Topics in Signal Processing



(a) plenoptic camera raw sensor data after demosaicing and devignetting

(b) light-field data structure

Fig. 1. Example of the raw sensor data and its corresponding LF image data

compression standards. Besides, since the different views of the LF image are quite similar to each other, the self-similarity compensated prediction and estimation (similar to intra block copy [9]) are usually employed to exploit the correlations among various views. However, the self-similarity based LF image compression methods are not flexible enough to fully exploit the correlations among various views.

The other kind of methods, named as the pseudo sequence based LF image compression [10], attempts to decompose the LF image into multiple views and tries to utilize the inter prediction in video coding standards to exploit the temporal correlations among various views. This kind of methods has the potential to achieve a very good coding efficiency since the inter prediction in video coding standards such as HEVC is quite efficient. However, since the LF image is very dense and with multiple views, the current works, which are all based on very coarse reference structures, are unable to make full use of the correlations among various views. How to organize the various views with an efficient coding order and a flexible reference structure remains a very serious problem.

Therefore, in this paper, we follow the approaches of the pseudo sequence based method to decompose the LF image into multiple views. Then all the views are efficiently represented using a 2-D coding structure. Around the 2-D coding structure, this paper has made the following key contributions to fully exploit the high correlations among various views.

- Efficient encoding order. We first divide all the views in the 2-D coding structure into four quadrants, and encode the four quadrants one after another. In this way, except for the views in the border of two quadrants, each quadrant can be considered as independent of each other and thus the reference buffer size can be reduced as much as possible. Besides, inside each quadrant, the views are proposed to be encoded in a hierarchical coding structure in both horizontal and vertical directions to fully exploit the correlations among various views.
- Distance based reference frame selection and motion vector (MV) scaling. For reference frame selection, we propose to use the distance between the current view and its reference views instead of the picture order count (POC) difference as the criterion for selecting better reference views for each inter view. The reference frames having the smaller distances with the current view will be put in the front of the reference frame lists. Besides, for MV scaling, the distance instead of the POC difference is also utilized as the scaling factor to improve the precision of the MV predictors.
- Optimal bit allocation. We try to optimize the bit allocation for the proposed 2-D hierarchical coding structure through a two-pass coding scheme. In the first pass, we collect some statistics on the quality dependencies between different views, and derive the optimal λ ratios of various views in the proposed 2-D coding structure. The λ here is the Lagrange Multiplier, which determines the bitrate and distortion of each view. And in the second pass, the optimal λ ratios will be applied to the encoding process to optimize the coding efficiency.

Part of the work has already been published in [11]. In this paper, we give a more theoretical analysis of the proposed 2-D hierarchical coding framework, and we also provide an optimal bit allocation algorithm to further significantly improve the LF image coding efficiency. The experimental results show that the proposed algorithm can lead to three times of bitrate savings compared with the previous work.

This paper is organized as follows. Section II gives an overview of the related work. Section III will provide a detailed description of the proposed 2-D hierarchical coding structure with a limited number of frames. The distance based reference frame selection and MV scaling will also be described in details in this section. Section IV will introduce the bit allocation algorithm for the proposed 2-D hierarchical coding structure in details. The experimental results on the proposed algorithms will be shown in Section V. Section VI concludes the whole paper.

II. Related work

As mentioned in Section I, the LF image compression methods can be mainly divided into two categories. The first

kind is the self-similarity based methods. The self-similarity here means that various views of the LF image are with very high correlations. The self-similarity based methods can be further divided into two groups. One group of them [12] proposed to use the so-called "image B-coder" by introducing the entire inter prediction scheme in HEVC into the context of intra prediction to make full use of the self-similarity. In the "image B-coder", the neighboring blocks belonging to same Coding Tree Unit (CTU) were set as list0 and the neighboring blocks belonging to the left CTU were set as list1. Both the uni-directional and bi-directional predictions are applied and the inter prediction directions are determined through the rate distortion optimization (RDO) for each block. Moreover, Li et al. [13] further extended this work to response for the 2016 ICME LF image compression challenge through investigating various rate-distortion ratios.

The other group of the self-similarity based methods tries to use the self-similarity compensated prediction and estimation to exploit the high correlations among various views. The self-similarity compensated prediction is very similar to the intra block copy (IBC) [9] in HEVC screen content coding [14]. The main difference is that the IBC uses the hash-based distortion metric in the motion estimation (ME) process while the sum of the absolute difference (SAD) is utilized in the ME process of the self- similarity compensated prediction. Conti et al. [15] first introduced the concept of self-similarity to H.264/AVC to compress the LF image in a more efficient way. This work was then extended to HEVC to take advantage of the flexible partitions [8]. In [8], the self similarity vectors were carefully predicted and encoded according to the characteristics of the LF image. Moreover, in [16], the bi-directional self-similarity compensated prediction and estimation were proposed to further improve the coding efficiency. In [17], the template matching was used to find multiple similar blocks to get a combined prediction block to improve the prediction accuracy.

The second kind of the LF image compression methods is the pseudo sequence based methods. The essence of the pseudo sequence based methods is decomposing the LF image into multiple views and compressing them like a sequence. The pseudo sequence based methods can also be further divided into two groups. The first group is the transform based methods. After decomposing the LF image into multiple views, 3D-Discrete Cosine Transform (3D-DCT) [18] [19] was applied to exploit the spatial redundancy as well as the temporal redundancy. Besides 3D-DCT, 3D-Discrete Wavelet Transform (3D-DWT) [20] [21] was also used to exploit the correlations among various views to better improve coding efficiency. Moreover, Elharar et al. [22] proposed a hybrid compression scheme which first applied a 2D DWT to each micro-image followed by a 2D DCT applied to sets of DWT coefficients from neighboring micro-images.

Along with the development of image and video coding standards, it has been demonstrated that the latest video coding standards with the block-based motion estimation and compensation are more efficient than the transform based methods. Therefore, the other group of the pseudo sequence based methods is based on the inter prediction of the newest video coding standard HEVC. For example, Perra and Assuncao [23] proposed to divide the LF image into multiple tiles and organize all these tiles sequentially into a sequence. Then the correlations among different frames were exploited using the HEVC inter prediction. Besides, Liu *et al.* [10] first proposed to decompose the LF image into multiple views according to the relative positions of the different micro-lenses. Then the multiple views are organized into a sequence according to their spatial position relationship. However, the coding orders, the reference frame management, and the bit allocation among various views are considered in a very coarse way thus the inter dependency among different views has not been fully exploited. Therefore, the rate-distortion (R-D) performance of the LF image compression scheme is far from the best.

Comparing the above mentioned two kinds of methods, Viola et al. [24] reported that the self-similarity based LF image compression methods cannot achieve the comparable performance with the pseudo sequence based methods due to their inflexibility to exploit the correlations among various views, especially in low bitrate case. Therefore, the proposed method in this paper also follows the approaches of the pseudo sequence based methods.

III. The Proposed 2-D Hierarchical Structure

The proposed 2-D hierarchical coding structure will be introduced in two aspects in the following two subsections. In the first subsection, we will introduce the basic concept of the proposed 2-D hierarchical coding structure within a limited number of reference frames. Then in the second subsection, we will describe the distance based reference frame selection and MV scaling algorithms under the 2-D hierarchical coding structure.

A. Basic concept of the proposed 2-D hierarchical structure

The LF image is first decomposed into multiple views following the approaches in [10]. Then some corner views, which are not beneficial for the overall LF image quality, are deleted from the original structure. After these steps, for the test LF images generated by Lytro as shown in [25], we can derive 165 views in total and organize all the views into a 2-D coding structure as shown in Fig. 2. It should be noted that the LF image generated by Lytro with the resolution of 7728×5368 is only used as an example to introduce the proposed 2-D hierarchical coding structures. The proposed algorithm can be easily extended to other kinds of LF images with various resolutions and a different number of views.

Then we try to find the view that has the highest correlation with the other views and code it as an intra frame. Here we define the correlation between the current view and all the other views as the average frame difference per pixel between the current view and the other views. Table I shows the top three smallest average differences per pixel between the current view and the other views for various LF images. From the table, we can see that the average frame difference per pixel of the center view is the top one smallest for most test images and it is always within the top three smallest. Therefore, in our scheme, the center view is assigned with

TABLE I
THE TOP THREE SMALLEST AVERAGE DIFFERENCES BETWEEN THE CURRENT VIEW AND THE OTHER VIEWS FOR VARIOUS SEQUENCES

Test images	Top 1		Top 2		Top 3	
	POC	Diff	POC	Diff	POC	Diff
I01	70	9.27	0	9.27	83	9.28
I02	70	11.40	0	11.42	82	11.49
I03	0	10.92	70	10.95	83	10.95
I04	70	7.64	71	7.67	0	7.68
I05	0	8.92	69	8.92	82	8.93
I06	0	4.33	95	4.34	83	4.35
I07	83	6.29	0	6.31	71	6.33
I08	0	4.38	83	4.39	95	4.41
I09	0	12.73	83	12.75	95	12.77
I10	0	5.12	70	5.13	83	5.19
I11	0	9.64	70	9.67	83	9.69
I12	0	10.54	70	10.58	83	10.62

POC 0 and coded as an intra frame. The other views are assigned the POC sequentially from top left to bottom right as shown in Fig. 2 and coded as inter B frames. Note that the POC here is just a symbol to represent each view instead of the display order in usual videos.

After the POC is assigned to each view, then comes the determination of the coding orders of all the views and the reference relationships among various views. Here, we have not used the RDO theory to optimize the reference frame management. There are mainly two reasons for us to make such a choice. First, it has been shown in [30] that the RDO based reference frame management algorithm can only achieve negligibly better R-D performance compared with the 1-D hierarchical coding structure for the natural test sequences. Second, with the RDO based reference frame management, the encoder complexity will increase by several times under the 1-D hierarchical coding structure. Therefore, in this paper, we extend the 1-D hierarchical coding structure to construct a 2-D hierarchical coding structure to make full use of the inter correlations among different views instead of using the RDO based reference frame management. When designing the 2-D hierarchical coding structure, we try to simultaneously achieve a quite good coding efficiency and use a relatively small reference buffer.

As we know, in the 1-D hierarchical coding structure, the depth first coding order can lead to the least reference buffer size [26]. For example, as shown in Fig. 3, the encoding order of a typical 1-D hierarchical coding structure with group of pictures (GOP) size 16 is 0, 16, 8, 4, 2, 1, 3, 6, 5, 7, 12, 10, 9, 11, 14, 13, 15. And the minimum size of the reference buffer is 5. The situation in the 2-D hierarchical coding structure is quite similar. To reduce the reference buffer size as much as possible, we first divide all the frames into four quadrants as shown in Fig. 2. Then each quadrant is coded in clockwise order from the top left quadrant, the top right quadrant, the bottom right quadrant, to the bottom left quadrant. In this way, except for the frames in the border of two quadrants, each quadrant can be considered as an independent one. Therefore, the reference frames belonging to only one quadrant can be pop out of the reference frame buffer as soon as possible to keep a relatively small reference buffer without influencing the coding efficiency.

Inside each quadrant, the depth first coding order will be used for both the horizontal and vertical directions to make full use of the correlations among various views. Take the top left quadrant as an example. The detailed encoding order can be seen from Fig. 4. In Fig. 4, the number inside each rectangle means the encoding order of each view. In both horizontal and vertical directions, we will follow the encoding order of 0, 6, 3, 5, 4, 2, 1. To be more specific, we will first encode the $0th$ row and the $0th$ column. Then according to the hierarchical coding structure in the vertical direction, the $6th$ row will be coded and followed by the $3rd$ row. Finally, we will encode the $5th$, $4th$, $2nd$, and $1st$ row. Inside each row or column, the order of 0, 6, 3, 5, 4, 2, 1 will also be used. It should also be noted that to guarantee the smallest reference buffer size, the coding order of the other three quadrants will be symmetrical to that of the top left quadrant. To be more specific, the encoding order for each row of the top right and bottom right quadrants will be from right to left, and the encoding order for various rows of the bottom right and bottom left quadrants will be from bottom to top. Take the row with frames 19, 20, 21, 22, 23, 24 as an example, the encoding order of the row will be from right to left as 24, 21, 23, 22, 20, 19. In this way, the frame 24 will only need to be stored in the reference buffer for the frame 23 and thus such a scheme can keep the reference frame buffer as small as possible.

After the encoding order is determined, then we will select the reference frames for each frame. As shown in Fig. 2, all the views are divided into four groups according to their frequencies to be referenced. The frequency means the times a reference frame can be referenced by other frames.

- The frames with the red block. This kind of frames is the most frequently referenced frames. They are always stored in the reference buffer until the end of encoding of the current quadrant. All the frames including the to-be-encoded red block frames in the current quadrant can take the red block frames as reference. In the current quadrant, the existence of these frames can guarantee that all the frames have a good prediction.
- The frames with the green block. This kind of frames is the second most frequently referenced frames. They will be referenced by the frames belonging to the current row in the same quadrant. For example, besides the red frames, the frame 26 can also take the frames 25 and 28 as references.
- The frames with the yellow block. This kind of frames will only be referenced by the frame encoded immediately after them in the same quadrant. For example, the frame 27 can take the frame 26 as reference.
- The frames with the black block. This kind of frames such as the frame 27 belongs to the non-reference frames.

It should be noted that since we are using a row-based coding order, to save the reference buffer size, the vertical references are much less than the horizontal references for most frames. For example, when encoding the frame 29, the immediate above frame 16 is unavailable under the proposed 2-D hierarchical coding structure.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2017.2725198, IEEE Journal of Selected Topics in Signal Processing

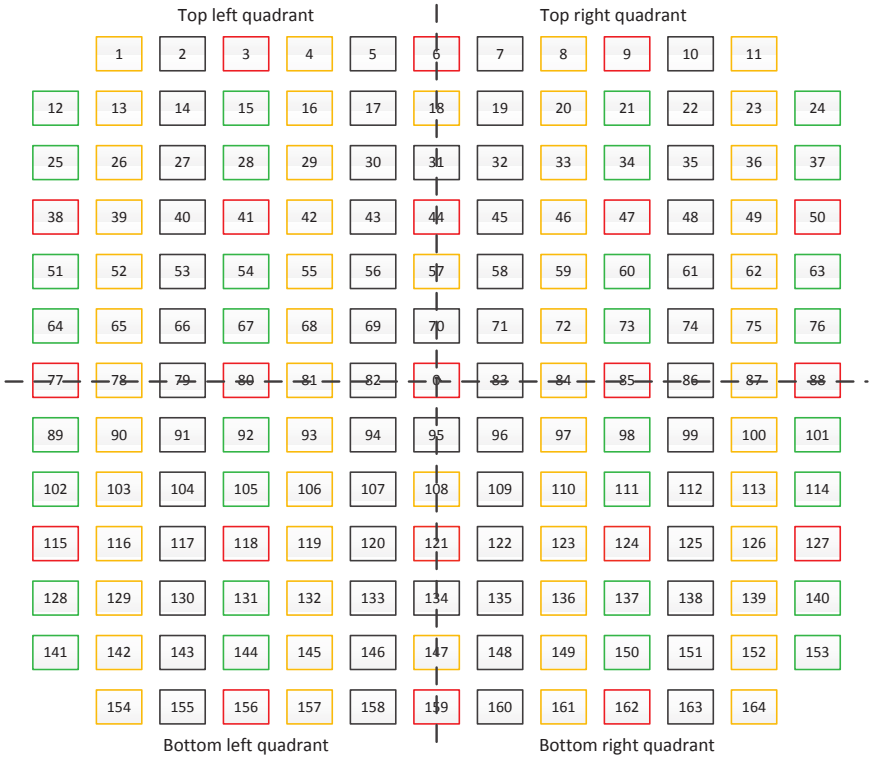


Fig. 2. This figure depicts the 13×13 views (excluding the 4 corner views) to be compressed. The views are assigned picture order count 0-164 shown in the figure. The views are divided into four quadrants.

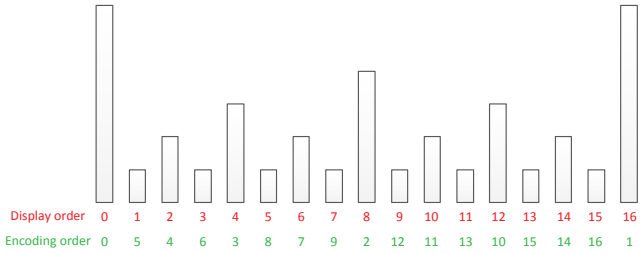


Fig. 3. The 1-D hierarchical coding structure with GOP size 16

Then we will analysis the reference buffer size needed to achieve such a reference frame structure. In the top left quadrant, according to our careful analysis, we find that the maximum number of reference frames appears when encoding the non-reference frames such as the frame 27. We should not only store the eight frequently used red reference frames in the reference buffer, but also the frames 26 and 28. Therefore, maximum 10 reference frames are needed for the top left quadrant. For the top right and bottom right quadrants, the situation is more complicated. As HEVC provides the constraint that the reference frames of the next frame can only be chosen from the reference frames of the current frame and the current frame itself [27], the frames 77 and 80 should always be stored in the reference buffer because they will be used as the reference frames for the bottom left quadrant. Therefore, maximum 12 reference frames are needed for the top right and bottom right quadrants. The situation of the bottom left quadrant is the same as that of the top left quadrant, for which maximum

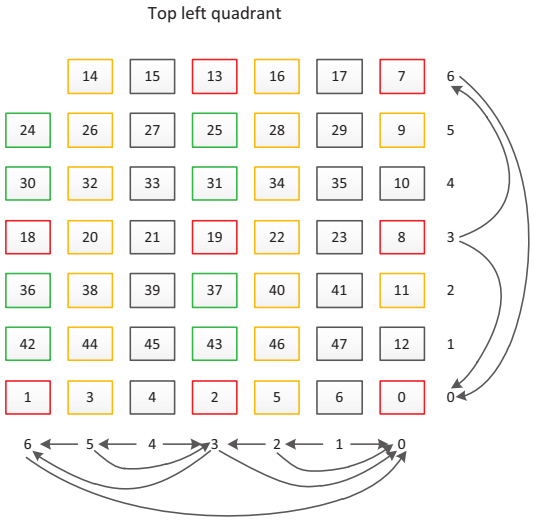


Fig. 4. The coding order of the top left quadrant

10 reference frames are needed. Therefore, in summary, the reference buffer size needed for encoding the entire pseudo sequence is 12.

B. Distance based reference frame selection and MV scaling

According to the analysis in the last subsection, we know that the maximum number of reference frames is 12. If we apply all these reference frames for both list0 and list1, the encoder will traverse all the reference frames to obtain the best

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2017.2725198, IEEE Journal of Selected Topics in Signal Processing

prediction block, which may increase the encoding complexity dramatically. Besides, the large reference index introduced by a large number of reference frames may also increase the overhead bits. Therefore, in this section, we first propose a distance based reference frame selection algorithm to reduce the overhead bits and decrease the encoding complexity. Then since the spatial positions of various views may also have influences on the MV scaling process in the merge [28] and advanced motion vector prediction modes, we also propose a spatial-coordinates-based MV scaling to further improve the coding efficiency.

In the 1-D hierarchical coding structure, the frames with the smaller POC differences are put in a relatively earlier position of the reference lists since they are nearer to the current frame and have larger possibilities to be referenced. However, in the proposed 2-D hierarchical coding structure, the POC is just a symbol to represent each view so the POC difference cannot reflect the distance between two frames. For example, as shown in Fig. 2, the POC difference between the frames 18 and 6 is 12, which is larger than the POC difference between the frames 18 and 15. However, the distance between the frames 18 and 6 is much smaller than that between the frames 18 and 15. Therefore, we need to first obtain the distances among various views before selecting the suitable reference frames instead of using the POC differences. In our implementation, we establish a coordinate system to derive the spatial coordinates of all the views and then use the spatial coordinates to calculate the distances among various views. The spatial coordinate of the most top left position of Fig. 2 is set as $(0, 0)$, and the right and down directions are set as positive. For example, the spatial coordinates of the frame 0 and the frame 1 is $(6, 6)$ and $(1, 0)$, respectively. Then we can derive the correspondence between the POC and the spatial coordinate of each frame (x, y) as follows.

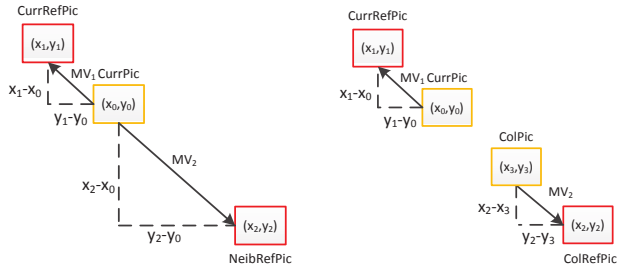
$$x = \begin{cases} 6 & if\ POC = 0 \\ POC\%13 & else\ if\ POC \leq 11 \\ (POC+1)\%13 & else\ if\ POC \leq 82 \\ (POC+2)\%13 & else\ if\ POC \leq 153 \\ (POC+3)\%13 & otherwise \end{cases} \quad (1)$$

$$y = \begin{cases} 6 & if\ POC = 0 \\ POC/13 & else\ if\ POC \leq 11 \\ (POC+1)/13 & else\ if\ POC \leq 82 \\ (POC+2)/13 & else\ if\ POC \leq 153 \\ (POC+3)/13 & otherwise \end{cases} \quad (2)$$

After the spatial coordinates are determined, we can easily calculate the distance between the frames (x_1, y_1) and (x_2, y_2) through the euclidean distance.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (3)$$

Then we will construct both the list0 and list1 according to the distances between the current frame and its reference frames. In the 1-D hierarchical coding structure, the forward (smaller POC compared with the current frame) and backward (larger POC compared with the current frame) reference frames are put into list0 and list1, respectively. Similarly, under the 2-D hierarchical coding structure, we should first define



(a) spatial case (b) temporal case

Fig. 5. Distance based MV scaling

the forward and backward directions. In our implementation, the above frames are all treated as the forward frames, and the below frames are all treated as the backward frames. For example, for the frame 17, the frames 1 to 16 are considered as the forward frames, and all the other frames including the frame 0 are considered as the backward frames. According to the above introduction, the totally available reference frames for the frame 17 are frames 16, 15, 6, 3, 38, 41, 44, 77, 80, and 0. If we set the number of reference frames in both lists as 4, according to the distances between the current frame and its reference frames in increasing order, the reference frames in list0 will be 16, 6, 15, and 3, and the reference frames in list1 will be 44, 41, 0, and 80.

Besides the reference frame selection, the spatial coordinates may also have significant influences on the MV scaling operations in both merge and advanced motion vector prediction modes. The MV scaling operations are performed when the spatial neighboring blocks or temporal co-located blocks are pointing to a different reference frame from the current block. The MV scaling can be divided into two kinds: the spatial and temporal MV scaling. In 2-D hierarchical coding structure, we should perform MV scaling based on the distance in x and y directions separately instead of POC.

The detailed processes are presented in Fig. 5. For the spatial case, the spatial coordinate of the current frame is (x_0, y_0), the spatial coordinate of the current reference frame is (x_1, y_1), and the spatial coordinate of the reference frame of the neighboring block is (x_2, y_2). The MV of the current block is $(MV_{1,x}, MV_{1,y})$, and the MV of the neighboring block is $(MV_{2,x}, MV_{2,y})$. Assuming that the motions among various frames are uniform, the spatial MV scaling process can be derived as follows.

$$MV_{1,x} = \frac{MV_{2,x}}{x_2 - x_0} \times (x_1 - x_0) \quad (4)$$

$$MV_{1,y} = \frac{MV_{2,y}}{y_2 - y_0} \times (y_1 - y_0) \quad (5)$$

It should be noted that the values of $x_2 - x_0$ and $y_2 - y_0$ are zero when the current frame and the neighboring reference frame are in the same row or column. In this case, the spatial scaling operations are not applied.

For the temporal case, except for the current frame and its reference frame, there are temporal co-located frame and its corresponding reference frame whose spatial coordinates

are (x_3, y_3) and (x_2, y_2), respectively. The MV of the current block is $(MV_{1,x}, MV_{1,y})$, and the MV of the co-located block is $(MV_{2,x}, MV_{2,y})$. Assuming that the motions among various frames are uniform, the temporal MV scaling process can be derived as follows.

$$MV_{1,x} = \frac{MV_{2,x}}{x_2 - x_3} \times (x_1 - x_0) \tag{6}$$

$$MV_{1,y} = \frac{MV_{2,y}}{y_2 - y_3} \times (y_1 - y_0) \tag{7}$$

Similar to the spatial MV scaling process, the values of $x_2 - x_3$ and $y_2 - y_3$ are zero when the co-located frame and the co-located reference frame are in the same row or column. In this case, the temporal scaling operations are not applied.

IV. The Proposed Optimal Bit Allocation Schemes

In the proposed 2-D hierarchical coding structure, the LF image is decomposed into multiple views, and the multiple views are organized into a pseudo sequence to better exploit the correlations among multiple views. Therefore, minimizing the distortion of the LF image under the bitrate constraint becomes equivalent to minimizing the distortion of the decomposed pseudo sequence with the same bitrate constraint. As mentioned in the previous section, all the frames are divided into four groups according to their frequencies to be referenced. In Fig. 2, the frames with the red rectangle are referenced the most while the frames with the black rectangle are non-reference frames. The quality of the frames which are referenced the most has the largest influence on the quality of the whole sequence and should be assigned the largest number of bits, while the quality of the non-reference frames has the smallest influence on the quality of the whole sequence and should be assigned the least number of bits. In this section, we will investigate the influence of various frames on the whole sequence, and then propose an optimal bit allocation algorithm to assign each frame with the suitable number of bits.

In the recent researches, Li et al. [29] [30] proposed that the Lagrange Multiplier λ instead of the quantization parameter (QP) was the key factor to determine the bits and distortion for each frame. In other words, to assign a suitable number of bits to each frame is equivalent to determining a suitable λ_i for each frame. Therefore, the optimization target of the proposed bit allocation algorithm is to determine a suitable λ_i for each frame to optimize the overall quality of all the frames under the target bits constraint,

$$\min_{\lambda_1,\lambda_2,\ldots,\lambda_N} \sum_{i=1}^{N} D_i, s.t. \sum_{i=1}^{N} R_i \le R_t \tag{8}$$

where the λ_i is the λ for the frame i, which determines the optimization target of the frame i. The D_i and R_i are the distortion and bits for the frame i, respectively. N is the total number of frames in the pseudo sequence. Note that, different from the traditional approach [31] to optimize the bit allocation, the proposed method enrolls λ_i for each picture as the optimization parameter. The constrained problem can be converted to the following unconstrained problem,

$$\min_{\lambda_1,\lambda_2,\ldots,\lambda_N} \sum_{i=1}^{N} D_i + \lambda \sum_{i=1}^{N} R_i \tag{9}$$

where the λ is the Lagrange Multiplier of the optimization problem, and determines the optimization target of the whole pseudo sequence. In the following, we will demonstrate that the λ is equal to the λ_i of the non-reference frames.

To solve (9), the common used Lagrange method is applied. Then we can derive the following equation by setting the derivative of the total cost to 0,

$$\frac{\partial \sum_{i=1}^{N} D_i}{\partial \lambda_j} + \lambda \frac{\partial \sum_{i=1}^{N} R_i}{\partial \lambda_j} = 0, \quad j = 1, 2, \ldots N \tag{10}$$

Since the λ_j is related to the distortion D_j, (10) can be converted to the following formula,

$$\frac{\partial \sum_{i=1}^{N} D_i}{\partial D_j} \cdot \frac{\partial D_j}{\partial \lambda_j} + \lambda \frac{\partial \sum_{i=1}^{N} R_i}{\partial D_j} \cdot \frac{\partial D_j}{\partial \lambda_j} = 0, \quad j = 1, 2, \ldots N \tag{11}$$

As the quality of the current frame will only have influence on itself and its subsequent frames, equation (11) can be rewritten in another form by ignoring the term $\frac{\partial D_j}{\partial \lambda_j}$,

$$\frac{\partial \sum_{i=j}^{N} D_i}{\partial D_j} + \lambda \frac{\partial \sum_{i=j}^{N} R_i}{\partial D_j} = 0, \quad j = 1, 2, \ldots N \tag{12}$$

Then both $\frac{\partial \sum_{i=j}^{N} D_i}{\partial D_j}$ and $\frac{\partial \sum_{i=j}^{N} R_i}{\partial D_j}$ can be written as the sum of the influence to itself and its subsequent frames,

$$\frac{\partial \sum_{i=j}^{N} D_i}{\partial D_j} = 1 + \frac{\partial \sum_{i=j+1}^{N} D_i}{\partial D_j} \tag{13}$$

$$\frac{\partial \sum_{i=j}^{N} R_i}{\partial D_j} = \frac{\partial R_j}{\partial D_j} + \frac{\partial \sum_{i=j+1}^{N} R_i}{\partial D_j} \tag{14}$$

Since the λ_j is the slope of the R-D curve for each frame, we can obtain the following formula,

$$\frac{\partial R_j}{\partial D_j} = \frac{1}{\frac{\partial D_j}{\partial R_j}} = -\frac{1}{\lambda_j} \tag{15}$$

By substituting (15), (13), and (14) into (12), we can obtain the equation as follows,

$$(1 + \frac{\partial \sum_{i=j+1}^{N} D_i}{\partial D_j}) + \lambda(-\frac{1}{\lambda_j} + \frac{\partial \sum_{i=j+1}^{N} R_i}{\partial D_j}) = 0 \tag{16}$$

Here we define the R-D cost of each frame J_i as

$$J_i \triangleq D_i + \lambda R_i, \quad i = 1, 2, \ldots N \tag{17}$$

It should be emphasized that for each frame, to calculate the R-D cost, the Lagrange Multiplier used is the sequence level λ

instead of the λ_i for each specified frame. Then we can solve (16) as

$$\lambda_j = \frac{\lambda}{(1+\frac{\partial \sum_{i=j+1}^{N} J_i}{\partial D_j})} \triangleq \frac{\lambda}{1+\omega_j} \tag{18}$$

To make the physical interpretation of (18) more obvious, we rewrite (18) as

$$\frac{\lambda_i}{\lambda_j} = \frac{1+\omega_j}{1+\omega_i} \tag{19}$$

Equation (19) means that the λ of different frames should be inversely proportional to its influence to the R-D cost of the whole sequence to achieve the optimal R-D performance. The larger the influence of the current frame on the whole sequence, the smaller the λ of the current frame will be. The small λ will result in a small distortion and thus a large number of bits assigned to the current frame. As we all know, a large number of bits assigned to the frame having a far-reaching influence on the whole sequence is beneficial the overall performance.

Then the only problem becomes how to determine the term w_j in (18). We know that the non-reference frames have no influence to the subsequent frames since they will not be referenced by any frames. Therefore, for the non-reference frames, the value of w_j will be equal to 0. From (18), we can easily see that sequence level λ is equal to the λ_j of the non-reference frames [32]. This work explicitly demonstrates that the sequence level λ is equal to the λ_j of the non-reference frames.

For the reference frames, we use a two-pass scheme similar to the methods in [31] to optimize the bit allocation problem. In the first encoding pass, we use the simple λ ratio between various frames in [10] as the initial bits ratio to generate the anchor R-D cost J_{j0} and the distortion D_{j0}. Then we change the QP and λ of the frame j to generate the test R-D cost J_{j1} and distortion D_{j1}. Both the J_{j0} and J_{j1} only include the R-D cost of the frames coded after the frame j since the frames coded before the frame j will not change. It should be noted that the λ used to calculate the R-D cost is the sequence level λ according to (17). Besides, the QP of the frame j is decreased by 5 to guarantee that the frame j will be with an obvious distortion change ΔD_j [31]. The number 5 is chosen according to our experience. It is neither too small to make a difference on the R-D cost of the entire pseudo sequence, nor too large to be not within a normal range. In this way, we can easily obtain the value of delta R-D cost ΔJ_j after changing the distortion of the test frame with ΔD_j. Then we can obtain an approximate value for ω_j.

$$\omega_j \approx \frac{\Delta J_j}{\Delta D_j} = \frac{J_{j1} - J_{j0}}{D_{j0} - D_{j0}} \tag{20}$$

In the second encoding pass, we can use the ω_j obtained in the first pass to calculate the λ_j of each frame according to (18). The QP is then determined through the following equation according to experience [33],

$$QP = 4.3281 \times \log(\lambda) + 14.4329 \tag{21}$$

After the determination of the λ and QP for each frame, we can finish the entire encoding process through RDO.

V. Experimental results

A. Simulation setup

The proposed pseudo sequence based 2-D hierarchical coding structure for LF image compression is implemented in HM-16.7 [34]. The previous pseudo sequence based LF image compression method proposed in [10] is used as the anchor to demonstrate the effectiveness of the proposed algorithm. Since the bitrates generated by various LF image compression algorithms are different, the Bjontegaard-Delta rate (BD-rate) [35] is used to measure the performance for a fair comparison. We use both the Y-PSNR and YUV-PSNR between the original LF image and the reconstructed LF image shown in [4] as an objective quality metric. The test images used are also specified in [4].

Since an optimal bit allocation algorithm is proposed for the 2-D hierarchical coding structure, we use the following method to guarantee that the optimization targets of the proposed algorithm and the anchor are the same. We first generate the anchor using the LF image compression algorithm in [10] with intra frame QPs 15, 20, 25, and 30. Then the λ of the non-reference frames is recorded and used as the λ of the non-reference frames for the proposed algorithm. The λs of the reference frames are then calculated using (18).

The performance of the proposed algorithm will be introduced in the following subsections. We will first present some experimental results on the overall framework. Then we will analyze the benefits from the following three aspects: the proposed distance based reference frame selection, spatial-coordinates-based MV scaling, and the optimal bit allocation algorithms.

B. The performance of the overall framework

The performance of the overall framework is shown in Table II. From Table II, we can see that the proposed algorithm can provide on average 18.3% and 19.0% bitrate savings in Y and YUV components, respectively. For the sequence I06-rawLF, the R-D performance improvements can be as high as 28.4% and 28.7% accordingly. It should also be mentioned that the bitrate savings brought by the proposed algorithm are quite consistent. We can achieve significant bitrate savings for all the test sequences without even one exception. The experimental results obviously demonstrate that the proposed LF image compression framework can well exploit the correlations among various views, and thus can significantly improve the R-D performances for both the Y and YUV components. Besides, note that the performance improvement provided by the overall framework consists of three parts: the distance based reference frame management, the distance based MV scaling, and the optimal bit allocation algorithm. The performance improvement shown in Table II is not provided individually by the proposed 2-D hierarchical coding structure since the optimal bit allocation algorithm is not enabled in the anchor. We will introduce the performances of proposed 2-D hierarchical structure and the optimal bit allocation individually in the next section.

For the encoding complexity, since the proposed 2-D hierarchical coding structure uses less reference frames compared

TABLE II
THE PERFORMANCE OF THE OVERALL FRAMEWORK

test images	Y-BDrate	YUV-BDrate	Enc. Time	Dec. Time
I01	–20.5%	–21.7%	94.3%	102.5%
I02	–16.4%	–18.2%	93.8%	98.1%
I03	–22.7%	–23.8%	93.9%	99.1%
I04	–16.6%	–16.0%	94.7%	100.1%
I05	–19.6%	–21.1%	91.7%	98.0%
I06	–28.4%	–28.7%	92.3%	100.6%
I07	–17.0%	–17.9%	92.5%	97.8%
I08	–12.9%	–12.7%	92.2%	100.6%
I09	–22.0%	–23.1%	94.1%	99.2%
I10	–15.1%	–17.1%	96.1%	103.8%
I11	–6.5%	–6.1%	93.1%	100.1%
I12	–8.0%	–8.7%	93.7%	102.7%
Average	–18.3%	–19.0%	94.0%	101.5%

TABLE III
THE PERFORMANCE OF THE DISTANCE BASED REFERENCE FRAME SELECTION

test images	Y-BDrate	YUV-BDrate	Enc. Time	Dec. Time
I01	–6.4%	–6.5%	94.2%	99.1%
I02	–4.6%	–4.6%	94.6%	97.1%
I03	–4.7%	–4.7%	93.7%	97.2%
I04	–2.8%	–2.1%	93.8%	95.6%
I05	–5.6%	–5.6%	93.4%	96.9%
I06	–13.8%	–13.7%	92.5%	97.5%
I07	–8.2%	–8.3%	93.7%	96.7%
I08	–12.2%	–12.1%	93.6%	99.5%
I09	–1.7%	–1.7%	95.1%	100.2%
I10	–9.2%	–9.1%	94.7%	100.1%
I11	–1.1%	–1.0%	95.4%	101.5%
I12	0.9%	0.7%	95.8%	99.7%
Average	–5.8%	–5.7%	94.3%	99.2%

TABLE IV
THE PERFORMANCE OF THE DISTANCE BASED MV SCALING

test images	Y-BDrate	YUV-BDrate	Enc. Time	Dec. Time
I01	–0.5%	–0.4%	102.1%	100.2%
I02	–0.6%	–0.6%	102.0%	98.9%
I03	–0.6%	–0.5%	100.6%	98.7%
I04	0.0%	0.0%	100.9%	97.5%
I05	–0.8%	–0.7%	100.7%	97.8%
I06	–0.5%	–0.5%	100.0%	94.7%
I07	–0.8%	–0.8%	100.7%	98.1%
I08	–0.1%	0.0%	102.1%	98.6%
I09	–0.3%	–0.3%	102.0%	99.5%
I10	–0.1%	–0.2%	101.0%	97.9%
I11	–1.9%	–1.9%	101.8%	99.7%
I12	–0.9%	–0.9%	101.8%	100.0%
Average	–0.6%	–0.6%	101.2%	98.3%

with the anchor for some frames, the proposed algorithm presents less encoding time compared with the anchor. Note that the complexity of the first coding pass to calculate the optimal bit allocation ratios among various views is excluded since we mainly try to show the encoding time of the proposed 2-D hierarchical coding structure. The complexity of the first coding pass can be quite high since we need to encode the sequence multiple times to obtain the optimal bit allocation ratios. For the decoding complexity, since almost no changes are applied to the decoder, the proposed algorithm shows similar decoding time compared with the anchor.

C. The performance and analysis of the separate three parts

1) The proposed distance based reference frame selection: The performance of the proposed distance based reference frame selection combined with the 2-D hierarchical coding structure is illustrated in Table III. From Table III, we can see that an average of 5.8% and 5.7% bitrate savings can be achieved through the proposed distance based reference frame selection. The proposed algorithm can achieve about $1/3$ of the total performance improvement. The performance improvement mainly comes from the fact that the distance based reference frame selection method can always guarantee a reference frame with relatively small distance with the current frame to be used as a real reference frame. Besides, compared with the previous pseudo sequence based LF image compression method, the proposed reference frame management algorithm can not only improve the R-D performance, but also save the reference frames to be stored in the reference frame buffer. The size of the reference buffer is only 12 through the effective management of the proposed algorithm. However, it can be as large as 51 under the algorithm in [10]. In summary, the proposed algorithm can achieve a better use of the reference frame buffer within a limited number of reference frames.

2) The proposed distance based MV scaling: The performance of the proposed distance based MV scaling individually is illustrated in Table IV. From Table IV, we can see that an average of 0.6% and 0.6% bitrate savings on Y and YUV components can be achieved through the proposed distance based MV scaling. Since a LF image is with very dense views, most MVs between different views are very small. That is the reason why the performance improvement brought by the proposed distance based MV scaling method is limited. The performance of the combined distance-based reference selection and MV scaling algorithms are illustrated in Table V. From Table V, we can see that an average of 6.5% and 6.5% R-D performance improvement can be obtained by combining these two algorithms. The experimental results obviously demonstrate that the combination of these two algorithms can lead to even better R-D performance.

3) The proposed optimal bit allocation algorithm: The performance of the proposed optimal bit allocation algorithm on the coding structure of the anchor is shown in Table VI. From Table VI, we can see that the proposed bit allocation algorithm can achieve an average of 13.2% and 14.0% on the coding structure of the anchor. Also, through the comparison of Table V and Table II, we can see that the proposed bit allocation algorithm can achieve also over 10% coding gains under the proposed 2-D hierarchical coding structure. The experimental results obviously demonstrate that the optimal bit allocation algorithm can achieve very significant bitrate savings under both coding structures.

Some typical examples of the QPs of various views are shown in Fig. 6. First, we can see from Fig. 6 that the QP of the first encoded view is much lower than the other views. Under the proposed 2-D reference structure, the first encoded view can be referenced by all the views in all four quadrants. Therefore, the coding quality of the first view has a far-

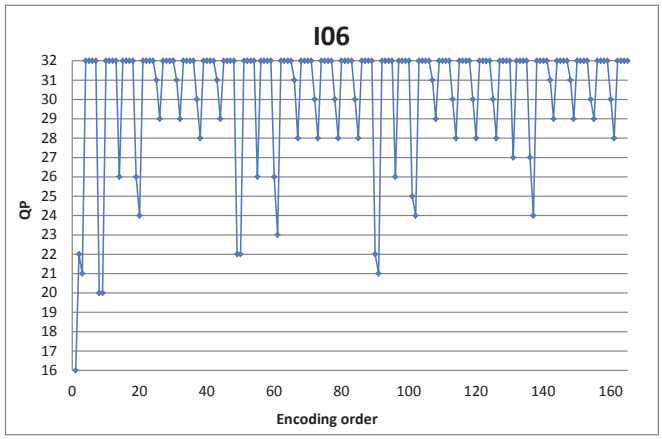
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2017.2725198, IEEE Journal of Selected Topics in Signal Processing

TABLE V
THE PERFORMANCE OF COMBINED DISTANCE BASED REFERENCE FRAME SELECTION AND MV SCALING

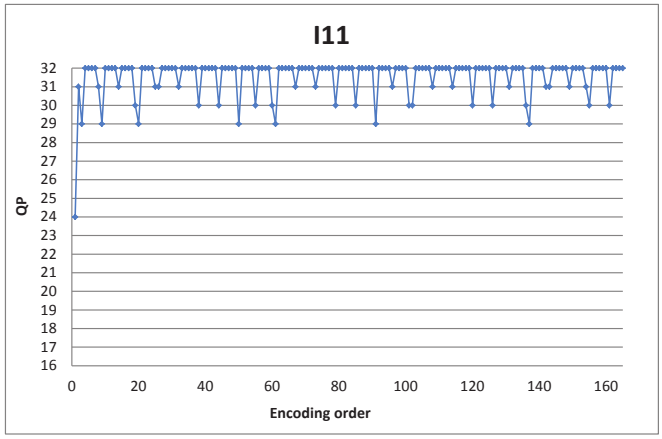
test images	Y-BDrate	YUV-BDrate	Enc. Time	Dec. Time
I01	–7.1%	–7.1%	95.4%	98.4%
I02	–5.4%	–6.1%	95.1%	98.1%
I03	–5.4%	–5.5%	94.8%	99.9%
I04	–3.4%	–2.6%	94.9%	97.7%
I05	–6.5%	–6.4%	94.4%	99.0%
I06	–14.2%	–14.2%	94.0%	98.6%
I07	–8.6%	–8.7%	95.2%	99.5%
I08	–12.9%	–12.7%	94.3%	98.4%
I09	–2.6%	–2.6%	96.0%	100.1%
I10	–9.8%	–9.7%	96.2%	98.7%
I11	–2.0%	–1.9%	96.5%	101.6%
I12	–0.4%	–0.6%	96.6%	99.7%
Average	–6.5%	–6.5%	95.4%	99.8%

TABLE VI
THE PERFORMANCE OF THE OPTIMAL BIT ALLOCATION ALGORITHM

test images	Y-BDrate	YUV-BDrate	Enc. Time	Dec. Time
I01	–13.9%	–15.2%	106.7%	108.8%
I02	–10.9%	–12.4%	106.1%	106.1%
I03	–17.7%	–18.6%	104.5%	106.5%
I04	–12.9%	–12.8%	105.3%	107.4%
I05	–13.7%	–15.1%	101.7%	101.5%
I06	–19.9%	–20.1%	101.5%	98.3%
I07	–10.2%	–10.7%	101.9%	102.9%
I08	–13.1%	–13.7%	101.0%	100.0%
I09	–12.5%	–14.2%	103.5%	102.5%
I10	–18.9%	–18.8%	105.9%	108.3%
I11	–8.6%	–8.9%	100.7%	99.6%
I12	–6.6%	–7.1%	102.6%	103.1%
Average	–13.2%	–14.0%	103.6%	104.1%



(a) I06



(b) I11

Fig. 6. The QPs of various views after bit allocation

reaching influence on the whole test sequence. That is why the first view should be assigned the largest number of bits and coded using a much smaller QP compared with the other views. Second, we can obviously see from Fig. 6 that the optimal bit allocation algorithm can assign reasonable QPs to various views according to the characteristics of various LF images. The characteristic of the LF image I11 is complex and the correlations between among various views are less compared with the LF image I06. Therefore, the differences of QPs of various views in the LF image I11 are much smaller than that in the LF image I06. The LF image characteristic adaptation feature of the proposed bit allocation algorithm is the reason why it can bring so significant R-D performance improvement.

D. Some examples of the R-D curves

Some examples of the R-D curves are shown in Fig. 7. From these R-D curves, we can see that the proposed distance based reference frame selection and MV scaling can bring obvious R-D performance improvements compared with the anchor. It can also be seen from these R-D curves that the distance based reference frame selection and MV scaling can mainly reduce the bitrate instead of increasing the PSNR and the performance improvements mainly come from the low bitrate case. We can also see from these figures that the proposed optimal bit allocation algorithm can improve the R-D performance significantly no matter in the low bitrate or high bitrate cases. The proposed optimal bit allocation algorithm can bring consistent bitrate savings in both low bitrate and high bitrate cases.

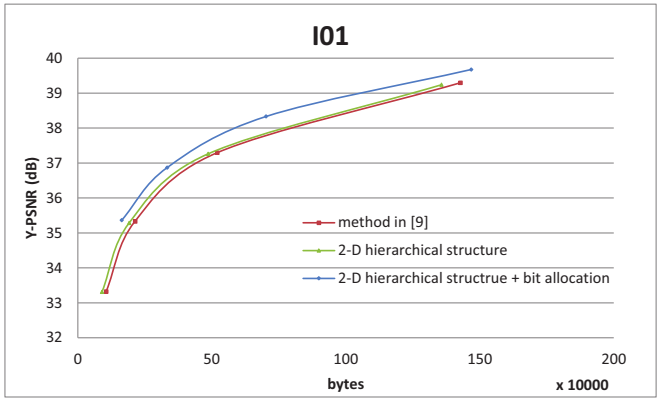
E. The comparison with the JPEG under the fixed compression ratios

Since the bitrates shown in Fig. 7 vary significantly from one image to another, we have also adjusted the QP for each test image to achieve compression ratios 10:1, 20:1, 50:1, and 100:1 to compare with JPEG for a more convenient comparison for the future work. Table VII summarizes the QP setting and the BD-PSNR of the proposed algorithm compared with JPEG. From Table VII, we can see that the proposed 2-D hierarchical coding structure combined with the optimal bit allocation algorithm can achieve an average of $4.641dB$ BD-PSNR improvement compared with JPEG. Also, according to our observation, the BD-PSNR improvement is much more significant in low bitrate case compared with high bitrate case.

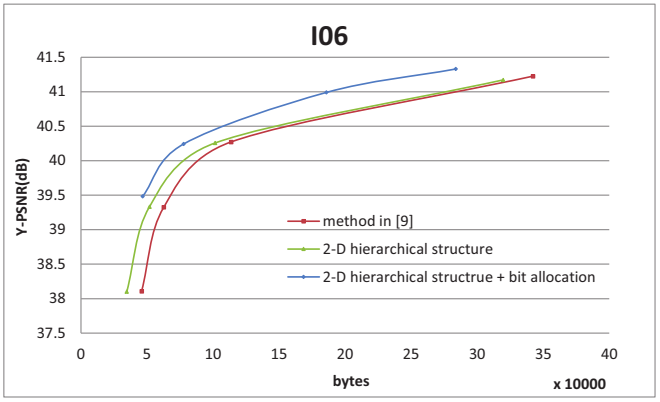
F. Summary of the experimental results

- The proposed 2-D hierarchical coding structure, as well as the optimal bit allocation algorithm, can well exploit the correlations among various views and bring over

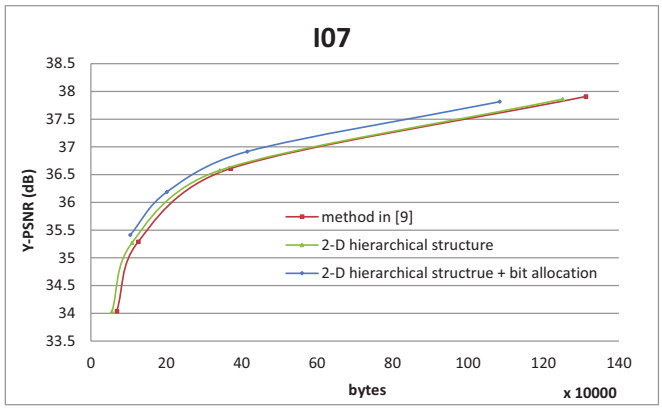
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTSP.2017.2725198, IEEE Journal of Selected Topics in Signal Processing



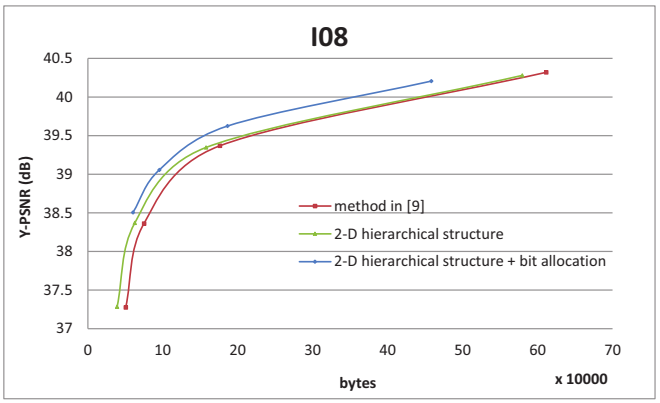
(a) I01



(b) I06



(c) I07



(d) I08

Fig. 7. Some examples of the R-D curves

TABLE VII
THE BASE QP SETTING AND BD-PSNR OF THE PROPOSED ALGORITHM COMPARED WITH JPEG

test images	Base QPs	BD-PSNR (dB)
I01	9, 12, 16, 22	5.091
I02	10, 14, 19, 24	5.524
I03	10, 14, 19, 24	4.658
I04	8, 12, 16, 21	4.305
I05	10, 13, 16, 20	4.542
I06	5, 8, 10, 13	4.039
I07	7, 11, 14, 19	3.024
I08	5, 8, 10, 14	3.722
I09	9, 13, 16, 22	6.581
I10	7, 10, 14, 20	3.473
I11	10, 14, 17, 19	5.467
I12	8, 12, 16, 21	5.262
Average		4.641

18% bitrate savings compared with the previous pseudo sequence based LF image compression method.

- The proposed distance based reference frame selection combined with the 2-D hierarchical coding structure can bring about 6% R-D performance improvements.
- The proposed distance based MV scaling method can bring about 0.6% bitrate savings.
- The proposed optimal bit allocation algorithm for the 2-D hierarchical coding structure can bring over 10% R-D performance improvements for the LF-image compression due to its LF image content adaptive feature.

VI. CONCLUSION

In this paper, we propose a novel pseudo sequence based 2-D hierarchical coding structure for light-field image compression. The proposed 2-D hierarchical coding structure mainly has three key contributions. First, we propose a 2-D hierarchical coding structure with a limited number of reference frames. To be more specific, we divide all the views into four quadrants, and all the views are encoded one quadrant after another to reduce the reference buffer size as much as possible. Inside each quadrant, all the views are encoded hierarchically in both horizontal and vertical directions to fully exploit the correlations between different views. Second, we propose to use the distance between the current view and its reference views instead of the picture order count as the criterion for selecting better reference frames for each inter view. The distance between different views is also applied to the motion vector scaling process to obtain more accurate motion vector predictors. Third, an optimal bit allocation algorithm for the proposed 2-D hierarchical structure is proposed to further exploit the inter correlations among various views and improve coding efficiency. The whole scheme is implemented in the reference software of High Efficiency Video Coding. The experimental results demonstrate that the proposed novel pseudo sequence based 2-D hierarchical structure can achieve

on average 18.3% bit-rate savings compared with the previous pseudo sequence based LF image compression method.

In this work, both the coding order and the reference structure are determined according to our experience. Only the bit allocation is optimized using the rate distortion optimization theory. In the future, we will try to optimize the combinations of the coding order, the reference structure, and the bit allocation, using the fundamental rate distortion optimization theory. Besides, since two-pass encoding is needed to obtain the specific bit allocation for each sequence, we will try to follow the approaches in [36] and [37] to derive a one-pass method to achieve an optimal bit allocation in the future.

REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42.

[2] L. Xu, L. Fang, W. Cheng, K. Guo, G. Zhou, Q. Dai, and Y. Liu, "FlyCap: Markerless Motion Capture Using Multiple Autonomous Flying Cameras," *ArXiv e-prints*, Oct. 2016.

[3] Light Field Tool box 0.4. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4

[4] M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira, and P. Schelkens, "ICME 2016 Grand Challenges: Light-Field Image Compression," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016.

[5] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, Feb 1992.

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[7] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[8] C. Conti, L. D. Soares, and P. Nunes, "HEVC-based 3D holoscopic video coding using self-similarity compensated prediction," *Signal Processing: Image Communication*, vol. 42, pp. 59–78, March 2016.

[9] X. Xu, S. Liu, T. D. Chuang, Y. W. Huang, S. M. Lei, K. Rapaka, C. Pang, V. Seregin, Y. K. Wang, and M. Karczewicz, "Intra block copy in HEVC screen content coding extensions," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. PP, no. 99, pp. 1–11, 2016.

[10] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[11] L. Li, Z. Li, B. Li, D. Liu, and H. Li, "Pseudo sequence based 2-D hierarchical reference structure for light-field image compression," in *arXiv preprint arXiv:1612.07309*, 2016.

[12] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 7, pp. 1308–1319, July 2016.

[13] Y. Li, R. Olsson, and M. Sjöström, "Compression of unfocused plenoptic images using a displacement intra prediction," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[14] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 50–62, Jan 2016.

[15] C. Conti, J. Lino, P. Nunes, L. D. Soares, and P. L. Correia, "Spatial prediction based on self-similarity compensation for 3D holoscopic image and video coding," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 961–964.

[16] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *IEEE International Conf. on Multimedia and Expo - ICME*, July 2016, pp. 1– 4.

[17] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. M. M. Rodrigues, S. Faria, C. Pagliari, E. Silva, and L. D. Soares, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in *IEEE International Conf. on Multimedia and Expo - ICME*, July 2016, pp. 1– 4.

[18] R. Zaharia, A. Aggoun, and M. McCormick, "Adaptive 3D-DCT compression algorithm for continuous parallax 3D integral imaging," *Sig. Proc.: Image Comm.*, vol. 17, no. 3, pp. 231–242, 2002.

[19] A. Aggoun, "A 3D DCT compression algorithm for omnidirectional integral images," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, May 2006, pp. II–II.

[20] A. Aggoun and M. Mazri, "Wavelet-based compression algorithm for still omnidirectional 3D integral images," *Signal, Image and Video Processing*, vol. 2, no. 2, pp. 141–153, 2008.

[21] A. Aggoun, "Compression of 3D integral images using 3D wavelet transform," *Journal of Display Technology*, vol. 7, no. 11, pp. 586–592, Nov 2011.

[22] E. Elharar, A. Stern, O. Hadar, and B. Javidi, "A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform," *Journal of Display Technology*, vol. 3, no. 3, pp. 321–325, Sept 2007.

[23] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2016, pp. 1–4.

[24] I. Viola, M. Rerabek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," in *2016 32nd Picture Coding Symposium*, Dec. 2016.

[25] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.

[26] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *2006 IEEE International Conference on Multimedia and Expo*, July 2006, pp. 1929–1932.

[27] R. Sjoberg, Y. Chen, A. Fujibayashi, M. M. Hannuksela, J. Samuelsson, T. K. Tan, Y. K. Wang, and S. Wenger, "Overview of HEVC high-level syntax and reference picture management," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1858–1870, Dec 2012.

[28] P. Helle, S. Oudin, B. Bross, D. Marpe, M. O. Bici, K. Ugur, J. Jung, G. Clare, and T. Wiegand, "Block merging for quadtree-based partitioning in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1720–1731, Dec 2012.

[29] B. Li, H. Li, L. Li, and J. Zhang, "λ domain rate control algorithm for high efficiency video coding," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3841–3854, Sept 2014.

[30] L. Li, B. Li, H. Li, and C. W. Chen, "λ domain optimal bit allocation algorithm for high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2016.

[31] H. Li, B. Li, and J. Xu, "Rate-distortion optimized reference picture management for high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1844–1857, Dec 2012.

[32] Y. Gao, C. Zhu, S. Li, and T. Yang, "Layer-based temporal dependent rate-distortion optimization in random-access hierarchical video coding," in *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*, Sept 2016, pp. 1–6.

[33] K. Andersson, P. Wennersten, R. Sjoberg, J. Samuelsson, J. Strom, P. Hermansson, and M. Pettersson, "Non-normative HM encoder improvements," Document JCTVC-W0062, San Diego, USA, Feb 2016.

[34] High Efficiency Video Coding test model, HM-16.7. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/

[35] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Document VCEG-M33, Austin, Texas, USA, April 2001.

[36] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M.-T. Sun, "Inter-frame dependent rate-distortion optimization using lagrangian multiplier adaption," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, June 2015, pp. 1–6.

[37] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M. T. Sun, "Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 117–129, Jan 2016.

Li Li (M'17) received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2011 and 2016, respectively. He is now a postdoc researcher in University of Missouri-Kansas City.

His research interests include image/video coding and processing. He received the Best 10% Paper Award at the 2016 IEEE Visual Communications and Image Processing (VCIP) Conference.

Zhu Li (M'02-S'07) is now an Associate Professor with the Dept of Computer Science & Electrical Engnieering (CSEE), University of Missouri,Kansas City, and directs the Multimedia Computing & Communication (MC2) Lab. He received his PhD in Electrical & Computer Engineering from Northwestern University, Evanston in 2004. He was AFRL Faculty Fellow at the US Air Force Academy in Summer 2016, 2017. Sr. Staff Researcher/Sr. Manager with Samsung Research America's Multimedia Standards Research Lab in Richardson, TX, 2012-2015, Sr. Staff Researcher/Media Analytics Group Lead with FutureWei (Huawei) Technology's Media Lab in Bridgewater, NJ, 2010 2012, and an Assistant Professor with the Dept of Computing, The Hong Kong Polytechnic University from 2008 to 2010, and a Principal Staff Research Engineer with the Multimedia Research Lab (MRL), Motorola Labs, from 2000 to 2008. His research interests include image/video processing and compression, machine learning, as well as video adaptation, media network optimization.

He has 25 issued or pending patents, 90+ publications in book chapters, journals, conference proceedings and standard contributions in these areas. He is an IEEE senior member, associated editor (2015) for IEEE Trans.on Multimedia, and associated editor (2016) for IEEE Trans on Circuits & System for Video Technology, associated editor (2015) for Journal of Signal Processing Systems (Springer), steering committee member of IEEE ICME, elected member (2014-2017, 2017-2020) of the IEEE Multimedia Signal Processing (MMSP) Tech Committee, Steering Committee Chair (2016-) of the IEEE Multimedia Communication Technical Committee (MMTC), He received the Best Paper Award at the IEEE Int'l Conf on Multimedia & Expo (ICME), Toronto, 2006, and the Best Paper Award at the IEEE Int'l Conf on Image Processing (ICIP), San Antonio, 2007.

Bin Li (M'14) received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2008 and 2013, respectively.

He joined Microsoft Research Asia (MSRA), Beijing, China, in 2013 and now he is a Researcher. He has authored or co-authored over 20 papers. He holds over 10 granted or pending U.S. patents in the area of image and video coding. He has more than 30 technical proposals that have been adopted by Joint Collaborative Team on Video Coding. His current research interests include video coding, processing, and communication.

Dr. Li received the best paper award for the International Conference on Mobile and Ubiquitous Multimedia from Association for Computing Machinery in 2011. He received the Top 10% Paper Award of 2014 IEEE International Conference on Image Processing. He has been an active contributor to ISO/MPEG and ITU-T video coding standards. He is currently the Co-Chair of the Ad Hoc Group of Screen Content Coding extensions software development.

Dong Liu (M'13) received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2004 and 2009, respectively.

He joined USTC as an Associate Professor in 2012. He was a Member of Research Staff with the Nokia Research Center, Beijing, China, from 2009 to 2012. He has authored or co-authored over 30 papers in journals and international conferences. His research interests include image/video coding, multimedia signal processing and data mining.

Dr. Liu received the 2009 IEEE Transactions on Circuits and Systems for Video Technology Best Paper Award, and the Best 10% Paper Award at the 2016 IEEE Visual Communications and Image Processing (VCIP) Conference.

Houqiang Li (M'10-S'12) received the B.S., M.Eng., and Ph.D. degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 1992, 1997, and 2000, respectively, where he is currently a Professor with the Department of Electronic Engineering and Information Science.

His research interests include video coding and communication, multimedia search, image/video analysis. He has authored and co-authored over 100 papers in journals and conferences. He served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2010 to 2013, and has been with the Editorial Board of the Journal of Multimedia since 2009. He was the recipient of the Best Paper Award for Visual Communications and Image Processing (VCIP) in 2012, the recipient of the Best Paper Award for International Conference on Internet Multimedia Computing and Service (ICIMCS) in 2012, the recipient of the Best Paper Award for the International Conference on Mobile and Ubiquitous Multimedia from ACM (ACM MUM) in 2011, and a senior author of the Best Student Paper of the 5th International Mobile Multimedia Communications Conference (MobiMedia) in 2009.