ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 4, No. 3, 2009, *pp*. 211-223

Rate-Complexity Scalable Multi-view Image Coding with Adaptive Disparity-Compensated Wavelet Lifting

Pongsak Lasang [1], Chang-su Kim [2] and Wuttipong Kumwilaisak [1]

[1] Electronics and Telecommunication Department, Faculty of Engineering, King Mongkut's University of Technology, Thonburi, Thailand

[2] Electrical Engineering Department, Faculty of Engineering, Korea University, Korea

(*Received January 10, 2009, accepted April 2, 2009*)

Abstract. This paper presents a novel algorithm for rate-complexity scalable multi-view image coding using an adaptive disparity-compensated (DC) wavelet lifting scheme. First, image regions of multi-view images are prioritized by counting matching points. The proposed algorithm selects either Haar, 5/3, or our proposed multiple picture reference DC wavelet lifting adaptively. The selection criterion is based on the bit budget constraint, the complexity budget constraint, and the priorities of image regions. Then, the low-pass and high-pass subbands, obtained from the DC wavelet lifting, are further decomposed by a spatial wavelet transform. The resulting wavelet coefficients are entropy encoded with the SPIHT codec. Experimental results show that the proposed algorithm provides an efficient adaptive framework for multi-view image coding.

Keywords: Multi-view image coding, rate-complexity scalability, wavelet lifting, disparity-compensation, image feature matching.

1. Introduction

Recently, multi-view image coding has been extensively researched due to its broad variety of applications, including virtual tours on the Internet, three-dimensional navigation, presentation, and medical diagnosis [1-15]. Since a set of multi-view image consists of several ordinary images, it requires a huge amount of data for storage and transmission. A typical multi-view image, however, contains considerable redundancies among image views, since they represent the same scene from different viewpoints. The objective of multi-view image coding is to achieve high compression ratio, reducing these redundancies, while maintaining high image quality.

Various algorithms have been proposed to encode multi-view image data. Magnor *et al.* [1] proposed a multi-view image coding algorithm based on texture coding and model-aided prediction. 3-D scene geometry is utilized for their algorithm. Tong and Gray [2] examined the DC predictive coding for multi-view images, which employs a block-based correspondence matching scheme. In [3, 6], the light field was taken into account for multi-view image coding. In [12, 13, 15], geometric structures in multi-view images, such as depth map, epipolar geometry, 3-D voxel model, were used for multi-view image coding. Also, there have been several attempts to utilize the wavelet transform in video coding and multi-view image coding [3, 6, 9, 10], especially after the success of the motion compensation with wavelet lifting, which guarantees the invertibility at the synthesis side [16]. The main advantages of wavelet compression are its scalability and high energy compaction [16, 21]. For multi-view image coding, Girod *et al.* [3, 6] used the 4-D Haar and 5/3 wavelet lifting schemes to perform the disparity compensation (DC) for light field compression. Anantrasirichai *et al.* [9, 10] employed the hybrid of Haar and 5/3 wavelet lifting in the block-based DC. In their work, the selection of DC mode is based on the minimum sum of absolute differences (SAD) criterion. However, they restricted reference image views only to adjacent views. In other words, their structure does not allow the usage of several views or combination of them for DC. Moreover, their work does not exploit the features, used in 3-D scene reconstruction, when performing the multi-view image coding.

In general, when encoding multi-view image contents, only the rate (or bit budget) constraint is used as the encoding constraint. However, in practice, different encoders may have different encoding capabilities. An encoder with low computational capability cannot choose a complex algorithm to encode multi-view

images in a limited time. In contrast, an encoder with high capability has more degrees of freedom in choosing a complex coding algorithm to achieve better reconstructed multi-view image quality. To be useful in practice, a multi-view image encoder should be flexible enough to adjust its complexity adaptively.

In this work, we propose a novel algorithm for rate-complexity scalable multi-view image coding using an adaptive DC wavelet lifting scheme. Fig. 1 shows the overall architecture of the proposed algorithm. The proposed algorithm combines an image region prioritization scheme with an adaptive DC wavelet lifting technique to flexibly adjust both the bit rate and the encoding complexity. First, image regions are prioritized. The priorities of image regions are derived from a feature matching method. In this paper, the Harris corner detector [19] is used to find the matching points. Image regions with higher densities of matching points are assigned higher priorities in the DC wavelet lifting.

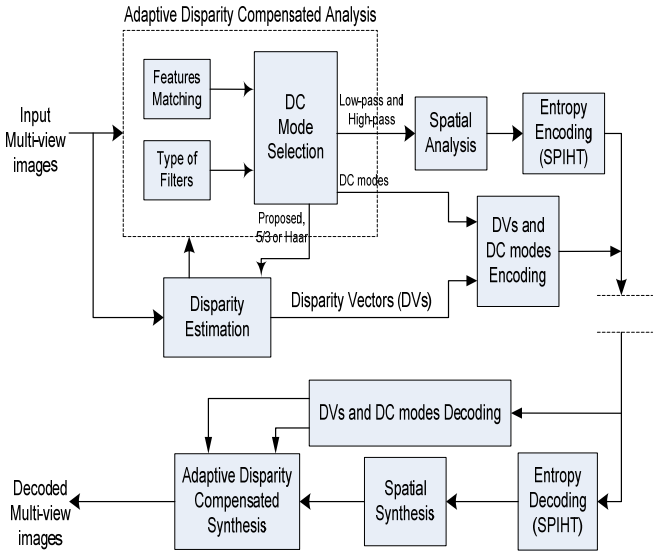


Fig. 1: The architecture of the proposed multi-view image coding algorithm.

Then, for each region, we select the mode of DC wavelet lifting among three modes: Haar, 5/3, or the proposed multiple picture reference modes. The mode is selected to achieve the minimum distortion given the constraint on the target bit rate, the encoding complexity, and the region priority. After the DC wavelet lifting is applied to reduce the redundancy among views, the discrete wavelet transform is applied to decompose low-pass and high-pass subbands, respectively. Finally, the wavelet coefficients are entropy coded with the SPHIT codec to generate the bitstream. Disparity vectors and DC modes are also entropy encoded by the arithmetic coding and sent to the decoder.

The remainder of the paper is organized as follows. Section 2 explains the image prioritization via image feature matching, and Section 3 describes the DC wavelet lifting modes. Section 4 briefly discusses the entropy codec. Section 5 proposes a mode selection scheme, which supports rate-complexity scalability. Section 6 provides experimental results. Finally, the conclusions are drawn in Section 7.

2. Image Region Prioritization via Image Feature Matching

In general, multi-view images taken from real life scenes with multiple cameras contain a huge amount of redundant information. We attempt to prioritize image regions so that a region with less redundancy has a higher priority and is encoded with a more precise prediction scheme. On the other hand, a region with more redundancy is assigned a lower priority, since it can be effectively encoded even with a simple prediction scheme.

We determine the priorities by counting the number of interesting points. Interesting points in an image view can be matched to interesting points in the other views. Notice that the regions, which have dense matching points, can be used to compute a depth map in 3-D reconstruction [19].

To find interesting points, the Harris corner detector is employed [19]. It detects the local variation of intensity in both *x* and *y* directions. Let $I(x, y)$ denote the intensity of the pixel at position (x, y). Also, let I_x and I_y denote the partial derivatives of $I(x, y)$ with respective to *x* and *y*, respectively. The local

structure matrix is defined as

$$L = w_G \times \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \tag{1}$$

where w_G is a Gaussian filter. The above equation can be written in another form as

$$L = \begin{bmatrix} \hat{I}_x^2 & \hat{I}_x \times \hat{I}_y \\ \hat{I}_x \times \hat{I}_y & \hat{I}_y^2 \end{bmatrix}, \tag{2}$$

where $\hat{I}_x = \sqrt{w_G} \times I_x$ and $\hat{I}_y = \sqrt{w_G} \times I_y$, which are the Gaussian smoothed versions of the partial derivatives of the local intensity.

The local structure matrix L in (2) is symmetric. Thus, it can be diagonalised by a matrix transformation. The resulting diagonal matrix has two non-negative eigenvalues λ_1 and λ_2. If both eigenvalues are large, the intensity has strong variation in two orthogonal directions, which indicates that (x, y) is probably a corner pixel. Therefore, we compute

$$\begin{aligned} M &= \lambda_1 \lambda_1 - k(\lambda_1 + \lambda_1)^2 \\ &= \det(L) - k(trace(L))^2 \end{aligned}, \tag{3}$$

where k is a small number. In this work, k is set to 0.04 as suggested in [19]. If *M* is larger than a threshold, the pixel at (x, y) is declared as an interesting point. The threshold is set to 400.

Matching points are determined as follows. First, given an interesting point at (x_0, y_0) in the reference view (the first view), we build up a window of 15×15 pixels around the center (x_0, y_0). Then, in each of the other views, we build up a window of the same size with center (x_0', y_0'), which is the position of the interesting point nearest to (x_0, y_0). We compute the sum of squared differences (SSD) between the corresponding pixels within the windows. Then, we move the center of the second window to the other interesting points within the search radius of 2 pixels. The center of the window yielding the minimum SSD is selected as the matching point of (x_0, y_0). We later use these matching points to select the mode of DC wavelet lifting adaptively.

3. Multiple Picture Reference Disparity Compensated Wavelet Lifting

Lifting schemes can be used to construct discrete wavelet transforms [20]. The subband decomposition into high-pass and low-pass components is achieved with a sequence of prediction and update steps in the lifting structure. Suppose that we have N image views. We divide this group of image views into even views, X_{2i}, and odd views, X_{2i+1}, which are similar and highly correlated in general. In the context of multi-view image coding, the disparity estimation and compensation can be effectively integrated into the prediction and update steps.

Figs. 2 and 3 show the first level decompositions of the Haar and 5/3 DC wavelet lifting, respectively. The Haar DC wavelet lifting is performed by using only a single adjacent view as a reference view, whereas the 5/3 DC wavelet lifting uses two adjacent views. Specifically, the Haar lifting uses image view $i-1$ or image view $i+1$ to reduce the redundancy in image view i, while the 5/3 lifting uses both of them. On the other hand, our proposed lifting mode enables us to use more than two reference

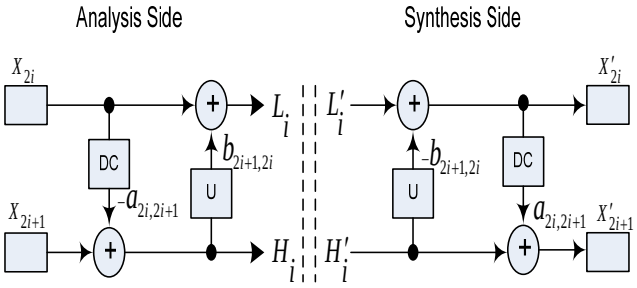


Fig. 2: The first level decomposition of the Haar DC wavelet lifting.

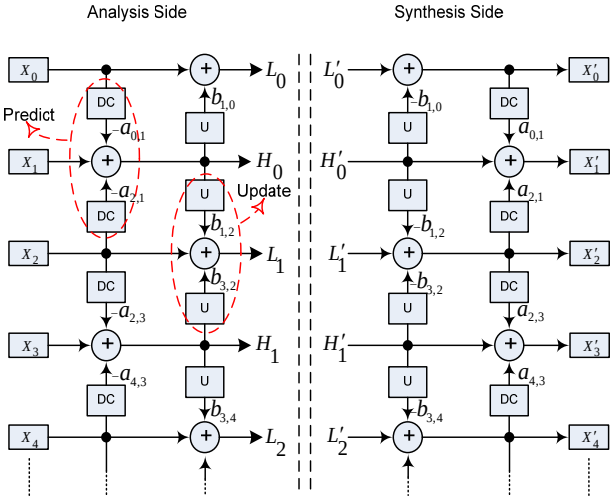


Fig. 3: The first level decomposition of the 5/3 DC wavelet lifting.

image views. To predict image view i with the new lifting mode, we use image views $i-1$, $i+1$, $i-3$, $i+3$, $i-5$, $i+5$, …. In other words, an even view is predicted from odd views, and an odd view is predicted from even views. In this way, it is guaranteed that all image views can be recovered at the synthesis side of the wavelet lifting.

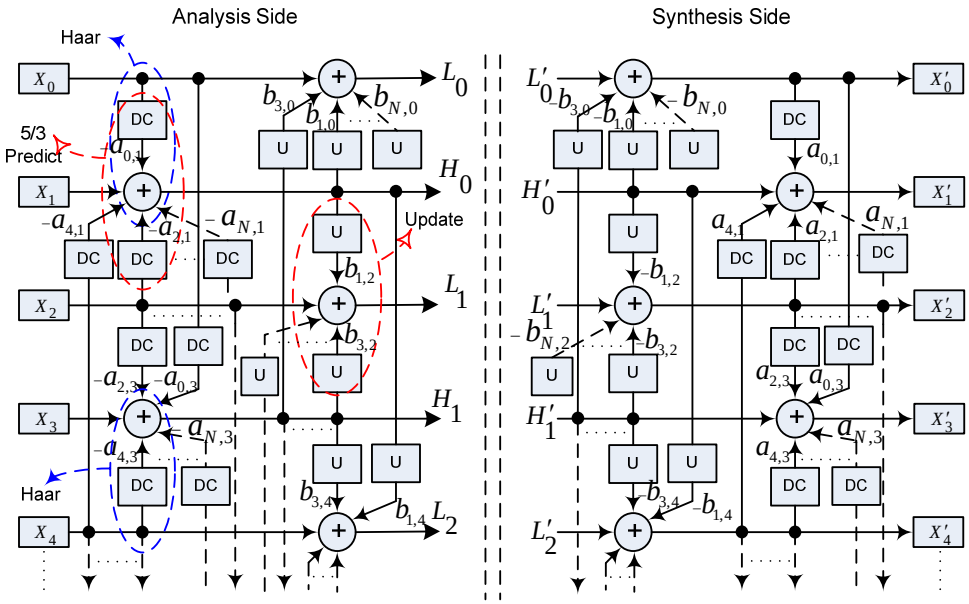


Fig. 4: The unified structure of DC wavelet lifting for multi-view image coding.

Fig. 4 shows the unified structure of DC wavelet lifting for multi-view image coding, which includes the Harr, 5/3, and our proposed lifting modes as special cases. In the unified structure, the number of reference

views can be flexibly chosen to obtain the best prediction. It also does not restrict reference views to adjacent image views only. In other words, non-adjacent views can be used as references as well. Suppose that a multi-view image consists of $N/2$ even views and $N/2$ odd views. Then, there are $\sum_{i=1}^{N/2}\binom{N/2}{i}=2^{N/2}-1$ possible choices for selecting reference image views for DC. The *i*-th low-pass (L_i) and high-pass (H_i) components can be written as

$$\begin{aligned} H_i = X_{2i+1} &- a_{2i,2i+1}\mathrm{DC}(X_{2i},\hat{d}_{2i+1\to 2i}) \\ &- a_{2i+2,2i+1}\mathrm{DC}(X_{2i+2},\hat{d}_{2i+1\to 2i+2}) \\ &- a_{2i-2,2i+1}\mathrm{DC}(X_{2i-2},\hat{d}_{2i+1\to 2i-2}) \\ &- a_{2i+4,2i+1}\mathrm{DC}(X_{2i+4},\hat{d}_{2i+1\to 2i+4}) \\ &- a_{2i-4,2i+1}\mathrm{DC}(X_{2i-4},\hat{d}_{2i+1\to 2i-4}) \\ &- a_{2i+6,2i+1}\mathrm{DC}(X_{2i+j},\hat{d}_{2i+1\to 2i+6}) \\ &- \dots \end{aligned} \tag{4}$$

and

$$\begin{aligned} L_i = X_{2i} &+ b_{2i-1,2i}\mathrm{DC}(H_{i-1},-\hat{d}_{2i-1\to 2i}) \\ &+ b_{2i+1,2i}\mathrm{DC}(H_i,-\hat{d}_{2i+1\to 2i}) \\ &+ b_{2i-3,2i}\mathrm{DC}(H_{i-2},-\hat{d}_{2i-3\to 2i}) \\ &+ b_{2i+3,2i}\mathrm{DC}(H_{i+1},-\hat{d}_{2i+3\to 2i}) \\ &+ b_{2i-5,2i}\mathrm{DC}(H_{i-3},-\hat{d}_{2i-l\to 2i}) \\ &+ b_{2i+5,2i}\mathrm{DC}(H_{i+2},-\hat{d}_{2i+1\to 2i}) \\ &+ \dots \end{aligned} \tag{5}$$

Table 1: Scaling factors in the prediction and update steps in different lifting modes. (N_r denotes the number of reference image views.)

Lifting modes	$a_{m,n}$	$b_{m,n}$
Haar mode	1	$\frac{1}{2}$
5/3 mode	$\frac{1}{2}$	$\frac{1}{4}$
Proposed mode	$\frac{1}{N_r}$	$\frac{1}{2N_r}$

where $\hat{d}_{n\to m}$ denotes the set of disparity vectors from image view *n* to image view *m*. $\mathrm{DC}(X_m,\hat{d}_{n\to m})$ is the disparity compensated image of X_n, which is obtained from the reference view X_n using $\hat{d}_{n\to m}$. The scaling factors $a_{m,n}$ and $b_{m,n}$ are used in the prediction and update steps, respectively. All the subscripts, denoting view indices, are restricted between 0 and *N*-1, where *N* is the number of image views.

Note that, for the Haar and 5/3 wavelet lifting schemes, $a_{m,n}$ is set to 1 and 0.5, and $b_{m,n}$ is set to 0.5 and 0.25, respectively. For our proposed lifting mode, we adopt the weighting scheme in [18]. Specifically, $a_{m,n}$ is set to the inverse of the number of reference views, and $b_{m,n}$ is set to $a_{m,n}/2$. Table 1 summarizes the scaling factors in the lifting modes. Fig. 5 illustrates the reference image views used in DC with Haar, 5/3, and our proposed lifting modes, when a number of views in a set of multi-view image are equal to five.

The complexity of DC wavelet lifting is determined by the block matching process. Suppose that the search range of each block is equal to $R \times R$ pixels, the block size is $N \times N$ pixels, and the image size is $M \times M$ pixels. The complexity, when we use only a single reference view, requires $(2R+1)^2 M^2$ operations to perform block matching process. Therefore, if we use *k* reference views, the total complexity per block is equal to $k(2R+1)^2 M^2$ operations. In this work, we treat $(2R+1)^2 M^2$ operations as one complexity unit. Therefore, we require k complexity units when we use *k* reference views.

Different DC wavelet lifting modes require different levels of complexity. A multi-view image encoder can choose the mode based on its computational capability. For example, if the encoder has low processing power, it should choose simple DC wavelet lifting modes, which require a small number of complexity units. On the other hand, if the encoder is powerful, it can choose more complex DC lifting modes.

4. Entropy Coding of Motion Vectors, Modes, and Residue Error

The DC wavelet lifting decomposes multiple image views into low pass subbands and high pass subbands, reducing the inter-view redundancy. Each subband, however, still contains spatial redundancies, and thus it is decomposed further by the bi-orthogonal discrete wavelet transform (DWT), as shown in Fig. 6.

The DWT coefficients are then compressed using the SPIHT coder [21], which is based on the set partitioning in hierarchical trees and encodes the coefficients in the bit-plane order from most significant bits to least significant bits. In general, the low pass subband images from the DC wavelet lifting contain more important information to human visual system than the high pass subband images. Therefore, we assign more bits to encode the DWT coefficients for a low pass subband image than those for a high pass subband image. If the total bit budget allocated to a low pass subband is equal to $B_{budget,L}$, the total bit budget allocated to a high pass subband will be equal to $B_{budget,H} = \kappa \times B_{budget,L}$, where $0 \le \kappa \le 1$.

The output bitstream of the SPIHT codec is transmitted to the decoder. At the decoder side, the SPIHT decoder obtains the low pass and the high pass subband images, which are, in turn, used to reconstruct the multi-view image at the synthesis side of the DC wavelet lifting.

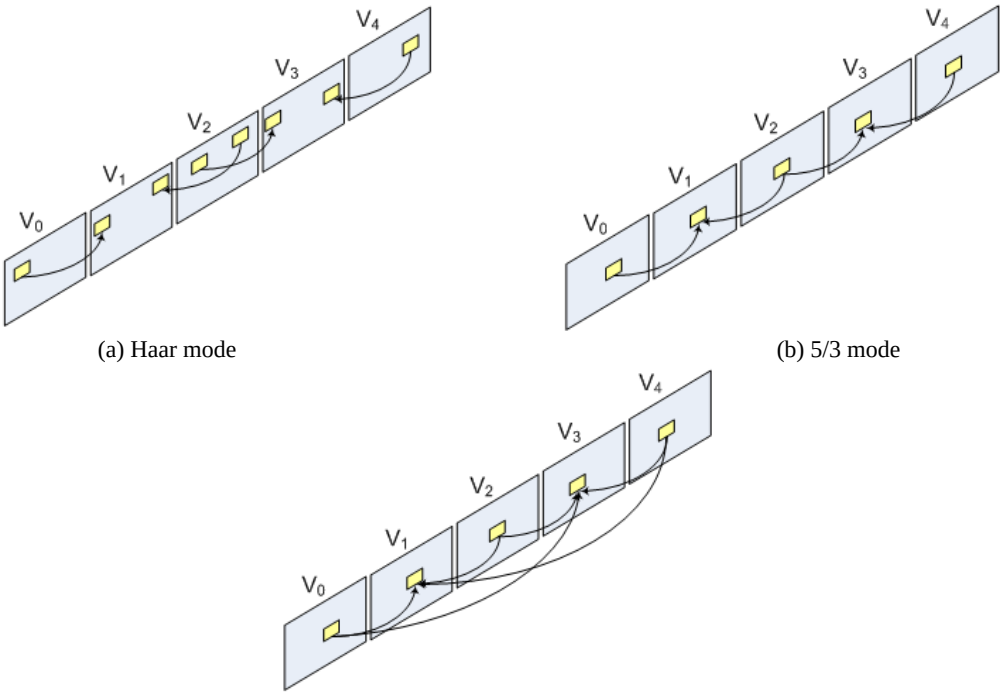
There are two parameters that should be also transmitted from the encoder to the decoder: disparity vectors and DC lifting modes. In this work, we encode disparity vector differences (DVD) and DC lifting modes using arithmetic coding. The disparity vector of the current block is predicted from the median of the disparity vectors of the left block, the top block, and the top right block. Then, we compute DVD, which is the difference between the disparity vector of the current block and the predicted one. We represent DVDs and DC lifting modes by code numbers, as shown in Table 2 and Table 3, respectively. The code numbers are then encoded using the arithmetic coder in [22] to obtain a bit stream. At the decoder, disparity vectors and DC lifting modes are decoded and then used to reconstruct the multi-view images.

Table 2: The code numbers for DVDs.

DVDs	Code numbers
0	0
1	1
-1	2
2	3
-2	4
3	5
…	…

Table 3: The DC lifting modes and the corresponding code numbers. Haar_left and Haar_right denote the Haar mode when the left side image and the right side image are used as the reference, respectively.

Lifting modes	Code numbers
5/3 mode	0
Haar_left mode	1
Haar_right mode	2
Proposed mode	3



(a) Haar mode

(b) 5/3 mode

(c) Proposed mode

Fig. 5: Illustration of the reference image views in the Haar, 5/3 and our proposed lifting modes, when a number of image views are equal to five.

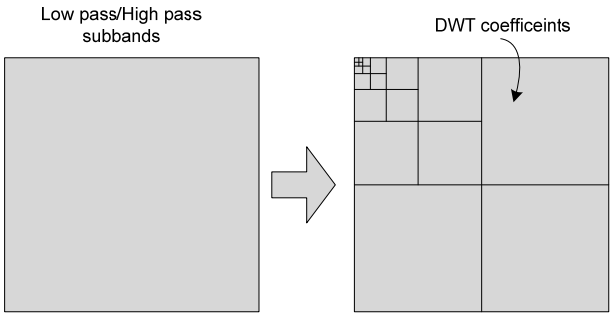


Fig. 6: After the DC wavelet lifting, the low pass and high pass subbands still contain spatial redundancies. Thus, each subband image is decomposed further by the spatial DWT.

5. Mode Selection with Scalable Rate and Complexity

5.1. Problem Formulation

The DC wavelet lifting modes determine the rate and the quality of reconstructed multi-view images, and the complexity of a DC mode is proportional to a number of reference frames in the compensation procedure. The mode should be selected to maximize the quality of reconstructed images, subject to both the rate and complexity constraints. Let $D_i(M_i, B_i)$ denote the distortion of the i th disparity compensated block, when it is compensated with mode M_i at bit rate B_i. The mode M_i is selected from {Haar_left, Haar_right, 5/3, Proposed}, and the bit rate B_i includes the bits for encoding the mode, disparity vectors, and residue errors. Also, let $C_i(M_i)$ denote the complexity to encode block i with mode M_i. Notice that the complexities of Haar, 5/3, and the proposed modes are 1, 2, and $N/2$ complexity units, where N is the number of multiple views.

Our objective is to obtain the highest image quality subject to both the rate and complexity constraints. It can be formally stated as

Problem: For block i, search for the DC mode M_i and the bit rate B_i that minimize the distortion

$$D_i(M_i, B_i) \tag{6}$$

subject to the constraints

$$B_i \le \gamma_i \times B_{budget}, \tag{7}$$

and

$$C_i(M_i) \le \alpha_i \times C_{budget}, \tag{8}$$

where B_{budget} and C_{budget} are the bit budget and the complexity budget, and γ_i and α_i are the priority factors of block i for the bit budget and the complexity budget, respectively. The priority factors can be determined based on the number of matching points within the block.

This constrained optimization problem can be solved by minimizing the Lagrangian cost function, given by

$$\Omega(\lambda_{bits}, \lambda_{comp}) = D_i(M_i, B_i) + \lambda_{bit} B_i + \lambda_{comp} C_i(M_i), \tag{9}$$

where λ_{bit} is the Lagrangian multiplier for the bit budget, and λ_{comp} is the Lagrangian multiplier for the complexity budget.

5.2. Problem Solution

To find out the exact solution that minimizes (9), we need the relationship between the bit budget, the complexity, and the distortion. However, the accurate estimation of the relationship requires a huge amount of computations. This is because we need consider all possible modes and bit rates.

Therefore, in this work, we instead obtain an approximate solution by constructing universal rate-distortion and complexity-distortion curves from a training set of multi-view images. These trained rate-distortion and complexity- distortion curves approximate the rate-distortion and complexity-distortion characteristics of incoming multi-view images. These trained curves are shown in Figs. 7 and 8, which were obtained from three sets of multi-view images: *Hotel*, *Marker*, and *Teddy*.

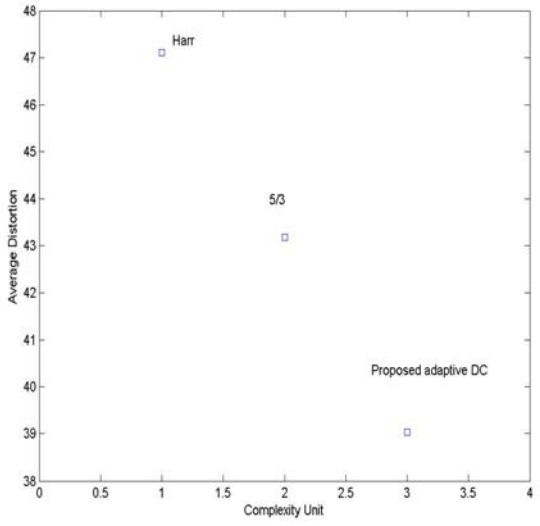


Fig. 7: The complexity-distortion plot of different DC wavelet lifting modes.

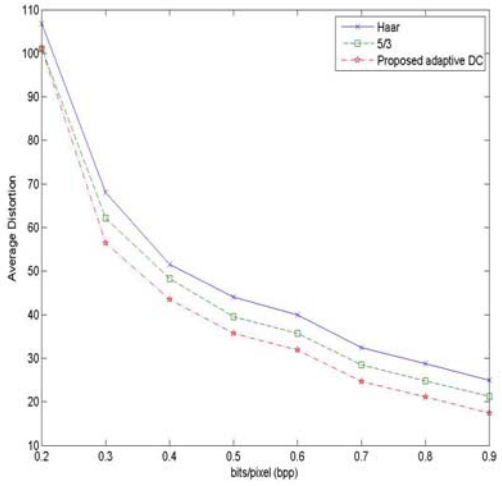


Fig. 8: The rate-distortion curves of different DC wavelet lifting modes.

Using these trained curves, we select the DC wavelet lifting mode and the bit rate in two steps. First, given the complexity budget, we select the mode achieving the lowest average distortion from the complexity-distortion curve. Then, with the selected mode, we choose the rate-distortion curve to determine the target bit rate. For example, suppose that there are totally five views in a set of multi-view images. If the complexity budget of a block is equal to 3 units, the proposed adaptive DC mode is selected, since it provides the lowest average distortion within the budget. Then, the rate-distortion curve corresponding to the proposed adaptive DC mode is selected. Finally, given the bit budget constraint, the optimal bit rate is obtained from the rate-distortion curve.

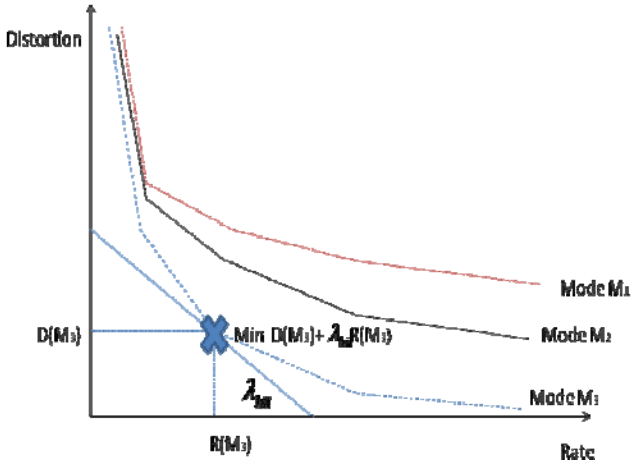


Fig. 9: The example of rate-distortion curves for different DC wavelet lifting modes. In this example, the optimal bit rate, corresponding to a Lagrangian multiplier λ_{bit}, is found.

The detailed algorithm can be stated as follows.

Algorithm: **(Rate-Complexity Optimized Mode Selection to Achieve the Highest Picture Quality)**

Step 1: Given the total bit budget and complexity budget for an entire frame, distribute them to all blocks by considering the priority factors γ_i and α_i.

Step 2: For each block in the frame, peform

Step 2.1: Given the complexity budget, select the mode, providing the minimum average distortion, from the trained complexity-distortion plot. Let M_i^* denote the selected mode.

Step 2.2: Choose the trained rate-distortion curve of the mode M_i^*. Use the bi-section algorithm in [23], to search for the optimal bit rate corresponding to a Lagrangian multiplier λ_{bit}. Let B_i^* denote the optimal bit rate. Fig. 9 illustrates the process of determining the optimal bit rate, assuming the selected mode in Step 2.1 is M_3.

Step 2.3: Encode the block using M_i^* and B_i^*.

Please acknowledge collaborators or anyone who has helped with the paper at the end of the text.

6. Experimental Results

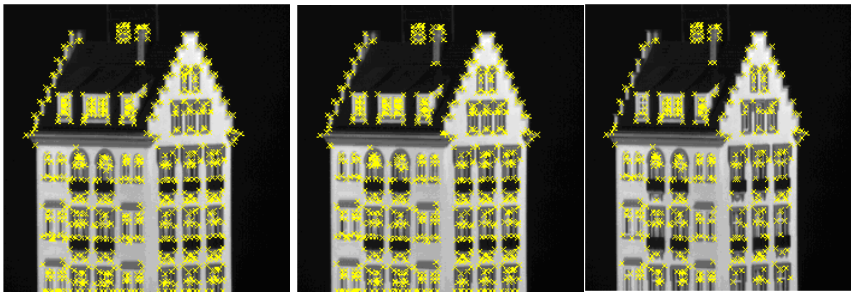
To evaluate the performance of the proposed rate-complexity scalable algorithm, we experiment on both synthetic and real multi-view images. The synthetic sequence is the *hotel* image. The real sequences are the *marker* and *teddy* images. These images consist of five views and have the size of 512 x 512 pixels. The performance of the proposed algorithm is evaluated in terms of DC prediction efficiency and PSNR. We compare the PSNR performances of the proposed algorithm under different rate and complexity budgets. Note that we use the block size of 4x4 pixels and the integer-pel search accuracy for all DC modes. The ratio of the bit allocation in low and high subbands is equal to 3 to 2. In other words, κ defined in Section 4 is set to 0.6.

As mentioned in Sections 2 and 5, the regions with high density of matching points are given higher priorities. Fig.10 shows the computed matching points of the *hotel* image. Fig. 10 (c) depicts the matching points obtained from Fig. 10 (a) and Fig. 10 (b). We divide image blocks to two classes: the class corresponding to high density matching points and the class corresponding to low density matching points. In the following tests, the complexity priority factor α_i for blocks with high matching point densities is set to be twice higher than the other blocks with low densities. The bit priority factor γ_i is set to be the same for all blocks. To be more specific, α_i is equal to 2 and 1 for blocks with high and low matching point densities, respectively. γ_i is set to be 1 for all blocks.

Figs. 11 and 12 show the high pass subbands of the *hotel* and the *teddy* images, respectively. They represent residue errors after the DC prediction. The normalized residue energy is defined as the squared sum of residue pixel errors, divided by the number of pixels. We see that residue errors are reduced, when we

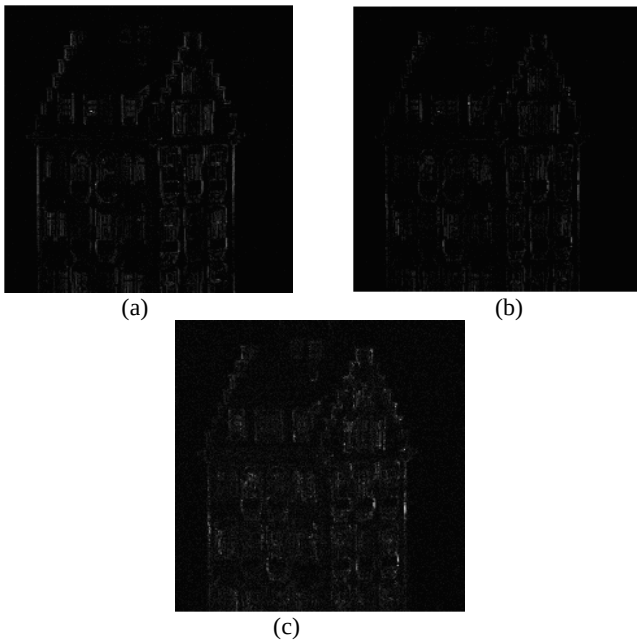
allow a higher complexity encoder to exploit more complex DC wavelet lifting modes. Fig. 13 compares the subjective quality of the *teddy* sequence under different complexity budgets. We can conclude from these experimental results that, if the encoder increases its complexity, the better performance can be achieved.

Figs. 14, 15, and 16 show the average PSNRs over all image views for the *hotel*, *teddy* and *marker* images, respectively, in terms of bits per pixel (bpp) under different complexity budgets. The complexity budgets for scenario 1 to scenario 5 are 16384, 24576, 32768, 40960, and 49152 units per image, respectively, which correspond to the average complexity budgets of 1.0 unit, 1.5 units, 2 units, 2.5 units, and 3.0 units per block. Scenario 1 has the lowest complexity budget, 1.0 unit per block, in which the only possible DC mode is the Haar mode [9,10]. On the other hand, in scenarios 2~5, the mode for a block can be selected from the Haar, 5/3, and the proposed modes, depending on the given complexity budget. It is observed that the better performance can be achieved when the encoder uses bigger complexity and rate budgets. Note that, at low bit rates, the performance gains are not significant even though we increase the complexity budgets. This is because the additional bits for more disparity vectors in the advanced modes become a heavier burden at low bit rates. Figs. 14, 15, and 16 indicate that the proposed algorithm can support different bit budgets and complexity budgets efficiently and adaptively. Note that bit rates in Figs. 14, 15, and 16 include all the bits for encoding wavelet coefficients, disparity vectors, and lifting modes.



(a) (b) (c)

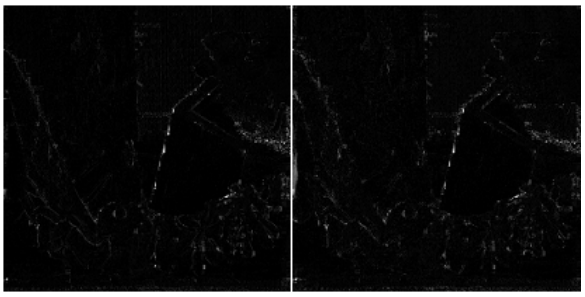
Fig. 10: (a) Interesting points from the first view (726 points). (b) Interesting points from the second view (730 points). (c) Matching points between the first and the second views (488 points).



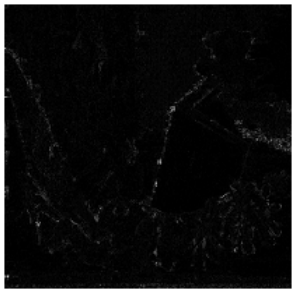
(a) (b)

(c)

Fig. 11: A high pass subband of the *hotel* image: (a) Haar mode (normalized energy = 19.79, complexity = 1 unit per block), (b) 5/3 mode (normalized energy = 10.341, complexity = 2 units per block), and (c) the proposed adaptive wavelet lifting (normalized energy = 5.095, complexity = 3 units per block).

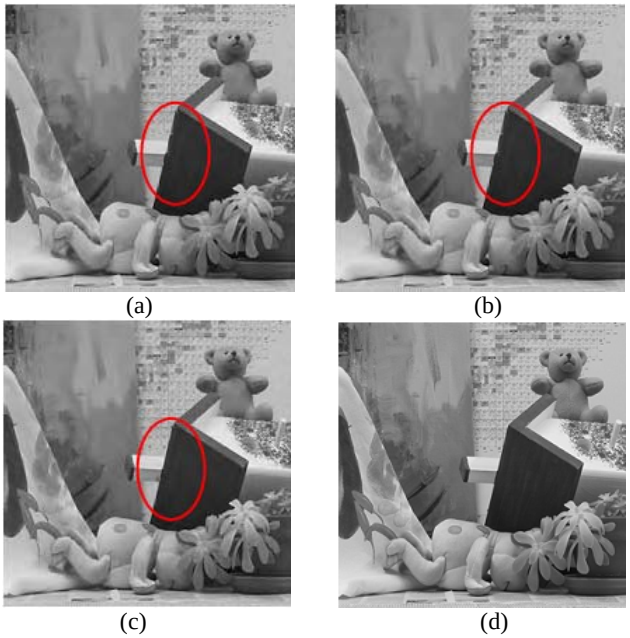


(a) (b)



(c)

Fig. 12: A high pass subband of the *Teddy* image. (a) Haar mode (normalized energy = 27.98, complexity = 1 unit per block), (b) 5/3 mode (normalized energy = 13.92, complexity = 2 units per block) and (c) the proposed adaptive wavelet wavelet lifting (normalized energy = 7.94, complexity = 3 units per block).



(a) (b)

(c) (d)

Fig. 13:Reconstructed *teddy* images at the bit rate of 1.0 bpp. (a) Haar mode (complexity = 1 unit per block), (b) 5/3 mode (complexity = 2 units per block), (c) the proposed adaptive wavelet lifting (complexity = 3 units per block), and (d) the original image.

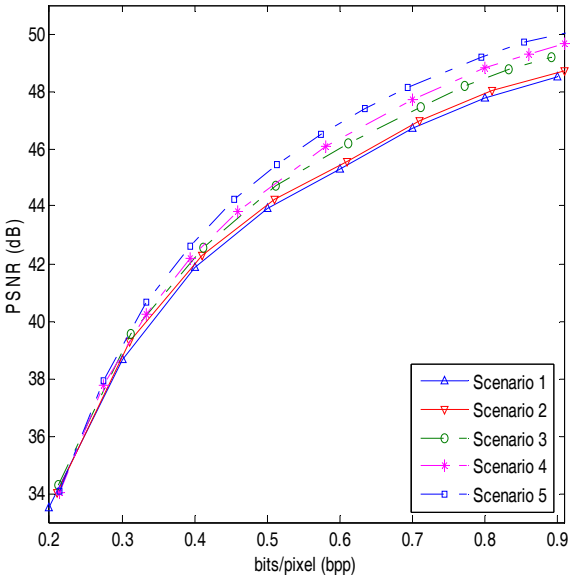
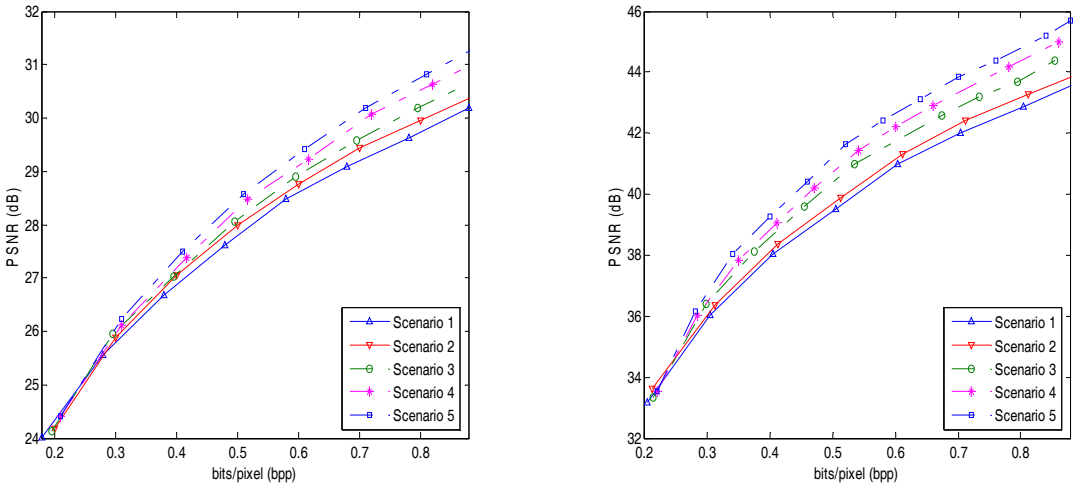


Fig. 14: The PSNR performances on the *hotel* image. In scenarios 1~5, the average complexity budgets are 1.0, 1.5, 2.0, 2.5, and 3.0 units per block, respectively.



(a) (b)

Fig. 15: The PSNR performances on the *teddy* image(Fig a) and *marker* image(Fig b). In scenarios 1~5, the average complexity budgets are 1.0, 1.5, 2.0, 2.5, and 3.0 units per block, respectively.

7. Conclusions

We proposed a rate-complexity scalable framework using adaptive DC wavelet lifting technique for multi-view image coding. The proposed algorithm smartly selects a proper DC mode among the Haar, 5/3, or our proposed wavelet lifting for each block subject to the complexity budget and bit budget constraints. The higher PSNR performance can be achieved by increasing the complexity, the bit rate, or both. Therefore, this prototype scheme is very scalable, which can be applied in various applications. The simulation results on test images demonstrated that the proposed algorithm can support the rate-complexity scalability efficiently.

8. References

[1] M. Magnor, P. Ramanathan, and B. Girod. Multi-view coding for image-based rendering using 3-D scene geometry, IEEE Trans. Circuits Syst. *Video Technol*.. 2003, **13**(11): 1092-1106.

[2] X. Tong and R. M. Gray. Coding of multi-view images for immersive viewing. *in Proc. ICASSP'00.* 2000, **4**: 1879-1882.

[3] B. Girod, C. L. Chang, P. Ramanathan, and X. Zhu. Light field compression using disparity-compensated lifting. *in Proc. ICASSP'03*. 2003, **4**: 760-763.

[4] M. Fierl and B. Girod. Coding of multi-view image sequences with video sensors. *in Proc. ICIP*. 2006, pp.609-612.

[5] M. Fierl, A. Mavlankar, and B. Girod. Motion and disparity compensated coding for multi-view video, IEEE Trans. *Circuit Syst. Video Technol*.. 2007, **17**(11): 1474-1484.

[6] C.-L. Chang, X. Zhu, P. Ramanathan, and B. Girod. Light field compression using disparity-compensated lifting and shape adaptation, IEEE Trans. *Image Proc*.. 2006, **15**(4): 93-806.

[7] A. Smolic and P. Kauff. Interactive 3-D video representation and coding technologies. *Proc. IEEE*. 2005, **93**(1): 98-110.

[8] R. S. Wang and Y. Wang. Multi-view video sequence analysis, compression, and virtual viewpoint synthesis. *IEEE Trans*. *Circuit Syst. Video Technol*.. 2000, **10**(4): 397-410.

[9] N. Anantrasitichai, C. N. Canagarajah, and D. R. Bull. Lifting-based multi-view image coding. *in Proc. ISCAS.* 2005, **3**: 2092-2095.

[10] N. Anantrasitichai, C. N. Canagarajah, and D. R. Bull. Multi-view image coding with wavelet lifting and in-band disparity compensation. *in Proc. ICIP*. 2005, **3**: 33-36.

[11] X. San, H. Cai, J.-G. Lou, and J. Li. Multi-view image coding based on geometric prediction. *IEEE Trans. Circuit Syst. Video Technol*.. 2007, **17**(11): 1536-1548.

[12] Y. Morvan, D. Farin, and H. N. Peter. Incorporating depth-image based view-prediction into H.264 for multiview-image coding. *in Proc. ICIP*. 2007, **1**: 205-208.

[13] J. Lu, H. Cai, J.-G. Lou, and J. Li. An epipolar geometry-based fast disparity estimation algorithm for multi-view image and video coding. *IEEE Trans. Circuit Syst. Video Technol*.. 2007, **17**(6): 737-750.

[14] F. Yang, Q. Dai, and G. Ding. Multi-view image coding based on multi-terminal source coding. *in Proc. ICASSP.* 2007, **1**: 1037-1040.

[15] Y. Gao and H. Radha. Multi-view image coding using 3-D voxel models. *in Proc. ICIP*. 2005, **2**: 257-260.

[16] A. Secker and D. Taubman. Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting. *in Proc. ICIP*. 2001, **2**: 1029-1032.

[17] L. Luo, J. Li, S. Li, Z. Zhuang, and Y. Q. Zhang. Motion compensated lifting wavelet and its application in video coding. *in Proc. ICME.* 2001, pp. 365-368.

[18] N. Mehrseresht and D. Taubman. Adaptively weighted update steps in motion compensated lifting based scalable video compression. *in Proc. ICIP*.2003, **3**: 771-774.

[19] C. J. Harris and M. Stephens. A combined corner and edge detector. *in AVC 4th Proc.*. 1988, pp. 147-151.

[20] W. Swelden. The lifting scheme: A construction of second generation wavelets. *SIAM Journal of Mathematical Analysis.* 1998, **29**(2): pp. 511-546.

[21] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. Circuits Syst. Video Technol*.. 1996, **6**(3): 243-250.

[22] A. Moffat, R. M. Neal, and I. H. Witten. Arithmetic coding revisited. *ACM Trans. Information Systems.* 1998, **16**(3): 256-294.

[23] A. Ortega. *Optimization Techniques for Adaptive Quantization of Image and Video under Delay Constraints.* Ph.D. thesis, Dept. Electrical Engineering, Columbia University, New York, NY, 1994.