Anantrasirichai, N., Canagarajah, C. N., & Bull, D. (2005). Lifting-based multi-view image coding. 2092 - 2095. 10.1109/ISCAS.2005.1465031

Link to published version (if available):
10.1109/ISCAS.2005.1465031

Link to publication record in Explore Bristol Research
PDF-document

Lifting-Based Multi-View Image Coding

N. Anantrasitichai, C. Nishan Canagarajah, and David R. Bull
Department of Electrical & Electronic Engineering
University of Bristol
Bristol BS8 1UB UK

***Abstract*—A number of lifting-based video coding schemes have been recently proposed for scalable video coding. In this paper, we present a novel multi-view image codec based on a wavelet lifting scheme. The proposed lifting scheme with disparity compensated channel filtering is very efficient in terms of compressions performance, memory requirements and implementation. We propose a number of enhancements to the basic scheme, such as hybrid prediction, adaptive weighing in update step and overlapped block disparity compensation which yield significant improvements in rate distortion performance. Experimental results show image quality gains of up to 1.5 dB and 1.2 dB against using well established methods such as the block-matching Haar and 5/3 wavelet lifting respectively.**

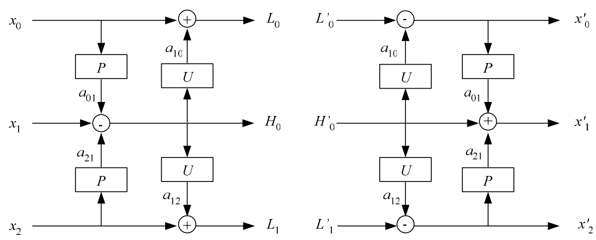
I. INTRODUCTION

In the recent past, the visual communication systems, including multi-view communication, have dramatically grown and have become useful in various applications. The multi-view system is capable of providing realistic immersive experience rendering them useful in many applications such as remote surveillance, navigation, medical imaging, telerobotics, entertainment and virtual reality. However, the amount of information generated by a 3D imaging system has limited many commercial applications, due to a manifold increase in bandwidth over existing monoscopic video. Thus, the efficient compression algorithms are vital to reduce the size of data without sacrificing the perceived image quality.

Multi-views of a scene, taken from two or more cameras so as to gather the depth information, typically contain significant correlation. Many algorithms have been introduced, such as in [1],[2]. In [1], the technique to code sum and difference of the two images in a stereo pair was proposed, while the three-dimensional discrete cosine transform coding of stereo image was presented in [2]. The multi-view images could be regarded as a special type of image sequences so that the compression methods which exploit temporal redundancy could be adapted to reduce the spatial geometric redundancy present in multi-view images. The concept of disparity compensation, which established correspondence between similar areas in a stereo pair using binocular disparity information, is considered to be one of the most successful compression techniques for multi-view sequences. In this approach, one view is coded independently of the others, while the remaining view is predicted from this view based on disparity estimation/compensation.

Recently, wavelet scheme has been proposed for image/video coding. This approach is very efficient in terms of implementation and scalability properties. In this work, we extend this approach for multi-view image compression to remove spatial, temporal and geometric redundancies present in such sequences. Moreover, the performance of the wavelet scheme can be enhanced by using an advanced prediction scheme to minimize correlation between views/frames. In temporal axis, the motion compensated temporal filtering (MCTF) was proposed by Ohm [3]. The block-based motion compensation is applied to three-dimensional discrete wavelet transform (3-D-DWT) and showed that a non invertible transform will affect coding gain if there are numerous disconnected pixels between blocks. This could be a significant problem in block-matching in disparity field due to some occlusion. Accordingly, the wavelet lifting scheme has been attractive, due to the invertible property. In [4], a non-linear lifting framework for temporal wavelet transform has been introduced which is shown to be superior to the MCTF. Luo et al. have implemented a motion compensated lifting structure by using bi-directional motion estimation in each lifting step of temporal direction [5]. Secker and Taubman have developed the highly scalable video coding approach [6]. It is based on the lifting with adaptive motion compensation. However, they have ignored the possibility of occlusion which is common in multi-view image system.

The implementation of the wavelet lifting scheme for multi-view image compression is novel and has not been investigated previously. In this paper, we present a framework which exploits spatial correlation as well as high geometric correlation present among multi-view images based on the wavelet lifting scheme. The paper is organized as follows: Section II briefly explains the fundamentals of wavelet lifting scheme and disparity compensation. The proposed codec is described in Section III. The experimental results are presented in Section IV with conclusions and future works in Section V.



(a) (b)

Figure 1. Wavelet lifting scheme for three-view images. (a) Analysis side. (b) Synthesis side.

II. DISPARITY COMPENSATED WAVELET LIFTING SCHEME

The lifting scheme, one of various techniques exploited to construct wavelet bases or to factor existing wavelet filters into basic building blocks, was firstly introduced by Sweldens [7]. Basically, the forward wavelet lifting method decomposes wavelet transforms into a set of stages. The operation starts with a split step, which divides the data set, x, into two groups. Normally, odd element, x_{2k+1}, and even element, x_{2k}, are chosen, because they are supposed to be similar and contain largest correlation. The next step is prediction where one element is used to predict other elements in the data. Then, the high-pass residual signal, H_i, generated by subtracting the predicted element from the original element, will contain very little energy thereby achieving significant compression. Finally an update step combines residual data from the previous process to reduce the effect of aliasing in low-pass signal, L_i.

This scheme can be simply adapted to the three-view image system. Its analysis and synthesis diagrams are illustrated in Fig. 1 (a) and (b) respectively, where x_0, x_1 and x_2 represent image in left view, middle view and right view, while a_{mn} is the coefficient map from view m to view n. In traditional filtering such as Haar and 5/3 transform, a_{mn} is fixed to specific values, namely it equals to 0.5 in the predict step and 0.25 in the update step for 5/3 wavelet and equals to 1 in the predict step and 0.5 in the update step for Haar.

If no advanced prediction schemes are exploited in lifting process, i.e. $P(x)=1$ and $U(x)=1$, the low-pass transformed images will suffer from the ghosting artifacts on account of uncompensated shifting of the corresponding data between channels. Furthermore, the high-pass subband will contain considerable energy compromising compression efficiency. This problem might be avoided by enhancing the prediction step within the lifting scheme. Hence, in the context of multi-view image system, the disparity estimation can be usefully applied to the lifting algorithm. This disparity model should be employed without sacrificing the invertible property of lifting transform. The disparity compensated wavelet lifting scheme can be mathematically written as follows.

Defined $D^c_{m\to n}(x)$ is the disparity compensated function predicting for data x contained in image view n by utilizing the information from image view m and generates disparity vectors, $d_{m\to n}$. The low-pass and high-pass components can be written as follows;

$$H_0 = x_1 - a_{01}\cdot D^c_{0\to1}(x_0) - a_{21}\cdot D^c_{2\to1}(x_2) \quad (1)$$

$$L_m = x_{2m} + a_{(1)(2m)}\cdot D^c_{1\to2m}(H_0) \quad m=0,1 \quad (2)$$

The reverse transforms are shown in (3) and (4).

$$x'_{2m} = L'_m - a_{(1)(2m)}\cdot \widetilde{D}^c_{1\to2m}(H'_0, d_{1\to2m}) \quad m=0,1 \quad (3)$$

$$x'_1 = H'_0 + a_{01}\cdot \widetilde{D}^c_{0\to1}(x'_0, d_{0\to1}) + a_{21}\cdot \widetilde{D}^c_{2\to1}(x'_2, d_{2\to1}) \quad (4)$$

where x' identifies lossy mode of x and $\widetilde{D}^c_{m\to n}(x, d_{m\to n})$ represents perfect inverse disparity compensation. Noticeably, in lossless situation, $L_m = L'_m$ and $H_0 = H'_0$, the invertible ability of lifting scheme is preserved.

III. PROPOSED SCHEME

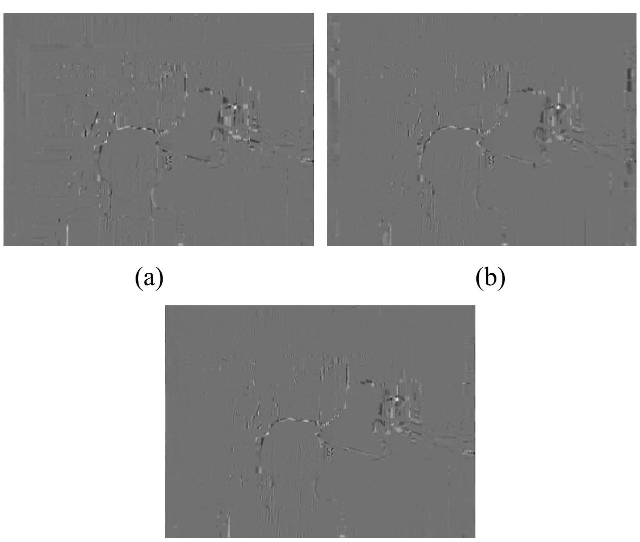
A. Hybrid Prediction

Adapting the lifting-based motion transform to multi-view image system so as to remove the redundancy among channel views is not conventional. Unlike the successive frame along a single-view video, the pictures simultaneously captured from multiple cameras are not too similar, since cameras are positioned at various angles to facilitate the 3D scene capture. Therefore, the choice of wavelets filters will be an important parameter in this approach. The longer filters such as the 9/7 biorthogonal wavelet transform are not suitable, since they may lead to poor prediction in occluded areas and also introduce numerous ghosting artifacts. Though, smaller filter such as 5/3 wavelet transform introduces fewer ghosting artifacts, it is still unable to cope with the occlusion absolutely. In contrast, much smaller filters such as the two tap Haar will be ideal for dealing with occlusion but are not suitable for achieving high compression. This presents a challenge in terms of an optimal wavelet for multi-view image/video compression. Hence, the hybrid wavelet lifting is proposed as one of the suitable approaches for multi-view image/video coding.

In the hybrid approach, the correlations among consecutive views are efficiently removed by 5/3 wavelet lifting scheme, while the occlusion regions are processed by Haar filters. The occlusion areas in one of the intermediate view will normally be uncovered/present in another side view. Therefore, instead of employing 5/3 transform for the whole image; the occluded areas would be better predicted by using Haar transform in the areas of the image in which they are visible. However, the process for detecting the occlusion has to be combined. To avoid the transmission overhead of additional bits, the local information is exploited to identify which filter is used for each block. However, such local information appearing at encoder and decoder has much opportunity to be different from each other because of lossy conditions, e.g. data stream truncation and channel loss. So, the filter used for a specific block at encoder and decoder may disagree that cause the worse predicting results. On the other hand, side information has to be transmitted or stored to achieve reliable occlusion detection at the encoder and decoder. This later approach is related to the bi-directional prediction in video standards, e.g. MPEG-2, H.264 and etc. where information utilized to indicate the prediction modes is included in bit stream. In the standards, the displacement between the original and predicted values (such as sum of absolute difference (SAD)) is used to determine the most suitable frame for prediction.

In our proposed scheme, the SAD cost after disparity compensation is employed to determine the optimal views for occlusion compensation. Following the block-based disparity estimation process, two best matching blocks for each block in the intermediate view are found, one by predicting from left view, P_{ij}^{L}, and one by predicting from right view, P_{ij}^{R}. Hence, for each block, the two disparity vectors and SAD are recorded. To get the displacement of 5/3 wavelet scheme, the new SAD cost is merely calculated by summing up the absolute values of the difference between the original image and the average of the existing predicted values, $P_{ij}^{5/3} = 0.5 \cdot (P_{ij}^{L} + P_{ij}^{R})$. It is noticeable that this approach does not require any additional complex computations. The minimum among the three SAD values identifies the type of filters to be used for each block, either Haar with left view, Haar with right view or 5/3 wavelet.

After selecting the optimum filter, therefore, a_{01} and a_{21} matrices do not contain only one specific value but depend



(a) (b) (c)

Figure 2. High-pass images of *Head* test image. (a) Haar lifting (normallized energy =48.89). (b) 5/3 wavelet lifting (normallized energy =36.37). (c) Hybrid lifting (normallized energy =35.76).

whether that block belongs to clear or occluding areas. In the other words, it indicates which filter is used for a given block. For instance, if the block in row i^{th} column j^{th} of intermediate view, x_1, can be seen in all views, then $a_{01}(i, j)$= $a_{21}(i, j)$= 0.5 since it is belong to 5/3 wavelet. Alternatively, if it is visible in only one view such as x_0, then $a_{01}(i, j)$=1, $a_{21}(i, j)$=0 and is filtered by Haar. The high-pass images from block-based disparity compensated Haar, 5/3 wavelet and hybrid lifting scheme with half-pel accuracy are shown in Fig.2. It can be seen that the residual image from hybrid prediction contains less energy than Haar or 5/3 filtering alone.

B. Adaptive Weighing in Update Step

In this section a modification to the update step is proposed. In the update step, the residual from disparity compensated prediction is added to the even images. It is possible to reduce noise and aliasing if and only if the disparity is accurately captured. The erroneous matching tends to introduce ghosting artifacts in low-pass signal. These artifacts appear as the repeat of object boundaries that contain high energy in the high-pass image. In other words, if a block in high-pass image has high energy, it implies that blocks around object boundary will lead to ghosting artifacts.

This visual problem also occurs in lifting based video coding schemes. Various adaptive algorithms for update step have been proposed by using the local information to avoid the extra coded bits [8],[9]. In the same location of a block, it is not strictly necessary to have the same factor in update step between the encoder and decoder, since it is not used in further prediction. In this paper, we use the normalized energy of the high-pass signal by assuming its relationship directly to ghosting artifacts in low-pass images, i.e. the coefficients of $a_{(1)(2m)}$ are given by normalized energy of $D^{c}_{1\to 2m}(H_0)$ at encoder and $\widetilde{D}^{c}_{1\to 2m}(H'_0, d_{1\to 2m})$ at decoder.

C. Overlapped block disparity compensation

Although, the block-based matching is straightforward to implement, one disparity vector representing the whole pixel in a block may not be sufficient to describe the actual disparity. This tends to produce blocking artifacts which are discontinuities introduced at block borders in the predicted image. They lead to annoying horizontal and vertical edges which are highly visible for the human eye, especially at low bit rates.

To overcome this problem, Overlapped Block Disparity Compensation (OBDC) is employed here to reduce blocking artifacts and disparity compensation error. Moreover, it has the advantages that no change in search range is required and no extra side bits are generated. This is achieved by linearly combining the predictions generated by using multiple disparity vectors, including a block's disparity vector and its neighbors. The OBDC is exactly adapted from the Overlapped Block Motion Compensation (OBMC) [10]. The OBDC could be applied to the predict step without obliterating the invertible property of lifting scheme.

IV. Experimental Results

In this section, simulation results from the proposed scheme described in section III are presented. The hybrid prediction is exploited to increase the accuracy in prediction step. The adaptive weighing factors in update step are identified by normalized energy of high-pass signal in that block, namely $a_{(1)(2m)}$ will be equal to 0.5, 0.25, 0.125 or 0 if such normalized energy is less than 0.25, 0.5, 0.75 and 1 respectively. The performance of this proposed invertible approach was compared to that of the Haar and 5/3 transform with block-based disparity compensation without the OBDC.

The simulations were conducted with two standard multi-view test images; Claude1 and Claude4. Although the sequences contain four views, three consecutive views were selected for investigating the performance of the proposed three-view image codec. Fig. 3 (a) and (b) illustrate the average luminance (Y) PSNR of the reconstructed Claude1 and Claude4 images respectively. The chrominance components (U,V) are compressed separately but share the desired target bits. The proposed scheme does not show much improvement in very low bit rates, but shows gains of to 1.3 dB and 1 dB over Haar and 5/3 wavelet respectively at higher rates. This is because, at low bit rate, most target bits are spent for coding side information, which includes the extra bits used for identifying the choice of filters in the hybrid prediction scheme, instead of being used for coding texture of image.

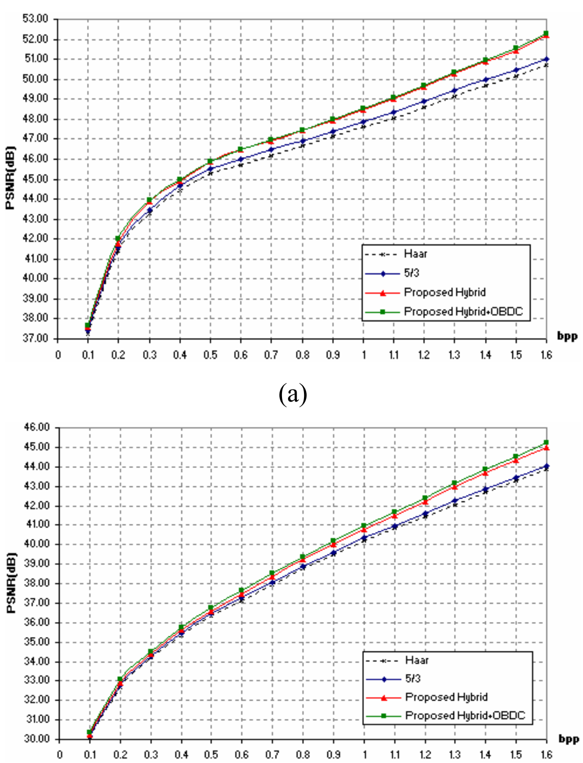


Figure 3. The results of proposed codec comparing to block-based disparity compensated 5/3 lifting. (a) *Claude1*. (b) *Claude4*.

Subsequently, the OBDC with weighting determined by a Raise Cosine window is implemented in the prediction step. The objective results are included in Fig. 3 (a) and (b). The subjective quality of reconstruction is significantly improved. However, OBDC does not affect much in Claude1, because it contains mostly uniform texture in the background.

V. Conclusions

A novel multi-view image coding based on lifting is proposed. By exploiting the wavelet lifting scheme with block-based disparity compensation, improved coding performance is reported. A hybrid prediction was proposed to deal with occlusions which are often present in multi-view images. We proposed a simple computationally efficient occlusion detection and view selection scheme based on the SAD criteria. Furthermore, an adaptive weighing in update step is included to reduce ghosting artifacts. An overlapped block disparity compensation approach is also proposed to reduce blocking artifacts. The results show that proposed approach can lead to significant improvements in both objective and subjective quality of the reconstructed image. It is clear that the lifting based approach has a number of advantages in multi-view coding including gain improvements of up to 1.5 dB and 1.2 dB over a conventional Haar and 5/3 transform block-matching codec respectively. In future, the codec will be extended to a scalable multi-view image and video system.

References

[1] M.G. Perkins, "Data Compression of Stereopairs," IEEE Trans. Communications, vol. 40, no. 4, pp. 684-696, 1992.

[2] I.Dinstein, G.Guy, J.Rabany, J.Tzelgov, A.Henik, "On Stereo Image Coding," IEEE Proc. ICPR, pp. 357-359, Nov 1988.

[3] J. Ohm, "Three dimensional subband coding with motion compensation," IEEE Trans. IP, vol.3, pp 559-571, Sep 1994.

[4] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional Lifting Schemes for Motion Compensated Video Compression," IEEE Proc. ICASSP'01, pp. 1793-1796, May 2001.

[5] L. Luo, J. Li, S. Li, Z. Zhuang and Y. Zhang, "Motion Compensated Lifting Wavelet and Its Application in Video Coding," IEEE Proc. ICME, pp. 481-484, Aug 2001.

[6] A. Secker and D. Taubman, "Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression," IEEE Trans. IP. vol. 12, no. 12, pp. 1530-1542, Dec 2003.

[7] W. Swelden, "The Lifting Scheme: A Construction of Second Generation Wavelets," SIAM Journal of Mathematical Analysis, vol. 29, no. 2, pp. 511-546, Mar 1998.

[8] G. Piella and H.J.A.M. Heijmans, "Adaptive Lifting Schemes With Perfect Reconstruction," IEEE Trans. Signal Processing, vol. 50, no. 7, pp.1620-1630, Jul 2002.

[9] N. Mehrsereht, D. Taubman, "Adaptively Weighted Update Steps in Motion Compensated Lifing Based Scalable Video Compression", IEEE proc. ICIP, vol.3, pp.771-774, Sept 2003.

[10] M.T. Orchard, G.J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," IEEE Trans. IP, vol. 3, no. 5, pp. 693 – 699, Sept 1994.