Anantrasirichai, N., Canagarajah, C. N., & Bull, D. R. (2005). Multi-view image coding with wavelet lifting and in-band disparity compensation. In IEEE International Conference on Image Processing 2005 (ICIP 2005) Genova, Italy. (Vol. III, pp. 33 - 36). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/ICIP.2005.1530321

# Multi-View Image Coding with Wavelet Lifting and In-Band Disparity Compensation

N. Anantrasirichai, C. Nishan Canagarajah and David R. Bull
Department of Electrical & Electronic Engineering
University of Bristol
Bristol UK

***Abstract*—This paper presents a novel framework to achieve scalable multi-view image coding. As open loop operation, the wavelet lifting scheme for geometric filtering has been exploited to overcome the limitation of SNR scalability and to attain view scalability. The essential key for achieving the spatial scalability is the in-band prediction. It removes correlations among subbands level-by-level via shift-invariant references obtained by Overcomplete Discrete Wavelet Transforms (ODWT). Additionally, the proposed disparity compensated view filtering is allowed to exploit the different filters and estimation parameters for each resolution level. The experiments show comparable results at full resolution and the significant improvement at coarser resolution over the conventional spatial prediction scheme.**

***Keywords-component; multi-view, disparity compensation, in-band prediction, ODWT, scalability***

## I. INTRODUCTION

Nowadays, the visual technologies have been dramatically developed, especially three-dimensional (3D) image system which pertains to many applications such as remote surveillance, medical imaging, telerobotics, entertainment and virtual reality. Basically, 3D images/videos are constructed from a series of multi-view sequences by capturing a target scene via multi-cameras which are located at different points of view. This system results in a manifold increase in bandwidth over existing monoscopic channel or requires much more space to store the whole raw data. Fortunately, the clear context of one view appearing as shift position from other views, which is known as disparity, causes redundant data in sequence. This geometric correlation could be removed by efficient compression algorithms. For example, in [1], the furthest right view is independently coded and the rest are predicted by using this reconstructed view. Due to the finite viewing and occlusion area, this method may yield poor performance over the independent coding. The other proposed method is to construct the intermediate view by utilizing the disparity information from left and right views [2]. The disparity estimation errors in this method could be exacerbated after interpolation. In this paper, we proposed an efficient algorithm to deal with occlusions which yields better performance.

Not only is high compression required in multi-view image system, but various highly desirable features are also compulsory to support the heterogeneous instruments and networks. That is, multi-view image coding should provide scalabilities as provided in advance video coding. The scalable image/video encoding allows specific subsets of the bitstream with different transport and presentation properties. Likewise, the bitstream from a flexible multi-view image codec should be embedded with SNR scalability and spatial or resolution scalability. Thus, the efficient compression algorithms with scalable properties are vital to reduce data size without sacrificing the perceived quality and to support currently changeable circumstances. Recently, the discrete wavelet transform (DWT) has become an attractive choice to achieve such requirements.

In temporal axis, the novel method called Motion Compensated Temporal Filtering (MCTF) has been introduced [3]. It is very satisfying because of its inherent provision of the spatial-temporal-SNR scalability. Based on the wavelet lifting scheme, it can simply be achieved by splitting successive image frames into two sequences and then use one sequence to predict another. The predicting sequence is updated by combining the residuals which are derived from the prediction step. Clearly, the update step introduces a motion compensated temporal low-pass filter, which is completely beyond conventional compression concept [4]. It contains the flexibility in the number of decomposition levels and the choice of filters. Moreover, it operates as open-loop prediction, so temporal drift problem could be resolved and error resilience is improved. Thus, in the same way, the disparity compensated view filtering (DCVF) is proposed here for the multi-view image compression.

However, exploiting wavelet transform after motion compensation limits the decoding efficiency at smaller spatial resolution, because spatial drift problems. To solve this, the direct prediction in low-low subband of each resolution level is the simplest approach [5]. Unfortunately, it is affected by the frequency aliasing due to down-sampling and, moreover, it generates more motion information over the spatial prediction. Consequently, in-band motion estimation has been investigated by predicting subband-to-subband, but the shift variance as a result of the decimation process in the wavelet transform frequently causes the inefficient estimation in high-pass subbands. The solution to this problem has been firstly introduced by using Low-Band-Shift method [6]. This approach generates all possible subbands of reference frames by shifting one pixel in horizontal, vertical and diagonal directions. The more advance techniques avoid performing the inverse discrete wavelet transform (IDWT). They calculate the

0-7803-9134-9/05/$20.00 ©2005 IEEE

shift invariance of critically sampled DWT by appropriate prediction filters [7], [8].

In this paper, we present a novel compression framework to achieve scalable multi-view image coding. The difficulty of multi-view prediction from occlusion is dealt by the hybrid disparity prediction which has been proposed in [9]. The scalabilities are achieved by operating the predict step in wavelet domain.

The rest of paper is organized as follows: Section II briefly explains the fundamentals of the in-band disparity estimation and compensation (DE/DC) and DCVF. The proposed codec is described in Section III. The experimental results are presented in Section IV followed by conclusions and future works in Section V.

## II. In-Band Disparity Compensated View Filtering

The proposed in-band DCVF scheme has been developed from disparity compensated lifting scheme, but the predict step is performed after the spatial transformation. First of all, each view of multi-view images is iteratively decomposed into various subbands. By operating independently in each level, multi-view images are split into two groups; odd-order views, $x_{2k+1}$, and even-order views, $x_{2k}$. The next step is the predict step which is to let one element represent the whole data, therefore the other element has to be predicted. The high-pass signal, $H_k$, the residual of subtracting the predicted element from the original element, will contain very little energy thereby achieving significant compression. Finally an update step combines such residual data to reduce the effect of aliasing in low-pass signal, $L_k$. In a wavelet tree, $H_k$ should be encoded more coarsely and might be dropped at a receiver if bit rate is constrained. In this manner, the output bitstream of the proposed encoder achieves scalability.

Obviously, the predict step operates in wavelet domain; however, to predict high-pass subband with high-pass subbands is a difficult problem. Unlike predicting the low-pass subband, after subsampling the estimation of high-pass subbands encounters a big difference between the original signal and the reference signal. However, the predicted high-pass subbands could be achieved through the perfect estimation if the subsampling process is omitted after filtering [10], called Overcomplete Discrete Wavelet Transforms (ODWT). This approach can be generated locally within the decoder and it is only necessary to encode only through critically sampled coefficients, i.e. the image size is not enlarged. Currently, a few research groups have been concerned about ODWT construction [6], [7], [8], [10], and only the prediction filters method has absolutely achieved spatial scalability [7],[8]. Methods in [6] and [10] require the inverse quantization and IDWT before generating the ODWT. These schemes are limited when the resolution scalability is required, because the high-frequency subbands of finer resolution levels may be truncated at the decoder. That is, the results of inversion are merely the estimation. Conversely, the prediction filters method generates all other possible high-pass subbands (non zero phase) from unshift (zero phase) low-pass and high-pass subbands without inverting. By level-by-level calculation with prediction filters $\mathbf{F}$, one-dimension complete-to-ODWT (CODWT) for level $k$ can be written as follows [7].

$$\mathbf{w}^k_{\mathrm{ODWT}} = \begin{bmatrix} \mathbf{w}^k_1(z) \\ \mathbf{w}^k_2(z) \\ \vdots \\ \mathbf{w}^k_{2^k-1}(z) \end{bmatrix} = \begin{bmatrix} \mathbf{P}^1(z) \\ \mathbf{P}^2(z) \\ \vdots \\ \mathbf{P}^k(z) \end{bmatrix} \mathbf{w}^k_0(z) \qquad (1)$$

where $\mathbf{w}^b_a(z) = \left[L^b_a(z)\ H^b_a(z)\right]^{\mathrm{T}}$ with polyphase $a$ at resolution level $b$.

$$\mathbf{P}^{u+1}(z) = \begin{bmatrix} \mathbf{F}^{u+1}_0(z) \\ \mathbf{F}^{u+1}_1(z) \\ \vdots \\ \mathbf{F}^{u+1}_{2^u-1}(z) \end{bmatrix} \text{ for every u, } 0 = \mathrm{u} < \mathrm{k}. \qquad (2)$$

and the prediction filters are recursively calculated as follows.

$$\mathbf{F}^{u+1}_p = F^u_{4\lfloor p/2 \rfloor, b_0}(z)\,\mathbf{I} + z^{-(1-b_0)} F^u_{4\lfloor p/2 \rfloor, 1-b_0}(z)\,\mathbf{F}^1_0(z) \qquad (3)$$

$$\mathbf{F}^1_0(z) = \begin{bmatrix} z\widetilde{H}_0(z)\widetilde{G}_0(z) - \widetilde{H}_1(z)\widetilde{G}_1(z) & \widetilde{H}_1(z)\widetilde{H}_1(z) - z\widetilde{H}_0(z)\widetilde{H}_0(z) \\ z\widetilde{G}_0(z)\widetilde{G}_0(z) - \widetilde{G}_1(z)\widetilde{G}_1(z) & \widetilde{H}_1(z)\widetilde{G}_1(z) - z\widetilde{H}_0(z)\widetilde{G}_0(z) \end{bmatrix} \qquad (4)$$

where $p = \sum_{j=0}^{l-1} b_j 2^j$, $b_j = \{0, 1\}$ and $\widetilde{H}$, $\widetilde{G}$ are the low- and high-pass analysis filters respectively with polyphase 0 = even and 1 = odd.

After the construction of the ODWT, each resolution level $u$ will be combined with the critically-sampled subbands $LH^u_{(i,j)}, HL^u_{(i,j)}, HH^u_{(i,j)}$, $0 \le i, j < 2^u$ and $LL$ at coarsest resolution level. The full search of the disparity estimation can be run in a level-by-level fashion. The best matching is found by jointly minimizing the distortion for each triplet of blocks from the $LH^u_{(0,0)}$, $HL^u_{(0,0)}, HH^u_{(0,0)}$ of the current view that correspond to the same spatial domain location of a triplet of blocks from $LH^u_{(i,j)}$, $HL^u_{(i,j)}$, $HH^u_{(i,j)}$ of the reference views. For the coarsest resolution level (containing only $LL$ subband), the estimation process is done separately.

Clearly, the proposed DCVF is allowed to exploit the different filter and estimation parameters for each resolution level. So, the in-band prediction creates the degree of freedom for compact scalable multi-view image compression across spatial resolution.

## III. Proposed Scheme

Our proposed multi-view image encoder is composed of in-band DE/DC, as well as geometric wavelet lifting scheme. As illustrated in Fig. 1, the block-based three-view image encoder starts the operation by performing spatial wavelet transform with total $R$ levels to each view and gaining subbands $s$, where $s = \{\mathcal{L}^R, \mathcal{H}^u\}$, $\mathcal{H} = \{HL, LH, HH\}$, $\mathcal{L} = \{LL\}$ and $1 \le u \le R$. Then, the transformed images, $X^s_v$, $v$ ={$l$ : left, $m$ : intermediate, $r$ : right view}, are filtered along the geometric direction. As explained in Section 2, three high-pass subbands of each resolution level are predicted together and the same disparity vectors (DV) are generated, while the low-pass subband which belongs to the coarsest resolution level is

predicted separately. The DE/DC enhances the predict step in lifting scheme by employing both ODWT coefficients $X_l^{os} = \mathbf{w}^u_{\mathrm{ODWT},l}$ of left reference images, and $X_r^{os} = \mathbf{w}^u_{\mathrm{ODWT},r}$ of right reference images, to predict the intermediate view $X_m^s$. It produces the predicted subband, $P_{m-l}^s$ and disparity vectors $d_{m-l}^s$ when predicting from left view, and $P_{m-r}^s$, $d_{m-r}^s$ when predicting from right view. The block size of the coarser resolution level will be smaller than that of the successive finer resolution level by factor of two. The results are better prediction without excess DVs compared to the conventional spatial estimation.

As mentioned in our previous paper [9], adapting the MCTF to multi-view image system is not conventional. The pictures captured from multiple cameras usually contain occlusions in the background that is not comfortable to predict. Thus, the choice of wavelet filters is an important parameter in the disparity compensated view filtering. We have proposed to remove the correlations among consecutive views by 5/3 wavelet lifting scheme, while the occlusion areas are processed by Haar filters. The filter selector is applied to identify which filter is exploited to a specific block and generate the coefficients $a_{m-l}^s$ and $a_{m-r}^s$ depending upon the chosen filter used for the predicted views $P_{m-l}^s$, and $P_{m-r}^s$, respectively. However, because of non ideal filters and disparity failure, the coding of the residual by the DE and the corresponding DV data is not necessarily superior to the coding of the wavelet coefficients themselves. This mainly applies to the subbands of high frequency whose coefficients are smaller compared to those of the low-pass subband. Hence, the band selector is applied to choose between the predicted high-pass and the original high-pass subbands. If the instant original high-pass subband contains less energy, $B_m^s$ will be a zero matrix; otherwise, it would be equal to $P_m^s$.

Subsequently, the left view and the right view are updated to obtain the low-pass images, $L_l^s$ and $L_r^s$ respectively. The function of the update block is to combine the results of DE/DC that are estimated with information of intermediate view. This consequence should be the inverse direction of $d_{m-l}^s$ for updating left view or $d_{m-r}^s$ for updating right view. Owing to occlusion, the best matching in the update step is not equal to the inversion of such DVs. Unfortunately, the erroneous matching tends to introduce ghosting artifacts in low-pass signal. Therefore, the adaptive weighing scheme is exploited to suppress the coefficients in a block that contains high energy [11], i.e. the coefficients of $a_{l-m}^s$ and $a_{r-m}^s$ are given by normalized energy of $D_{v\to m}^c(H_m^s)$, where $D_{v\to m}^c(x)$ represents the disparity compensated function predicting for data $x$ contained in view $v$ by utilizing the information from image view $m$.

Further advance technique is included, i.e. deblocking filter in DC. One characteristic of block-based coding is the production of visible block structure on account of discontinuity at block boundary. It increases the high frequency contained in high-pass subbands that seriously harms the context coding performance. The appropriate deblocking filter should reduce blocking artifacts while maintaining the sharpness. Thus, the algorithm that the filter weighs all pixels in a block is not much suitable for employing in high-pass subbands, which contain the edge information of image (e.g. overlapped block compensation).

To preserve image sharpness, the true edges should be left unfiltered as much as possible while filtering artificial edges to reduce their visibility. In our proposed codec, the Adaptive deblocking filter [12] is applied only in high-pass subbands. The basic idea is that if a relatively large absolute difference between samples near a block edge is measured, it is quite likely a blocking artefact and should therefore be reduced. However, if the magnitude of that difference is so large, the edge is more likely to reflect the actual behaviour of the source picture and it should not be smoothed over.

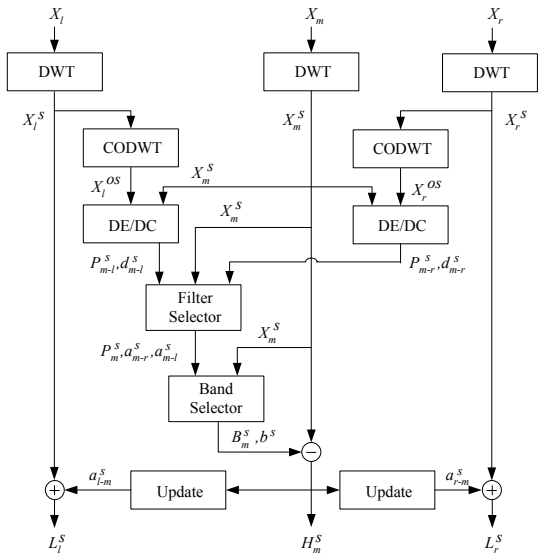
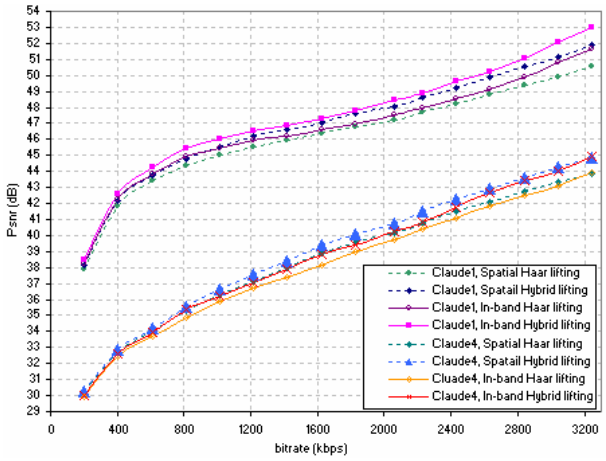



Figure 1. The forward in-band DCVF of thee-view image system

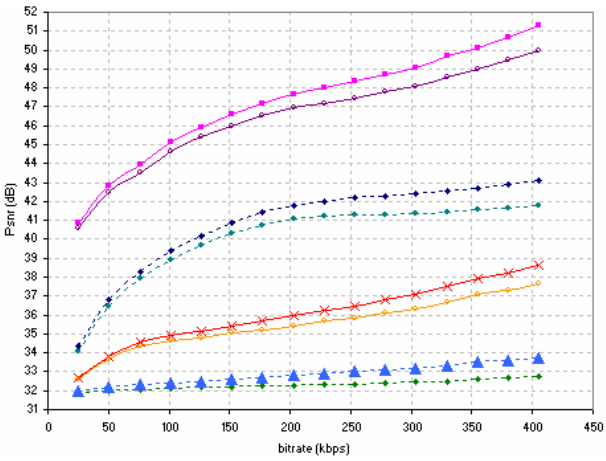
## IV. Experimental Results

In this section, the experimental results from the proposed scheme described in section 3 are presented. The simulations were conducted with two standard multi-view test images: Claude1 and Claude4. Both pictures are YUV data in CIF format. Although the sequences contain four views, three consecutive views were selected for investigating the performance of the proposed three-view image codec. The DWT was implemented with biorthogonal 9/7 filter up to 3 decomposition levels in luminance component and up to 2 decomposition levels in chrominance components. Block sizes of 16x16, 8x8 and 4x4 are chosen for in-band estimating in level 1, 2 and 3 respectively, while the size of 16x16 is used for spatial estimation. The outputs of geometric lifting schemes were encoded by EBCOT coding (Embedded Block Coding with Optimized Truncation). The chrominance components (U,V) were compressed separately but shared the desired target bits. Figure 2 (a) and (b) illustrate the average luminance (Y) PSNR of the reconstructed images in full resolution and quad-resolution respectively. Experimentally, the occlusion problem is dealt with the hybrid lifting scheme and the performance gains up to 1-1.3 dB compared to the conventional Haar lifting

scheme in both predicting in spatial domain and wavelet domain. The image quality at coarser resolution has been much improved when applying the in-band DE/DC, especially in high bit rates. However, at full-resolution, if an image contains many details that cause high energy in high-pass subbands, the objective results of in-band prediction may not be better than that of conventional spatial prediction.




(a)




(b)

Figure 2. The results of proposed codec comparing between predicting in spatial domain and wavelet domain which were exploited with Haar lifting and Hybrid lifting schemes (a) full-resolution. (b) 1/4-resolution.

## V. CONCLUSIONS

A novel multi-view image coding based on wavelet lifting scheme with in-band disparity compensation is proposed in this paper. The hybrid prediction is exploited so as to deal with the occlusion problem. This prediction is operated in wavelet domain, thus the resolution scalability is achieved efficiently. The SNR scalability is embedded in EBCOT coding. As a result, the proposed codec is scalable for high proficient multi-view image compression. Our experiments show comparable results at full resolution and much improvement at coarser resolution.

## REFERENCES


[1] M. E. Lukacs, "Predictive Coding of Multi-Viewpoint Image Sets," in ICASSP, pp. 521-524, 1986.

[2] J. Liu and R. Skerjanc, "Construction of Intermediate Pictures for a Multiview 3D System," in Stereoscopic Displays and Applications III, vol. 1669, pp.10-19, SPIE, 1992.

[3] V. Bottreau, M. Benetiere, B. Felts, B. Pesquet-Popescu, "A Fully Scalable 3D Subband Video Codec," Proc. 2001 International Conference on Image Processing , Vol 2 , pp. 1017 – 1020, 7-10 Oct. 2001.

[4] 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.

[5] A. Nosratinia and M. Orchard, "A multi-resolution framework for backward motion compensation," VCIP'1995, pp.190-200, San Jose, CA, February 1995.

[6] P. Hyun-Wook and K. Hyung-Sun, "Motion Estimation Using Low-Band-Shift Method for Wavelet-Based Moving-Picture Coding," IEEE Trans. on IP, vol.9, no.4, pp. 577-587, April 2000.

[7] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. Schaar, J. Cornelis and P. Schelkens, "In-Band Motion Compensated Temporal Filtering," Signal Processing: Image Communication 19, pp. 653 – 673, 2004.

[8] G. Van der Auwera, A. Munteanu, P. Schelkens and J. Cornelis, "Bottom-up motion compensated prediction in the wavelet domain for spatially-scalable video", IEE Electronics Letters, vol. 38, no. 21, pp. 1251-1253, Oct. 2002.

[9] N. Anantrasitichai, C. Nishan Canagarajah, and David R. Bull, "Lifting-Based Multi-View Image Coding," accepted to ISCAS'2005.

[10] C. Mayer and S. Albert, "Spatially Scalable Coding with In-band Prediction," SPIE Image and Video Communications and Processing, Santa Clara, CA, USA, Jan 2003.

[11] 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.

[12] P. List, A. Joch, J. Lainema, G. Bjøntegaard, and M. Karczewicz, "Adaptive deblocking filter," IEEE Trans. Circuits Syst. Video Technol., vol. 13, pp. 614–619, July 2003.