
Publisher's PDF, also known as Version of record

Link to published version (if available): 10.1109/TCSVT.2009.2031371

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
In-Band Disparity Compensation for Multiview Image Compression and View Synthesis

Nanttheera Anantrasirichai, Member, IEEE, Cedric Nishan Canagarajah, Member, IEEE, David W. Redmill, and David R. Bull, Senior Member, IEEE

Abstract—This paper presents a novel framework to achieve scalable multiview image compression and view synthesis. The open-loop wavelet-lifting scheme for geometric filtering has been exploited to achieve signal-to-noise ratio scalability and view-type scalability (mono, stereo, or multiview). Spatial scalability is achieved by employing in-band prediction which removes correlations among subbands (level-by-level) via shift-invariant references obtained by overcomplete discrete wavelet transforms. We propose a novel in-band disparity compensated view filtering approach, akin to motion compensated temporal filtering, for achieving a scalable multiview codec. In our codec, hybrid prediction is proposed to deal with occlusions, and a novel cost function in dynamic programming (DP) for disparity estimation is introduced to improve view synthesis quality. Experiments show comparable results at full resolution and significant improvements at coarser resolutions, compared to a conventional spatial prediction scheme. View synthesis efficiency is extensively improved by utilizing disparity estimation from the proposed DP approach.

Index Terms—Disparity compensation, dynamic programming, multiview image, occlusion, overcomplete discrete wavelet transform, scalability.

I. INTRODUCTION

VISUAL TECHNOLOGIES, especially 3-D imaging systems, have found many applications in remote surveillance, medical imaging, telerobotics, entertainment and virtual reality. 3-D images/videos are typically constructed from a series of multiview sequences captured from multiple cameras located at different positions. This results in very high raw data rates, or storage space compared to monoscopic systems. Fortunately, the clear context of one view appearing as shift position from other views, which is known as disparity, leads to significant redundancies in such sequences. This geometric correlation can be removed by efficient prediction and compression algorithms. For example, in [1], the furthest right view is independently coded and the remaining views are predicted using this reconstructed view. Due to the finite viewing and occlusion area, this method may perform poorly compared to the independent coding. Moreover, propagation of transmission errors can result from the recursive decoder structure exploited in the first view coding. Another approach [2] constructs intermediate views by utilizing the disparity information from the left and the right views. However, the errors from disparity estimation can be exacerbated after the interpolation. In this paper, we propose an efficient hybrid prediction algorithm to deal with the occlusions and yield better performance.

Multiview systems require both high compression and a variety of other features to support heterogeneous systems and networks. Multiview image coding should provide scalability mechanisms similar to those provided by scalable video codecs. Scalable image/video encoding allows extraction of specific bitstream subsets with different transport and presentation properties. Recently, the discrete wavelet transform (DWT) has become an attractive choice to meet these requirements. In this paper, we extend the wavelet scheme for image/video coding to remove spatial, temporal and geometric redundancies present in multiview sequences. To exploit temporal information, a method based on motion compensated temporal filtering (MCTF) has been introduced [3]–[6], which inherently facilitates the spatial-temporal-signal-to-noise ratio (SNR) scalability. However, these systems typically ignore the possibility of occlusion which is common place in multiview image system. This problem can be solved with our proposed hybrid prediction scheme [7] and enhanced disparity estimation using dynamic programming [8].

Based on the wavelet lifting scheme, MCTF contains flexibility in the number of decomposition levels and the choice of filters. Moreover, it operates as open-loop prediction, so temporal drift problem is eliminated and error resilience is improved. However, reconstructed images for the limited display size still suffer from the drift problem. Direct prediction at the smaller size could eliminate this difficulty by using in-band motion estimation which predicts subband-to-subband [9]. However, the shift variance as a result of the decimation process in the wavelet transform frequently causes the inefficient estimation in high-pass subbands. This problem can be avoided by omitting the subsampling process after filtering leading to an overcomplete discrete wavelet transforms (ODWT) [10]–[13]. In this paper, a novel in-band disparity compensated...
view filtering (I-DCVF) is proposed for multiview image compression. The proposed in-band prediction can remove the redundancy between similar subbands in different views. It improves the decoding efficiency at smaller spatial resolutions. Moreover, blocking artifacts are eliminated by using the in-band prediction, since the boundaries are filtered out during inverse DWT [14].

View synthesis is a key requirement in multiview systems. This feature needs reliable disparity information, hence disparity estimation becomes an important element for 3-D visualization. Due to occlusions, imperfect camera calibrations, and imperfect light balance, accurate disparity information is hard to achieve. An acceptable compromise is to ensure viewers feel the different depth of each object and also feel comfortable. In this paper, we use the dynamic programming (DP) to estimate the global optimum disparity/depth fields.

The DP for estimating the disparity between stereo images was first introduced in [15] and it shows better performance over traditional matching schemes. Moreover, the DP can deal with the proper disparity estimation and the occlusion detection simultaneously. A novel cost function proposed in [8] is used for searching the minimum-cost path, and then the performance of view synthesis at numerous spatial resolutions is investigated.

In this paper, we present a compression/synthesis framework to achieve scalable multiview image coding with the following novel contributions: 1) a hybrid prediction scheme to deal with the presence of occlusions; 2) a novel cost function in the DP for disparity estimation in order to obtain depth and occlusion information simultaneously thereby improving view synthesis; 3) an I-DCVF to achieve spatial, SNR and view-type scalabilities (mono, stereo, or multiview); and 4) a combination of disparity map and low-pass subband to reach an acceptable compromise between compression and view synthesis performances.

The rest of the paper is organized as follows. Section II briefly explains the fundamentals of the wavelet lifting scheme and its adaptation to multiview environments. The proposed codec is described in Section III. Subsequently, in Section IV, the coding system combines the disparity map obtained by using the DP in order to improve view synthesis. The experimental results are presented in Section V followed by conclusions and future work in Section VI.

## II. Geometric Lifting Scheme

The lifting scheme introduced by Sweldens [16], [17] is a technique for constructing or factoring wavelet filters into basic building blocks. A related structure called ladder networks was proposed earlier by Bruekers [18]. Basically, the forward wavelet lifting method decomposes the wavelet transforms into a set of stages. The operation starts with a split step, which divides the data set into groups, normally composed of odd and even samples. The next step is the prediction where one group is used to predict the other group. Then, the high-pass residual signal, $H_r$, generated by subtracting the predicted element from the original element, will contain very little energy thereby achieving significant compression. Subsequently, an update step combines residual data from the previous process to reduce the effect of aliasing in low-pass signal, $L_0$. Obviously, in a wavelet tree, $H_r$ should be encoded more coarsely and can be dropped at a receiver if bit rate is constrained. In this manner, the output bitstream of the proposed encoder achieves (SNR, temporal) scalability.

### A. MCTF to DCVF

In video coding, the wavelet transform is applied to filter data in the intra-frame and also along the temporal direction, called 3-D discrete wavelet transform. The performance of prediction and update steps in the temporal direction can be improved by utilizing motion compensation leading to the MCTF algorithm. Ohm has proposed 2-tap Haar filter for MCTF by using the block-based motion compensation and showed that a non invertible transform will affect coding gain if there are numerous disconnected pixels between blocks [19].

In [4], a nonlinear lifting framework for temporal wavelet transform has been introduced which is shown to be superior to the conventional temporal filtering. Luo et al. have employed a bidirectional motion estimation in each lifting step of temporal direction [5], while the longer filters have been introduced with a higher coding gain for video compression by Golwelkar and Woods [20]. Secker and Taubman [6] have developed the highly scalable video coding approach. It is based on the lifting with adaptive motion compensation. However, they have ignored the possibility of occlusion which is a common problem in multiview imaging systems.

The use of lifting scheme in view direction without advanced prediction scheme will result in the ghosting artifacts on account of uncompensated shift of corresponding data between views. Furthermore, the high-pass subband will contain considerable energy thereby compromising compression efficiency. In the context of multiview image system, disparity estimation can be usefully applied to the lifting algorithm. Obviously, the view-type scalability is supported by this geometric wavelet transform [7], [21]–[23]. A number of views are selectively determined at the receiver according to the type of display modes, mono, stereo, or multiviews. Then, only the selected views are decoded and displayed.

Errors in disparity information come from two sources; the presence of noise in images and the invisibility of corresponding points in reference images. The latter case is dependent on the location of cameras. Three or more cameras are often able to compensate such areas and the disparity compensated function is then modified to include the known occlusion information. The forward transform of the DCVF are displayed in Fig. 1. The disparity compensation and the inverse disparity compensation are utilized to improve the performance of prediction and update step respectively. The occlusion detection is also used to reduce the disparity mismatch error. In Section IV, we propose disparity estimation based on the DP that simultaneously detects occlusions.

### B. Types of Wavelet Filter for Multiview Imaging

Adapting the wavelet lifting scheme to multiview differs from MCTF with single-view video. The pictures simultaneously captured from multiple cameras are usually less similar.
Therefore, the choice of wavelets filters will be an important parameter in geometric view filtering.

We have investigated three filters: Haar (short-length filter), 5/3 wavelet (medium-length filter) and 9/7 wavelet (long-length filter), for two types of multiview sequences: narrow-baseline and wide-baseline geometries. The simulations use conventional block-based disparity estimation with 8 block size to multiview test sequences, Head (five views with size 384 × 288) and Santa (nine views with size 320 × 240). The 5-view Head sequence is extended symmetrically at the furthest left and right view until it contains totally nine views when the 9/7 wavelet transform is applied. For the wide-baseline geometry, only odd views or nonconsecutive views are used. This increases the presence of occlusions. The performance of each filter type is illustrated in Fig. 2. The 9/7 wavelet gives significantly lower performance for multiview sequences with either narrow-baseline or wide-baseline geometry. It utilizes information from nine views that generates a series of ghosting artifacts around object boundaries. Moreover, the disparity estimation error could be exacerbated after many steps of lifting. Although the 5/3 wavelet outperform other filters, the results still contain errors around occlusion areas which are not visible in both reference views. Better results in such occlusion areas are predicted with only the visible view reference. For the wide-baseline test which presents more occlusions, the Haar filter yields performance close to that of 5/3 wavelet. This suggests that the Haar filter deals well with the occlusion problems.

III. MULTIVIEW IMAGE CODEC WITH IN-BAND DISPARITY COMPENSATED VIEW FILTERING

The proposed I-DCVF scheme has been developed from the disparity compensated lifting scheme, with the prediction step is performed after the spatial transformation. The direct subband-to-subband prediction makes disparity estimation at coarse resolution more precise. Additionally, the proposed hybrid prediction scheme enhances the disparity estimation, since it can compensate for the effect of the occlusion [24].

Our proposed geometric wavelet lifting scheme with in-band hybrid prediction is illustrated in Fig. 3. The encoder starts the operation by performing spatial wavelet transform up to R levels to each view and gaining subbands, where s = \{L\, L\, \hat{H}\}, \hat{H} = \{H\_L, H\_H, H\_H\}, L = \{L\_L\} and 1 ≤ u ≤ R. Then, the transformed images, X_u, are filtered along the geometric direction. The wavelet coefficients of the reference views are converted to ODWT coefficients by using the complete-to-overcomplete discrete wavelet transform (CODWT) method [25], [26]. The disparity estimation/disparity compensation (DE/DC) enhances the prediction step in the lifting scheme by employing both ODWT coefficients X_2k+2, v = \{\_L, \_H\} of (2k+2)th reference images, and X_2k+1, v = \{\_L, \_H\} of (2k+1)th reference images, to predict the intermediate view X_2k+1. It produces the predicted subband, P_v(2k+1)→2k+2, and disparity vectors d_{(2k+1)→2k+2}. The filter selector is applied to provide the hybrid prediction on a block wise basis to generate the coefficients d_{(2k+1)→2k+2} depending upon the chosen filter \(P_{v(2k+1)}\). However, because of non ideal filters and disparity failure, the coding of both the residual and the corresponding digital video (DV) is not necessarily superior to coding of the wavelet coefficients themselves. This mainly applies to the subbands of high frequency, since the residual results of high-pass prediction sometimes have higher energy than the original high-pass subbands. Hence, the band selector is applied to choose between the predicted high-pass and the original high-pass subbands. If the original high-pass subband contains less energy, \(B_{(2k+1)}\) will be a zero matrix, otherwise, it would be equal to \(P_{(2k+1)}\).

The even views are updated subsequently to obtain the low-pass images L_\{2k\}. 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 $d_{l+1}^{k_+1}$ for updating left view or $d_{l+1}^{k_+2}$ for updating right view. If there is occlusion, the best match in the update step is not equal to the inversion of such DVs. Unfortunately, the erroneous matching tends to be the inverse direction of $d_{l+1}^{k_+1}$ for updating left view or $d_{l+1}^{k_+2}$ for updating right view. If there is occlusion, the best match in the update step is not equal to the inversion of such DVs. Unfortunately, the erroneous matching tends to introduce ghosting artifacts in the low-pass signal. Therefore, an adaptive weighing scheme is exploited to suppress the coefficients in a block that contains high energy [27], [28]. The subbands of the coarsest resolution ($L^k$, $H^k$) are further decomposed, and then coded by using the embedded block coding with optimized truncation (EBCOT) [29].

The decoder operates as a mirror of the encoder. By subband-to-subband, the odd views are decoded by subtracting the high-pass signal. Then, the ODWT and the prediction coefficients of the selected bands and filters are generated. Using disparity vectors, the predicted signal of the even views is reconstructed via the inverse DE/DC process. The high-pass signal is then added to generate the even views. Ultimately, all reconstructed subbands are combined together and using the inverse DWT.

A. In-Band Disparity Compensated View Filtering

In the I-DCVF process, each view of multiview images is iteratively spatially decomposed into various subbands. Then, the geometric wavelet lifting scheme with the disparity estimation and compensation is applied independently to each subband.

After constructing ODWT, each resolution level $n$ will be combined with the critically sampled subbands $L_{k_+1}^n$, $H_{k_+1}^n$, $H_{k_+2}^n$, $0 \leq i, j < 2^n$ and $L_L$ at coarsest resolution level. The full search of the disparity estimation process is done separately. The block size of the coarsest resolution level will be smaller than that of the successive finer resolution level by factor of two. This achieves better prediction compared to the conventional spatial estimation and requires equal or less DV.

The proposed DCVF provides a compact scalable multiview image compression across multiple spatial resolutions. Users are allowed to exploit different filter and estimation parameters for each resolution level. Therefore, the disparity maps can be exploited in order to improve view synthesis, which will be discussed in Section IV.

B. Hybrid Prediction Scheme

From Section II-B, we know that the longer filters are not suitable for multiview system, since they may lead to poor prediction in occluded areas and also introduce numerous ghosting artifacts. Although, smaller filter gives fewer ghosting artifacts, it is still unable to cope with the occlusion. In contrast, much smaller filters such as the two tap Haar will be ideal for dealing with occlusion but they are not suitable for achieving high compression [20]. This presents a challenge in terms of an optimal wavelet for the multiview image/video coding. Hence, hybrid wavelet lifting is proposed here as a suitable approach for multiview image/video coding.

In the hybrid approach, 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 views 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. However, the process for detecting the occlusion has to be defined and side information has to be transmitted or stored to achieve reliable occlusion detection at the encoder and decoder.

The mean square error (MSE) results of three standard test multiview images are tabulated in Table I to demonstrate the results of the hybrid lifting scheme compared to the traditional Haar and 5/3 wavelet transforms by applying the 8 × 8 block-based disparity estimation. The sum-of-absolute-difference (SAD) cost after disparity compensation is employed to determine the optimal views for occlusion compensation. Note that using the SAD criteria basically gives the good compression in terms of objective results with low complexity. Following the block-based disparity estimation process, two best matching blocks for block $(i, j)$ in the odd view are found, one by forward predicting from $(2k_i + 1)\text{th even view}$, $P_{2k_i+1}^{i,j}$, and another one by backward predicting from $(2k_i+2)\text{th odd view}$, $P_{2k_i+2}^{i,j}$. 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. The minimum among the three SAD values identifies the type of filters to be used for each block, either Haar with $(2k_i)\text{th}$ view, Haar with $(2k_i+1)\text{th}$ view or $5/3$ wavelet. Note that the MSE can be used as criterion to find the best matching, but it produces higher complexity [30].
TABLE I

<table>
<thead>
<tr>
<th>Sequence</th>
<th>MSe of geometric high-pass images</th>
<th>Hybrid (SAD)</th>
<th>Hybrid (occlusion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santa</td>
<td>11.56</td>
<td>9.31</td>
<td>7.19</td>
</tr>
<tr>
<td>Head</td>
<td>18.14</td>
<td>16.72</td>
<td>10.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence</th>
<th>MSe of disparity maps</th>
<th>Hybrid (SAD)</th>
<th>Hybrid (occlusion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santa</td>
<td>7.32</td>
<td>5.54</td>
<td>5.14</td>
</tr>
<tr>
<td>Head</td>
<td>4.24</td>
<td>3.14</td>
<td>3.19</td>
</tr>
</tbody>
</table>
The synthesized intermediate views. Correct disparity, thereby generating texture and position errors information. Therefore, conventional disparity estimation, e.g., objects needs to be stored or transmitted with the texture construction, disparity or depth information of the interested information. (c) In-band prediction.

Fig. 8. Enlarge images at 40 kb/view. (a) Original image. (b) Spatial prediction. (c) Block-based 3-level in-band prediction (PSNR $Y_{12} = 38.54$). (d) DP 3-level in-band prediction (PSNR $Y_{12} = 37.82$).

A. Disparity Estimation by Dynamic Programming

The matching pixels along each scanline are searched via the DP with a novel cost functions proposed by the authors of [8]. The key idea comes from the three possible disparity values of each particular pixel, which are equal to, more or less than that of the consecutive pixel. The first case usually occurs in the no-occlusion areas, while the last two cases possibly occur in the occlusion areas. Hence, three costs are defined to identify occlusion is fixed [35]. In this paper, we propose a simple but effective cost function by exploiting confidence information from other cameras. The errors from the matching blocks are employed to calculate the error function. A ceiling is imposed on the error function so that very large individual errors will not dominate the accumulated cost used for defining the global path in DP approach. The original contributions here also include the algorithms for multiview image extension and for combining horizontal and vertical scanning.

\[ C_1 = \frac{1}{\lambda} (1 - E_i) + E_r, \quad C_2 = \frac{1}{\lambda} (1 - E_i) + E_l, \quad C_3 = E_i + E_l \]

where $C_1$ and $C_2$ are the occlusion costs of the pixel, which are invisible in left view and right view respectively, and $C_3$ is the cost of the pixel in no-occlusion areas. These costs are expressed as follows:

\[ E_i = \frac{c_1^2(d)}{1 + \lambda \cdot c_2^2(d)} \]

where $c_1 = \sum_{k \in \text{occlusion}} a_k$, $\phi_k = (\rho^L_k - \rho^R_k)^2$, $a_k$ is a weighing coefficient that is inversely proportional to the distance from the pixel $i$, and $\rho$ is the intensity of pixel $i$. The $B_{L}$ and $B_{R}$ are the baseline between the current camera and the closest left and the closest right cameras respectively, while $\lambda$ is the occlusion parameter.

Finally, the least accumulated cost of the last pixel in a scanline is selected to identify the optimum path. After tracking back along the optimum path, the estimated disparities
are generated and the occlusion areas are simultaneously marked at the pixels where the occlusion cost dominates.

The incorrect path might appear at the occlusion areas composed of similar details of the neighboring nonocclusion areas. However, if more cameras are available, this problem could be eliminated by using information from other reference views as following sections.

1) Multi-view Image Extension: To extend the three-view disparity estimation to the general multiview disparity estimation for a linear configuration, the error from all the reference views are compared and the minimum is selected. Better prediction is achieved by exploiting the information from other view references. However, these error functions are adapted to the cost $C_1$ only, whilst the cost $C_2$ and $C_3$ are still calculated from the closest left and the closest right views. This is because the smaller error might lead the mistaken path in occlusion areas, i.e., $C_1$ would dominate instead of $C_2$ or $C_3$.

2) Combining Horizontal and Vertical Scanning: The proposed scheme is extended to the planar camera configuration. The estimated disparity from scanning in one camera axis can be used to modify the cost for each node of the scanning process in another camera axis by pre-marking the occlusion areas. After vertical scanning, for example, the possible occlusion regions of the horizontal scanning are marked from the vertically estimated disparity map, and then the costs $C_1$ and $C_2$ of such regions are modified with the proportion $\alpha$ as follows: $C_{\text{new}} = \alpha C_1$ and $C_{\text{new}} = C_2$, where $0 < \alpha < 1$.

As a result, the disparity and the occlusion in this horizontal scanning are more reliable than the one that does not exploit information from the vertical scanning. Moreover, the result of the vertical scanning where $C_2$ dominates is subsequently employed to replace the result of the horizontal scanning which is marked as the occlusions.

The results of the proposed disparity estimation by DP are shown in Fig. 4. For the Head sequence, the percentage of the pixels whose absolute disparity error are greater than 1 is 4.05% which is slightly better than the DP proposed in [36], but it is worse than some algorithms as a comparison in [37] (for Santa sequence, the proposed scheme achieves 3.77% error, which the DP in [36] achieves 3.85%). However, our proposed algorithm is obviously simpler by exploiting only intensity constraint. Moreover, the results of the estimation are good enough to use with proposed I-DCVF schemes thereby achieving both compression and view synthesis performance which will be discussed in the following sections.

B. Low-Pass Subband Prediction Improvement

As discussed in the previous section, the low-pass subband prediction does not suffer from the decimation process in wavelet decomposition. Therefore, the disparity estimation for low-pass subband can be directly applied without CODWT. However, the too small size of image might diminish the estimation performance. We suggest operating the disparity estimation at the full resolution, and then the disparity results are subsampling into the proper size associating with the resolution level. Afterward, we make it discrete associating with the block size by using median value of disparities within a particular block. However, if the frequency of the maximum value is greater than one quarter of a number of pixels in a block, this value is selected thereby leading the clear and sharp edges of front objects. The estimated disparity map is transformed to discrete version in order to reduce the number of disparity vectors. Note that the high-pass subband estimation is still operated as explained in Section III-A. The comparison between the block-based estimation and the DP in Fig. 5 shows that the estimation by using the DP does not deteriorate the quality of reconstructed images. The estimated disparity maps of the Head sequence are improved as displayed in Fig. 4(h)–(i) for the quarter resolution and the full resolution, respectively. The estimated disparity maps of the conventional block-based matching present noisy disparities in homogenous areas and also the errors because of unit block size with the position of object boundary. Interestingly, the results of Breakdancers show the improvement of the DP at low-bit rate. This is because the estimated depth of the DP is significantly smoother than that of the block-based estimation thereby achieving lossless compression.

V. EXPERIMENTAL RESULTS

This section presents simulation results for the proposed scheme. Hybrid prediction is exploited to increase the accuracy in prediction step. Adaptive weighing factors in update step are identified by normalized energy of high-pass signal in that block, namely $\alpha_d(i,j,k,l)$ 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 simulations were conducted with three standard multiview test images; Breakdancers, Head, and Santa, containing eight, five, and nine views, respectively. Nonconsecutive views were selected for investigating the performance of the proposed image codec, while the rest views were used for investigating view synthesis efficiency. The DWT was implemented with biorthogonal 9/7 filter up to a total of five decomposition levels in luminance component and up to four decomposition levels in chrominance components. While the size of 8 × 8 is used for spatial estimation. Estimation was performed using one the luminance component, and then the prediction information was used for chrominance components. Residual data was encoded with EBCOT coding. The chrominance components (U, V) were compressed separately but included within the desired target bits.

A. Decomposition Levels of In-Band Prediction

In this section, the performance of in-band prediction at several decomposition levels is investigated with three multiview images which contain the different natures and contexts. The Head sequence comprises many edges and occlusion areas, while the Santa sequence is a doll containing smooth edges. The geometry of these two sequences is parallel, but the geometry of the Breakdancers is nonparallel. The luminance component was decomposed into 1–4 levels, and then the in-band prediction was applied. Subsequently the residuals of low-pass prediction results were further decomposed until five levels were achieved before applying
Table II

<table>
<thead>
<tr>
<th>Level</th>
<th>Head Santa Breakdancers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
<tr>
<td>0*</td>
<td>31.04</td>
</tr>
<tr>
<td>1</td>
<td>30.96</td>
</tr>
<tr>
<td>2</td>
<td>30.88</td>
</tr>
<tr>
<td>3</td>
<td>30.80</td>
</tr>
<tr>
<td>4</td>
<td>30.45</td>
</tr>
</tbody>
</table>

*Decomposition level of 0 indicates the spatial prediction.

Fig. 9: Rate-distortion performance of (a) half-resolution images and (b) quarter-resolution images.

Fig. 10: Quarter-resolution reconstructed middle views of Head at 30 kb/view. (a) Block-based spatial prediction (PSNR Y = 34.87). (b) DP spatial prediction (PSNR Y = 35.94). (c) Block-based 3-level in-band prediction (PSNR Y = 35.94). (d) DP 3-level in-band prediction (PSNR Y = 36.35).

B. Compression Performance

Fig. 6 shows the comparison between predicting in spatial domain and in wavelet domain at 3-level decomposition. The results of the prediction in spatial domain and in wavelet domain are comparable. The in-band prediction improves the performance at high-bit rate, but gives a poorer quality at lower bit rates where more bits are used for coding side information, including the choice of critically sampled subbands. The examples of the reconstructed views are shown in Fig. 7. The performance of the block-based estimation is similar to that of the proposed DP. The prediction of high-pass subbands plays an important role in increasing the compression efficiency.

The direct prediction in high-pass subbands gives a good edge result and no blocking artifacts [14]. Fig. 8 depicts the subjective results of Head sequence. At the same bit rate, the spatial prediction produces visually blocking artifacts, especially along the diagonal edges.
C. Spatial Scalability

The performance of the proposed scheme is compared to that of conventional spatial prediction by considering the image quality of the reduced-size images. The evaluation has been done by calculating the PSNR values of the reconstructed low-pass images at the same level of the decomposed original images. The average PSNR is displayed in Fig. 9 and the subjective results of the quarter size are shown in Fig. 10. Noticeably, the in-band prediction outperforms the spatial prediction but it does not improve much at very low-bit rate as a result of side information. The in-band prediction at level 1 gains similar results of the quarter size as the spatial prediction, since the fractional values disparity vectors of low-pass subband are rounded to integer values. It is greatly affected at the object boundaries in which the discontinuity of depth visibly presents. Consequently, the optimal level for predicting subband-to-subband should be equal or higher (more decomposition) than the required resolution at the receiver. However, the quality of the reconstructed images at the full-resolution should not be deteriorated.

D. View Synthesis Performance

The view synthesis performance was investigated at full resolution and also coarser resolutions by taking off one of the reference views to be used for view synthesis assessment. The remaining views were then used for disparity/depth estimation by the traditional block-based estimation and the proposed DP. At the decoder, the disparity/depth is used with the decoded textures for synthesizing the intermediate views. The disparity vectors from the low-pass subband prediction are utilized. They are linearly interpolated for the finer resolution and subsampled for the coarser resolution. The textures of the reference views are mapped to the intermediate views by disparity information. The textures of the regions attaining the disparity consistency are an average of textures from both nearest reference views, while the holes are filled with the texture of the visible view. However, the problem of filling the holes probably occur if one of the reference views is the furthest left or the furthest right view of multiview sequence. Such areas might get the incorrect disparity, because no corresponding areas in other views. We synthesize the intermediate view in these positions by filling the tricky holes with the disparity values similar to the back objects as described in Fig. 11.

1) View Synthesis at Full Resolution: The evaluation was made by comparing the virtual views to the original views. The average PSNR are shown in Fig. 12 and the examples of the subjective results are displayed in Fig. 13. Comparing between the in-band and spatial prediction, the spatial prediction outperform at full resolution, since the disparities are directly estimated at the full resolution. The in-band prediction generates high mistakes due to the linear interpolation of the disparity/depth maps. The Head results synthesized from the in-band prediction show distorted curves at the depth discontinuity, e.g., the books located behind the head or the lamp.

Comparing between the block-based estimation and the DP, the disparity/depth maps estimated by the DP can significantly improve the view synthesis performance, especially at...
Fig. 14. Rate-distortion performance of the compression and view synthesis for Head at quarter resolution.

Fig. 15. Intermediate view of Head between the left and middle views at quarter resolution. (a) Block-based spatial prediction. (b) DP spatial prediction. (c) Block-based 3-level in-band prediction. (d) DP 3-level in-band prediction.

Fig. 16. Rate-distortion comparison of various resolutions of disparity map for Head.

2) View Synthesis at Coarse Resolution: In the limited display size, the view synthesis process exploited the small reconstructed image, and the disparity/depth maps are upsampled (for the case that the available disparity/depth map contains higher resolution) or are interpolated (for the case that the available disparity/depth map contains lower resolution) into the proper resolution. Figs. 14 and 15 show the objective and subjective results of view synthesis of the Head sequence at quarter resolution respectively. The average PSNR of the spatial prediction and 3-level in-band prediction by block-based matching are comparable, while the in-band prediction by proposed DP greatly improves the view synthesis at small image size. This is the result of the more accurate disparity estimation of the DP than that of the block-based algorithm. Note that the experimental results of the Santa and Breakdancers sequences show the similar trend of rate-distortion performance in which the DP improves the synthesis performance up to 3 dB.

E. Compression Versus View Synthesis

The best algorithm for multiview image coding is a trade-off between the required data size and quality of virtual view construction. Some disparity estimation algorithms support the compression requirement but are not suitable for view synthesis, e.g., block-based matching. Some enhanced approaches offer the correct depth results but require more data which compromises compression efficiency. In this paper, the hybrid in-band prediction and the DP have been proposed to improve the low-pass subband prediction. The proper filters are selected so as to increase the geometric coding performance, while the estimated disparities/depths present more precisely thereby achieving view synthesis.

The relationship between the compression and the view synthesis was investigated by considering the bit allocation for coding texture and disparity/depth information. Basically, the disparity/depth is compressed with lossless coding, while the textures are compressed with lossy coding. The bits allocated for disparity/depth coding can be adjusted by varying the resolution of the disparity/depth maps. Theoretically, the higher resolution disparity/depth maps give the better quality of the synthesized results; therefore, the performance of the view synthesis can be improved by using the dense disparity/depth maps. However, the denser disparity/depth map needs higher amount of bits for coding which may lead poor quality of textures for each particular bit target.

The simulation results are illustrated in Fig. 16 with various disparity/depth resolutions. It clearly proves that the higher resolution disparity/depth maps generate better quality of the view synthesis. However, this benefit is achieved only in high-bit rate because the dense disparity/depth maps require the large space for storing or transmitting. Interestingly, the highest compression performance achieved when using the 4 or 8-pixel resolution. The quality of the reconstructed views with the coarser-resolution disparity maps, e.g., 16 and 32-pixel resolution, decreases since it produces the poorer prediction which causes high energy at the high-pass views and deteriorates the efficiency of entropy coding. From the simulation of these three multiview test sequences, it can be concluded that the 8-pixel resolution gives the best results for both compression and view synthesis.
The authors would like to thank the University of Tsukuba, Ibaraki, Japan, for supporting the multiview test sequences Head and Santa. The authors are also grateful to Microsoft for providing the Breakdancers sequence.

VI. CONCLUSION

A novel in-band disparity compensation for the multiview image coding based on wavelet lifting scheme has been proposed in this paper. Hybrid prediction is exploited to deal with the occlusion problem. This prediction operates in wavelet domain efficiently providing resolution scalability. Additionally, the proposed I-DCVF can exploit different filters and estimation parameters for each resolution level. Perceived view synthesis quality was enhanced by using the DP for disparity estimation and filter selector in low-pass subband prediction. SNR scalability was provided by using EBCOT coding. As a result, the proposed codec is both scalable and efficient at full resolution and significant improvement at coarser factory view synthesis. Our experiments showed comparable and the viewers do not feel the difference. In contrast, the performance of the view synthesis is significantly improved (up to 1.5 dB).

The in-band prediction is applied to the geometry lifting domain. The simulation results show the improvement of smaller scale of reconstructed images, especially at very small size (up to 4.5 dB). By truncating the high-pass subbands, the information for reconstructing the lower resolution image from the in-band prediction does not lose as that from the spatial prediction. Therefore, these features highly benefit the very small display. However, the disparity information employed for view synthesis is gained from only the low-pass subbands which cause the poor quality of full-resolution virtual views. Consequently, the disparity map from the DP approach is applied to the low-pass prediction instead of the block matching. The view synthesis is highly improved as illustrated both the subjective results and the objective results (up to 1.4 dB at full resolution and 1.84 dB at lower resolution). The main drawback of the proposed I-DCVF is that the CODWT process is required to generate the overcomplete high-pass subbands at the decoder. Consequently, the complexity of the codec is higher than the conventional MCTF.

ACKNOWLEDGMENT

REFERENCES


video coding, multiview processing, and distributed video coding.


Cedric Nishan Canagarajah (M’95) received the B.A. and Ph.D. degrees in electrical and electronic engineering, both from the Department of Engineering, University of Cambridge, Cambridge, U.K., in 1991 and 1995, respectively.

He is currently a Professor of Multimedia Signal Processing at the Department of Electrical and Electronic Engineering, University of Bristol, Bristol, U.K. He was previously a Research Assistant and Lecturer at the University of Bristol, where he investigated DSP aspects of mobile radio receivers. He has been involved in a number of the European Union (EU) fifth and sixth framework programme projects where the team has been developing novel image/video processing algorithms. He has published two books and more than 160 papers. His research interests include image and video coding, image segmentation, content-based video retrieval, 3-D video, and image fusion, and he has received support from the EU and the Engineering and Physical Sciences Research Council for his work in these areas.

Cedric Nishan Canagarajah is a member of the Engineering and Physical Sciences Research Council Peer Review College.

David W. Rodnivill received the B.A. and Ph.D. degrees in electrical and informational sciences, both from the Department of Engineering, University of Cambridge, Cambridge, U.K., in 1991 and 1995, respectively.

He is currently a Lecturer at the Department of Electrical and Electronic Engineering, University of Bristol, Bristol, U.K. His research interests include low-bit rate image and video coding for noisy channels, as well as multimedia image/video processing.