Level-by-Level Adaptive Disparity Compensated Prediction in Wavelet Domain for Stereo Image Coding

Shigao Li(✉) and Liming Jia

School of Mathematic & Computer Science, Wuhan Polytechnic University,
Wuhan 430023, Hubei, China
sg51@163.com

Abstract. Disparity compensation prediction and transform coding are incorporated into a hybrid coding to reduce the bit-rate of multi-view images. However, aliasing and inaccurate displacement impair the performance of disparity compensation, especially in wavelet domain. In this paper, we propose a level-by-level adaptive disparity compensated prediction scheme for scalable stereo image coding. To get spatial scalable feature, wavelet transform is first applied to the target image of a stereo image pair. A separable 2-D filter applied to the reference image is optimized for each resolution layer by minimizing the energy of the prediction high-bands of the target image. To form a multi-resolution representation, similar processes are then applied to the low-band image pairs generated by the prior resolution layer iteratively. Experimental results show that the proposed scheme can provide significant coding gain compared to other scalable coding scheme.

Keywords: Disparity compensation · Image compression · Stereo image coding · Wavelet image coding

1 Introduction

Stereoscopic imaging systems are extensively applied in photogrammetry, entertainment and machine vision. Especially in the field of digital photogrammetry, stereo image pairs are used to generate Digital Elevation Model (DEM). However, a mass of image data bring a challenge to image storage and transmission. Especially, how to cater to the capacity of wireless channel for those stereo sensors set on satellites is a strenuous task. Compression techniques are usually used to solve the problem. Stereo image compression techniques have thus received attention of many researchers [1-6].

It's well known that there exists high correlation between two view images. The correlation (or redundancy) can be removed by inter-view prediction using disparity estimation (DE) and disparity compensation (DC). How to perform efficiently DE and DC gets the key of coding methods and many different methods are thus developed. In [2-3], spatial domain DC-based methods are proposed. In general, these methods can be briefed as follows. First, one image of an image pair as a reference is used to

---

This work was supported by the Chinese Natural Science Foundation (61201452).

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 428–437, 2015.
DOI: 10.1007/978-3-319-23234-8_40

predict another image called target image and a residue image is then generated. Second, a transform coding is applied to the reference and the residue independently. In last two decades, discrete wavelet transform (DWT) applied to image and video compression has achieved a great success [7-8]. A typical coding method (or coder) is EBCOT [9]. It's remarkable that wavelet-based methods can produce scalable code stream. However, DC in spatial domain violates the resolution scalability.

In [1], M.Kaaniche et al. proposed a novel vector lifting schemes (VLS), in which the DC process is incorporated into the decomposition procedure of target images. Because the information of the reference image directly joins the lifting process of target images level by level, no residue images are generated. Experiment results in [1] indicate that VLS achieves a significant coding gain compared to the conventional lifting scheme of wavelet. Due to the DC in multiple resolution levels, VLS can produce scalable code stream.

To get a scalable video coding, in-band prediction compensation were proposed by many researchers [10]. Nantheera Anantrasirichai applied the scheme to multi-view image coding [11]. The in-band compensation directly predicts the sub-bands of target images by using the sub-bands of reference images level by level based on the relation of subbands with different phases [10]. Although in-band compensation provides the scalable feature, aliasing introduced by the critically-sampled DWT and inaccurate displacement estimation impair the performance of DC/MC. In [12], a scheme called in-scale compensation were proposed to scalable compression for video. Due to the fixed analysis filters are applied to generated prediction compensation frames, in-scale compensation are similar with in-band compensation in the essence.

In recent years, adaptive interpolation filter (AIF) and adaptive loop filter (ALF) are introduced into video compression by many researchers [13-15]. In these AIF methods [13-14], filters have been designed to eliminate aliasing and to interpolate subpixel data so that more exact displacement vectors can be obtain. Thanks to the adaptivity, AIFs provide significant coding gains compare to fixed interpolation filter.

In this paper, we propose a multi-layer adaptive disparity compensation scheme for scalable stereo image coding. To get spatial scalable feature, 2-D wavelet transform (2D-DWT) is first applied to the target image of a stereo image pair. The reference image rather than subbands is used to predict the target subbands at each resolution layer. That is to say, the prediction is performed between the reference image (or the low-frequency approximation at low resolution layers) and the high-subands of the target image at corresponding level. Although the in-scale motion compensation scheme for spatially scalable video coding is similar with ours, we adopt adaptive prediction filters instead of fixed analysis filters. In addition, considering occulsion between reference images and target images, a piecewise selection procedure is designed to exclude those pixels in occlusion region. Alternatively, original coefficients of DWT are reserved and encoded. When compared to the VLS, because a better scheme and more accurate displacement vectors are adopted, the proposed scheme obtain less high-frequency sub-band energy of target image.

This paper is organized as follows. In Section 2, we introduce the general layer-based prediction compensation formulation. In the fact, the in-band scheme and the in-scale are specific instance. The proposed structure is described and the problem to be solved is formulated in Section 3. The fourth section is devoted to occlusion culling. Experimental results show that the proposed outperforms other related schemes in terms of coding efficiency in Section 5. Section 6 provides brief conclusions.

2 Layer-Based Prediction Compensation Formulation

Many researchers applied DWT to video compression. In general, DWT is first applied to multi-frame group at temporal direction. And 2D-DWT is applied to each sub-band frame generated at the first step. This scheme is called t+2D scheme. However, when this scheme is used in scalable video coding, drifting impairs the coding efficiency [16]. To avoid the drifting, in-band MC/DC and in-scale MC are proposed. Generally, to get a scalable scheme, these methods carry out the prediction compensation level by level. A general decomposition can be depicted as Fig.1. The compensation process can be given by

$$\hat{d}_{\mathrm{n}}^{t} = d_{\mathrm{n}}^{t} - \mathcal{P}_n\big(\mathcal{DC}_{n-1}(I_{n-1}^{r})\big)$$

$$\cdots\cdots$$

$$\hat{d}_{\mathrm{j}}^{t} = d_{\mathrm{j}}^{t} - \mathcal{P}_j\left(\mathcal{DC}_{j-1}(I_{j-1}^{r})\right)$$

$$\cdots\cdots$$

$$\hat{d}_1^{t} = d_1^{t} - \mathcal{P}_1\big(\mathcal{DC}_0(I_0^{r})\big) \tag{1}$$

where $\mathcal{DC}$ denotes disparity compensation operations used to form the disparity compensated image (DCI) $\tilde{I}_{\mathrm{j-1}}^{\mathrm{t}}$ as a prediction of $I_{\mathrm{j-1}}^{\mathrm{t}}$ and $\mathcal{P}_j$ denotes the prediction compensation operations used to predict the high-frequency sub-bands at the j-th level. Varied $\mathcal{P}_j$ forms various scheme. For instance, the $\mathcal{P}_j$ used in in-band MC based on low-bands shifting (LBS) [17] consist of inverse DWT, over-complete DWT by translating. Being similar with in-band MC, in-scale MC adopts fixed predictors.

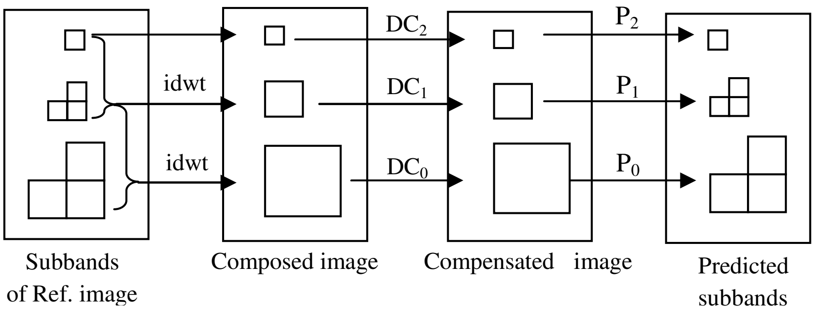

Fig. 1. Level-by-level Prediction Compensation

Fractional-pixel disparity accuracy can be adopted to enhance the correlation between target images and prediction images generated from reference images in the DC process. However residual translation errors still exist and impair the performance of compensation in bands because of fixed filters. In order to improve the coding efficiency by considering the non-stationary statistical properties resulting from residual translation errors and illumination differences between two different views, adaptive prediction filters are introduced in this paper. In the next section, we'll give the proposed scheme and the design of adaptive predictors $\mathcal{P}_j$'s for each level.

3 Proposed Scheme

3.1 Separable Level-by-Level Adaptive Disparity Compensation Scheme

This subsection presents the proposed level-by-level lifting scheme by using adaptive predictors. The proposed scheme can be depicted as Figure 2. At first, the conventional 2D-DWT is applied to target images at the j-th level and three high-frequency sub-bands $d_{\mathrm{V},j}^t$, $d_{\mathrm{H},j}^t$, $d_{\mathrm{D},j}^t$ and a low-frequency approximation image I_j^t are generated. To exploit the correlation between reference images and target images, three prediction compensation processes are carried out for the three high-frequency sub-bands independently. As a result, three new high-frequency sub-bands, i.e. prediction residue images, are formed. The decomposition process of target images can be formulated as follows,

$$\begin{aligned}
I_j^t &= \mathcal{A}_{LL}\left(I_{j-1}^t\right)\\
d_{\mathrm{V},j}^t &= \mathcal{A}_{\mathrm{LH}}\left(I_{j-1}^t\right)\\
d_{\mathrm{H},j}^t &= \mathcal{A}_{HL}\left(I_{j-1}^t\right)\\
d_{\mathrm{D},j}^t &= \mathcal{A}_{HH}\left(I_{j-1}^t\right)\\
\hat{d}_{\mathrm{V},j}^t &= d_{\mathrm{V},j}^t - \mathcal{P}_{V,j}\left(\mathcal{DC}_{j-1}\left(I_{j-1}^r\right)\right)\\
\hat{d}_{\mathrm{H},j}^t &= d_{\mathrm{H},j}^t - \mathcal{P}_{H,j}\left(\mathcal{DC}_{j-1}\left(I_{j-1}^r\right)\right)\\
\hat{d}_{\mathrm{D},j}^t &= d_{\mathrm{D},j}^t - \mathcal{P}_{D,j}\left(\mathcal{DC}_{j-1}\left(I_{j-1}^r\right)\right)
\end{aligned} \tag{2}$$

where $\mathcal{A}_{LL}$, $\mathcal{A}_{\mathrm{LH}}$, $\mathcal{A}_{HL}$, $\mathcal{A}_{HH}$ denotes the analysis operators of 2D-DWT, and $\mathcal{P}_{V,j}$, $\mathcal{P}_{H,j}$, $\mathcal{P}_{D,j}$ denotes the adaptive predictors for $d_{\mathrm{V},j}^t$, $d_{\mathrm{H},j}^t$, $d_{\mathrm{D},j}^t$ respectively. $\hat{d}_{\mathrm{V},j}^t$, $\hat{d}_{\mathrm{H},j}^t$, $\hat{d}_{\mathrm{D},j}^t$ denotes the new high-frequency sub-bands. To simplify the problem, separable and symmetric filters can be adopted for adaptive predictors $\mathcal{P}_{V,j}$, $\mathcal{P}_{H,j}$, $\mathcal{P}_{D,j}$.

At last, 2D-DWT are applied to reference images, new approximation reference image and three high-frequency sub-bands are produced. Similar operations can be applied to the low-frequency approximation images of the reference image and target image for further decomposition.

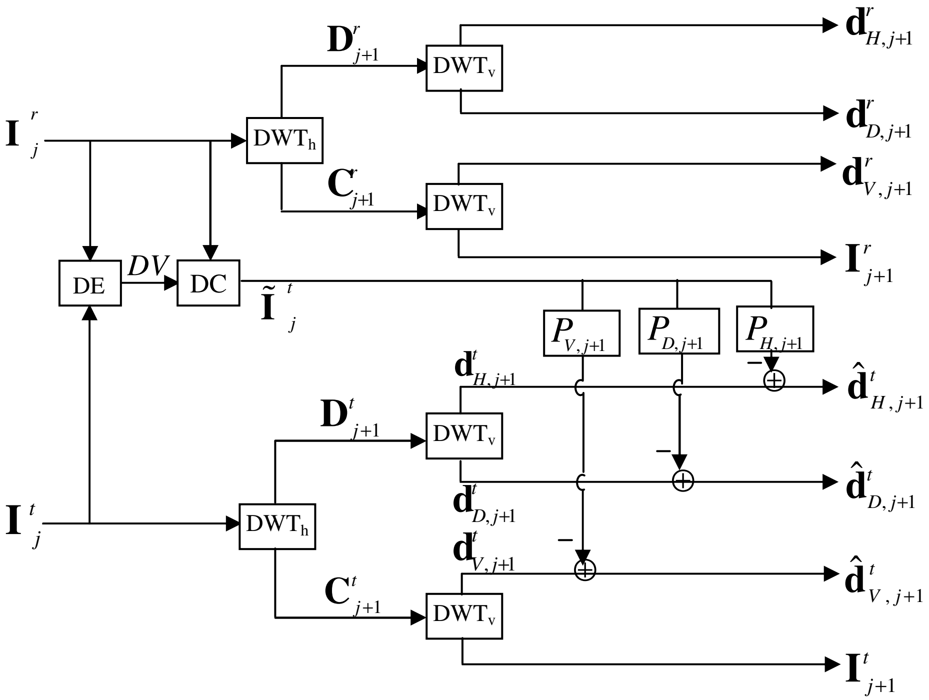

Fig. 2. Adaptive disparity compensation scheme

3.2 Separable Adaptive Prediction Filters for Target High-Bands

It's critical to obtain those adaptive predictors in the last subsection. However, we should strike a balance between prediction performance and complexity. Therefore, in this sub-section, separable and symmetric filters are designed for those adaptive predictors. The separable prediction processes can be formulated as follows:

$$
\begin{aligned}
\mathrm{C}_j^t &= \mathcal{A}_L^h\left(I_{j-1}^t\right)\\
\mathrm{D}_j^t &= \mathcal{A}_H^h\left(I_{j-1}^t\right)\\
\tilde{\mathrm{C}}_j^t &= \mathscr{q}_{j,L}^h\left(\mathcal{DC}_{j-1}\left(I_{j-1}^r\right)\right)\\
\widetilde{\mathrm{D}}_j^t &= \mathscr{q}_{j,H}^h\left(\mathcal{DC}_{j-1}\left(I_{j-1}^r\right)\right)\\
\hat{d}_{\mathrm{V},j}^t &= d_{\mathrm{V},j}^t - \mathscr{q}_{j,V}^v\left(\tilde{C}_j^t\right)\\
\hat{d}_{\mathrm{H},j}^t &= d_{\mathrm{H},j}^t - \mathscr{q}_{j,H}^v\left(\widetilde{D}_j^t\right)\\
\hat{d}_{\mathrm{D},j}^t &= d_{\mathrm{D},j}^t - \mathscr{q}_{j,D}^v\left(\widetilde{D}_j^t\right)
\end{aligned}
\tag{3}
$$

where q's denote the one dimensional prediction operators whose superscript 'h' denotes horizontal direction and 'v' denotes vertical direction. It's clear that 5 prediction vectors should be designed for each level. We can estimate the coefficients of the prediction vectors by minimizing the energy of the prediction error

$$\hat{q}^h_{j,L} = \arg\min_{q^h_{j,L}} \left\{ \left\| C^t_j - \tilde{C}^t_j \right\|^2 \right\}$$

$$\hat{q}^h_{j,H} = \arg\min_{q^h_{j,H}} \left\{ \left\| D^t_j - \tilde{D}^t_j \right\|^2 \right\}$$

$$\hat{q}^v_{j,H} = \arg\min_{q^v_{j,H}} \left\{ \left\| \hat{d}^t_{H,j} \right\|^2 \right\}$$

$$\hat{q}^v_{j,V} = \arg\min_{q^v_{j,V}} \left\{ \left\| \hat{d}^t_{V,j} \right\|^2 \right\}$$

$$\hat{q}^v_{j,D} = \arg\min\nolimits_{q^v_{j,D}} \left\{ \left\| \hat{d}^t_{D,j} \right\|^2 \right\}. \tag{4}$$

4 Occlusion Culling

In occlusion regions, there exists no correlation between reference images and target images. In this case, the proposed scheme will bring a bad effect in the prediction. Clearly, the wavelet transform should be used to exploit intra-view correlation instead. To correct the prediction method when necessary, conventional 2D-DWT is incorporated into the proposed scheme. In order to make a proper choice of prediction, a piecewise decision method is proposed. That is to say, the pixels in the same block use a same prediction method. As a result, only one bit is necessary to record the choice of an entire block. Bit '0' expresses the prediction result of the proposed scheme is used and bit '1' denotes the result of 2D-DWT is used. The choice of Block k can be determined as the following formulation

$$\text{f(x)}=\begin{cases} 0, & \sum_{(x,y)\in\mathcal{B}(k)} \left(\hat{d}^t_j(x,y)\right)^2 \leq \sum_{(x,y)\in\mathcal{B}(k)} \left(d^t_j(x,y)\right)^2 \\ 1, & \text{else} \end{cases}. \tag{5}$$

To seek a balance between accuracy and bitrate cost, different block sizes are used in different decomposition levels. As the subbands in the first level have the least energy weight, a block size of 16×16 is used in the first decomposition level. A block size of 8×8 is used in the second decomposition level. For higher decomposition level, smaller block size is used. Due to the effect of low-pass filters, the correlation between left and right views at the higher level is very high. As to the decomposition levels higher than the third, the prediction result of the proposed scheme is always used. Assuming a three-level decomposition are used, the cost of recording the flag for each block is negligible (about 0.0117BPP). If a entropy encoder is used, lower bitrate can be obtained.

After culling those pixels within occlusion regions labeled with bit '1', prediction filters q's are recalculated again to obtain more accurate results.

5 Experiments and Results

This section designs several experiments to test the mention-above algorithm. Following Kaaniche et al. (2009), we use 5/3 wavelet. In the following experiments, we used the MQ-coder to encode the transform coefficients, which is used in EBCOT as a part of JPEG2000. We have compared the proposed method with other two compression methods. The first is independent encoding method using JPEG2000. We use OpenJPEG software (version 1.3) that is an open source JPEG2000 codec. Another one is VLS proposed by Kaaniche et al. (2009).

We have used three image pairs called "Tile1", "Tile2" and "Tile3" with the size of 512×512 extracted from a big image pairs, which are downloaded from http://www.isprs.org/data/ikonos_hobart/default.aspx and derived from the high-resolution satellite Ikonos. Tile1 and Tile2 are extracted from urban areas while Tile3 is extracted from the region covered with vegetation.

A fixed block size of 8x8 and a simple block matching method are adopted in disparity estimation. We adopt displacement vectors with integer pixel accuracy. To lower the bitrate of side information, the DPCM and arithmetic coding method are used to encode the parallax vector.

5.1 The Performance of Prediction

The experiments of this subsection are designed mainly to compare the proposed to the VLS proposed by Kaaniche et al.. The variances of subbands of target images are estimated. As shown in Table 1-3, the proposed scheme obtains smaller subband variances than the VLS. It's clearly that the proposed scheme can improve the prediction performance.

Table 1. The variance of subbands of the target image of Tile1

Level	VLS			Proposed		
	LH	HL	HH	LH	HL	HH
1	6632.2	6603.4	5916.4	5308.2	6195.3	5509.0
2	19423.9	18749.7	29494.3	12070.1	13417.4	24497.4
3	34794.3	32244.3	61611.9	17813.5	17238.7	44839.1

Table 2. The variance of subbands of the target image of Tile2

Level	VLS			Proposed		
	LH	HL	HH	LH	HL	HH
1	7805.0	7321.2	7811.6	6207.8	6919.3	7181.7
2	19061.5	16574.2	43542.4	13107.7	12300.0	37343.9
3	26520.5	24355.4	55883.4	13505.3	15330.0	41814.3

Table 3. The variance of subbands of the target image of Tile3

Level	VLS			Proposed		
	LH	HL	HH	LH	HL	HH
1	1408.8	1342.7	1456.3	1207.1	1276.8	1388.4
2	3521.2	3013.5	5398.7	2276.8	2298.5	4643.4
3	6379.8	5123.7	10732.5	2923.2	2877.7	7469.1

5.2 The Performance of Lossy Compression

In this subsection, several experiments are designed to compare the compression performance. Fig. 3, Fig. 4 and Fig. 5 present the test results of the images abovementioned respectively. The PSNR expresses the joint peak signal-to-noise ratio that is calculated by the following formulation.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{(2^{\mathrm{BP}}-1)^2}{(MSE^{(t)}+MSE^{(r)})/2}\right) \tag{6}$$

where BP is the bit number of each sample, $MSE^{(t)}$ and $MSE^{(r)}$ expresses the mean square error of the reconstruct of reference and target images respectively.

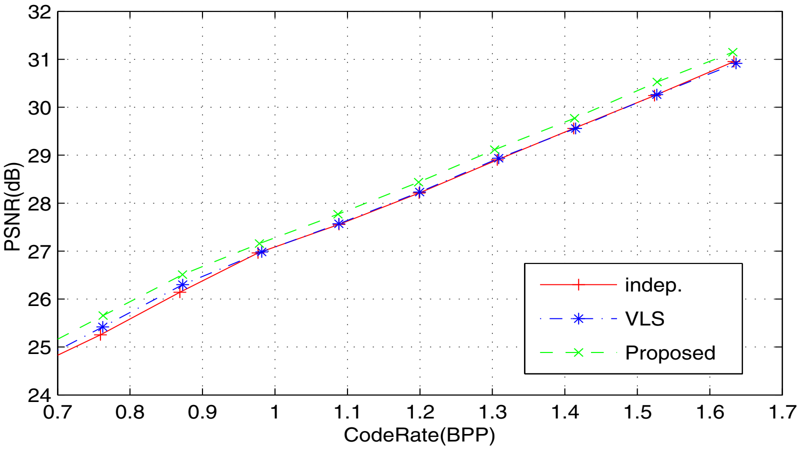

Fig. 3. Comparison performance of the proposed, VLS and independent compression for Tile 1

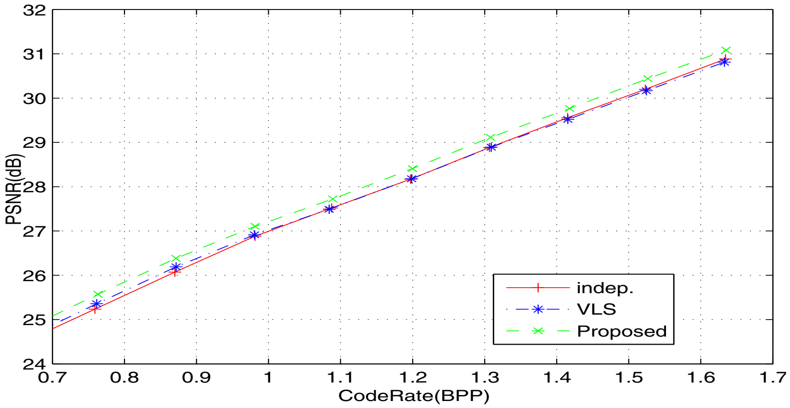

Fig. 4. Comparison performance of the proposed, VLS and independent compression for Tile 2

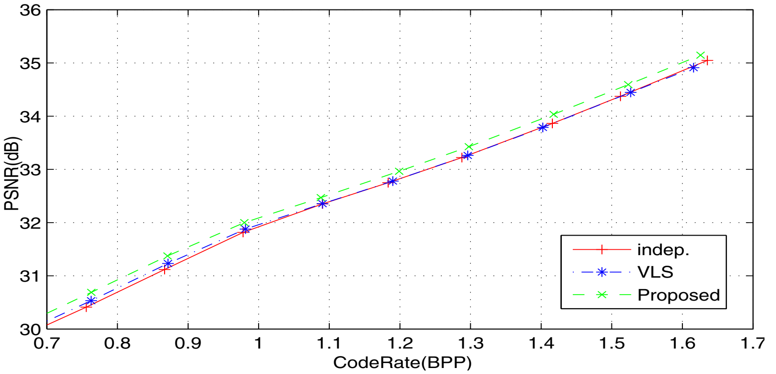

Fig. 5. Comparison performance of the proposed, VLS and independent compression for Tile 3

As shown in Fig. 3-5, under the condition of low bitrate (less than 1 BPP), joint compression methods including VLS and the proposed significantly are superior to the independent compression method. The lower the code-rate is, the more obvious the advantage is. Under the condition of high bit rate, the proposed can maintain the advantage while the VLS gets no better than the independent method. Compared to the VLS, the proposed obtains a coding gain of about 0.2 dB.

6 Conclusions and Future Work

This paper proposed a level-by-level adaptive disparity compensated prediction scheme for scalable stereo image coding. To get spatial scalable feature, 2-D wavelet transform is first applied to the target image of a stereo image pair. A separable 2-D filter applied to the reference image is optimized for each resolution layer by minimizing the energy of the high-bands of the target image. To form a multi-resolution representation, similar processes are then applied to the low-band image pairs generated by the prior resolution layer iteratively. And a piecewise decision method is used to choose the better prediction result from the proposed scheme and 2D-DWT. Experimental results show that the proposed scheme can provide significant coding gain compared to other scalable coding scheme.

Because the proposed scheme are easy extended to much more views, in the future, we'll use this scheme to compress multiview images.

References

1. Kaaniche, M., et al.: Vector Lifting Schemes for Stereo Image Coding. IEEE Trans. Image Process **18**(11), 2463–2475 (2009)
2. Perkins, M.G.: Data Compression of Stereopairs. IEEE Trans. Communications **40**(4), 684–696 (1992)
3. Moellenhoff, M.S., Maier, M.W.: Transform coding of stereo image residuals. IEEE Trans. Image Processing **7**(6), 804–812 (1998)
4. Jiang, Q., Lee, J.J., Hayes, M.H.: A wavelet based stereo image coding algorithm. In: International Conference on Acoustics, speech, and signal processing, Phoenix, pp. 684–696. Arizona State University, Arizona (1999)
5. Boulgouris, N.V., Strintzis, M.G.: A Family of Wavelet-Based Stereo Image Coders. IEEE Trans. Circuits Syst. Video Technol. **12**(10), 898–904 (2002)
6. Edirisinghe, E.A., Nayan, M.Y., Bez, H.E.: A wavelet implementation of the pioneering block-based disparity compensated predictive coding algorithm for stereo image pair compression. Signal Processing: Image Communication **19**, 37–46 (2004)
7. Sweldens, W.: The lifting scheme: A construction of second generation wavelets. SIAM J. Math. Anal. **29**(2), 511–546 (1998)
8. Daubechies, I., Sweldens, W.: Factoring wavelet transforms into lifting steps. J. Fourier Anal. Appl. **4**(3), 247–269 (1998)
9. Taubman, D.S.: High performance scalable image compression with EBCOT. IEEE Trans. Image Proc. **9**, 1158–1170 (2000)
10. Andreopoulosa, Y., Munteanu, A.: In-band motion compensated temporal filtering. Signal Processing: Image Communication **19**, 653–673 (2004)
11. Anantrasirichai, A., Canagarajah, C.N.: In-Band Disparity Compensation for Multiview Image Compression and View Synthesis. IEEE Transactions On Circuits And Systems For Video Technology **20**(4), 473–484 (2010)
12. Xiong, R., Xu, J., Wu, F., Li, S.: In-scale motion aligned temporal filtering. In: IEEE International Symposium on Circuits and Systems, pp. 3017–3020 (2007)
13. Ye, Y., Motta, G.: Enhanced Adaptive Interpolation Filters for Video Coding. Data Compression Conference, Snowbird, UT, USA, pp. 24–26 (2010)
14. Vatis, Y., Ostermann, J.: Adaptive Interpolation Filter for H.264/AVC. IEEE Transactions On Circuits And Systems For Video Technology **19**(2), 179–187 (2009)
15. Yoo, Y.J., Seo, C.W., Han, J.K.: Enhanced Adaptive Loop Filter for Motion Compensated Frame. IEEE Transactions On Image Processing **20**(8), 2177–2188 (2011)
16. Xiong, R., Xu, J., Wu, F., Li, S.: Studies on spatial scalable frameworks for motion aligned 3D wavelet video coding. In: VCIP 2005, Beijing, vol. 5960, pp. 189–200 (2005)
17. Park, H.W., Kim, H.S.: Motion estimation using lowband-shift method for wavelet-based moving picture coding. IEEE Trans. Image Process. **9**(4), 577–587 (2000)