1D DIRECTIONAL DCT-BASED STEREO RESIDUAL COMPRESSION

Walid Hachicha [1], *Azeddine Beghdadi* [1], *Faouzi Alaya Cheikh* [2]

[1]L2TI, Institut Galilée, Université Paris 13, Sorbonne Paris Cité, France
[2]The Norwegian Colour and Visual Computing Lab, Gjøvik University College, Norway
walid.hachicha@univ-paris13.fr, azeddine.beghdadi@univ-paris13.fr, faouzi.cheikh@hig.no

ABSTRACT

A novel stereo image compression method based on disparity-compensated residual image structure analysis is presented in this paper. The residual image resulting from the subtraction of the target image by the predicted one is exploited in this scheme. Unlike natural images, for which 2D-orthogonal transforms such as 2D-DCT provides good energy compaction in the frequency domain, stereo residual images exhibit directional and sparse components making them difficult to compress with classical methods. We propose to exploit the directional properties of the residual image, to compress it more efficiently, by using 1D Directional Discrete Cosine Transform (1D-DDCT). This 1D transform is applied to non-overlapping blocks of the residual image, along four directions: i.e. horizontal, vertical, positive diagonal and negative diagonal. The experimental results show that the proposed method, on some test images, outperforms the considered state-of-the-art methods.

Index Terms— Stereo image coding, Disparity compensation, Directional DCT, Embedded algorithm.

1. INTRODUCTION

The rapid development of stereo technology acquisition and 3D visualization has generated tremendous interest in many applied fields of research such as tele-medicine, video games, robotics, 3DTV and 3D cinema. To accommodate efficiently all these applications, compression schemes are required to ensure efficient storage and rapid transmission. A stereoscopic pair is a couple of images of the same scene representing two different perspectives. Conventional monoscopic compression techniques [1, 2] can be applied to each view separately, at the cost of doubling the bitstream size. Obviously, it is the natural way to compress stereo images. But it is inefficient because it does not exploit the significant redundant inter-view information.

One of the popular approaches to compress stereoscopic images content is the disparity-compensated (DC) residual coding by analogy with the well known motion compensated residual (MC) used in video coding. One of the two views is independently coded as a reference, and the residual one corresponding to the difference between the original and the predicted target image, is then encoded. The predicted target image is estimated by exploiting the correlation between the two views of the stereo pair. This could be achieved by estimating the disparity map representing the displacement field on the projected stereo pair images. This disparity map is generally coded using lossless scheme.

To encode the residual image efficiently, it is important to analyze the spatial distribution of the pixel intensity. The residual image characteristics depend on many factors related not only to the 3D scene contents and the occluded objects but also on the adopted encoding scheme for the reference image. Consequently, it is difficult to derive a model that could account for all these parameters. Furthermore, the residual image contains sparse and peaked components due to occlusion regions and mismatched pixels along edges. It contains also intensity values near zero resulting from good disparity estimation. Fig .1 illustrates the disparity-compensated residual image which exhibits a directional structure.

(a) (b)

Fig. 1: (a) Right view and (b) absolute residual image (scaled by 8 for visualization) of the 'Room' stereo pair ($w = 8$).

In this paper, we present a novel stereo residual coder which takes into account the residual image structure by using 1D-DDCT applied to non-overlapping blocks.

The rest of this paper is organized as follows. Section 2 gives an overview of some existing stereo image compression methods. Section 3 describes the main stages of the proposed coder. In section 4, we present experimental results. Finally, conclusions and possible further research directions are summarized in Section 5.

2. RELATED WORKS

In this section, we review some related stereo image compression techniques. Perkins [3] introduced for the first time the stereo compression that consists of coding the sum and the difference of the stereo pair. This idea is based on the uncorrelated sum and difference of stereo pair images having the same first and second-order statistics. In [4], Yamaguchi *et al.* introduced a modification to the sum-difference approach by shifting one of the images in the horizontal direction to the point where the correlation between the images of the stereo pairs reaches its maximum. The difference between these two aligned pairs is then encoded. This approach assumes that the objects in the scene have in general the same disparity, which is conflicting with 3D structure of the scene. Disparity-compensated based methods aim to exploit the redundant information in the stereo pair due to the binocular dependency. Several approaches have been proposed in this direction [5–16]. Disparity compensation methods based on block matching was first introduced by Lukacs [5]. Perkins [6] has proposed a disparity-compensated transform domain predictive (DCTDP) coding; the compensated-residual image is performed in transform domain. The idea seems to be excellent but in practice the coder do not give a good result due to the three-level uniform quantization. Boulgouris and Strintzis [12] proposed an embedded coder in which the reference and the residual images are decomposed using discrete wavelet transform (DWT) and EZ coded [17]. Following the same idea, a morphological embedded wavelet coder has been proposed by Ellinas [15]. These two methods support progressive transmission; however, they do not take into account the intrinsic properties of residual image. The work by Moellenhoff and Maier [11], shows that the residual image has two specific properties. First, most of the pixels have nearly-zero values. Second, the remaining content consists mainly of narrow vertical edges. They propose to transform the stereo residuals with the 2D-DCT and a DWT. Finally, a selective quantization scheme is applied. In [14], authors treated occluded blocks differently from blocks that are well estimated. A mixed transform (MT) is used which consists of a Haar transform of 3 levels for occluded blocks and 2D-DCT for others. The DCT coefficients are grouped into a wavelet decomposition to line up with the Haar-transformed coefficients and then embedded coded. Although the MT coder gives a good performance, it is worth to point out that non-occluded blocks can contain high frequencies near edges, making thus 2D-DCT not suitable in this case. Other techniques have elaborated more sophisticated disparity compensation methods. For example in the method described in [7], an overlapped disparity compensation approach using adaptive windows with variable shapes is used. Authors proposed also an efficient bit allocation between the reference and the residual image using dynamic programming. Other model-based algorithms based on the use of the Markov Random Fields (MRF) theory for disparity compensation have been also considered in the literature. The disparity estimation is expressed as a maximum a-posteriori probability (MAP) problem [8, 9]. These MRF-based methods produce smooth and dense disparity map which requires a large portion of the bitstream. Moreover, disparity compensation based on edginess information and on object/feature based methods have been used in [10, 13, 16]. Tzovaras *et al.* [10] proposed an object based coding approach using motion and disparity information. The DE is performed using a pixel-based hierarchical dynamic programming algorithm. Jiang and Edirisinghe [13] combine the existing block based stereo matching techniques with an object/feature based technique. In [16], a template based on inter-prediction method is proposed, to encode images with a fully pattern matching based scheme.

For 3D sequences, stereo and Multi-view video (MVC) [18] extensions of H.264/MPEG-4 AVC Standard are developed where the disparity is estimated using a block-based scheme similar to standard motion vector estimation methods.

3. THE PROPOSED CODER

The aim of this work is to design a coder for disparity-compensated (DC) stereo residual image. Fig. 2 describes schematically the proposed coder. Initially, the reference image is coded and decoded using any existing image coder. Second, block matching algorithm is performed in order to estimate the correspondence between blocks of the stereo pair. The reference image is then divided into $k \times k$ blocks and each block of the reference image is matched with an approximating block in the target image by minimizing a cost function such as the sum of square difference (SSD) or the sum of absolute difference (SAD). Given a disparity map, the target image is predicted from the decoded reference one. Then the difference between these two images is considered as the residual image. It is well known

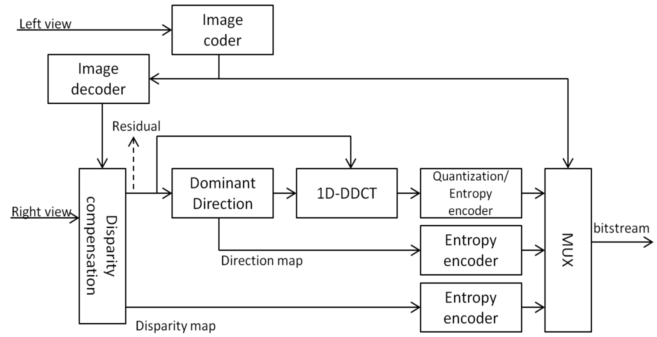


Fig. 2: Block diagram of the proposed encoder.

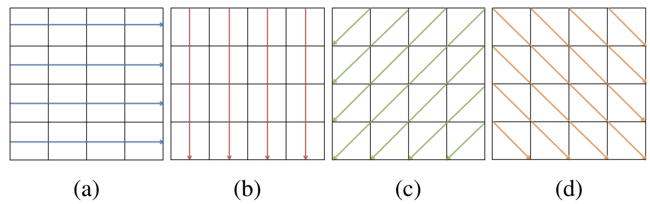


Fig. 4: The four directions: (a) H, (b) V, (c) PD, (d) ND.

that occlusion may result either from finite viewing area or depth discontinuity. In our case, only the occluded finite view is considered, as illustrated in Fig. 1b in the right side of the absolute residual image. We denote by $w \geq k$ the width of the occluded finite view area. Two sub-images are constructed to distinguish non-occluded area from occluded one. Each sub-image is divided into $k \times k$ block. 1D-DDCT is applied to non-occluded sub-image. Whereas, 2D orthogonal transform is applied to occluded blocks. In the rest of this paper, we refer to the non-occluded sub-image as the residual image.

3.1. Directional DCT

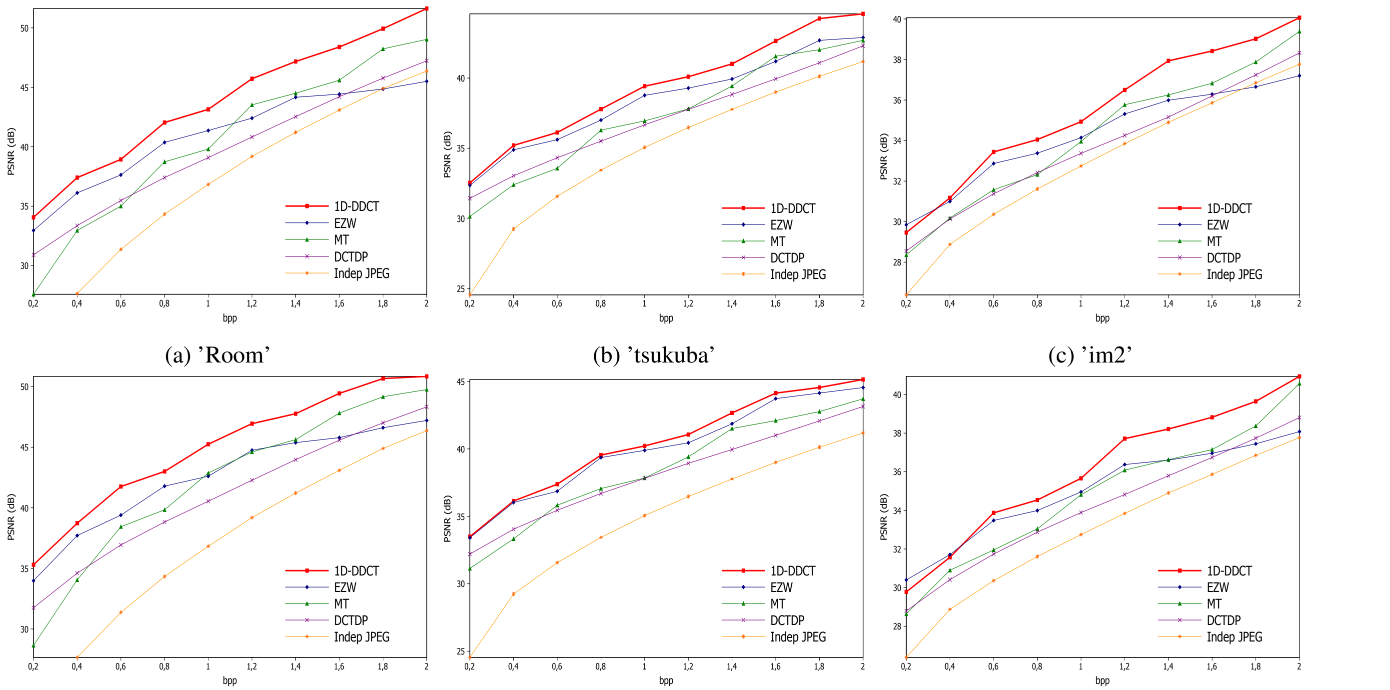
The 2D-DDCT is introduced in [19], based on the computation of two successive 1D-DCT. The first one is calculated along a specific direction, and then a horizontal 1D-DCT is performed. A sized N vector $\{v_0, ..., v_{N-1}\}$ is transformed into the vector $\{V_0, ..., V_{N-1}\}$ according to the following formula:

$$V_q = \sum_{p=0}^{N-1} \alpha_p v_p \cos[\frac{\pi}{N}(p+\frac{1}{2})q],\ \alpha_p = \begin{cases} \frac{1}{\sqrt{2}} & if\ \ p=1, \\ 1 & Otherwise, \end{cases} \quad (1)$$

It is shown that 2D-DDCT based coder provides a better coding performance for natural image blocks. However, this may not be the optimal solution for 1D structure in the residual. Therefore, the 1D-DDCT along a specific direction with high correlated pixels can perform better energy compaction than 2D-DDCT. The 2D-DDCT consists in applying two successive 1D-DCT. The first one is performed along the dominant direction in the block, while the second one is applied along the rows in the rearranged pattern. The correlation of residuals in such direction (the second 1D-DCT) is relatively low. Thus, in our work, we choose to apply the 1D-DDCT. We consider only four dominant directions: horizontal (H), vertical (V), positive diagonal (PD) and negative diagonal (ND). Figure 4 illustrates the defined directions.

3.2. Prediction of the dominant direction

Dominant edge direction prediction is a crucial step in our coding process. It is straightforward to consider only 4 directions instead



(a) 'Room' (b) 'tsukuba' (c) 'im2'

(d) 'Room' (e) 'tsukuba' (f) 'im2'

Fig. 3: The Rate-Distorsion curves of the right image for different stereo images. For (a), (b) and (c) the left images are JPEG coded at quality 50 and for (d), (e) and (f) the left images are JPEG coded at quality 75.

of the 8 defined in [19]. This process is realized within two stages. First, given a residual block $\mathcal{B}$, $[\log_2(k)]$ level wavelet decomposition ($[.]$ is the floor function) is performed to determine if the direction is horizontal, vertical, or diagonal, using the following equation:

$$dir = \arg \max_{s \in \{1,2,3\}} \sum_{j=1}^{[\log_2(k)]} \sum_{i=1}^{n_s} W_{i,j}^s, \tag{2}$$

where dir correspond to the predicted dominant direction, s is the direction index, $W_{i,j}^s$ is the i^{th} wavelet coefficient at scale j and direction s and n_s is the number of wavelet coefficient at each subband. The residual block $\mathcal{B}$ is decomposed at each scale into three directions (HL, LH and HH) i.e. horizontal, vertical and diagonal edges. In Eq .(2), the three directions are mapped into $\{1, 2, 3\}$ and the dominant direction corresponds to the index with highest energy. Nevertheless, this stage is not able to predict the (PD) and (ND) directions if $dir = 3$. In this case, a second stage is required. This step needs, for each diagonal block, the construction of a point cloud and then we compute its orientation. The point cloud is formed by pixels having value equal to 1 after thresholding, as follows:

$$\delta_{i,j} = \begin{cases} 1 & if \ |b_{i,j}| \geq \lambda \max(\mathcal{B}), \\ 0 & Otherwise, \end{cases} \tag{3}$$

where $b_{i,j}$ is the intensity value of the residual block $\mathcal{B}$ at (i, j) position, $\max(\mathcal{B})$ is the absolute maximum of $\mathcal{B}$ and $\lambda \in]0, 1[$ is a parameter to control locally the threshold value. The orientation θ is given by solving the following equation:

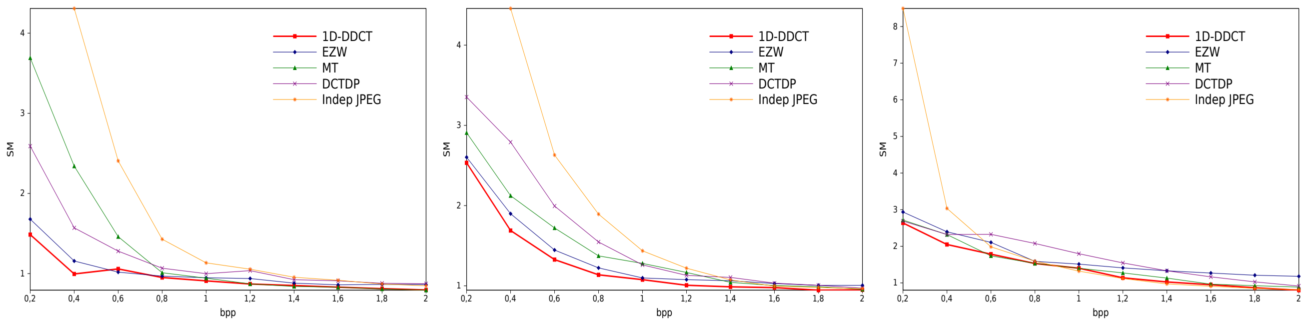
$$(\sigma_x^2 - {\sigma_y}^2) \tan(2\theta) = 2\sigma_{xy}, \tag{4}$$

where (x, y) is a couple of vectors which contain the coordinate of the points in the cloud, σ_x , σ_y and σ_{xy} are respectively, the variance of x, the variance of y and the covariance of x, y. We note that, θ is positive when σ_{xy} is positive. Thus, if $\sigma_{xy} \geq 0$ then the dominant direction is (PD) else (ND).

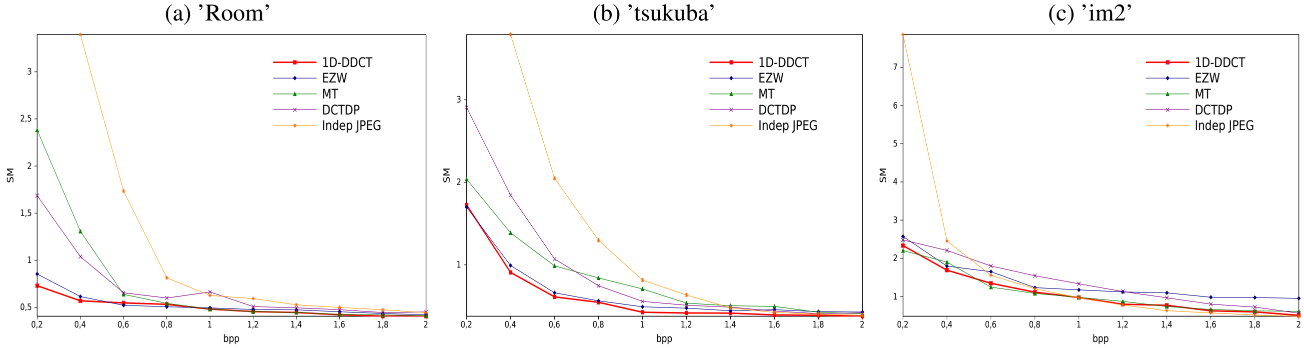
4. EXPERIMENTAL RESULTS

In this section, we report experimental evaluation of the proposed coder. First, the compression performance of the coder is evaluated. Second, the coder complexity is discussed. Two metrics are used for the evaluation, a perceptual metric 'SM' [20] designed for stereo image and the peak-signal-to-noise-ratio PSNR is used. The strength of the 'SM' metric is that it models the binocular suppression by estimating the local contrast [21] and the binocular Just-Noticeable-Difference [22]. In this study, the performance of the proposed coder is evaluated by analyzing the rate-distortion behavior. Three images are used in the experiments. 'Room' [1] image which is mostly used in the literature, 'tsukuba' from middleburry website [2] and 'im2', from LIVE 3D IQA database [3]. The experiments is performed only on gray level images, for that, the Matlab rgb2gray function is used to convert color to gray image. λ is fixed empirically to 0.5 (a little change of the λ value do not influences the performance of the coder) and k to 8. The SSD cost function is used for disparity map estimation. Three-level Haar-DWT is used to transform occluded blocks and to predict the dominant direction. Transform coefficients are grouped into a wavelet decomposition sub-band style structure and coded using the EZ algorithm to align with the state-of-the-art coders.

[1] http://www.duke.edu/web/isis/gessler/topics/stereo.htm
[2] http://vision.middlebury.edu/stereo/data/
[3] http://live.ece.utexas.edu/research/quality/live_3dimage_phase1.html



(a) 'Room' (b) 'tsukuba' (c) 'im2'



(d) 'Room' (e) 'tsukuba' (f) 'im2'

Fig. 5: The Rate-Distorsion curves for different stereo images. The 'SM' metric is used as a distorsion. For (a), (b) and (c) the left images are JPEG coded at quality 50 and for (d), (e) and (f) the left images are JPEG coded at quality 75.

	'Room'	'tsukuba'	'im2'	Avg
Parameters (D, w)	(12,8)	(16, 8)	(16, 8)	-
MT [14]	2.89	1.96	1.26	2.03
EZW [12]	2.45	0.40	1.07	1.30
DCTDP [6]	3.98	2.30	1.91	2.73

Table 1: The average decibel gain obtained by our coder relative to the considered disparity-compensated residual coders.

	'Room'	'tsukuba'	'im2'
Size	256×256	384×288	640×360
Time (s)	0.028	0.042	0.089

Table 2: Average of decoding time in second.

Figures 5 and 3 show the performance of the proposed coder. The left image is used as reference and JPEG coded at quality 50 and 75. The PSNR is computed using the MSE of the right image and the 'SM' metric accounts for the stereo quality. For all stereo pairs, the proposed coder outperforms the considered methods. For visual comparison, the results for 'Room' right view coded at 0.4 *bpp*, are displayed on Fig. 6 where the left image is JPEG coded at quality 75. It is shown that the reconstructed right view contains less artifacts than the ones obtained by the other considered methods.

Furthermore, table 1 shows that the proposed coder offers a significant average gain compared to the residual compensated methods, for a bit rate ranging from 0.2 to 2 *bpp* for the considered stereo pairs. The use of the EZ algorithm for quantizing the obtained coefficient is not very convincing. Indeed, it is known that the EZ algorithm yields excellent performance when using a dyadic image representation, which is not the case of the used transform and yet the coder is effective.

The MT, the EZW, the DCTDP and the proposed coder are implemented using C++ and simulation is conducted using machine equipped by an Intel Core i7 3.2 GHz CPU. The main complex task of the encoder is the disparity estimation, which is a common task for the considered methods. The remaining tasks complexity is similar to the decoder. Computational time of the decoder is given in Tab.2. It is demonstrated that our decoder offers a good computational performance, around 0.028 second in average for a bit rate in range 0.2 to 2 *bpp* for 'Room' image. Thus, it can be deployed in real-time applications. The complexity of the algorithm depends on two tasks, the adopted transformation, and the quantization process. The 1D-DDCT is less complex than the 2D-DCT and the DWT. The EZ algorithm influences the computational encoding/decoding time because it depends on the maximum value of coefficients which determines the number of dominant and subordinate passes. In addition, our coder performs for each block a DWT to predict the dominant edge direction. It is worth noting that the proposed coder do not use the fast implementation of the DCT, thus there is still of room to reduce its complexity.

5. CONCLUSION

In this paper, a novel approach for residual stereo image compression has been proposed. The coder is based on 1D-DDCT, which is well adapted for the 1D anisotropic edge structure of the residual. The coder has a good performance compared to the state-of-the-art. Moreover, the coder is characterized by its non complex computational load due to the unidirectional based transform despite the use of a non-optimal implementation of the DCT. An embedded algorithm is adopted to quantize the 1D-DDCT coefficients. In addi-

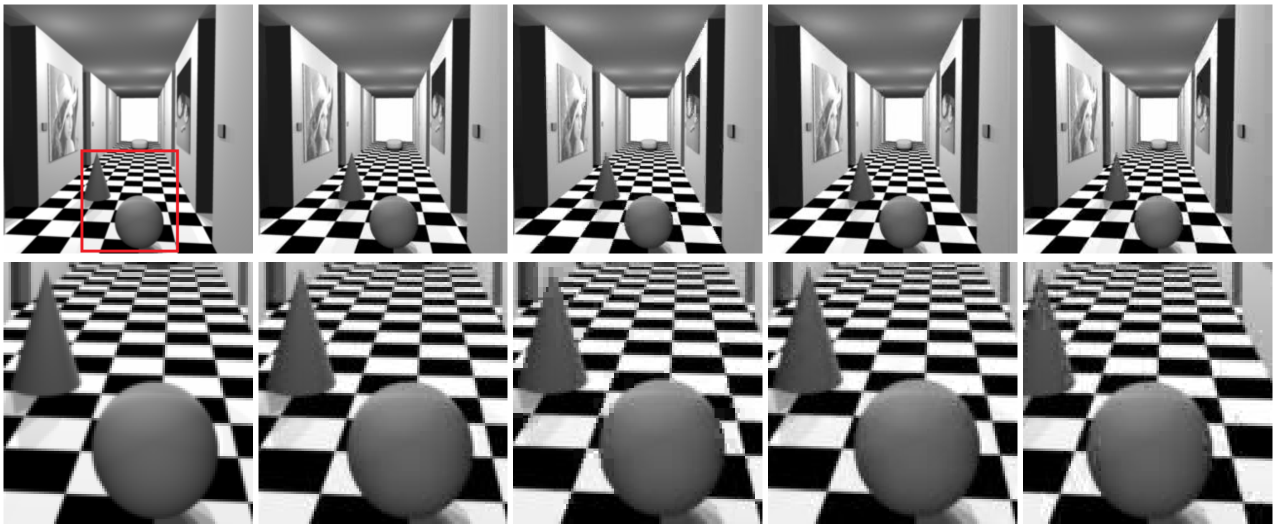


Fig. 6: Visual results of the right image of 'Room' stereo image coded at 0.4 *bpp* with red rectangle zoomed region, the left view is JPEG coded at 75 quality. From left to right : Original image, the proposed coder, the MT coder [14], the EZW coder [12] and the DCTDP coder [6].

tion, in this work, the block matching disparity estimation algorithm is used, which is not well adapted for a good inter-view decorrelation. In future work, more sophisticated disparity estimation algorithm will be investigated and other perceptual 3D factors will be also incorporated in the design of the coder.

6. REFERENCES

[1] G.K. Wallace, "The jpeg still picture compression standard," *Communications of the ACM*, pp. 30–44, 1991.

[2] D.S. Taubman, M.W. Marcellin, and M. Rabbani, "Jpeg2000: Image compression fundamentals, standards and practice," *JEI*, vol. 11, no. 2, pp. 286–287, 2002.

[3] M. Perkins, *Data compression of the stereopairs*, Ph.D. thesis, Stanford University CA, 1988.

[4] H. Yamaguchi, Y. Tatehira, K. Akiyama, and Y. Kobayashi, "Stereoscopic images disparity for predictive coding," *ICASSP*, pp. 1976–1979, 1989.

[5] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets," *ICASSP*, pp. 521–524, 1986.

[6] M. Perkins, "Data compression of stereopairs," *IEEE Trans on Communications*, vol. 40, no. 4, pp. 684–696, 1992.

[7] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Trans on CSVT*, vol. 10, no. 2, pp. 194 –200, 2000.

[8] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Trans on PAMI*, vol. 14, pp. 910–927, 1992.

[9] W. Woo and A. Ortega, "Stereo image compression with disparity compensation using the mrf model," in *SPIE VCIP*, 1996, pp. 28–41.

[10] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequences using joint 3-d motion/disparity compensation," *IEEE Trans on CSVT*, vol. 7, no. 2, pp. 312–327, 1997.

[11] M. S. Moellenhoff and M. W. Maier, "Transform coding of stereo image residuals," *IEEE Trans on Image Processing*, vol. 7, no. 6, pp. 804 –812, jun 1998.

[12] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Trans on CSVT*, vol. 12, no. 10, pp. 898–903, 2002.

[13] J. Jiang and E. A. Edirisinghe, "A hybrid scheme for low bit-rate coding of stereo images," *IEEE Trans on Image Processing*, vol. 11, no. 2, pp. 123–134, 2002.

[14] T. Frajka and K. Zeger, "Residual image coding for stereo image compression," *Optical Engineering*, vol. 42, no. 1, pp. 182–189, 2003.

[15] J. N. Ellinas and M. S. Sangriotis, "Stereo image compression using wavelet coefficients morphology," *IVC*, vol. 22, no. 4, pp. 281–290, 2004.

[16] L. F. R. Lucas, N. M. M. Rodrigues, E. A. B. da Silva, and S. M. M. de Faria, "Stereo image coding using dynamic template-matching prediction," in *EUROCON*, 2011, pp. 1–4.

[17] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans on Image Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[18] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the h.264/mpeg-4 avc standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626–642, 2011.

[19] B. Zeng and J. Fu, "Directional discrete cosine transforms for image coding," in *ICME*. IEEE, 2006, pp. 721–724.

[20] W. Hachicha, A. Beghdadi, and F.A Cheikh, "Stereo image quality assessment using a binocular just noticeable difference model," in *ICIP*, Melbourne, Australia, 2013.

[21] W. Hachicha, A. Beghdadi, and F.A Cheikh, "Combining depth information and local edge detection for stereo image enhancement," in *EUSIPCO*, Bucharest, Romania, 2012, pp. 250–254.

[22] Y. Zhao, Z. Chen, C. Zhu, Y-P. Tan, and L. Yu, "Binocular just-noticeable-difference model for stereoscopic images," *Signal Processing Letters, IEEE*, vol. 18, no. 1, pp. 19 –22, jan. 2011.