BLOCK DEPENDENT DICTIONARY BASED DISPARITY COMPENSATION FOR STEREO IMAGE CODING

G. Dauphin, M. Kaaniche, A. Mokraoui

L2TI, Institut Galilée, Université Paris 13 Sorbonne Paris Cité
99, Avenue Jean-Baptiste Clément 93430 Villetaneuse, France
{gabriel.dauphin, mounir.kaaniche, anissa.mokraoui}@univ-paris13.fr

ABSTRACT

With the recent advances in stereoscopic display technologies, there is a growing demand for designing efficient stereo image compression techniques. For this reason, a great attention should be paid to the disparity/estimation process used to generate the residual image. In this paper, we propose to improve the disparity compensation process in a typical closed-loop-based stereo image coding scheme. A new formulation of this process, based on a block dependent dictionary, is developed. More specifically, the main idea aims to link together the disparities yielding similar compensations and assign a common disparity candidate to each subset of disparities. Experimental results have shown the interest of the proposed method in terms of bitrate saving and quality of reconstruction.

***Index Terms*—** Stereoscopic image, disparity estimation, disparity compensation, compression, closed-loop structure.

1. INTRODUCTION

Stereo images correspond to a pair of views, called left and right images, obtained by recording the same scene from two slightly different view angles. These images play a key role in various applications fields such 3D-cinema, 3D-TV, robotics, video conferencing and tele-medicine [1]. As a result, a large amount of stereo data has been generated which needs to be efficiently encoded for storage and transmission purpose.
To this end, the intuitive coding approach consists of encoding separately each view by applying still image compression techniques. However, since the left and right images present similar contents, state-of-the-art methods resort to disparity compensation-based coding approaches [2]. Indeed, most pixels of a given view can be found in the other one thanks to a horizontal displacement referred to as a disparity map. Once the disparity map is available, most of the reported stereo coding schemes proceed as follows [3–5]. First, one image (for example the left one) is selected as the reference image. Then, the other one, called target image, is predicted from the reference one using the estimated disparity field. This step is known as disparity compensation. After that, the difference between the original right image and the predicted one, called residual image, is generated. Finally, the reference image and the residual one as well as the disparity map are encoded. Therefore, one of the key steps that has a great impact on the performance of stereo image coding schemes concerns the disparity estimation/compensation task. For this reason, the disparity estimation problem has been extensively studied in the literature [6, 7]. Generally, block-based or pixel-based approaches can be used [8–10]. While pixel-based approaches compute the disparity for each pixel of the image leading to a dense map, the block-based ones assume that the disparity is a blockwise constant. Among them, the block-matching technique has been widely used for video and stereo compression purpose [4, 11].
Based on the fact that different disparity maps can yield similar predicted right images (i.e. similar disparity-compensated images), we propose in this paper to exploit such additional redundancies that may exist among the disparities. More precisely, we provide an algorithm yielding a Block dependent Dictionary (BD) at the encoder and the decoder side. To this respect, disparities yielding similar compensations are linked together and have a common disparity candidate. Then, these subsets of disparities are labeled in the decreasing order of their likelihood.
The remainder of this paper is organized as follows. In Section 2, we recall some basic facts on the standard residual-based stereo coding scheme as well as the disparity estimation problem. In Section 3, we describe the proposed BD-based disparity compensation approach. Finally, in Section 4, experimental results are given and some conclusions are drawn in Section 5.

2. DISPARITY-COMPENSATED BASED STEREO IMAGE CODING

2.1. Standard coding structure

As mentioned before, the basic idea behind most of the reported stereo image coding methods consists of encoding a reference image I^l, a residual one I^e and a disparity map d.

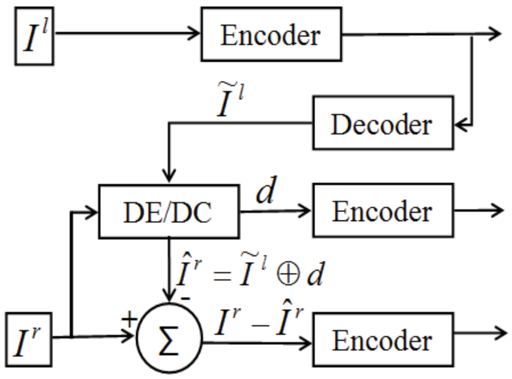


Fig. 1. Standard closed-loop-based stereo coding scheme.

Generally, we can distinguish two main structures according to the disparity estimation/compensation process.
The first one, referred to as open-loop structure, performs the disparity estimation/compensation (DE/DC) stage on the *original* reference image I^l at the encoder side. However, at the decoder side, the decoded reference image $\tilde{I}^l$ will be used during the disparity compensation stage. Thus, the main drawback of this structure is that the encoder and decoder side do not use the same reference image which makes it sub-optimal.
To alleviate this problem, a closed-loop structure has been developed [4, 8]. Thus, as shown in Fig. 1, the reference image (i.e. the left one I^l) is firstly encoded and decoded. Then, the disparity estimation/compensation stage is performed on the decoded left image $\tilde{I}^l$ to generate the residual image $I^e = I^r - \hat{I}^r$. We should note that previous works have shown the interest of this structure compared to the open-loop one in terms of reconstruction error [4]. For this reason, we focus in this paper on a closed-loop-based stereo image coding architecture.

2.2. Disparity estimation/compensation process

The disparity estimation/compensation stage plays a crucial role in designing efficient joint stereo image coding approaches. For instance, in the context of coding application, block-based ones (more specifically the block-matching technique) are preferred since a single disparity value is assigned to all the pixels within the block. This technique consists firstly in partitioning the target image (i.e. the right one I^r) into nonoverlapping blocks of size $b_x \times b_y$, which are then compared to the candidate blocks located in a search area S of the decoded left image (since a closed-loop structure is considered). Finally, the disparity map $d^m(i,j)$ for a current block in I^r is obtained by minimizing a given cost functional:

$$d^m(i,j) = \arg\min_{d \in S} \|I^r(i,j) - [\tilde{I}^l \oplus d](i,j)\| \tag{1}$$

where (i,j) are the spatial coordinates associated with the top leftmost pixel in the block, $[\tilde{I}^l \oplus d](i,j)$ represents the disparity compensated block associated to the possible candidate disparity value d, and $\|\cdot\|$ is the mean square root of the current block. Note that $I^r(i,j)$ and $[\tilde{I}^l \oplus d](i,j)$ are matrices of components $I^r(i+m, j+n)$ and $\tilde{I}^l(i+m, j+n+d)$ where $0 \le m \le b_x - 1$ and $0 \le n \le b_y - 1$.

3. BUILDING A BLOCK DEPENDENT DICTIONARY

The proposed coding scheme processes the right image as a set of nonoverlapping blocks. For each block, we build a dictionary that contains only candidate disparities which would yield images being sufficiently different from one to another. Contrary to the conventional disparity estimation method where both views are used to generate the disparity map, the proposed dictionary is built based *only* on the decoded left image I^l.
Note that the decoder has access to the same decoded image $\tilde{I}^l$ and can therefore build the same BD using the same algorithm. As a consequence, the only information to be transmitted is the index of the selected candidate disparity. Moreover, blocks having smooth areas will yield dictionaries of smaller size, where disparities are linked to a small number of candidate disparities.

3.1. Problem statement

The block dictionary, denoted by $\mathcal{A}$, is defined as a set of subsets of disparities $(\mathcal{D}_c)_c$, an application Φ transforming the possible disparity values into candidate disparities, and an application Ψ assigning labels to the candidate disparities:

$$\mathcal{A} = \left((\mathcal{D}_c)_{c \in \{1,\ldots,C\}}, \Phi, \Psi\right) \tag{2}$$

where C is the number of candidates, $\mathcal{D}_c$ is used for partitioning the range of disparities into subsets of disparities yielding similar predictions in a given block of the right image, and Φ is used to identify the specific candidate of the subset $\mathcal{D}_c$.
From a clustering view point, candidates should be seen as centers of subsets $(\mathcal{D}_c)_c$. Thus, $\Phi^{-1}(d)$ refers to the *subset* $\mathcal{D}_c$ containing d and whose center is $\Phi(d)$, and $\Psi(d)$ is the index that will be transmitted instead of the disparity d.
We are looking forward to find a dictionary while minimizing its size C subject to the following constraints:

- **C1** The set of subsets $(\mathcal{D}_c)_{c \in \{1,\ldots,C\}}$ is a partition of the range of possible disparities $\{0,\ldots,D-1\}$

$$\begin{cases} \bigcup_{c=1}^{C} \mathcal{D}_c = \{0,\ldots,D-1\} \\ \forall c \neq c',\ \mathcal{D}_c \bigcap \mathcal{D}_{c'} = \emptyset \end{cases} \tag{3}$$

- **C2** Each subset is bounded by a ball centered at $\Phi(d)$ and of radius T.

$$\forall c, \exists d,\ \text{such that } \mathcal{D}_c = \Phi^{-1}(d) \text{ and its center is } \Phi(d)$$
$$\forall d,\ \left\| [\tilde{I}^l \oplus d](i,j) - [\tilde{I}^l \oplus \Phi(d)](i,j) \right\| \le T \tag{4}$$

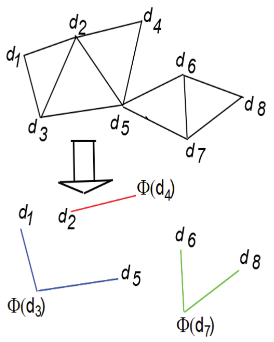
where T is a parameter representing the upperbound of the distortion. This parameter achieves a trade-off between the quality of the predicted image and the coding cost of the index.

- **C3** $\Psi(d)$ is a non-increasing sequence of the size of $\mathcal{D}_c$ containing d:

$$\forall d, d', \text{ if } |\Phi^{-1}(d)| \leq |\Phi^{-1}(d')|, \text{ then } \Psi(d) \geq \Psi(d') \quad (5)$$

where $|\cdot|$ refers to the size of a given set.

This constraint indicates that the greater the size of subsets $|\mathcal{D}_c|$ is, the more likely this disparity candidate will be selected.
Figure 2 illustrates the objective of building a BD for an example of eight possible disparity values $\{d_1, \ldots, d_8\}$.



$$\begin{aligned} d_3 &= \Phi(d_1) = \Phi(d_3) \\ &= \Phi(d_5) \\ d_4 &= \Phi(d_2) = \Phi(d_4) \\ d_7 &= \Phi(d_6) = \Phi(d_7) \\ &= \Phi(d_8) \end{aligned}$$

$$\downarrow$$

$$\begin{aligned} |\Phi^{-1}(d_3)| &= 3 \\ |\Phi^{-1}(d_4)| &= 2 \\ |\Phi^{-1}(d_7)| &= 3 \end{aligned}$$

$$\downarrow$$

$$\begin{aligned} \Psi(d_3) &= 1 \\ \Psi(d_4) &= 3 \\ \Psi(d_7) &= 2 \end{aligned}$$

Fig. 2. Up: example of graph connecting disparities d_p and d_q when $\|[\tilde{I}^l \oplus d_p](i,j) - [\tilde{I}^l \oplus d_q](i,j)\| \leq T$. Down: example of three clusters with centers d_3, d_4, d_7. Right: example of sorted labels $\Psi(d_3), \Psi(d_4), \Psi(d_7)$ with respect to the size of the clusters $|\Phi^{-1}(d_3)|, |\Phi^{-1}(d_4)|, |\Phi^{-1}(d_7)|$.

In this example, all disparities are connected with a black line on the upper graph of the figure. The lower graph displays with different color lines the subsets of disparities that fulfill the constraints mentioned above. Finally, on the right side of the figure, we give the values of Φ and Ψ corresponding to this configuration.

3.2. Proposed algorithm

We propose in this part a new solution for finding the BD. The following procedure is performed for each block:

- **S1** For each disparity $d \in \{0, \ldots, D-1\}$, if d is too far from any center, then there is a new candidate which, at first, is considered as the center of the subset containing only itself.

$$\text{If } \|[\tilde{I}^l \oplus d](i,j) - [\tilde{I}^l \oplus \Phi(d)](i,j)\| > T, \quad \text{then } \begin{array}{l} C := C+1 \\ \Phi(d) = d \end{array} \quad (6)$$

- **S2** Link each disparity with the nearest center without changing the subset of all centers $\{\Phi(0), \ldots, \Phi(D-1)\}$

$$\Phi(d) = \underset{d' \in \{\Phi(0), \ldots, \Phi(D-1)\}}{\arg\min} \|[\tilde{I}^l \oplus d](i,j) - [\tilde{I}^l \oplus d'](i,j)\|$$

- **S3** Choose the best center for each subset of disparities without changing these subsets of disparities $\Phi^{-1}(d)$

$$\Phi(d) = \underset{d'' \in \Phi^{-1}(d)}{\arg\min} \sum_{d' \in \Phi^{-1}(d)} \|[\tilde{I}^l \oplus d''](i,j) - [\tilde{I}^l \oplus d'](i,j)\|^2$$

- **S4** Remove each pair of centers (d, d') that is too close to each other

  for each center d
    for each center d'
      if $\|[\tilde{I}^l \oplus \Phi(d)](i,j) - [\tilde{I}^l \oplus \Phi(d')](i,j)\| \leq T$,
      then remove center d'

- **S5** Repeat steps S1, S2, S3 and S4 as long as the dictionary is still evolving in a non-periodic fashion.

- **S6** Find Ψ such that constraint $C3$ is satisfied using a sorting process.

It is important to note that, even though there is not a unique Ψ that fulfills (5), we can guarantee that the BD will be the same at the encoder and the decoder side since they use the same sorting strategy.

3.3. BD-based compensation

Once the BD is built and therefore Φ and Ψ have been defined, we need to transmit for each block in I^r an index $s(i,j)$ given by:

$$s(i,j) = \Psi\Big(\underset{d \in \{\Phi(0), \ldots, \Phi(D-1)\}}{\arg\min} \Big\| I^r(i,j) - [\tilde{I}^l \oplus d](i,j) \Big\| \Big) \quad (7)$$

Thus, the predicted block $\hat{I}^r(i,j)$ is obtained thanks to the index s as follows:

$$\hat{I}^r(i,j) = [\tilde{I}^l \oplus \Psi^{-1}(s)](i,j) \quad (8)$$

where $\Psi^{-1}(s)$ refers to the disparity centering the subset labeled as s by Ψ.

Stereo Images	Block Dictionary Bitrate	PSNR	Block Matching Bitrate	PSNR
DEIMOS_8	0.045	25.49	0.068	25.56
DEIMOS_13	0.060	24.91	0.072	24.96
DEIMOS_67	0.037	30.72	0.064	30.99
JISCT_Synth	0.073	23.76	0.083	23.86
CMU_Fruit	0.060	18.36	0.086	18.56
SYNTIM_Angle	0.030	29.18	0.054	29.78

Table 1. Bitrate (in bpp) of the disparity maps and the distortion (in terms of PSNR) of their corresponding disparity-compensated left images.

	BD-Compensation vs BM-Compensation		BD-Compensation vs Independent	
Stereo Images	low	middle	low	middle
DEIMOS_8	0.35	-0.04	2.24	0.61
DEIMOS_13	0.12	0.01	1.22	0.87
JISCT_Synth	0.17	-0.06	0.12	0.42
SYNTIM_Angle	0.02	-0.26	0.01	-0.46

Table 2. Average PSNR difference at low and middle bitrates. The gain of the BD technique w.r.t the BM-based residual coding scheme and the Independent one.

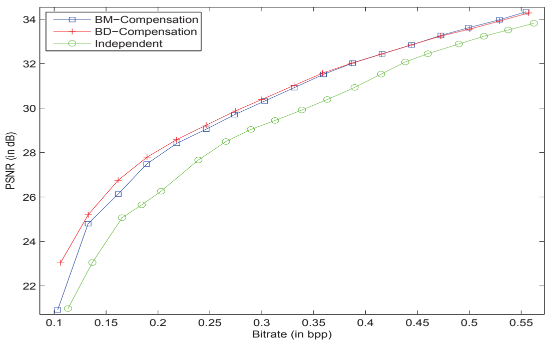


Fig. 3. PSNR (in dB) versus the bitrate (in bpp) for the stereo pair "DEIMOS_8".

4. EXPERIMENTAL RESULTS

The simulations are performed on different stereo images downloaded from various standard databases such as CMU[1], JISCT[2], SYNTIM database[3], and DEIMOS[4] [12]. The disparity estimation techniques are performed by using a 8×8 block size. Concerning the proposed BD-based approach, the parameter T is set to 20.6. In what follows, we illustrate the impact of the proposed method on the disparity compensation stage as well as the overall stereo image compression performance. To this end, we firstly compare it with the standard block-matching technique in terms of the quality of the predicted right image. Recall that this latter corresponds to the disparity compensated left image used generally in residual-based stereo image coding schemes.

Table 1 gives the bitrate of the resulting disparity maps as well as the PSNR of the predicted right images. Note that these results are obtained by using the original left image as a reference and the disparity maps are encoded with the arithmetic encoder. As it can be seen in Table 1, while the two methods lead to similar quality of reconstruction, our approach presents the advantage of reducing the bitrate of the disparity maps.

In addition, we have also evaluated these methods in the context of a typical closed-loop-based stereo image coding scheme. As mentioned before, the reference image is firstly encoded and decoded by using the 9/7 wavelet transform retained in the lossy compression mode of the standard JPEG2000 [13]. Then, the disparity compensation steps based on the Block-Matching (BM) technique and the proposed BD one are performed to generate the residual images. Finally, the 9/7 transform is also applied to the obtained residual images. These schemes, denoted in what follows by "BM-Compensation" and "BD-Compensation", will also be compared with the "Independent" scheme where the 9/7 transform is applied separately to the left and right images (without disparity estimation/compensation process).

Fig. 3 illustrates the variations of the average PSNR versus the average bitrate of the stereo pair "DEIMOS_8" for these different methods. It can be noticed that the BD-based approach leads to an improvement compared with the classical BM-technique, especially at low bitrate, where the improvement can reach 1.5 dB. We have also evaluated these methods with the Bjontegaard metric [14]. The results are provided in Table 2 for low and middle bitrates corresponding respectively to the four bitrate points $\{0.11, 0.13, 0.22, 0.25\}$ and $\{0.42, 0.47, 0.53, 0.58\}$ bpp. It can also be observed that our method clearly outperforms the independent coding scheme. Moreover, compared to the BM-based residual coding scheme, our method results in an average gain of about 0.02-0.35 dB.

5. CONCLUSION

In this paper, we have focused on the disparity estimation/compensation stage to improve the performance of residual-based stereo image coding schemes. To this end, a new formulation based on a block-dependent dictionary has been addressed. Experimental results have shown the benefits of the proposed approach in the context of stereo image coding.

[1] http://vasc.ri.cmu.edu/idb/html/stereo/
[2] http://vasc.ri.cmu.edu/idb/html/jisct/
[3] http://perso.lcpc.fr/tarel.jean-philippe/syntim/paires.html
[4] http://www.deimos-project.cz/tag/stereo

6. REFERENCES

[1] I. Feldmann, W. Waizenegger, N. Atzpadin, and O. Schreer, "Real-time depth estimation for immersive 3D videoconferencing," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, Tampere, June 2010, pp. 1–4.

[2] M.G. Perkins, "Data compression of stereo pairs," *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 684–696, April 1992.

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

[4] N. V. Boulgouris and M. G. Strintzis, "A family of wavelet-based stereo image coders," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 898–903, October 2002.

[5] R. Darazi, A. Gouze, and B. Macq, "Adaptive lifting scheme-based method for joint coding 3D-stereo images with luminance correction and optimized prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, April 2009, pp. 917–920.

[6] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frames stereo correspondance algorithms," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 7–42, 2002.

[7] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 993–1008, 2003.

[8] W. Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 194–200, 2000.

[9] I. Daribo, M. Kaaniche, W. Miled, M. Cagnazzo, and B. Pesquet-Popescu, "Dense disparity estimation in multiview video coding," in *IEEE International Workshop on Multimedia Signal Processing*, Rio de Janeiro, Brazil, October 2009, pp. 1–6.

[10] M. El Gheche, J.-C. Pesquet, J. Farah, M. Kaaniche, and B. Pesquet-Popescu, "Proximal splitting methods for depth estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2011, pp. 853–856.

[11] S. Wang and H. Chen, "An improved algorithm of motion compensation MPEG video compression," in *IEEE International Vehicle Electronics Conference*, China, September 1999, vol. 1, pp. 261–264.

[12] M. Klima, K. Fliegel, P. Pata, S. Vitek, M. Blazek, P. Dostal, L. Krasula, T. Kratochvil, V. Ricny, M.Slanina, L. Polak, O. Kaller, and L. Bolecek, "DEIMOS-an open source image database," *Radioengineering*, vol. 20, no. 4, pp. 1016–1023, December 2011.

[13] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Kluwer Academic Publishers, Norwell, MA, USA, 2001.

[14] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," Tech. Rep., ITU SG16 VCEG-M33, Austin, TX, USA, April 2001.