Improving block-matching algorithm by selecting sets of disparities minimizing distortion for stereoscopic image coding

Aysha Kadaikar, Gabriel Dauphin and Anissa Mokraoui
L2TI, Institut Galilée, Université Paris 13 Sorbonne Paris Cité
99 avenue Jean-Baptiste Clément 93430 Villetaneuse, France

***Abstract*—This paper deals with the estimation of a blockwise disparity map in the context of stereoscopic image coding. This disparity map is used to predict one view using the other view as reference. It is generally computed according to the Block Matching algorithm which achieves good performance in terms of bitrate-distortion. Namely, disparities are selected amongst a search area by minimizing a local distortion metric modelling the quality of the prediction. Note that the larger the search area is, the more often a better disparity can be chosen and the lower the global distortion is. On the other hand, using large search areas yield disparity maps containing a higher number of different disparities, and such disparity maps are generally encoded with a larger bitrate. In this paper two algorithms are proposed, they enable to compute collections of good search areas among which one can find a specific search area minimizing the distortion at a given bitrate. Simulations results confirms the benefits of both algorithms.**

I. Introduction

The number of applications, related to 3D contents, has been growing in the last decades [1], [2] as they allow a greater immersion in the scene to the viewers. Stereoscopic images can yield a perception of a 3D scene when the left view is seen by the left eye and the right view is seen by the right eye. Such perception arises as objects have a slightly different location on the right view than that of the left view. This small displacement is called the disparity, it is not constant as its value is related to the perception of depth.

Coding a stereoscopic image can be carried out by processing seperately each view. Better performance is obtained by making use of cross-view redundancies, one view being predicted using the other view. This approach is known as the disparity compensated coding scheme, it takes the following steps.

(i) One view (here the left one) is taken as the reference one, it is independently coded.
(ii) As a mean to predict the right view using the left view, a disparity map is estimated and losslessly encoded.
(iii) The prediction error between the right view and the predicted right view is called the residual image, it is encoded.

The prediction is achieved using a blockwise processing: both views are divided into non-overlapping blocks and the disparity map is the collection of all disparities, one for each block. This disparity map is usually encoded with an entropy-based coder, while the left view and the residual image are encoded with Discrete-cosine-transform-based coders.

Research has achieved improvements in bitrate-distortion performance when coding stereoscopic images by focusing mainly on three aspects: the estimation of the disparity map [3], [4], [5], the entropy-coding of the disparity map [6] and the choice of the transform applied to the left view and to the residual image [7], [8], [9]. Our work concerns the first aspect. In this context, a well-known algorithm for disparity map estimation is the Block-Matching Algorithm (BMA). For each block of the predicted view, a disparity is selected amongst a set of disparities by minimizing a local distortion metric (generally the Sum of Absolute Differences or the Sum of Squared Differences) computed between the right view and the predicted view. This set of disparities is called a search area as this selecting process is indeed searching on a small area of the left view a block that would be most similar to the given block of the right view. The larger the search area is and the better the quality of the predicted image is thanks to more adequate choices of disparities. But enlarging the search area can also result in a more expensive disparity map in terms of bitrate, as the range of the selected disparities may also increase. Generally search areas have a rectangular shape, our contribution is to consider any set of disparities contained in a larger rectangular shape, the choice of the set being based only on how it enables the BMA to have a good bitrate-distortion performance. The rest of the paper is organized as follows. Section II shows how selecting good search areas can be regarded as an optimization problem. Two algorithms are proposed in section III. Simulation results are discussed in section IV and finally section V concludes this work.

II. Selecting of good search area: an optimization problem

This section deals with the problem of selecting a search area in such a way as to minimize the distortion of the predicted view for a given bitrate associated to the disparity map. Let us first introduce some notations.

A. Notations

I_l represents the left view taken as the reference view and I_r the right view to be predicted. Both are of size $K \times L$ pixels and are divided into non-overlapping blocks of the same size.

$I_k(i,j)$ (with $k \in \{l,r\}$) represents the intensity of the pixel located at position (i,j) in the corresponding view. Each block of the right view is associated to a unique two-coordinates disparity $\underline{d}$ chosen amongst a set of disparities S containing $|S|$ different disparities. The predicted view is denoted $\widehat{I}_{r,D}$ as it depends on the disparity map D containing each selected disparity for each block. The set of all different disparities in D is denoted $\mathrm{Co}(D)$ as it can be regarded as the codomain of the mapping function D.

The BMA computes, for a given search area S, the best disparity map $D(S)$ in terms of global distortion of the predicted view, here measured as the Sum of Squared Differences:

$$D(S) = \underset{\{D|\,Co(D)\subset S\}}{\arg\min} \sum_{i=0}^{K-1}\sum_{j=0}^{L-1}\left(\widehat{I}_{r,D}(i,j) - I_r(i,j)\right)^2 \quad (1)$$

where $\{D|\,\mathrm{Co}(D) \subset S\}$ is the set containing all disparity maps whose codomain is included in S. This minimization is achieved in an efficient manner as for each block the disparity is selected by minimizing a local distortion metric regardless of the disparities selected for the other blocks. The global distortion of the predicted view depends also on S in the following way:

$$E(S) = \min_{\{D|\,Co(D)\subset S\}} \sum_{i=0}^{K-1}\sum_{j=0}^{L-1}\left(\widehat{I}_{r,D}(i,j) - I_r(i,j)\right)^2 \quad (2)$$

where $\{D|\,\mathrm{Co}(D) \subset S\}$ is the set containing all disparity maps whose codomain is included in S. The bitrate of the disparity map is approximated by the measure of its entropy :

$$H(S) = -\sum_{\underline{s}\in S} P(\underline{d}=\underline{s})\log_2\left(P(\underline{d}=\underline{s})\right) \quad (3)$$

where $P(\underline{d}=\underline{s})$ is the occurrence frequency of the disparity $\underline{s}$ in the map $D(S)$.

For practical reasons, the set S is selected among the subsets of a larger rectangular window denoted W and the set of all such subsets is denoted $\mathcal{P}(W)$.

B. Problem statement

In the coding context, finding the best search area S is minimizing the global distortion for a given bitrate b.

$$S = \arg\min_{\substack{S'\subset\mathcal{P}(W)\\ H(S')\leq b}} E(S') \quad (4)$$

Instead of this intractable optimization problem we address the following problem, where S is selected as the subset minimizing the global distortion being composed of N different disparities. The entropy constraint is no longer taken into account. To further reduce the complexity of the problem, we replace W by the set of different disparities contained in the disparity map associated to W, that is $W_0 = \mathrm{Co}(D(W))$:

$$S = \arg\min_{\substack{S'\subset\mathcal{P}(W_0)\\ |S'|=n}} E(S') \quad (5)$$

(5) can be used to yield a suboptimal solution of (4). Let us assume we can solve (5) for n ranging from 1 to $|W_0|$, we get a family of sets S_n having decreasing values of global distortion. Indeed increasing the number of disparities enable to choose adequate disparities and to reduce the global distortion. As the size of S_n increases, it is expected that the entropy increases, (note that it may not increase). A suboptimal solution of (4) is obtained by selecting the highest index n^* for which $H(S_{n^*}) \leq b$ and considering S_{n^*}.

The two proposed algorithms to solve (5) are described in the next section.

III. Proposed algorithms

This section discusses the complexity of finding the optimal solution of (5) and the proposed algorithms are presented.

A. Optimal solution of (5)

To solve (5), one method consists in processing the BMA on all possible sets $S \in \mathcal{P}(W_0)$ which are of size n disparities. The optimal solution is the set S for which the global distortion of the predicted image is minimized. Depending on the size of W_0 and the value of n, this can rapidly become a complex combinatorial problem for which this method would require a long processing time. Let us consider as an example, a rectangular window W containing $121 \times 3 = 363$ disparities, finding the set of 15 disparities minimizing the global distortion of the predicted image implies to process $\binom{363}{15} \approx 10^{26}$ sets. Even if this window was replaced by a small search area W_0 containing 33 disparities, we would still have to process $\binom{33}{15} \approx 10^9$ sets. To overcome this issue, two suboptimal algorithms are presented in the following subsections.

B. Selecting the set of disparities using BMA_H

The first algorithm, called BMA_H, consists in processing the BMA on the window W so as to find the disparity map $D(W)$. The disparities contained in $D(W)$ are sorted by decreasing order of their occurrence frequency in $D(W)$. We define S as the set containing the n first disparities, this set fullfills the contraint $S \in \mathcal{P}(W_0)$ and $|S| = n$, and it is the suboptimal solution of (5) proposed by this algorithm.

This algorithm is interesting in terms of numerical complexity as it requires only little more time processing than that when using the BMA on W. However when n is much lower than $|W_0|$, we should not expect good performance in terms of reducing the global distortion. Indeed assessing the utility of a given disparity in a given set S depends on the other disparities of S. It is likely that some of the disparities of an optimal set for n small may have little occurence frequency in $D(W)$ and thus be discarded when using BMA_H.

In an attempt to overcome this issue, the following method is being proposed.

C. Selecting the set of disparities using BMA_S

The second algorithm finds a suboptimal solution of (5) using a sequential reduction of the size of S from $|W_0|$ down to n by considering first $S = W_0$ and at each step by pruning one disparity, the one for which the global disparity is being the less increased.

Let us assume that with this algorithm, we have already found a set S of size $|S| = m$. There are m subsets of S of size $|S'| = m-1$, there are defined as $S\backslash\{\underline{s}\}$ for all possible disparity $\underline{s} \in S$. Selecting the best set S' is solved by

$$\begin{array}{rcl} s & = & \arg\min_{\underline{s}' \in S} E\left(S\backslash\{\underline{s}'\}\right) \\ S' & = & S\backslash\{\underline{s}\} \end{array} \quad (6)$$

Note that even though (6) suggests processing the BMA m times at each step and hence $|W_0|!$ times when using BMA_S to solve (5) for all possible values of n, it is actually possible to reduce the numerical complexity. Indeed at each step, it suffices to apply a modified version of the BMA. While processing each block this modified BMA keeps account not only of the best disparity and the lowest local distortion, but it also keeps account of the lowest local distortion in case this best disparity is being pruned. Once the whole image is processed, selecting the disparity to be pruned is achieved by minimizing the lowest increase of the global distortion for each disparity, using the collected information.

Note that this algorithm is not optimal as the minimization is computed on a restricted domain, namely $\{S' : \exists \underline{s} \in S \text{ and } S' = S\backslash\{\underline{s}\}\}$ instead of $\{S' : S' \in \mathcal{P}(W_0) \text{ and } |S'| = n\}$. The rationale is that if S is a good choice it is likely to have a good choice for S', but just as for BMA_H, it remains possible that a disparity of a set solving optimally (5) for n small is pruned by the BMA_S while processing sets of a larger size.

IV. Experimental results

This section presents some simulation results to evaluate the performance of the two proposed algorithms in comparison with the BMA in terms of their ability to predict the right view. We think that such simulations give a basic approximation of the performance that would be obtained with the complete disparity compensated scheme. Performance is given as the Peak Signal-to-Noise Ratio (PSNR) evaluated between the original and the predicted right views versus the bitrate of the disparity map in bits per pixel (bpp) approximated by the measure of its entropy as computed in (3).

Simulations are conducted on colourless stereoscopic images from the CMU-VASC database [10] using a window W containing disparities ranging horizontally from -60 to 60 and vertically from -1 to 1. Indeed simulations have shown that disparities having a negative horizontal component and a small vertical component could clearly improve performance. On one hand it seems that the human visual system is able to cope with such disparities and still perceive depth and on the other hand sterescopic cameras may not always be that precise. In the coding context, since such stereoscopic images are being used, coding schemes should take such redundancies into account.

A. Comparing BMA_H and BMA_S and the optimal solution on a small stereoscopic image

We first consider a small stereoscopic image of size 64×64 pixels extracted from the stereo set "sand". Simulations have been conducted using blocks of size 10×10 pixels. Fig. 1 shows the bitrate-distortion performance of both algorithms, using a dashed curve joining black circles for the BMA_H and a solid line joining red circles for the BMA_S. Both curves start from the same point corresponding to the BMA which yields a disparity map $D(W)$ containing 33 disparities. The circles of each curve show the performance for a given size of the search area S ranging from 33 to 1. Actually the simulation is not possible for sets yielded using the BMA_H for sizes below 2 as the BMA cannot process some blocks of the stereoscopic image for which none of the disparities, contained in the yielded sets, could be used. Such an error occurs for exemple when processing blocks on the right side with sets having only negative disparities. This experiment shows that the BMA_S achieves better performance than the BMA_H.

The blue cross in Fig. 1 shows the performance achieved with the optimal set of 5 disparities found after processing all possible sets of 5 disparities amongst the 33 disparities of W_0, meaning that $\binom{33}{5} = 237336$ sets have been processed! In this experiment, the optimal set appear to have exactly the same performance as the one found using the BMA_S for $n = 5$.

B. Comparing bitrate distortion performance using different sizes of blocks on the stereoscopic image "house1"

The next experiment compares the performance of both algorithms with the BMA on the stereoscopic image "house1". Simulations have been carried on using blocks of size 4×4 to 12×12 pixels. Results are shown in Fig. 2. The performance of the BMA_S is represented by the set of red curves in solid line joining circles while the performance of the BMA_H is represented by the set of black dashed curves joining circles. For each algorithm, each curve is obtained using a different size of blocks and each point of each curve corresponds to a different size of searching area. Note that for a given size of block, the curves representing the performance of both algorithm start from the same point as it corresponds to the BMA using this size of blocks. The dashed blue line joining squares represents the performance of the BMA where each squares corresponds to a specific size of blocks. Note that the BMA_S performs better than the BMA_H and BMA in terms of bitrate-distortion. For a given bitrate of the disparity map, the BMA_S predicts the right view with a greater precision as compared both to the BMA and the BMA_H. For example, at a bitrate of 0.09bpp, the BMA and the BMA_H both leads to a predicted image of quality 26.4dB while the BMA_S yields to 28.6dB resulting in a gain of 2.2dB.

C. Comparing different predicted right views using different sizes of blocks on the stereoscopic image "house1"

We compare the predicted view yielded by processing the BMA using the different search areas computed with the three algorithms at a bitrate of 0.09bpp. The original right view is shown on Fig. 3, the white square frames a specific area. This area is shown as close-ups of the three predicted views computed respectively with BMA, BMA_H, and BMA_S. One clearly sees that the window of the house is better predicted using the BMA_S.

Note that this higher perfomance is achieved using blocks of smaller size for BMA_S than those used for BMA_H and for BMA as these two algorithms would not have coped with such small blocks and the bitrate constraint of 0.09dB.

D. Comparing performance as measured by the Bjontegaard metric on "house1"

An average PSNR difference can be calculated between the BMA_S and the BMA on one hand, and between the BMA_S and the BMA_H on the second hand using the Bjontgaard's metric, [11]. The average PSNR difference is calculated for low and medium bitrates, with target bitrates measured in bpp and chosen as follows [0.06 0.07 0.08 0.09] and [0.1 0.25 0.40 0.55]. For each target bitrate, the point achieving the best performance in terms of PSNR under the bitrate constraint is retained for the computation of the Bjontegaard metric for each algorithm. The retained points for the three algorithms considering the target bitrates mentioned previously are represented in Fig. 5.

For example, considering a target bitrate of 0.55bpp, the best points selected for the three algorithms are identical and correspond to the very first point (at the top right) of each curve as this is the best point in terms of PSNR under this bitrate for all algorithms. This point is also considered as the best one in terms of PSNR for the target bitrate equal to 0.40bpp as the bitrate of this point (0.36bpp) is inferior to it.

At low bitrate, the average PSNR difference of the BMA_S as compared to the BMA is of 1.40dB and the average PSNR difference of the BMA_S as compared to the BMA_H is of 1.36dB. At medium bitrate, the average PSNR difference of the BMA_S as compared to the two other algorithms is of 0.59dB.

E. Comparing performance as measured by the Bjontegaard metric on nine stereoscopic images

Table. I contains the relative performance of the algorithms computed with the Bjontegaard metric at low bitrates and at medium bitrates. Simulations have been conducted on nine stereoscopic images from the CMU-VASC database which are: "whouse", "wdc2r", "toys", 'telephone", "sphere", "rubik", "mars1r", "house2" and "house1". Note the values computed in the former subsection have been reported on this table at the last line. The columns ΔBMA at low and medium bitrate give the average PSNR difference of the BMA_S as compared to the BMA. The columns ΔBMA_H at low and medium bitrate give the average PSNR difference of the BMA_S as compared to the BMA_H.

One clearly sees the improvement of the BMA_S as compared both to the BMA and the BMA_H in terms of bitrate-distortion performance. Note that the BMA_H achieves in most case better performance than the BMA.

TABLE I
GAIN OF THE BMA_S OVER THE BMA AND THE BMA_H

Stereo	**Low bitrate**		**Medium bitrate**	
Images	**ΔBMA**	**ΔBMA_H**	**ΔBMA**	**ΔBMA_H**
mars1r	0.30	0.10	0.35	0.20
whouse	0.33	0.21	0.35	0.24
sphere	0.37	0.20	0.47	0.37
wdc2r	0.39	0.12	0.47	0.29
telephone	0.48	0.29	0.46	0.42
rubik	0.52	0.36	0.22	0.22
toys	1.64	0.25	0.56	0.28
house2	1.81	1.21	1.56	1.54
house1	1.40	1.36	0.59	0.59

V. CONCLUSION

Two algorithms have been presented in this paper to take advantage of a large search area, allowing better prediction of the right view, while reducing the bitrate required to encode the disparity map. This is achieved by selecting an appropriate searching area and by processing the BMA with this specific searching area.

Simulation results conducted on several stereoscopic images have shown the benefits of both proposed algorithms as compared to the BMA with a particular advantage for the second one, the BMA_S.

In future works, the implementation of these two methods in a complete disparity compensated scheme will be investigated.

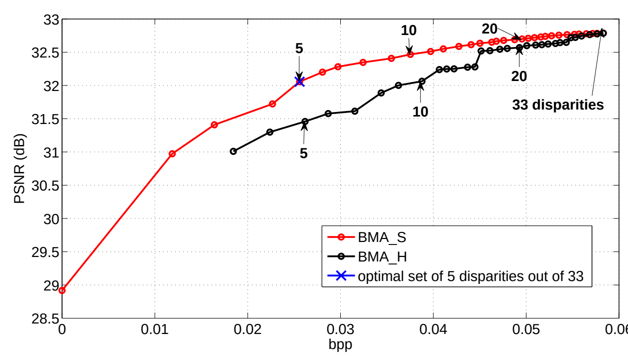

Fig. 1. Rate-Distortion performance on a small stereoscopic image extracted from "sand".

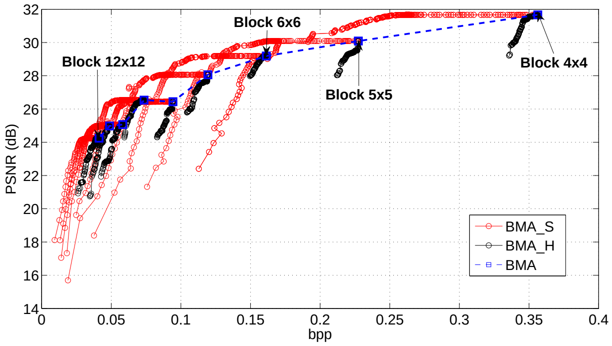

Fig. 2. Rate-Distortion performance on the stereoscopic image "house1".

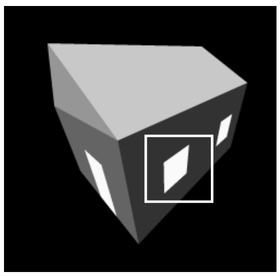

Fig. 3. "house1" original right image.

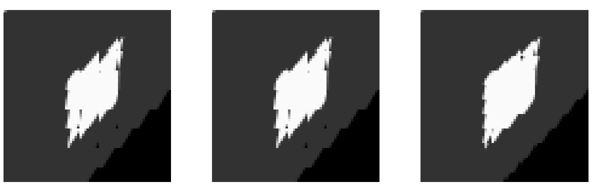

Fig. 4. Close-up of the predicted right view using the BMA (left figure), the BMA_H (middle figure) and the BMA_S (right figure).

REFERENCES

[1] N. A. Dodgson "Autostereoscopic 3D Displays," *IEEE Computer*, vol. 38(8), pp. 31–36, Aug. 2005.

[2] P. Kauff and O. Schreer "An immersive 3D video-conferencing system using shared virtual team user environments," *Proceedings of the 4th international conference on Collaborative virtual environments*, pp. 105–112, Sept. 2002.

[3] W. Woo, A. Ortega, "Stereo image compression based on disparity field segmentation," *SPIE Conference on Visual Communications and Image Processing*, vol. 3024, pages 391–402, 1997

[4] W. Woo, A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *Circuits and Systems for Video Technology, IEEE Transactions*, vol. 10(2), pp. 194–200, 2000.

[5] A. Kadaikar, G. Dauphin and A. Mokraoui, "Sequential block-based disparity map estimation algorithm for stereoscopic image coding," *Elsevier Journal, Signal Processing : Image Communication*, September 2015, DOI 10.1016/j.image.2015.09.007.

[6] J. Ellinas, M. S. Sangriotis, "Morphological wavelet-based stereo image coders," *IJournal of Visual Communication and Image Representation*, vol. 17(4), pp. 686–700, 2006.

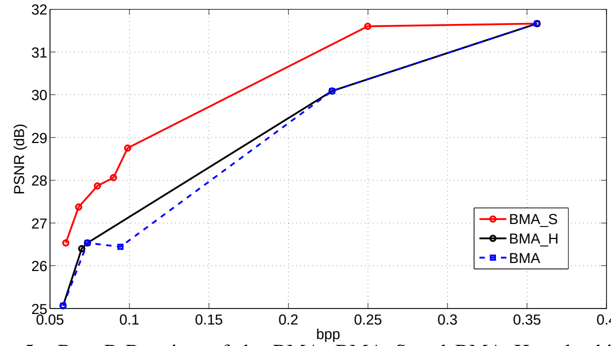

Fig. 5. Best R-D points of the BMA, BMA_S and BMA_H under bitrate constraints.

[7] M. Kaaniche, J. C. Pesquet, A. Benazza-Benyahia and B. Pesquet-Popescu, "Two-dimensional non separable adaptive lifting scheme for still and stereo image coding," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1298–1301, 2010.

[8] T. Frajka, K. Zeger, "Residual image coding for stereo image compression," *Optical Engineering*, vol. 42(1), pp. 182–189, 2003.

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

[10] "CMU Vision and Autonomous Systems Center (VASC) Image Database," "URL: http://vasc.ri.cmu.edu/idb/.

[11] G. Bjontegaard, "Calculation of average PSNR differences between RD," in *ITU-T SC16/Q6, 13th VCEG Meeting*, Austin, Texas, USA, April 2001, Document VCEG-M33.