ENTROPY-CONSTRAINED DENSE DISPARITY MAP ESTIMATION ALGORITHM FOR STEREOSCOPIC IMAGES

Aysha Kadaikar, Anissa Mokraoui, Gabriel Dauphin

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

ABSTRACT

This paper deals with the stereo matching problem to estimate a dense disparity map. Traditionally a matching metric such as mean square error distortion is adopted to select the best matches associated with disparities. However several disparities related to a given pixel may satisfy the distortion criterion although quite often the choice that is made does not necessarily meet the coding objective. An entropy-constrained disparity optimization approach is developed where the traditional matching metric is replaced by a joint entropy-distortion metric so that the selected disparities reduce not only the reconstructed image distortion but also the entropy disparity. The algorithm sequentially builds a tree avoiding a full search and ensuring good rate-distortion performance. At each tree depth, the M-best retained paths are extended to build new paths to which are assigned entropy-distortion metrics. Simulations show that our algorithm provides better results than dynamic programming algorithm.

Index Terms— Stereoscopic images, matching, disparity, entropy, optimization

1. INTRODUCTION

Stereoscopic image (3D image) is composed of an image pair captured by a stereo camera which provides the same scene for the right and left eye. Hence it requires twice the amount of information to be transmitted or stored compared to a traditional image (2D image). Therefore to compress efficiently stereoscopic image, the developed coding algorithms exploit inter-view redundancies since the stereo camera captures the same scene. The underlying idea consists in extracting the spatial displacements between the left and right image and estimating the disparity map. Its efficient estimation ensures a reliable reconstruction of the original stereoscopic image.

The state of the art performed on the estimating problem of the disparity map shows that many studies already addressed this problem and several stereoscopic matching algorithms have been deployed. The stereoscopic matching problem is generally formulated as the minimization problem over the overall image of an energy function (global cost) resulting in global approaches or several energy functions (local costs) resulting in local approaches. These methods, as summarized in [1, 2], differ in the choice of : (i) the primitives (e.g. pixels, interest points, segments, regions, edges) and their attributes (e.g. gray level, contrast, color components, segment position, segment orientation) to be matched ; (ii) the global cost of correspondence including the local matching costs which measure the dissimilarity between two corresponding primitives and the constraint cost for all correspondences (e.g. uniqueness, ordering, smoothness); (iii) the matching window size; (iv) the aggregation area ; and (v) the optimization method.

The main objective of the optimization methods is to minimize the global or local cost to ensure a high matching accuracy. The most naive method is the greedy search which performs an exhaustive search for the best matching. Due to its high computational load, this method has not been retained. Dynamic programming technique has been the first optimization method introduced in stereo matching context where smoothness constraints have been added to optimize matches in scan lines [3]. However among different developed versions, Veksler imposed smoothness in both horizontal and vertical directions with the objective of recovering the real disparity map [4]. Many other optimization methods such as relaxation [5], graph cut [6, 7] and belief propagation [8, 9] have been also exploited.

Among the stereoscopic matching approaches developed in literature, this paper focuses on pixel-based matching algorithms. Mean Square Error (MSE) and Mean Absolute Error (MAD) are usually used as matching criterion. Sometimes, for a given pixel it is possible to get not only one correspondence but also a set of correspondences satisfying the criterion. Nevertheless, some of these correspondences are more expensive than others in terms of bit-rate. To address this problem, we propose to replace the traditional matching metric by a joint entropy-distortion metric so that the selected disparities reduce not only the predicted image distortion but also the disparity entropy. This problem is formalized by the Lagrangian minimization where the cost function is exploited as the new matching metric. To avoid computational load related to a full search solutions, we rely on a tree which is sequentially constructed (i.e. a breadth-first search algorithm). At each tree depth, the algorithm retains the M-best paths and extends them in the next step.

The remainder of the paper is organized as follows. Section 2 introduces notations and assumptions, then states the optimization problem which is formulated as the Lagrangian minimization. The proposed entropy-constrained dense disparity map algorithm is then developed. Section 3 discusses simulation results. Section 4 concludes the work.

2. PROPOSED ALGORITHM

Assumptions and notations are introduced before describing the proposed optimization algorithm. Images of the left and right view of the stereoscopic image are assumed to be rectified. I_l and I_r represent respectively the left and right image of size $K \times L$. $I_r(i,j)$ (respectively $I_l(i,j)$) is the intensity of the pixel located at position (i,j) in I_r (respectively I_l).

In what follows, the proposed algorithm is described so that the estimated disparity map is related to the right view I_r and the left view I_l is used as reference image. $d(i,j)$ is the spatial displacement associated with the pixel $I_r(i,j)$.

2.1. Rate-distortion optimization problem

The problem addressed in this paper concerns the estimation of the disparity map denoted $\mathbf{d} = \{d(i,j) \text{ with } i = 0, ..., K-1;\ j = 0, ..., L-1\}$ that minimizes the global distortion cost of the reconstructed right image. It is expressed as:

$$E_{global}(\mathbf{d}) = \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} (\widehat{I}_r(i,j) - I_r(i,j))^2 \quad \text{with } \widehat{I}_r(i,j) = I_l(i, j + d(i,j)), \tag{1}$$

subjected to a given entropy constraint $H(\mathbf{d})$ being an estimate of the bit-rate used to encode disparity. This problem is formulated as the Lagrangian minimization:

$$\widehat{\mathbf{d}} = \text{argmin} J(\lambda, \mathbf{d}) = \text{argmin}(E_{global}(\mathbf{d}) + \lambda H(\mathbf{d})), \tag{2}$$

where λ is the Lagrange multiplier. Minimizing $J(\lambda, \mathbf{d})$ for any λ yields the points on the convex hull of all possible Rate-Distortion (R-D) points.

2.2. Entropy-constraint dense disparity map based on M-algorithm

This section deals with the optimization problem formulated in equation (2) where the main objective is to estimate an efficient dense disparity map associated with one view of the stereoscopic image in terms of entropy-distortion.

The underlying idea of the developed algorithm is related to the generic sequential decoding M-algorithm deployed in [10]. This algorithm has been also exploited in communication to estimate the transmitted data stream through a noisy channel according to the maximum likelihood criterion. This algorithm is a sub-optimal optimization method based on a tree-search technique parsing only a part of the tree. However many changes have been made to this algorithm to be adapted to stereo matching problem.

To reduce the computational load of the optimization problem, the proposed algorithm limits the matching search process while trying to ensure good performance in terms of entropy-distortion. A sliding matching window W of size N set on the pixel located at position (i,j) on the right image I_l is introduced.

The t-th depth of the tree (i.e. the number of pixels processed) depends on the row index i, the column index j and the number of columns L of the image I_r. It is expressed as follows:

$$t = i \times L + j \text{ with } i = 0, ..., K-1 \text{ and } j = 0, ..., L-1. \tag{3}$$

At $(t-1)$-th depth of the tree, assume that the M-best retained paths are sorted in a decreasing order according to the entropy-distortion cost $J_{t-1}^k(\lambda, d)$ given by:

$$J_{t-1}^k(\lambda, d) = E_{t-1}^k + \lambda H_{t-1}^k \quad \text{with } k = 1, ..., M, \tag{4}$$

where E_{t-1}^k is the cumulative distortion metric and H_{t-1}^k is the disparity entropy both associated with the k-th path at $(t-1)$-th depth. At this stage M-best disparity maps, denoted S^k, are retained:

$$S^k = \{d_1^k, d_2^k, ..., d_{t-1}^k\} \quad \text{with } \ k = 1, ..., M, \tag{5}$$

where d_l^k is the disparity of the k-th path at l-th depth.

On the next depth, i.e. t-th, each M selected path is then extended by N branches. Each branch is affected by a disparity equal to w (with $w = w_{min}, ..., w_{max}$ depending on the size of the sliding matching window W, i.e. $N = w_{max} - w_{min} + 1$) and a local distortion ($E_{b_t}{}^w$) given by:

$$E_{b_t}{}^w = ((I_r(i,j) - I_l(i, w+j))^2 \text{ with } w = w_{min}, ..., w_{max}. \tag{6}$$

The distortion of each of the $M \times N$ extended paths is then updated according to:

$$E_t^m = E_{t-1}^k + E_{b_t}{}^w \text{ for } m = 1, .., M \times N \quad \text{with } k = 1, ..., M \text{ and } w = w_{min}, ..., w_{max}. \tag{7}$$

For a given λ, the $J_t^k(\lambda, d)$ cost on the t-th depth is computed as given below:

$$J_t^k(\lambda, d) = E_t^k + \lambda H_t^k \text{ with } k = 1, ...M \times N, \tag{8}$$

where H_t^k is the entropy derived from the true probability distribution of disparities $(d_1^k, d_2^k, ..., d_t^k)$ of the k-th path until the t-th depth provided by

$$H_t^k = -\sum_{w=w_{min}}^{w_{min}} p_t^k(d=w) log_2(p_t^k(d=w)) \quad \text{for } k = 1, ...M \times N. \tag{9}$$

However this entropy cannot be calculated since it requires the true disparity distribution knowledge. Our contribution is to propose an estimate of these probabilities (i.e. $\{p_t^k(d = w)\}$) according to a finite mixture distribution represented as a sum of weighted discrete distributions as follows:

$$\widehat{p}_t^k(d = w|d_1^k, d_2^k, ..., d_t^k) = C_a \times p_a(d = w) + C_{exp} \times p_{exp}^k(d = w|d_1^k, d_2^k, ..., d_{t-1}^k) + C_c \times p_c(d = w|d = d_t^k), \quad (10)$$

where the coefficients C_a, C_{exp} and C_c satisfy the following condition:

$$C_a + C_{exp} + C_c = 1, \quad (11)$$

with $C_a = \frac{\beta a}{\beta a+b+c}$; $C_{exp} = \frac{b}{\beta a+b+c}$ and $C_c = \frac{c}{\beta a+b+c}$. These coefficients depend on the current depth, i.e. on the number of pixels processed and are parameterized as follows:

$$a = K \times L - t;\ b = t \text{ and } c = 1. \quad (12)$$

p_a is the probability density assumed to be a discrete uniform distribution on the selected matching window W given by:

$$p_a(d = w) = \frac{1}{N} \text{ with } w = w_{min}, ..., w_{max}. \quad (13)$$

β is a constant parameter smaller than 1. It provides a freedom degree to adjust the weight of p_a in the finite mixture distribution (in equation (10)).
The probability $p_{exp}^k(d = w|d_1^k, d_2^k, ..., d_{t-1}^k)$ is calculated from the retained disparities until the $(t-1)$-th depth (i.e. $d_1^k, d_2^k, ..., d_{t-1}^k$).
Note that "a" decreases linearly while "b" increases linearly. Indeed as the image is being processed and as t increases, we less and less need p_a. Therefore the estimation is more and more close to the true probability distribution.
The probability $p_c(d = w|d = d_t^k)$ is the probability related to the choice that the algorithm makes when it selects at depth t the branch with disparity w_c among other branches:

$$p_c(d = w|d_t^k = w_c) = \begin{cases} 1 \text{ if } w = w_c \\ 0 \text{ if } w \neq w_c \end{cases} \quad (14)$$

The J_t^k costs are then sorted in a decreasing order and M-best paths are retained. The M disparity maps (i.e S^k) are also updated. This process is iterated until scanning the complete reference image. Therefore the first path contains the best disparity map in terms of entropy-distortion. Figure 1 illustrates an example for $M = 2$ and $N = 5$.

The different steps of the proposed optimization algorithm are summarized below.

Algorithm 1: Entropy-distortion M-algorithm

Input: Left image I_l and right image I_r of size $K \times L$
Output: Estimated dense disparity map associated with I_r;

1. Set initial values: λ; M; w_{min}; w_{max}; β; $i = -1$ and $j = -1$;
2. Increment by 1 the row index i;
3. Increment by 1 the column index j;
4. Set the sliding matching window on the pixel $I_l(i, j)$;
5. Extend all M-best current paths of the tree to depth t;
6. Compute the distortions of $M \times N$ branches;
7. Update the distortions of the extended paths;
8. Estimate the disparity probabilities of each path;
9. Deduce the disparity entropy of each path;
10. Compute the entropy-distortion cost of each path;
11. Sort the paths in a decreasing order of rate-distortion cost;
12. Select among the $M \times N$ paths, the M-best paths;
13. Update the M disparity maps;
14. Start again from step 3 if $j < L$ otherwise continue;
15. Start again from step 2 if $i < K$ otherwise continue;
16. Select the best dense disparity map associated with I_r.

3. SIMULATION RESULTS

In this section, simulation results are presented to evaluate the performance of the proposed optimization algorithm. Comparisons are carried out with the dynamic programming algorithm provided by the computer vision system toolbox of Matlab [12]. This algorithm exploits not only block matching metric as the cost function but also constrains the disparities to change very slightly between adjacent pixels. Middlebury stereo dataset is used [11].

Simulations provided in this paper are performed on Poster stereoscopic image using im6.ppm (respectively im2.ppm) for the right (respectively left) view shown on by Figure 2. The spatial resolution of these images is equal to 383×435 pixels. In the provided results, the right view is reconstructed. The performance in terms of distortion is measured in terms of $PSNR$ calculated between the original and reconstructed images using the luminance component.

Figure 3 displays the reconstructed right view using our estimated disparity map with the following parameters: $M = 4$; $\beta = 0.02$; $\lambda = 1100000$ and $N = 30$ (with $w_{min} = -15, w_{max} = 14$). The evaluated $PSNR$ is equal to 33.97dB with a bit-rate equal to 2.88bpd (bits per disparity). Figure 4 is related to the reconstructed right view according to dynamic programming disparity map using the same window size as in our algorithm (i.e. $N = 30$) and a block size of 3×3 providing the best results in terms of PSNR. The evaluated $PSNR$ is equal to 23.15dB with 3.05bpd. For an equivalent bit-rate (2.88bpd and 3.05bpd), we obtain a gain of 10.82dB in terms of $PSNR$. This is confirmed by what we observe on the reconstructed images. Indeed we clearly see that the reconstruction of "newspaper" part in the Poster image is of lower quality for dynamic programming disparity map. This also happens for equivalent $PSNR$ (e.g. 23.81dB using our algorithm) although the bit-rate of our algorithm is lower (0.87bpd) than the dynamic programming (3.05bpd).

Figure 6 (respectively Figure 5) shows the estimated disparity map used to reconstruct the right view given by Figure 3 (respectively Figure 4). The distribution of disparities is completely different and is provided by Figures 8 and 9.

Figure 11 provides the $PSNR$ (in dB) as a function in bit-rate (in bpd) with the following parameters: $M = 4$; $\beta = 0.02$; $N = 30$; 60; 120 with symmetric ($[w_{min} = -\frac{N}{2}, w_{max} = \frac{N}{2} - 1]$) and positive ($[w_{min} = 0, w_{max} = N - 1]$) window. Each point of the curve is obtained with a specific value of λ. Even if the bit-rate involved by the dynamic programming technique is divided by 2 (i.e. 1.5bpd), a gain of 4.75dB in terms of $PSNR$ can still be reached.

For low entropies, without applying any smoothness constraint, the estimated disparity map is nevertheless smooth. An example is provided by Figure 7 in which the entropy is equal to 0.87bpd. This is also confirmed by the disparity distribution illustrated by Figure 10. Therefore for future investigations, it would be interesting to modify the estimated dense disparity map using a relevant division into blocks so as to improve the rate-distortion performance of our optimization algorithm. Another track would be to adapt our algorithm so that the entropy would be calculated on the difference between consecutive disparities rather than on disparities.

4. CONCLUSION

This paper addressed stereo matching problem where a pixel-based approach is adopted to estimate a dense disparity map. The optimization problem of selecting the best disparities in terms of entropy-distortion is formulated as the Lagrangian minimization. In order to reduce the computational load associated with a full search solutions, this optimization problem statement is solved according to the developed algorithm which sequentially builds a tree avoiding a full search and ensuring good rate-distortion performance. Indeed at each tree depth, only the M-best retained paths are extended to build new paths for which entropy-distortion metrics are assigned. Simulations performed on stereoscopic images clearly show the advantage of our algorithm compared to the dynamic programming technique in the particular context of coding.

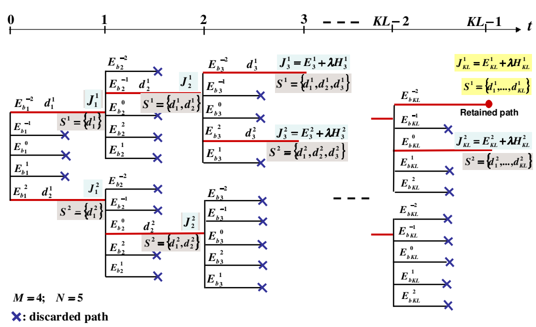

Fig. 1. Entropy-constrained dense disparity map estimation algorithm with $M = 2$ and $N = 5$.

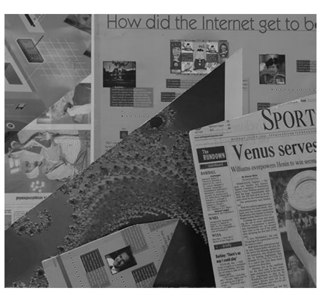

Fig. 2. Original right image.

Fig. 3. Reconstructed image using our estimated disparity map ($PSNR$=33.97dB and H= 2.88bpd).

Fig. 4. Reconstructed image using dynamic programming disparity map ($PSNR$=23.15dB and H=3.05bpd).

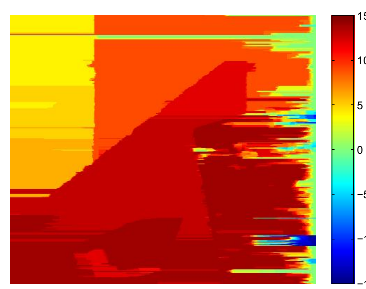

Fig. 5. Disparity map using dynamic programming (23.15dB; 3.05bpd).

Fig. 6. Estimated disparity map (33.97dB; 3.05bpd).

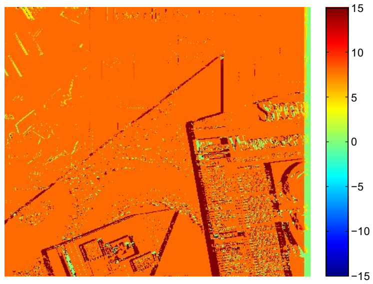

Fig. 7. Estimated disparity map (23.81dB; 0.87bpd).

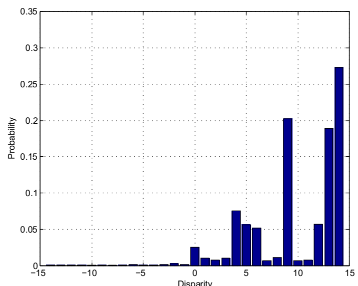

Fig. 8. Disparity distribution for dynamic programming.

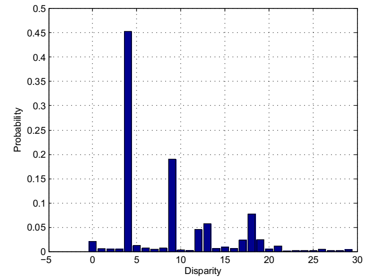

Fig. 9. Disparity distribution for the proposed algorithm (with 2.88bpd).

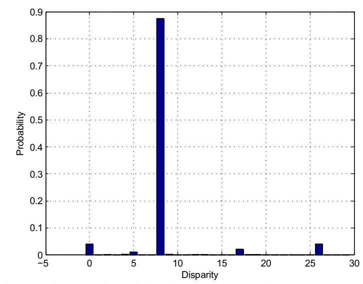

Fig. 10. Disparity distribution for the proposed algorithm (with 0.87bpd).

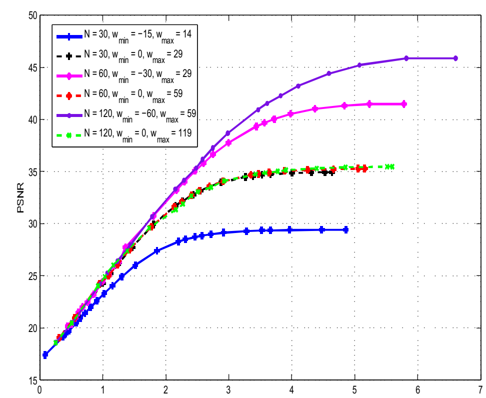

Fig. 11. Rate-Distortion optimization.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondance Algorithms," *International Journal of Computer Vision, IJCV*, vol. 47(1), pp. 7–42, Apr. 2002.

[2] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 25(8), pp. 993–1008, Aug. 2003.

[3] Y.-M. Ohta and T. Kanade, "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 7(2), pp. 139–154, Mar. 1985.

[4] O. Veksler, "Stereo Correspondence by Dynamic Programming on a Tree," *IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR*, vol. 2, pp. 384–390, San Diego, United States, Jun. 2005.

[5] N. M. Nasrabadi, "A Stereo Vision Technique Using Curve-Segments and Relaxation Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 14(5), pp. 566–572, May. 1992.

[6] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts.," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 23(11), pp. 1222–1239, Nov. 2001.

[7] M. Bleyer and M. Gelautz, "Graph-based surface reconstruction from stereo pairs using image segmentation," *Videometrics VIII*, vol. SPIE-5665, pp. 288–199, San Jose, United States, Jan. 2005.

[8] J. Sun, "Stereo Matching Using Belief Propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, vol. 25(7), pp. 787–800, Dec. 2003.

[9] Y.-C. Tseng, N. Chang and T.-S. Chang, "Low memory cost block-based belief propagation for stereo correspondance," *2007 IEEE International Conference on Multimedia and Expo, IEEE*, pp. 1415–1418, 1998.

[10] F. Jelinek, "Fast sequential decoding algorithm using a stack," *IBM J. Res.*, Develop. 13, 1969, pp. 675–685.

[11] "http://vision.middlebury.edu/stereo/data/".

[12] "Matlab toolbox: http://www.mathworks.fr/products/computer-vision/code-examples.html?file=%2Fproducts%2Fdemos %2Fshipping%2Fvision%2Fvideostereo.html"