

PROXIMAL SPLITTING METHODS FOR DEPTH ESTIMATION

Mireille El Gheche^{1,2}, *Jean-Christophe Pesquet*¹, *Joumana Farah*²,
Mounir Kaaniche^{1,3} and *Béatrice Pesquet-Popescu*³

¹ Université Paris-Est
LIGM and UMR-CNRS 8049,
77454 Marne-la-Vallée cedex, France
el.gheche@univ-mlv.fr
pesquet@univ-mlv.fr

² Department of Telecommunications,
Faculty of Engineering,
Holy-Spirit University of Kaslik,
P.O. Box 446, Jounieh, Lebanon
joumanafarah@usek.edu.lb

³ Telecom ParisTech,
Signal and Image Proc. Dept.
75014 Paris, France
kaaniche@telecom-paristech.fr
pesquet@telecom-paristech.fr

ABSTRACT

Stereo matching is an active area of research in image processing. In a recent work, a convex programming approach was developed in order to generate a dense disparity field. In this paper, we address the same estimation problem and propose to solve it in a more general convex optimization framework based on proximal methods. More precisely, unlike previous works where the criterion must satisfy some restrictive conditions in order to be able to numerically solve the minimization problem, this work offers a great flexibility in the choice of the involved criterion. The method is validated in a stereo image coding framework, and the results demonstrate the good performance of the proposed parallel proximal algorithm.

Index Terms— Stereo vision, disparity estimation, convex programming, variational methods, parallel proximal algorithm, proximity operator.

1. INTRODUCTION

A stereo vision system generates two views, called left and right images, by recording two slightly different view angles of the same 3D scene. An important task in stereo vision is to find corresponding pixels which result from the projection of the same 3D point onto the two image planes. The displacement between the pixel coordinates of these points is called disparity. This information plays a crucial role in many application fields such as 3D reconstruction, intermediate view synthesis, obstacle detection, and stereo/multiview image compression [1].

The disparity estimation problem has been extensively studied in computer vision [2]. Traditionally, disparity estimation algorithms are basically classified into two categories: local methods and global ones. Algorithms in the first category, where the disparity at each pixel depends only on intensity values within a local window, perform well in highly textured regions. However, they often produce noisy disparities in textureless regions and fail at occluded areas. These problems

can be reduced by using global methods which aim at finding the disparity field that minimizes a global energy function over the entire image. For this purpose, several energy minimization algorithms have been proposed. The most common approaches are dynamic programming [3], graph cuts [4] and variational methods [5]. While dynamic programming and graph cuts methods operate in a discrete manner, variational techniques work in a continuous space. Therefore, they possess the advantage of producing a disparity field with ideally infinite precision. Among these global approaches, it has been shown in [5] that variational-based disparity estimation methods are among the most competitive techniques, for their preservation of depth discontinuities. In [5], the dense disparity estimation problem has been formulated as a convex programming problem within a global variational approach. More precisely, a convex quadratic objective function has been employed and minimized subject to some appropriate convex constraints. The used optimization algorithm is able to handle a wide range of constraints. However, in order to satisfy the conditions of convergence of this algorithm, the cost function must be quadratic and strictly convex.

In the present work, we reformulate the problem introduced in [5] and we develop a parallel proximal algorithm which allows us to efficiently solve the problem over feasibility sets determined by multiple constraints modelling prior information. Our main contribution is to extend the approach in [5] to a more general framework by relaxing the limitations on the choice of the similarity measure. The proposed algorithm offers a great flexibility in the choice of the convex cost function to be minimized, which, for example, may be based on the ℓ_1 -norm, an ℓ_p -norm with $p > 1$ or the Kullback-Leibler divergence. The remainder of this paper is organized as follows. In Sec. 2, we present the stereo matching problem formulated as a constrained optimization problem. In Sec. 3, an efficient iterative solution is proposed based on a parallel proximal method. Experimental results, showing the accuracy of the resulting disparity maps as well as their applications to stereo image coding, are given in Sec. 4. Finally, some conclusions are drawn in Sec. 5.

2. PROBLEM STATEMENT

The process of stereo matching amounts to finding for each pixel in the left image I_L a corresponding pixel in the right image I_R . The matching problem is then equivalent to searching for the disparity field u which minimizes an error measure:

$$J(u) = \sum_{(x,y) \in \mathcal{D}} \phi(I_L(x,y) - I_R(x - u(x,y), y)), \quad (1)$$

where ϕ is assumed to be a proper lower-semicontinuous convex function from \mathbb{R} to $]-\infty, +\infty]$ and \mathcal{D} is the considered image domain. Similarity measures often used in literature are the Sum of Square Differences (SSD) and the Sum of Absolute Differences (SAD).

2.1. Cost function

Despite the convexity of ϕ , note that J is non-convex with respect to the field u . To circumvent this difficulty, we assume that an initial estimate \bar{u} of u is available and that the magnitude difference of the fields u and \bar{u} is small enough so that we can perform a Taylor expansion of the nonlinear term $I_R(x - u, y)$ around an initial estimate \bar{u} as follows:

$$I_R(x - u, y) \simeq I_R(x - \bar{u}, y) - (u - \bar{u}) I_R^x(x - \bar{u}, y), \quad (2)$$

where $I_R^x(x - \bar{u}, y)$ is the horizontal gradient of the disparity compensated right image. (For simplicity, we have dropped the dependence on (x, y) of u and \bar{u}). Inserting linearization (2) in (1), we end up with the following convex cost function:

$$J(u) = \sum_{(x,y) \in \mathcal{D} \setminus \mathcal{O}} \phi(T(x,y) u - r(x,y)) \quad (3)$$

where

$$T(x,y) = I_R^x(x - \bar{u}(x,y), y), \quad (4)$$

$$r(x,y) = I_R(x - \bar{u}(x,y), y) + \bar{u}(x,y) T(x,y) - I_L(x,y) \quad (5)$$

and \mathcal{O} denotes the occlusion area. Indeed, the occluded pixels, which correspond to points visible only in one image of the stereo pair need to be discarded [5].

Optimizing this criterion is an ill-posed problem as the components of T may locally vanish. Thus, to solve the problem in a reliable manner, it is necessary to incorporate additional constraints. This can be performed by formulating the problem within a set theoretic framework [6].

2.2. Convex constraints

The goal of set theoretic estimation is to obtain a feasible solution satisfying various constraints. Typically, each constraint is described by a closed convex set S_i with $i \in$

$\{1, \dots, m\}$, in a Hilbert image space \mathcal{H} . The intersection S of all the m sets, the feasibility set, constitutes the family of admissible solutions [5]. In this context, the stereo matching problem can be formulated as a constrained optimization problem where the cost function given by (3) is minimized over the feasibility set. We have then to

$$\text{Find } u \in S = \bigcap_{i=1}^m S_i \text{ such that } J(u) = \inf J(S). \quad (6)$$

Constraint sets can generally be modelled as lower level sets:

$$\forall i \in \{1, \dots, m\}, \quad S_i = \{u \in \mathcal{H} \mid f_i(u) \leq \delta_i\}, \quad (7)$$

where $f_i: \mathcal{H} \rightarrow \mathbb{R}$ is a convex function and $\delta_i \in \mathbb{R}$.

As mentioned before, the construction of convex constraints is derived from the properties of the field to be estimated. An example of possible prior knowledge is the range of the disparity values. We can restrict the variation of the values within the specific range $[u_{\min}, u_{\max}]$. This can be expressed by the following constraint set S_1 :

$$S_1 = \{u \in \mathcal{H} \mid u_{\min} \leq u \leq u_{\max}\}. \quad (8)$$

Most importantly, other constraints can be incorporated in order to enforce the smoothness of the disparity field in the homogeneous areas while preserving edges. First, we can use the Total Variation (TV) measure which attracted much attention in image recovery problems [7]. Hence, a total variation based regularization constraint amounts to impose an upper bound $\tau > 0$ on the TV:

$$S_2 = \{u \in \mathcal{H} \mid \text{TV}(u) \leq \tau\}. \quad (9)$$

Alternatively, we can adopt a wavelet domain approach to construct a smoothness constraint based on 2D separable wavelet coefficients. Inspired from the recent work in [5], we consider the Besov space $B_{1,1}^1$ as being appropriate to model images containing discontinuities. The convex set associated with a semi-norm of this space is given by

$$S_3 = \{u \in \mathcal{H} \mid \sum_{j \geq 1, k \in \mathbb{Z}^2, o \in \{H,V\}} |c_{j,k,o}^{\mathcal{B}}| \leq \kappa\} \quad (10)$$

where $\kappa > 0$ and the coefficients of $u \in \mathcal{H}$ in a wavelet basis or redundant frame are denoted by $(c_{j,k,o}^{\mathcal{B}}(u))_{k \in \mathbb{Z}^2, o \in \{H, V, D\}}$ being the orientation parameter and $j \in \mathbb{N}$ the resolution level.

3. PROPOSED ALGORITHM

3.1. Motivation

The objective of this section is to develop an iterative algorithm to solve the stereo matching problem as formulated by (6). The proximal method we propose allows us to find an optimal solution by using direct projections of the estimate onto the different constraint sets.

3.2. Optimization background

In Problem (6), let us assume that $\mathcal{H} = \mathbb{R}^K$ and that each constraint set S_i with $i \in \{1, \dots, m\}$ can be expressed as $L_i^{-1}(C_i)$ where C_i is a non-empty closed convex subset of \mathbb{R}^{N_i} and L_i is a matrix in $\mathbb{R}^{N_i \times K}$. Let us further assume that the projection onto each convex set C_i takes a closed form (in the sense that it can be computed in a finite number of operations). The considered optimization problem then takes the following generic form:

$$\underset{L_i u \in C_i, i \in \{1, \dots, m\}}{\text{minimize}} \quad J(u). \quad (11)$$

An algorithm allowing us to solve this problem is the algorithm described below. It basically consists of iterating computations of the proximity operator of J and projections onto convex sets.

We recall that the proximity operator of the proper lower-semicontinuous convex function J is $\text{prox}_J: \mathcal{H} \rightarrow \mathcal{H}: v \mapsto \text{prox}_J v$ where $\text{prox}_J v$ is the unique minimizer of $u \mapsto J(u) + \frac{1}{2} \|u - v\|^2$. This operator generalizes the notion of projection since $\text{prox}_{\iota_C} = P_C$ where ι_C is the indicator of a nonempty closed convex set C and P_C denotes the projection onto C . For more details, the reader is referred to [8].

3.3. PPXA+ algorithm

We propose to employ the PPXA+ (Parallel Proximal Algorithm) detailed in [9], and which constitutes a generalization of some recent parallel proximal algorithms [8, 10]. As shown in Algorithm 1, it consists of computing parallel projections onto the sets $(C_i)_{1 \leq i \leq m}$ and the proximity operator of a scaled version of J , followed by a certain averaging procedure taking into account the matrices $(L_i)_{1 \leq i \leq m}$.

Suppose that the following assumptions hold.

- $\sum_{i=1}^m L_i^\top L_i + \mathbb{I}$ is an invertible matrix.
- $\exists \tilde{u} \in \mathbb{R}^K$ such that $(\forall i \in \{1, \dots, m\}) L_i \tilde{u} \in \text{ri}(C_i)$ and $\tilde{u} \in \text{ri}(\text{dom } J)$.¹
- $(\lambda_n)_{n \in \mathbb{N}}$ is a sequence such that $(\forall n \in \mathbb{N}) \tilde{\lambda} \leq \lambda_{n+1} \leq \lambda_n < 2$, where $\tilde{\lambda} \in]0, 2[$.

Then, the sequence $(u_n)_{n \in \mathbb{N}}$ generated by Algorithm 1 converges to a solution to Problem (6) provided that such a solution exists.

4. SIMULATION RESULTS

Experiments were carried out on synthetic stereo images “corridor” and “teddy” taken from ².

¹dom J is the domain of J and the relative interior of a set C is denoted by $\text{ri}C$.

²<http://cat.middlebury.edu/stereo/scenes2005/>

[Initialization]

$$\begin{aligned} (w_1, \dots, w_m) &\in]0, +\infty[^m, \gamma > 0 \\ (z_{i,0})_{1 \leq i \leq m+1} &\in \mathbb{R}^{N_1} \times \dots \times \mathbb{R}^{N_m} \times \mathbb{R}^K \\ Q &= (\sum_{i=1}^m w_i L_i^\top L_i + \gamma \mathbb{I})^{-1} \\ u_0 &= Q(\sum_{i=1}^m w_i L_i^\top z_{i,0} + \gamma z_{m+1,0}) \end{aligned}$$

For $n = 0, 1, \dots$ **do**

For $i = 1, \dots, m$ **do**

$$| \quad p_{i,n} = P_{C_i}(z_{i,n})$$

end For

$$p_{m+1,n} = \text{prox}_{\frac{J}{\gamma}}(z_{m+1,n})$$

$$c_n = Q(\sum_{i=1}^m w_i L_i^\top p_{i,n} + \gamma p_{m+1,n})$$

For $i = 1, \dots, m$ **do**

$$| \quad z_{i,n+1} = z_{i,n} + \lambda_n (L_i(2c_n - u_n) - p_{i,n})$$

end For

$$z_{m+1,n+1} = z_{m+1,n} + \lambda_n (2c_n - u_n - p_{m+1,n})$$

$$u_{n+1} = u_n + \lambda_n (c_n - u_n)$$

end For

Algorithm 1: Projective version of PPXA+.

In order to show the performance of the proposed method using an ℓ_1 error measure and the constraint sets S_1 and S_2 (here designated by “DDE-L1”), it has been compared with a block-based disparity estimation (“BDE”) technique using the fixed size block matching method. For PPXA+ algorithm, we set $w_1 = 100$, $w_2 = 10$, $\gamma = 200$ and $\lambda_n \equiv 1.5$. We have also tested the approach recently proposed by Miled *et al.* (“DDE-L2”) which uses a cost function based on the ℓ_2 -norm and the same constraints. At this stage, it is worth pointing out that it was shown in [5] that the quality of the results provided by the latter approach is comparable with state-of-the-art methods, such as those based on graph cuts and dynamic programming [11]. Fig. 1 shows the generated disparity maps using the different mentioned methods. It can be seen that “DDE-L2” and “DDE-L1” produce smooth disparity maps. Furthermore, they are much less sensitive to noise than “BDE”. Now, if we focus on the two dense disparity estimation methods, it can be observed that our approach “DDE-L1” better preserves the discontinuities around object boundaries. Numerical results are given by computing the PSNR values between the estimated disparity map and the ground truth. In order to emphasize the benefit from using the proposed method, we also evaluated its performance when used in a stereo image coding application. A straightforward approach consists of independently coding each image by using existing still image coders. However, most of the existing works rely on disparity compensation techniques [1]. The first step in these techniques consists of estimating the disparity map. Then, one image is considered as a reference image and the other is predicted in order to generate a prediction error referred to as a residual image. Finally, the disparity

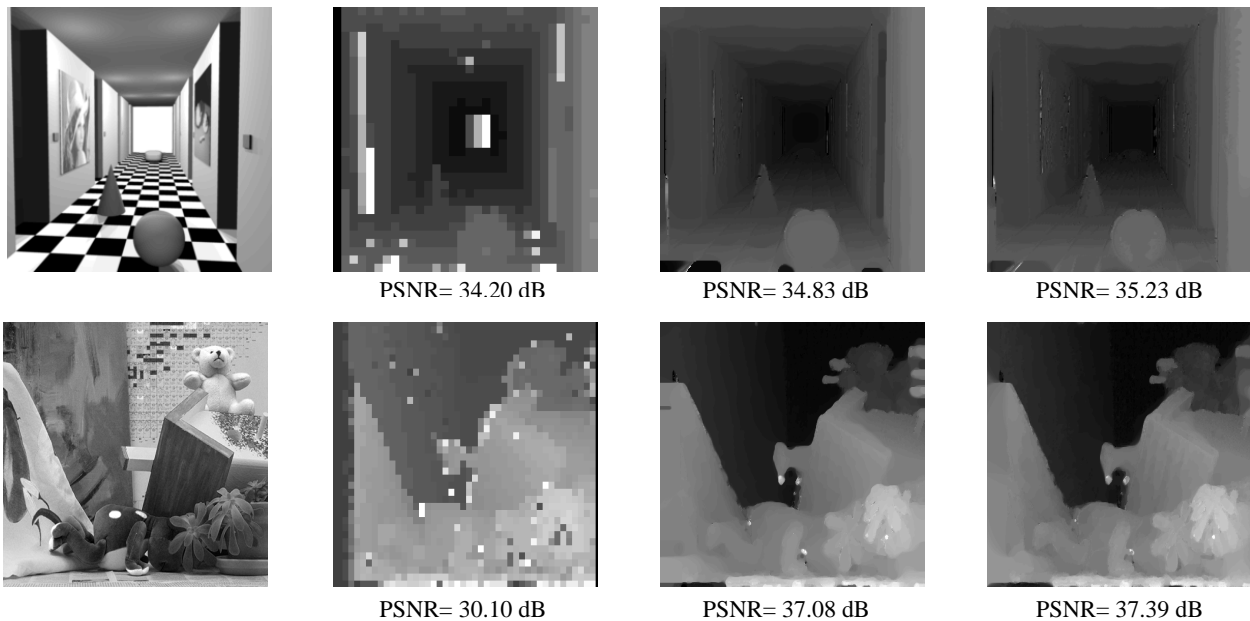


Fig. 1. Dense disparity maps for “corridor” and “teddy” stereo pair. From left to right: left image, “BDE” method, “DDE-L2” method, “DDE-L1” method .

field, the reference image and the residual one are encoded. In our experiments, the disparity maps are encoded by employing the coding module integrated in the H.264/AVC standard. Concerning the reference and the residual images, they are encoded by applying a 5/3 wavelet-like transform. Thus, disparity estimation constitutes a key step in stereo image coding. In this context, DDE raises an important problem: the cost of transmission/storage of the disparity field in terms of bitrate. In order to alleviate this problem, we apply a quadtree coding of the estimated dense maps [1]. Fig. 2 illustrating the scalability in quality of the reconstruction procedure, shows that the proposed method outperforms “DDE-L2” by 0.1-0.3 dB.

5. CONCLUSIONS

We have presented a new parallel proximal algorithm for estimating disparity maps which has the advantage of allowing the use of various convex similarity measures and constraints. In particular, we have shown the benefit in coding applications which can be drawn from using an ℓ_1 cost function instead of the ℓ_2 one employed in previous DDE methods [5].

6. REFERENCES

- [1] M. Kaaniche, W. Miled, B. Pesquet-Popescu, A. Benazza-Benyahia, and J.-C. Pesquet, “Dense disparity map representations for stereo image coding,” *IEEE International Conference on Image Processing*, Nov. 7–10, 2009.
- [2] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.
- [3] O. Veksler, “Stereo correspondence by dynamic programming on a tree,” in *IEEE Conference Proceedings of Computer Vision and Pattern Recognition*, San Diego, United States, 2005, vol. 2, pp. 384–390.
- [4] V. Kolmogorov and R. Zabih, “Computing visual correspondence with occlusions using graph cuts,” in *IEEE International Conference on Computer Vision*, Vancouver, BC, Canada, July 2001, vol. 2, pp. 508–515.

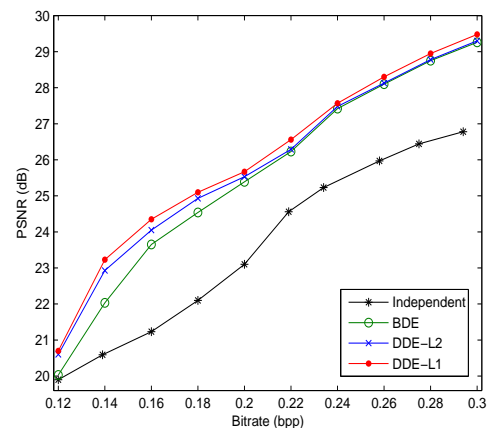


Fig. 2. PSNR (in dB) versus the bitrate (in bpp) after JPEG2000 encoding for the stereo pair “corridor”.

- [5] W. Miled, J.-C. Pesquet, and M. Parent, “A convex optimisation approach for depth estimation under illumination variation,” *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 813–830, April 2009.
- [6] P. L. Combettes, “The foundations of set theoretic estimation,” *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, Feb 1993.
- [7] P. L. Combettes and J.-C. Pesquet, “Image restoration subject to a total variation constraint,” *IEEE Trans. Image Process*, vol. 13, no. 9, pp. 1213–1222, Sep 2004.
- [8] P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. Springer-Verlag, New York, 2010.
- [9] J.-C. Pesquet, “A parallel inertial proximal optimization method,” http://www.optimization-online.org/DB_FILE/2010/11/2825.pdf, 2010.
- [10] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems,” <http://arxiv.org/abs/0912.3481>.
- [11] Y. Deng and X. Lin, “A fast line segment based dense stereo algorithm using tree dynamic programming,” in *Proc. Eur. Conf. Comput. Vis.*, Graz, Austria, May 2006, pp. 201–212.