Wavelet-constrained stereo matching under photometric variations

Wided Miled[1], Jean-Christophe Pesquet[2] and Michel Parent[3]

[1]École Nationale Supérieure des Télécommunications, Signal and Image Processing Department, 46 rue Barrault, 75634 Paris Cedex 13, France
[2]Université Paris-Est, Institut Gaspard Monge / UMR-CNRS 8049, 77454 Champs-sur-Marne, France
[3] INRIA Rocquencourt, Domaine de Voluceau, 78150 Le Chesnay Cedex, France

ABSTRACT

We propose a new method to address the problem of stereo matching under varying illumination conditions. First, a spatially varying multiplicative model is developed to account for photometric changes induced between both images in the stereo pair. The stereo matching problem based on this model is then formulated as a constrained optimization problem in which an appropriate convex objective function is minimized under convex constraints. These constraints arise from prior knowledge and rely on various properties of both disparity and illumination fields. In order to obtain a smooth disparity field while preserving discontinuities around object edges, we consider an appropriate wavelet-based regularization constraint. The resulting multi-constrained optimization problem is solved via an efficient block iterative algorithm which offers great flexibility in the incorporation of several constraints. Experimental results demonstrate the efficiency of the proposed method to recover illumination changes and disparity map simultaneously, making stereo matching very robust w.r.t. such changes.

Keywords: Stereovision, Illumination variations, Joint estimation, Convex optimization, Convex constraints, Wavelets, Regularization.

1. INTRODUCTION

Stereo matching is a basic task in computer vision with many interesting applications including video coding, view synthesis, robot navigation and object recognition. The main goal of stereo matching is to recover the depth information of a scene from a pair of left and right images taken from two different locations. It involves finding corresponding pixels, i.e. pixels resulting from the projection of the same 3D point onto the two image planes. The difference in location between corresponding pixels forms the so-called disparity map. In most of the classical disparity estimation techniques,[1] the scene is assumed to be Lambertian and, more generally, corresponding pixels in different views are assumed to have identical intensity values. However, in the presence of illumination changes often encountered in practice, this assumption, commonly referred to as the *brightness constancy* assumption, is violated which may largely reduce the efficiency of disparity estimation.

Photometric variations between the left and right images of a stereo pair can be caused by the presence of non-Lambertian surfaces, differences in cameras responses, shadows, image noise, etc. In order to compensate for the effects of these variations, a simple and commonly used approach is to pre-process the images before stereo matching. Indeed, techniques such as gradient filters, luminance balancing or histogram transform can be performed to reduce noise and enhance the contrast in the images. Based on the observation that two images of a stereo pair are statistically similar, the idea of the method proposed in[2] is to equalize their mean and their variance by a simple linear transformation applied to the luminance of one of the two images. As a

Further author information: (Send correspondence to Wided Miled)
Wided Miled: E-mail: wided.miled@telecom-paristech.fr, Telephone: +33 (0)1 45 81 71 49
Jean-Christophe Pesquet: E-mail: jean-christophe.pesquet@univ-paris-est.fr, Telephone: +33 (0)1 60 95 77 39
Michel Parent: E-mail: michel.parent@inria.fr, Telephone: +33 (0)1 39 63 55 93

pre-processing step, the histogram specification technique, which consists of mapping one image histogram to the other, have also been used to obtain images that are illumination independent. This step yields a new pair of stereo images, with identical intensity histograms, to which the disparity estimation is applied. Cox et al.[3] proposed a dynamic histogram warping that consists of directly matching histogram values and performing a global optimization via dynamic programming. Other approaches combine traditional stereo algorithms with robust correlation measures such as normalized cross-correlation and rank transform. Recently, mutual information (MI) has also been used as a stereo similarity measure, due to its robustness to many complex intensity transformations. Originally proposed for use in image registration,[4] MI has attracted much interest in the field of stereo vision for matching images affected by photometric changes. However, used with a local method,[5] the performance of MI degrades in the presence of Gaussian noise due to the small statistical power of the matching windows. This problem can be overcome by incorporating MI into a global algorithm and using an energy minimization framework,[6].[7]

An alternative approach to cope with the illumination inconsistency problem is to model illumination variations, so extending the brightness constancy assumption to account for brightness changes. In,[8] lighting changes were modelled by constant multiplicative and additive factors. The analysis performed in[3] on the JISCT stereo database through histogram examination reveals, however, that the affine model is not always accurate. Gennert[9] provided a more reliable model, showing that the intensities of two corresponding points are related by a spatially varying multiplicative term that is a function of surface orientation and reflectance models. Using this model, he developed a stereo algorithm based on a nonconvex cost function consisting of a linear combination of a brightness matching error, and of smoothness penalties on the disparity and the multiplication field. However, minimizing this functional based on solving the associated Euler-Lagrange equations is difficult and computationally expensive. Very recently, Zhang et al.[10] modelled illumination changes using an illumination ratio map where the intensity ratio of corresponding points in an image pair is computed. They formulated the problem using a Markov network and employed an iterative optimization method based on belief propagation to recover the illumination ratio map and disparity image at the same time. Although this approach achieves good results validating the efficiency of the multiplicative illumination model, the optimization method adopted in this work is computationally intensive.

In this paper, we propose a new method that extends our previous work,[11] to address the problem of stereo matching under varying illumination conditions. To deal with illumination variations, we replace the assumption of brightness constancy with a more realistic model that is close to the one used in.[9] The stereo matching problem based on this model is then formulated as a constrained optimization problem in which an appropriate convex objective function is minimized under convex constraints. These constraints arise from prior knowledge and rely on various properties of both disparity and illumination fields. To maintain the spatial smoothness in homogeneous regions while preserving discontinuities around contours of the disparity image, we use a regularizing constraint based on image transformed coefficients. Since wavelets provide a simple characterization of a wide class of regular function spaces, the regularization constraint is defined as a semi-norm of the disparity field in such an image space. This regularization is appropriate for modelling fields exhibiting both sharp discontinuities and homogeneous areas. The resulting multi-constrained optimization problem is solved via an efficient block iterative algorithm which allows to combine both wavelet and spatial domain constraints.[12] In addition, this algorithm is well suited for an implementation on parallel processors which helps in reducing the computational time.

The paper is structured as follows. Section 2 focuses on the modelling of photometric variations and describes the set theoretic formulation of the stereo problem. Section 3 addresses the construction of constraint sets exploiting available information on both disparity and illumination fields. In Section 4, we review the parallel block-iterative algorithm that will be employed to solve the convex programming problem. Section 5 presents results and comparisons and, finally, Section 6 concludes the paper.

2. PROBLEM FORMULATION

2.1 Modelling of photometric variations

The standard mathematical model used to compute the disparity map from a pair of stereo images is based on the brightness constancy assumption : two homologous points in the left and right views have identical intensity values. To take into account possible illumination change, we replace this assumption with the multiplicative model proposed by Gennert.[9] The intensities of two corresponding points are then related by a spatially varying multiplicative factor as follows :

$$I_r(x - u(s), y) = v(s)\ I_l(s)\ , \tag{1}$$

where I_l and I_r are the left and right images of a stereo pair, respectively, $s = (x, y)$ is a spatial position in either image, u denotes the horizontal disparity and v represents the multiplicative coefficient of the intensity illumination change. Based on this model, we can compute u and v by minimizing the following cost function based on the Sum of Squared Differences (SSD) metric:

$$\widetilde{J}(u, v) = \sum_{s \in \mathcal{D}} \ [v(s)\ I_l(s) - I_r(x - u(s), y)]^2\ , \tag{2}$$

where $\mathcal{D} \subset \mathbb{N}^2$ is the image support. This expression is non-convex with respect to the displacement field u. Thus, to avoid a non-convex minimization, we consider a Taylor expansion of the non-linear term $I_r(x - \bar{u}, y)$ around an initial estimate $\bar{u}$ as follows :

$$I_r(x - u, y) \simeq I_r(x - \bar{u}, y) - (u - \bar{u})\ \ \nabla I_r^x(x - \bar{u}, y)\ , \tag{3}$$

where $I_r^x(x - \bar{u}, y)$ is the horizontal gradient of the warped right image. Note that the initial value $\bar{u}$ may be obtained from a correlation based method or from previous iteration within an iterative process. This linearization leads to the following convex quadratic criterion:

$$\widetilde{J}(u, v) \simeq \sum_{s \in \mathcal{D}} [L_1(s)\ u(s) + L_2(s)\ v(s) - r(s)]^2\ , \tag{4}$$

where

$$\begin{aligned} L_1(s) &= \nabla I_r^x(x - \bar{u}(s), y), \\ L_2(s) &= I_l(s), \\ r(s) &= I_r(x - \bar{u}(s), y) + \bar{u}(s)\ L_1(s). \end{aligned}$$

Setting $w = (u, v)^\top$ and $L = [L_1, L_2]$, we end up with the following quadratic criterion to be minimized:

$$J_{\mathcal{D}}(w) = \sum_{s \in \mathcal{D}} [L(s)\ w(s) - r(s)]^2\ . \tag{5}$$

2.2 Set theoretic formulation

Minimizing the objective function (5) is an ill-posed inverse problem due to the fact that two variables $u(s)$ and $v(s)$ have to be determined for each pixel s and that the components of L may locally vanish. Thus, to convert this problem to a well-posed one, it is useful to incorporate additional constraints modelling prior knowledge and available information on the solution. In the field of computer vision, such constraints were most commonly formulated as additional penalty terms in the objective function. In this work, the problem is addressed from a set theoretic formulation, where each constraint is represented by a convex set in the solution space and the intersection of these sets constitutes the family of admissible solutions. The aim then is to find an admissible solution minimizing the given objective function. A formulation of this problem in a Hilbert image space $\mathcal{H}$ is therefore

$$\text{Find } w \in S = \bigcap_{i=1}^{m} S_i \text{ such that } J(w) = \inf J(S)\ , \tag{6}$$

where $J : \mathcal{H} \to]-\infty, +\infty]$ is a convex function and $(S_i)_{1 \le i \le m}$ are closed convex sets of $\mathcal{H}$. Without loss of generality, constraint sets can be represented as level sets :

$$\forall i \in \{1, \ldots, m\}, \qquad S_i = \{w \in \mathcal{H} \mid f_i(w) \le \delta_i\}\ , \tag{7}$$

where, for all $i \in \{1, \ldots, m\}$, $f_i : \mathcal{H} \to \mathbb{R}$ is a continuous convex function and $(\delta_i)_{1 \le i \le m}$ are real-valued parameters.

The advantage of the convex program (6) is that a wide range of constraints modelling prior information can be explicitly incorporated to the problem. In,[13] it was shown that many spatial and spectral constraints commonly used in image recovery problems can be modelled as closed convex sets of the form (7). Recently, there have been several attempts in formulating constraints in the wavelet domain. Indeed, as shown in,[12] convex wavelet constraints can be constructed and used with various spatial constraints to refine the feasibility set S in (6), yielding improved results for many image processing applications. A further advantage of the set theoretic formulation is to benefit from the availability of powerful optimization algorithms, e.g. the constrained quadratic minimization method developed in,[14] which allows the combination of constraints arising in both the spatial and the wavelet domain.

3. ROBUST STEREO MATCHING

In this section, we introduce the objective function to be minimized as well as the considered convex constraints on both fields to be estimated, namely the disparity u and the illumination field v.

3.1 Global objective function

The objective function to be minimized is the quadratic measure derived from the data model (1) linearized around an initial disparity estimate $\bar{u}$. In order to guarantee that the linearization in (3) is valid, it is important to deal with a consistent initial disparity field. As an alternative to the classical initialization that consists of using a pyramidal strategy to refine the initial field at each level of an image pyramid, we use, in this paper, a simpler block matching method with an affine illumination variation factor, which allows to estimate an initial illumination field $\bar{v}$ and an initial disparity estimate $\bar{u}$. More precisely, we compute

$$\bar{u}(x,y) = \arg\min_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{B}} \left[\beta_{x,y}(u)\ I_l(x+i,y+j) - I_r(x+i-u,y+j)\right]^2 , \tag{8}$$

where $\mathcal{U} \subset \mathbb{N}$ is the search disparity set, $\mathcal{B}$ corresponds to a matching block centered at the pixel (x,y) and $\beta_{x,y}(u)$ is the following least squares estimate of the illumination factor for block $\mathcal{B}$:

$$\beta_{x,y}(u) = \frac{\sum_{(i,j)\in\mathcal{B}} I_l(x+i,y+j) I_r(x+i-u,y+j)}{\sum_{(i,j)\in\mathcal{B}} I_l(x+i,y+j)^2}. \tag{9}$$

The estimate of the initial illumination variation constant is then given by:

$$\bar{v}(x,y) = \beta_{x,y}\big(\bar{u}(x,y)\big). \tag{10}$$

Furthermore, to cope with large deviations from the data model, occlusion points which are pixels only visible from one view of the stereo images have been detected and discarded in the expression of the similarity criterion. Occluded image areas are one of the biggest challenges in stereovision since they yield very large disparity errors. Egnal and Wildes[15] have provided comparisons of various approaches for finding occlusions. For the results in this paper, occlusion points are detected by enforcing the uniqueness and ordering constraints. Denoting by $\mathcal{O}$ the occlusion field, the objective function to be minimized becomes

$$J_{\mathcal{D}\setminus\mathcal{O}}(w) = \sum_{s\in\mathcal{D}\setminus\mathcal{O}} [L(s)\ w(s) - r(s)]^2. \tag{11}$$

According to the conditions of convergence of the algorithm we use (see Section 4), the objective function J must be strictly convex. However, since the components of L may vanish in (11) and $\mathcal{O}$ is usually nonempty, $J_{\mathcal{D}\setminus\mathcal{O}}$ is not secured to be strictly convex. We therefore introduce an additional strictly convex term as follows :

$$J(w) = \sum_{\mathrm{s}\in\mathcal{D}\setminus\mathcal{O}} [L(\mathrm{s})\, w(\mathrm{s}) - r(\mathrm{s})]^2 + \alpha \sum_{\mathrm{s}\in\mathcal{D}} |w(\mathrm{s}) - \bar{w}(\mathrm{s})|_2^2 , \tag{12}$$

where $\bar{w} = (\bar{u}, \bar{v})$ is an initial estimate as described above, $| \cdot |_2$ denotes the Euclidean norm in $\mathbb{R}^2$ and α is a positive constant that weights the second term relatively to the first in the right-hand side of (12). We emphasize that the primary role of the latter term is not to regularize the solution but to make J strictly convex, in compliance with the assumption required to guarantee the convergence of the proposed algorithm. We also note that we iteratively refine the initial estimate $\bar{w}$ by choosing the result from a previous estimate as the initial value of the next step. This allows an improvement of the quality of the solution while reducing the dependence of the final solution on the initial estimate.

3.2 Wavelet regularization constraint

Smoothness has received a great deal of attention in recent years not only in the field of stereo matching but also in image restoration, tomography, segmentation and optical flow estimation. The concept of regularization by a smoothing function was initially introduced by Tikhonov.[16] However, the Tikhonov regularization, by considering a quadratic function, tends to oversmooth discontinuities. In disparity estimation, discontinuities are steep transitions between smoothly varying regions and typically coincides with occluding contours of objects lying at different depths. In order to smooth isotropically inside homogeneous areas and preserve discontinuities around object edges, Total Variation (TV), which we have considered in a previous work,[17] has proven to be a very powerful feature. Initially introduced by Rudin, Osher and Fatemi,[18] this regularity measure has emerged as an effective tool to recover smooth images in variational image recovery,[19] which naturally motivates its extension to the field of variational stereo methods.[17,20]

In the present work, we adopt an alternative wavelet domain approach to construct a regularization constraint in the transformed domain. The key idea of this regularization is to suppose that the semi-norm of the disparity image u, expressed in terms of the wavelet coefficients of u in a smoothness space $\mathcal{X}$, is bounded. The convex constraint modelling this prior information assumes the following general form:

$$S_1 = \{(u, v) \in \mathcal{H} \ | \ \|u\|_{\mathcal{X}} \leq \kappa\}, \tag{13}$$

where $\kappa > 0$ and $\| \, . \, \|_{\mathcal{X}}$ is a semi-norm in the space $\mathcal{X}$. In order to deal with a regularization process that avoids smoothing across discontinuities, we have to consider a function space $\mathcal{X}$ that possesses particular smoothness characteristics. We consider here Besov spaces $B^{\sigma}_{p,q}$ $(0 < \sigma < \infty,\ 1 \leq p \leq \infty,\ 1 \leq q \leq \infty)$ as they are appropriate to model images that contain discontinuities. Moreover, norms in these spaces are equivalent to sequence norms for the image wavelet coefficients. Regularization in Besov spaces has been used with great success in the area of image denoising[12,21,22] and has been shown to guarantee the requested smoothness properties. In this work, by defining a convex set based on an orthonormal wavelet basis, we investigate a new wavelet regularization constraint which is useful in disparity estimation.

Let $W^{\mathcal{B}}$ be the 2-D wavelet transform in a separable wavelet basis $\mathcal{B}$. The wavelet coefficients of $u \in H$ are denoted by $(c^{\mathcal{B}}_{j,o,k}(u))_{j\in\mathbb{Z},k\in\mathbb{Z}^2,o\in\{1,2,3\}}$, where $j \in \mathbb{Z}$ is the resolution level and $o \in \{1, 2, 3\}$ is the orientation parameter. The maximum resolution level is denoted by l.

If the wavelet function possesses $N > \sigma$ vanishing moments, the norm in a Besov space $B^{\sigma}_{p,q}$ of an image $u \in \mathcal{H}$ is equivalent to the following norm of its wavelet coefficients:

$$\|u\|_{B^{\sigma}_{p,q}} \simeq \Big(\sum_{j\in\mathbb{Z}}\Big(\sum_{k\in\mathbb{Z}^2,o\in\{1,2,3\}} 2^{-j(\sigma p+p-2)}|c^{\mathcal{B}}_{j,k,o}(u)|^p\Big)^{q/p}\Big)^{1/q}. \tag{14}$$

Roughly speaking, the space $B^{\sigma}_{p,q}$, mainly determined by the regularity index σ and the tolerance index p, contains functions with σ derivatives in $L^p(\mathbb{R}^2)$. The third index q, providing a finer distinction in smoothness, may appear of secondary importance. One particular Besov space which plays an important role in image processing is the space $B^{\sigma}_{p,p}$ with $p = q$ and $1/p = \sigma/2 + 1/2$. The norm in this space reduces to a simple weighted p-norm of the wavelet coefficients and, by further discarding the diagonal coefficients where noise often dominates, the semi-norm in this space is given by:

$$\|u\|_{B^{\sigma}_{p,p}} \simeq \Big(\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}^2,o\in\{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u)|^p\Big)^{1/p}. \tag{15}$$

Motivated by the success of the space of functions of Bounded Variation (BV), the space of minimal smoothness $B^1_{1,1}$, by taking $p = \sigma = 1$, have also been widely considered. Indeed, this space is very close to the space BV, since $B^1_{1,1} \subset \mathrm{BV}$, and thus it is so appropriate to model images with sharp discontinuities. We propose here to use another semi-norm which is mathematically equivalent to a semi-norm in $B^1_{1,1}$, but which has been shown to yield better results.[11] Imposing an upper bound $\kappa_1 > 0$ on this semi-norm, expressed in terms of the wavelet coefficients of the image, restricts the solutions to the following convex set

$$S_1^{(0,0)} = \{(u, v) \in \mathcal{H} \mid \sum_{j\in\mathbb{Z}, k\in\mathbb{Z}^2} (\sum_{o\in\{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u)|^2)^{1/2} \leq \kappa_1\}. \tag{16}$$

Note that the upper bound κ_1 can be estimated from experiments on available databases.

The above constraint satisfies the properties of edge-preserving regularization. However, it suffers from the shift variance of the wavelet basis decomposition that arises due to the decimation of a factor 2 in the decomposition process. Shift invariance ensures that, when the image is translated in space, its wavelet coefficients are translated as well. For stereo matching, this property is essential because stereo image pairs can be considered as the shifted versions of each other. The shifted value with respect to each pixel is the disparity, which depends heavily on the position of the pixel in the image. In order to deal with translation variance, the continuous wavelet transform can be used. However, its high redundancy increases the computational cost. The shift invariant wavelet representation that we consider in this work amounts to computing the wavelet coefficients of 2^{2l} circular shifts of the original image.[23] In order to limit the computational complexity, rather than using a global smoothness constraint on the wavelet frame coefficients, we split this constraint into several constraints involving the semi-norm of the shifted image. In other words, we consider the following constraint

$$S_1 = \bigcap_{\mathrm{d}\in\{0,\ldots,2^l-1\}^2} S_1^{(\mathrm{d})}, \tag{17}$$

where

$$S_1^{(\mathrm{d})} = \{(u, v) \in \mathcal{H} \mid \sum_{j=1}^{l} \sum_{k\in\mathbb{Z}^2} (\sum_{o\in\{1,2\}} |c^{\mathcal{B}}_{j,k,o}(u^{(\mathrm{d})})|^2)^{1/2} \leq \kappa_1\}. \tag{18}$$

where $u^{(\mathrm{d})}$ is the shifted-by-d image. Obviously, for $\mathrm{d} = (0, 0)$, $u^{(\mathrm{d})}$ is the original image u and $S_1^{(\mathrm{d})}$ simply reduces to the convex set given in (16).

3.3 Spatial constraints

The employed optimization method can handle constraints arising in both the spatial and the wavelet domains. The wavelet regularization constraint being described above, we will now introduce spatial constraints as convex sets modelling prior knowledge on disparity and illumination change.

3.3.1 Disparity range constraint

The most common constraint on disparity is the knowledge of its range of possible values. Indeed, disparity values are nonnegative and often have known minimal and maximal amplitudes, denoted respectively by $u_{\min} \geq 0$ and $u_{\max}$. The associated set is

$$S_2 = \{(u, v) \in \mathcal{H} \mid u_{\min} \leq u \leq u_{\max}\} . \tag{19}$$

3.3.2 Tikhonov based regularization

In most scenes, brightness changes have slow variations in space. A constraint should therefore be imposed on the illumination coefficient v which is consistent with this behavior. A Tikhonov-like quadratic term is appropriate to recover this kind of locally smooth field. Hence, we impose an upper bound on the quadratic norm of the discrete gradient $\widehat{\nabla} v$ of v, so restricting the solution to the convex set

$$S_3 = \{(u, v) \in \mathcal{H} \mid |\widehat{\nabla} v|_2^2 \leq \tau_v\} , \tag{20}$$

where $\tau_v > 0$.

3.3.3 Illumination range constraint

It has been shown in,[9] through experiments with real images, that the illumination change v usually ranges between $v_{\min}$ and $v_{\max}$ where typical values are $v_{\min} = 0.8$ and $v_{\max} = 1.2$. The constraint set arising from this knowledge is

$$S_4 = \{(u, v) \in \mathcal{H} \mid v_{\min} \le v \le v_{\max}\} \ . \tag{21}$$

In summary, the problem of stereo matching robust to photometric variations can be formulated as jointly estimating the disparity and illumination fields which minimize the energy function (12) subject to the constraints $(S_i)_{1 \le i \le 4}$.

4. OPTIMIZATION ALGORITHM

The objective of this section is to develop a numerical solution to the stereo matching problem which has been formulated as a convex optimization problem. A parallel block iterative algorithm will be employed to efficiently minimize the quadratic objective function (12) over the feasibility set $S = \cap_{i=1}^{4} S_i$.

Let $N_r \times N_c$ be the size of the considered images. The solution space is the real Hilbert space $\mathcal{H} = \mathbb{R}^{N_r \times N_c} \times \mathbb{R}^{N_r \times N_c}$, endowed with the standard scalar product $\langle . \mid . \rangle$ and the associated Euclidean norm $\| \ . \ \|$. Let S_i be the nonempty closed and convex subset of $\mathcal{H}$ given by (7), where f_i is a continuous and convex function. For every $w \in \mathcal{H}$, f_i possesses at least one subgradient at w, i.e., a vector $g_i \in \mathcal{H}$ such that

$$\forall \ z \in \mathcal{H}, \qquad \langle z - w \mid g_i \rangle + f_i(w) \le f_i(z) \ . \tag{22}$$

The set of all subgradients of f_i at w is the subdifferential of f_i at w and is denoted by $\partial f_i(w)$. If f_i is differentiable at w, then $\partial f_i(w) = \{\nabla f_i(w)\}$. Fix $w \in \mathcal{H}$ and a subgradient $g_i \in \partial f_i(w)$, the subgradient projection $G_i w$ of w onto S_i is given by:

$$G_i w = \begin{cases} w - \dfrac{f_i(w) - \delta_i}{\|g_i\|^2} g_i, & \text{if } f_i(w) > \delta_i; \\ w, & \text{if } f_i(w) \le \delta_i. \end{cases} \tag{23}$$

The proposed algorithm activates the constraints by means of subgradient projections rather than exact projections. The former are much easier to compute than the latter, as they require only the availability of a subgradient (i.e. the gradient in the differentiable case). However, when the projection is simple to compute, one can use it as a subgradient projection. In our case, exact projections onto S_2 and S_4 are straightforwardly obtained whereas a subgradient projection onto S_3 can be easily calculated. For the constraint S_1, the expression of a subgradient projection can be deduced from the following proposition.[12]

PROPOSITION 4.1. *Let f and φ be two convex functions from $\mathcal{H}$ to $]-\infty, +\infty]$, such that*

$$f = \varphi \ \circ \ W^{\mathcal{B}} \ . \tag{24}$$

Suppose that there exist a point $w \in H$ such that φ is continuous at $W^{\mathcal{B}}(w)$, then

$$\partial f(w) = (W^{\mathcal{B}})^* \partial \varphi(W^{\mathcal{B}}(w)) \ , \tag{25}$$

where $(W^{\mathcal{B}})^* = (W^{\mathcal{B}})^{-1}$ due to the orthonormality of the basis $\mathcal{B}$.

We now proceed to the description of the proposed block iterative algorithm to simultaneously estimate the disparity u and illumination v.

ALGORITHM 4.1.

① *Set $n = 0$. Compute w_0 as*

$$w_0(s) = \begin{cases} (L(s)^\top L(s) + \alpha \ I_{2,2})^{-1}(L(s)^\top r(s) + \alpha \ \bar{w}(s)) & \text{if } s \in \mathcal{D} \setminus \mathcal{O}, \\ \bar{w}(s) & \text{otherwise.} \end{cases} \tag{26}$$

② *Take a nonempty index set* $\mathbb{K}_n \subseteq \{1, \dots, m\}$.

③ *For every* $i \in \mathbb{K}_n$, *set* $a_{i,n} = G_{i,n} - w_n$ *where* $G_{i,n}$ *is a subgradient projection of* w_n *onto* S_i *as in* (23).

④ *Set* $z_n = |\mathbb{K}_n|^{-1} \sum_{i \in \mathbb{K}_n} a_{i,n}$ *and* $\kappa_n = |\mathbb{K}_n|^{-1} \sum_{i \in \mathbb{K}_n} \|a_{i,n}\|^2$, *where* $|\mathbb{K}_n|$ *denotes the number of elements in* $\mathbb{K}_n$.

⑤ *If* $\kappa_n = 0$, *exit iteration. Otherwise, set*

- $b_n = w_0 - w_n$,
- c_n *such that*

$$c_n(s) = \begin{cases} (L(s)^\top L(s) + \alpha\ I_{2,2}) b_n(s) & \text{if } s \in \mathcal{D} \setminus \mathcal{O}, \\ \alpha\ b_n(s) & \text{otherwise,} \end{cases} \tag{27}$$

- d_n *such that*

$$d_n(s) = \begin{cases} (L(s)^\top L(s) + \alpha\ I_{2,2})^{-1} z_n(s) & \text{if } s \in \mathcal{D} \setminus \mathcal{O}, \\ \alpha^{-1} z_n(s) & \text{otherwise,} \end{cases} \tag{28}$$

- $\lambda_n = \kappa_n / \langle d_n, z_n \rangle$.

⑥ *Set* $\tilde{d}_n = \lambda_n d_n$, $\pi_n = -\langle c_n, \tilde{d}_n \rangle$, $\mu_n = \langle b_n, c_n \rangle$, $\nu_n = \lambda_n \langle \tilde{d}_n, z_n \rangle$ *and* $\rho_n = \mu_n \nu_n - \pi_n^2$.

⑦ *Set*

$$w_{n+1} = \begin{cases} w_n + \tilde{d}_n, & \text{if } \rho_n = 0,\ \pi_n \geq 0; \\ w_0 + (1 + \frac{\pi_n}{\nu_n}) \tilde{d}_n, & \text{if } \rho_n > 0,\ \pi_n \nu_n \geq \rho_n; \\ w_n + \frac{\nu_n}{\rho_n} (\pi_n b_n + \mu_n \tilde{d}_n), & \text{if } \rho_n > 0,\ \pi_n \nu_n < \rho_n. \end{cases} \tag{29}$$

⑧ *Increment* n *and go to step* ②.

This algorithm is well adapted to our need since it allows the combination of constraints arising in both the spatial and the wavelet domain. In addition, it has been shown in[14] that, due to its block iterative structure, this algorithm offers a lot of flexibility in terms of parallel implementation. In particular, several processors can be used in parallel to compute the subgradient projections on the different constraint sets $(S_i)_{1 \leq i \leq m}$. For instance, if 2^{2l} parallel processors are available, wavelet based constraints with respect to all possible circular shifts s in (18) can be processed simultaneously, leading to improved results while reducing the computational time. We finally note that the constraints considered in this work are separable in the sense that either u or v is constrained for each of them. This means that the associated subgradient projections will remain the other field (v or u) unchanged. This can be exploited to limit the memory load in the implementation of the algorithm, in particular in Step ③.

5. EXPERIMENTAL RESULTS

In this section we evaluate the performance of the proposed method using both synthetic and real data sets with varying illumination. We also compare our results with those obtained using the SSD block matching method with an affine illumination variation model, the standard normalized cross-correlation method (NCC) and the Semi-Global Matching (SGM) algorithm of Hirschmller.[7] To parameterize our method, the constant α in equation (12) was set to 50 and bounds on the constraint sets $(S_i)_{1 \leq i \leq 4}$ were fixed by calculating first the values of the associated convex functions on the initial disparity and illumination estimates and then choosing the appropriate bounds as a fixed ratio (40 %) of these values. If true disparity and illumination fields are available, exact bounds on the constraint sets can be computed directly from these known fields. For the choice of the wavelet basis, different families have been tested and it appeared that, when translation-invariant representations are used, Haar wavelets are well-suited for disparity map images with sharp discontinuities. Notice finally that three cycles of iterations were performed, for the considered stereo images, to refine the initial disparity and illumination fields as described in Section 3.1.

(a) (b) (c)

(d) (e) (f)

Figure 1. Results for the *Corridor* stereo pair. (a) Left image. (b) SSD with affine model. (c) Normalized cross-correlation. (d) SGM algorithm. (e) Our approach. (f) Ground truth.

Table 1. Comparative results on the *Corridor* and *Dolls* stereo pairs.

	Corridor		*Dolls*	
Technique	MAE	Err	MAE	Err
SSD with affine illumination	1.58	28	2.65	23
Normalized cross-correlation	1.54	26	1.73	21
SGM[7]	0.47	13	0.67	11
Our approach	**0.34**	**10**	**0.51**	**9**

5.1 Synthetic data with artificial illumination variation

We first demonstrate our method using the synthetic *Corridor* stereo pair from the University of Bonn (see Figure 1). To introduce a significant illumination variation in this stereo pair, we modified the right image by multiplying it with the Gaussian profile shown in Figure 2(a). As both true disparity and illumination fields are available for this data set, we evaluate the different results quantitatively by computing two error measures: the Mean Absolute Error (MAE) between computed and ground truth fields and the percentage of bad matching pixels (Err) with absolute error larger than 1.

(a) (b)

Figure 2. Illumination change maps for the modified *Corridor* stereo pair: (a) true map and (b) estimated map.

The computed disparity map and the ground truth are shown in Figure 1, along with the results from other methods. As we can see, our method is not affected by illumination changes and provides an accurate depth map. We also observe its robustness with respect to depth discontinuities. In particular, low textured areas corresponding to objects, ceiling and floor are smoothly estimated while edges remain sharp. This is a consequence of using the wavelet based regularization constraint. From the results in Table 1, it is clear that Hirschmller's approach[7] and our method outperform the other ones. However, unlike SGM, our approach allows us to estimate illumination changes. Figures 2(a) and 2(b) allow us to compare the recovered illumination field using our method and the ground truth. The mean absolute error for the illumination field estimation is 0.015.

5.2 Real images with ground truth and real exposure changes

In this experiment, we evaluate our method on real stereo datasets with real illumination changes. These datasets are available at the Middlebury stereo vision website*. Each dataset consists of 7 rectified views taken with three different exposures and from equidistant points along a line. Ground truth disparity maps, created by using the structured lighting technique of,[24] are provided for viewpoints 2 and 6 of each dataset. For the results in this paper, we only consider the stereo pair shown in Figure 3, named *Dolls.* As left and right input images, we use images 2 and 6 taken with exposure 2 and 1, respectively.

In Figure 3, the results provided by our method are presented and compared to those from other methods. We notice that both SSD with affine illumination and NCC perform poorly and incur large errors in the illumination incoherent regions. However NCC gives better results than SSD. Our approach gives more consistent results than the other methods. As expected, severe matching errors are greatly reduced by using the proposed illumination variation model. For the quantitative analysis, we see from the results reported in Table 1, that

*http://cat.middlebury.edu/stereo/scenes2005/

(a) (b) (c)

(d) (e) (f)

Figure 3. Results for the *Dolls* stereo pair. (a) Left image. (b) SSD with affine model. (c) Normalized cross-correlation. (d) SGM algorithm. (e) Our approach. (f) Ground truth.

(a) (b) (c)

(d) (e) (f)

Figure 4. Results on the *Shrub* stereo pair. (a) Left image. (b) Right image. Estimated disparity map using (c) SSD with affine model. (d) Normalized cross-correlation. (e) SGM algorithm. (f) Our approach.

the proposed method again leads to the best results. Its performance is however close to that of the SGM algorithm.

5.3 Real data

For the real image case, we show the results provided by the considered methods for a real image pair taken under real illumination variations. Figure 4 shows the left image of the *Shrub* stereo pair from the well-known JISCT database. Examination of this stereo pair reveals that the left and right image pair violated the constant image brightness assumption. Indeed, the noticed difference between corresponding intensity histograms indicates that this stereo pair incorporates a real illumination change.

We display in Figure 4 the results obtained with the proposed method and we show comparison with various methods. However, since no ground truth is available, the comparison is only visual. We notice that local methods give noisy results and are very sensitive to illumination changes while the SGM algorithm and our method allow to obtain a smooth disparity map with sharp depth discontinuities and they show good performance in the illumination inconsistency regions.

6. CONCLUSION

We have proposed a convex programming approach for the problem of stereo matching in the presence of photometric variations. To deal with these variations, we have developed a spatially varying multiplicative model that accounts for brightness changes between both images in the stereo pair. Within a convex set theoretic framework, a two-dimensional quadratic objective function was derived and efficiently minimized subject to convex constraints. Based on wavelet frames, we have investigated an edge-preserving regularization constraint on the disparity image. The resulting multi-constrained optimization problem is solved via a block iterative method that allows to combine both wavelet and spatial domain constraints. The experimental results reported in this paper on synthetic and real stereo pairs with variable illumination show that our method is robust and reliable.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer vision* **47**(1), pp. 7–42, 2002.

[2] J. Konrad and A. Mancini, "Robust quadtree-based disparity estimation for the reconstruction of intermediate stereoscopic images," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, **3995**, pp. 53–64, Jan. 1998.

[3] I. J. Cox, S. Roy, and S. L. Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," in *Proc. Inter. Conf. on Image Processing*, **2**, pp. 366–369, (Washignton, USA), Oct. 1995.

[4] P. A. Viola and W. M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision* **24**, pp. 137–154, Sep. 1997.

[5] G. Egnal, "Mutual information as a stereo correspondence measure," Technical Report MS-CIS-00-20, Computer and Information Science, University of Pennsylvania, Philadelphia, USA, 2000.

[6] J. Kim, V. Kolmogorov, and R. Zabih, "Visual correspondence using energy minimization and mutual information," in *International Conf. Computer Vision*, pp. 1033–1040, (Nice, France), Oct. 2003.

[7] H. Hirschmller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. Computer Vision and Pattern Recognition*, pp. 807–814, (Washington, DC, USA), June 2005.

[8] I. J. Cox, "A maximum likelihood N-camera stereo algorithm," in *Proc. International Conference on Computer Vision and Pattern Recognition*, pp. 733–739, (Seattle, WA), June 1994.

[9] M. A. Gennert, "Brightness-based stereo matching," in *Proc. International Conference on Computer Vision*, pp. 139–143, 1988.

[10] J. Zhang, L. McMillan, and J. Yu, "Robust tracking and stereo matching under variable illumination," in *Proc. International Conference on Computer Vision and Pattern Recognition*, **1**, pp. 871–878, (New York, USA), June 2006.

[11] W. Miled, J. C. Pesquet, and M. Parent, "Wavelet-constrained regularization for disparity map estimation," in *Proc. European Signal and Image Processing Conference*, (Florence, Italy), Sep. 2006.

[12] P. L. Combettes and J. C. Pesquet, "Wavelet-constrained image restoration," *International Journal of Wavelets, Multiresolution, and Information Processing* **2**, pp. 371–389, Dec. 2004.

[13] P. L. Combettes, "The convex feasibility problem in image recovery," in *Advances in Imaging and Electron Physics*, E. P. Hawkes, ed., **95**, pp. 155–270, New York: Academic Press, 1996.

[14] P. L. Combettes, "A block iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Transactions on Signal Processing* **51**, pp. 1771–1782, Jul. 2003.

[15] G. Egnal and R. P. Wildes, "Detecting binocular half-occlusions: empirical comparisons of five approaches," *IEEE Trans. Pattern Analysis and Machine Intelligence* **24**, pp. 1127–1133, Aug. 2002.

[16] A. Tikhonov and A. Arsenin, "Solution of ill-posed problems," *John Wiley and Sons* , 1977.

[17] W. Miled, J. C. Pesquet, and M. Parent, "Disparity map estimation using a total variation bound," in *Proc. 3rd Canadian Conference on Computer and Robot Vision*, pp. 48–55, (Quebec, Canada), June 2006.

[18] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259–268, 1992.

[19] P. L. Combettes and J. C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Transactions on Image Processing* **13**, pp. 1213–1222, Sep. 2004.

[20] N. Slesareva, A. Bruhn, and J. Weickert, "Optic flow goes stereo : A variational method for estimating discontinuity preserving dense disparity maps," in *DAGM-Symposium*, pp. 33–40, 2005.

[21] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Transactions on Image Processing* **7**, pp. 319–335, Mar. 1998.

[22] H. Choi and R. Baraniuk, "Multiple basis wavelet denoising using besov projections," in *Proc. International Conference on Image Processing*, **1**, pp. 595–599, (Kobe, Japan), Oct. 1999.

[23] J. C. Pesquet, H. Krim, and H. Carfantan, "Time invariant orthonormal wavelet representations," *IEEE Transactions on Signal Processing* **44**, pp. 1964–1970, Aug. 1996.

[24] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. Computer Vision and Pattern Recognition*, **1**, pp. 195–202, (Madison, WI), June 2003.