INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Joint Framework for Disparity and Surface Normal Estimation

Ramya Narasimha — Elise Arnaud — Florence Forbes — Radu Horaud

N° 7090

October 2009

Domaine 1

ISSN 0249-6399 ISRN INRIA/RR--7090--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE
INRIA
centre de recherche GRENOBLE - RHÔNE-ALPES

A Joint Framework for Disparity and Surface Normal Estimation

Ramya Narasimha , Elise Arnaud , Florence Forbes, Radu Horaud

Domaine : Mathématiques appliquées, calcul et simulation
Équipes-Projets PERCEPTION

Rapport de recherche n° 7090 — October 2009 — 17 pages

Abstract:
This paper deals with the stereo matching problem, while moving away from the traditional fronto-parallel assumption. We propose an algorithm that provides disparities in accordance with the surface properties of the scene under consideration. To do so, we carry out cooperatively both disparity and surface normal estimations by setting the two tasks in a unified Markovian framework. A novel joint probabilistic model is defined through two Markov Random Fields (MRF) to favor both intra field (within neighboring disparities and neighboring normals) and inter field (between disparities and normals) consistency. Geometric contextual information is introduced in the pair-wise Markovian regularizing term used in both MRFs. Segmentation and plane fitting procedures, usually performed as final steps to increase the quality results are here explicitly used in one of the MRF data terms. We then design an appropriate alternating maximization procedure based on standard Belief Propagation. We illustrate the performance of our approach on synthetic and real data. The results obtained are comparable to the state-of-the-art and show improvement in many cases.

Key-words: Stereo vision, Markov Random Fields, Belief propagation Alternating Maximization

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Modèle Markovian pour l'estimation conjointe de la disparité et des normales à la surface

Résumé :
Dans cet article, nous abordons le problème de stéréo correspondance, tout en levant l'hypothèse classique qui suppose que la scène observée est constituée de plans parallèles au plan focal. Un algorithme permettant d'estimer les disparités de la scène en accord avec les normales à la surface est proposé. Pour cela, les estimation des disparités et des normales sont réalisées de façon alternée et coopérative, et ces deux étapes sont unifiées dans un cadre markovien. Un modèle joint probabiliste est défini à partir de deux champs de Markov, afin d'assurer la cohérence intra-champs (entre les disparités voisines, et entre les normales voisines), et inter- champs (entre les disparités et les normales). L'information géométrique contextuelle est introduite dans le modèle de Markov au travers du terme de régularisation, identique pour les deux champs. Les étapes de segmentation et d'approximation par des plans, classiquement utilisées en post-processing pour améliorer les résultats, sont ici introduites explicitement dans les termes d'attache aux données des champs de Markov. La procédure de optimisation alternée est basée sur l'algorithme Belief propagation. La méthode est validée sur des données synthétiques et des données réelles.

Mots-clés : Stereo vision, Champs de Markov, Belief propagation, Optimisation alternée

Contents

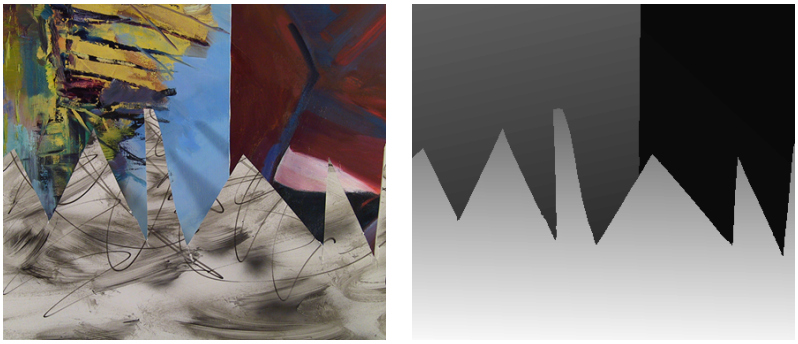
1 Introduction

Stereo matching has been one of the core challenges in computer vision for decades. The most recent algorithms show very good performance, as one can see on the Middlebury dataset benchmark. However, as outlined in [14, 15], most existing stereo algorithms have inherent *fronto-parallel assumption* in their modeling of the stereo correspondence problem. Such an assumption supposes that the scene under consideration can be approximated by a set of fronto-parallel planes (on which the disparity is constant) and thus biases the results towards staircase solutions. For example, for a stereo input of Figure 1(a)[1] with ground-truth Figure 1(b), the result obtained from a Belief Propagation algorithm without post-processing is Figure 1(c). In this paper, we propose an novel algorithm that provides surface consistent solutions, as illustrated in Figure 1(d).

1.1 Related Work

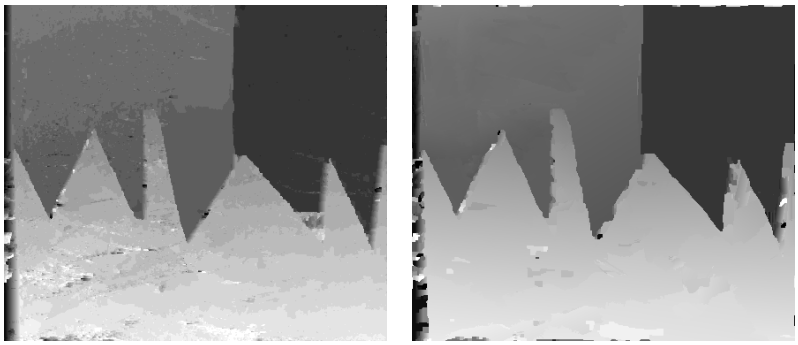
Few attempts have been made to move beyond the fronto-parallel assumption in stereo correspondance problem. Geometric constraints have been considered to account for more general surfaces *e.g.* slanted or curved surfaces. In [5], it is proposed to extend the classical correlation method to estimate both the disparity and its derivatives directly from the image data. They then relate these derivatives to differential properties of the surface such as those encoded by normals and curvatures. However, in this paper the authors acknowledge numerical instability when computing second-order or higher disparity derivatives. In [2][16], the authors propose to perform iteratively segmentation and

[1] Note that only the left image is shown



(a) Left reference image

(b) Disparity ground truth



(c) Staircase solution obtained by standard Belief Propagation

(d) Output using our approach

Figure 1: Sawtooth Image: staircase (c) versus surface consistent (d) disparity maps.

correspondence, so that each segmented region can be modeled as a slanted or curved surface. The limitation of this approach is that the proposed algorithm is likely to get stuck in local minima in case of untextured surfaces. Also they attempt to fit a plane or a spline to large surface patches and do not consider the geometrical constraints of the surface itself.

Beyond such correlation-based algorithms that do not guarantee local consistency of the disparity map, most of the recent methods are based on neighborhood regularization via Markov Random Field (MRF) modeling using Belief Propagation (BP) or Graph Cuts [3, 6, 19, 21]. In such a setting, slanted surfaces are usually recovered by post-processing the disparities using plane-fitting [10, 11, 20, 22, 25, 27]. The disparity plane parameters are optimized in the final step to obtain disparity maps. The regions for the plane fitting are obtained from an over-segmentation process based on color or grey values, so that disparity discontinuities are preserved. A good taxonomy of some of these approaches can be found in [18]. Although most of them perform very well on the Middlebury data set, these approaches implicitly assume fronto-parallel planes in the definition of their objective function. Even with additional post-processing, the latter cannot handle curved surfaces.

One way to explicitly take into account the differential geometric contextual information, in the MRF-based disparity estimation framework, is proposed in [14, 15]. The authors introduce the notion of geometric consistency between nearby matching pairs using both depth (position disparity) and surface normals. They measure the consistency of the normals by transporting them along the surface. They show that contextual information can be enriched with extra geometric constraints for stereo correspondence. As a consequence, they propose to use surface normals and curvatures to guide the disparity estimation towards a geometrically consistent map. In order to overcome numerical instability issues, they perform all the derivative computations in the depth space. As well as requiring the knowledge of the internal camera parameters, this algorithm precomputes the local suface normals. Torr et al. [23] present a framework incorporating higher order proirs to encode the surface properties. and optimize using a new QBPO algorithm.

In this paper, we propose to carry out cooperatively both disparity and surface normal estimations by setting the two tasks in a unified Markovian framework. We define a new joint probabilistic model based on the definition of two MRFs that are linked to encode consistency between disparities and surface properties.

1.2 Multiple Random Fields

The idea of using multiple MRFs to estimate one or more variables has been previously used in contexts such as: image segmentation, motion, boundary detection, including stereo. Bouthemy et al [9] suggested a simultaneous estimation of motion discontinuities and optical flow. Classification of edges based on their source of origin (depth, texture, motion, or color) was suggested by Geiger et al [8] under a similar MRF framework. Their work, however, uses depth, texture, etc., as cues rather than actually estimating them. In [12], Kolmogrov et al, combine color, contrast and depth to achieve background/foreground separation. In this paper, stereo information is used mainly to differentiate between occluded regions and background/foreground. Here, stereo disparity is used only as cue, along with color and contrast, for background/foreground segmentation. Sun et al [21] integrate three MRFs, i.e., disparity, line process and occlusion, for estimating disparity. In [17] stereo disparity is simultaneously extracted with boundary information using coupled-MRFs

In our case consideration of normal and disparity maps as two separate random fields increases the model's flexibility. For both MRFs, we include geometric contextual information in the pair-wise regularizing term, thus favoring a disparity solution consistent with the scene surfaces – possibly slanted and/or curved. The respective MRFs data terms are designed to extract data information that specifically impact both the disparity and normal fields. In particular, for normals, we propose a data term favoring proximity to a set of *observed* normals derived from an over-segmentation of the image into small regions and a plane fitting procedure, thus including explicitly these steps within the model. The surface properties are then approximated using surface normals in disparity space. These normals provide a reasonable approximation of the true surface. The proposed joint model results in a posterior distribution, for both the disparity and normal fields, which is used for their estimation according to a Maximum A Posteriori (MAP) principle using standard Belief propagation algorithms. The

alternating maximization procedure used for the MAP search leads to cooperative estimation and mutual improvement of disparity and normal accuracies. Moreover, our approach has the following advantages:

- it does not require the computation of high-order disparity derivatives,
- it embeds the estimation of surface properties in the Markovian model rather than refining the results using a post-processing step,
- the consideration of normal and disparity as two separate random fields increases modeling flexibility i.e, the energy terms for two fields can be made more dependent or independent according to the information to be incorporated,
- the use of a well based statistical estimation framework resulting in cooperative estimation and mutual improvement of both disparities and normals accuracy, and
- the algorithm does not depend upon the type of optimization used. In our case we use the standard Belief propagation to optimize both the fields.

The remaining of this document is organized as follows. Details of the proposed joint model are provided in section 2. The alternating estimation procedure is explained in section 3. Experimental results and concluding remarks are presented in section 4 and section 5 respectively.

2 Joint Disparity and Normal Model

We consider a finite set $\mathcal{S}$ of $p \times q$ pixels on a regular 2D-grid. The observed data are made of left and right images, $\mathbf{I}_L$ and $\mathbf{I}_R$, which are together referred to as $\mathbf{I}$. In our setting, the left image is taken as the reference image. We denote by $\mathbf{D} = \{D_{\mathbf{x}}, \mathbf{x} \in \mathcal{S}\}$ the unknown disparity values at each pixel $\mathbf{x} = (u, v)$. The $D_{\mathbf{x}}$'s are considered as random variables that take their values in a finite discrete set of L disparity labels $\mathcal{L}_L$. $\mathbf{D}$ is referred to as the disparity field or disparity map and takes its values in $\mathcal{D} = \mathcal{L}_L^{p \times q}$. Similarly, we consider a surface normal field $\mathbf{N} = \{\mathbf{N}_{\mathbf{x}}, \mathbf{x} \in \mathcal{S}\}$. Each $\mathbf{N}_{\mathbf{x}} = (N_u, N_v, N_d)$ takes its values from a discrete space $\mathcal{N}_K$ corresponding to $K = 162$ directions uniformly distributed on the unit sphere. As in [24], $\mathcal{N}_K$ is built from subdivisions of a uniform unit icosahedron. The normal configuration space is then denoted by $\mathcal{N} = \mathcal{N}_K^{p \times q}$.

To account for relationships between disparities and surface normals, we adopt a data conditional approach (see [13] for the rationale behind such an approach) and consider a joint conditional model $p(\mathbf{d}, \mathbf{n}|\mathbf{I})$. We use small letters $\mathbf{d}$ and $\mathbf{n}$ to denote specific realizations of the random fields $\mathbf{D}$ and $\mathbf{N}$. Defining such a joint model is equivalent to define the two conditional distributions $p(\mathbf{d}|\mathbf{n}, \mathbf{I})$ and $p(\mathbf{n}|\mathbf{d}, \mathbf{I})$. These two conditional distributions are enough to account for cooperation mechanisms between $\mathbf{D}$ and $\mathbf{N}$. Their use allows more flexibility in the modeling than a single complex joint distribution that would necessarily be more constrained. In particular, defining Markovian conditional models does not necessarily lead to a Markovian joint model. In addition, conditional models can easily be associated to estimation procedures that alternate naturally between disparity and normal estimation.

2.1 Normal conditional disparity model

We first specify the disparity distribution conditionally to the normal field and the observed data. The model is expressed as a Markov Random Field (MRF) with an energy function consisting of two terms, a data dependent term and a regularizing or interaction term. Our data term is set to be the normalized Sum of Absolute Differences (SAD) over a window between the pixels of the left and right images. Our interaction term is a symmetric modified version of the one presented in [15]. Although the interpretation is similar, we propose to include geometric information via surface normals considered as a separate random field. Expressing compatibility between the disparity and normal fields enables us to encode geometric constraints without computing disparity derivatives directly from the disparity field. Furthermore, we achieve results similar to [15] using only first order differential information but obtained from the normal field. More specifically, we define, for all $\mathbf{n} \in \mathcal{N}$, $p(\mathbf{d}|\mathbf{n}, \mathbf{I})$ as a MRF on $\mathcal{D}$, as follows:

$$p(\mathbf{d}|\mathbf{n}, \mathbf{I}) \propto \Phi_D(\mathbf{d}, \mathbf{I})\ \Psi(\mathbf{d}, \mathbf{n})\ . \tag{1}$$

2.1.1 Data Term for Disparity Field

The first term represents the **data term** and assigns a cost at each location $\mathbf{x}$ based on the intensity difference between the left and the right images. It is formulated as a robust function depending on two parameters λ and T specified in Section 4. $\mathbf{I}_L(\mathbf{x})$ (resp. $\mathbf{I}_R(\mathbf{x})$) denotes the pixel value at location $\mathbf{x}$ in the left (resp. right) image, we set:

$$\Phi_D(\mathbf{d}, \mathbf{I}) = \exp\Big(-\sum_{\mathbf{x}\in\mathcal{S}} \lambda \min\Big(\sum_{\tilde{\mathbf{x}}\in W_{\mathbf{x}}} |\mathbf{I}_L(\tilde{\mathbf{x}}) - \mathbf{I}_R(\tilde{\mathbf{x}} + d_{\mathbf{x}} e_2)|, 2T\Big)\Big)\ , \tag{2}$$

where $W_{\mathbf{x}}$ is a window of size $w \times w$ centered at $\mathbf{x}$ and $e_2 = (0, 1)$ is one of the canonical unit vectors. For color images, considering $\mathbf{I}_L(\mathbf{x})$ and $\mathbf{I}_R(\mathbf{x})$ as luminance values, we found that SAD worked well for most cases. However, more complex functions as described in [26] could be used to integrate the color properties.

2.1.2 Interaction Term

Our **interaction term** has then a standard pair-wise interaction form $\Psi(\mathbf{d}, \mathbf{n}) = \prod_{(\mathbf{x},\mathbf{y})} \psi(d_{\mathbf{x}}, d_{\mathbf{y}}, \mathbf{n})$ where $(\mathbf{x}, \mathbf{y})$ denotes neighboring pixels on the image grid. The term $\psi(d_{\mathbf{x}}, d_{\mathbf{y}}, \mathbf{n})$ specifies how the neighboring disparities interact but also encodes geometric constraints via consistency with the surface normal field $\mathbf{n}$. In the spirit of the co-planar model in [15], our interaction term favors neighboring disparities lying on the same planar surface. For a valid MRF energy term, we propose then an expression which is symmetric in $\mathbf{x}$ and $\mathbf{y}$:

$$\begin{aligned}
&\psi(d_{\mathbf{x}}, d_{\mathbf{y}}, \mathbf{n}) = \psi(d_{\mathbf{x}}, d_{\mathbf{y}}, \mathbf{n}_{\mathbf{x}}, \mathbf{n}_{\mathbf{y}}) = \\
&(1-\epsilon_D)\Bigg[\exp\Bigg(-\frac{|d_{\mathbf{y}} - d_{\mathbf{x}} - \frac{\partial d_{\mathbf{x}}}{\partial u}(u_{\mathbf{y}} - u_{\mathbf{x}}) - \frac{\partial d_{\mathbf{x}}}{\partial v}(v_{\mathbf{y}} - v_{\mathbf{x}})|}{\sigma_D}\Bigg) \\
&\exp\Bigg(-\frac{|d_{\mathbf{x}} - d_{\mathbf{y}} - \frac{\partial d_{\mathbf{y}}}{\partial u}(u_{\mathbf{x}} - u_{\mathbf{y}}) - \frac{\partial d_{\mathbf{y}}}{\partial v}(v_{\mathbf{x}} - v_{\mathbf{y}})|}{\sigma_D}\Bigg)\Bigg] + \epsilon_D\ ,
\end{aligned} \tag{3}$$

where $\mathbf{x} = (u_{\mathbf{x}}, v_{\mathbf{x}})$, $\mathbf{y} = (u_{\mathbf{y}}, v_{\mathbf{y}})$, ϵ_D and σ_D are scalar parameters for robustness. With $\mathbf{n_x} = (n_u, n_v, n_d)$, the disparity partial derivatives are computed from the normals as follows:

$$\frac{\partial d_{\mathbf{x}}}{\partial u} = -\frac{n_u}{n_d} \;\; ; \;\; \frac{\partial d_{\mathbf{x}}}{\partial v} = -\frac{n_v}{n_d}.$$

2.2 Disparity conditional normal model

Similarly, the disparity conditional normal model is defined as a pair-wise MRF of the form

$$p(\mathbf{n}|\mathbf{d}, \mathbf{I}) \propto \Phi_N(\mathbf{n}, \mathbf{d}, \mathbf{I}) \; \Psi(\mathbf{n}, \mathbf{d}) \; , \tag{4}$$

where the **interaction term** $\Psi(\mathbf{n}, \mathbf{d})$ is the same as in the disparity model above. The rationale is that $\Psi(\mathbf{d}, \mathbf{n})$ as defined by (3) (in Section 2.1.2) corresponds to consistency conditions which are joint between $\mathbf{d}$ and $\mathbf{n}$ and represents mutual interactions. As indicated, in (3), the dependence in $\mathbf{n}$ actually reduces to a dependence on $\mathbf{n_x}$ and $\mathbf{n_y}$. In addition, expression (3) encodes the desirable property that neighboring normals are not smoothed if neighboring disparities are too different.

2.2.1 Data term for Normal Field

The normal model differs then mainly in the form of the data term $\Phi_N(\mathbf{n}, \mathbf{d}, \mathbf{I})$. The idea is to focus on data information that can directly and specifically impacts the normal field values. The particularity is that for normals estimation, data information cannot be expressed in terms of $\mathbf{I}$ only but depends also on the current disparity field $\mathbf{d}$. Assuming a current $\mathbf{d}$, we compute an *observed* normal value $\mathbf{p_x}$, for each pixel $\mathbf{x} \in \mathcal{S}$, using a plane fitting procedure. A set of planes is found by first segmenting the image (in our case the left image $\mathbf{I}_L$) into small regions and then approximating the surface corresponding to $\mathbf{d}$ in each region by a plane. For a given region, the normal to its plane provides then the value of $\mathbf{p_x}$ for all $\mathbf{x}$ in the region. It follows a **data term** that favors small distances between the current normals $\mathbf{n}$ and the *observed* normals denoted by $\mathbf{p} = \{\mathbf{p_x}, \mathbf{x} \in \mathcal{S}\}$:

$$\Phi_N(\mathbf{n}, \mathbf{d}, \mathbf{I}) = \exp(-\sum_{\mathbf{x} \in \mathcal{S}} ||\mathbf{n_x} - \mathbf{p_x}||^2) \; . \tag{5}$$

This segmentation and plane fitting process is essential for two reasons:

1) when the regions are small enough, the resulting set of planes provides a reasonable approximation of the possibly curved disparity surface as a piece wise linear surface;

2) the regions are found according to the color/greylevel properties of the various objects in the scene so that the resulting normals are likely to reflect a number of discontinuities.

The above data term then prevents over smoothing of these discontinuities by balancing the regularizing term (3) which is performed across regions.

3 Alternating maximization

Our goal is to estimate realizations of $\mathbf{D}$ and $\mathbf{N}$ that are consistent with our joint probabilistic model and the observed stereo image pair. Ideally we are interested in finding the MAP (Maximum A Posteriori) estimates of $\mathbf{D}$ and $\mathbf{N}$,

$$(\mathbf{d}^{MAP}, \mathbf{n}^{MAP}) \quad = \quad \arg\max_{\mathbf{d},\mathbf{n}} p(\mathbf{d}, \mathbf{n}|\mathbf{I})\,.$$

However, this global optimization problem has in general no straightforward solution. Thus, we consider instead an iterative approach consisting in maximizing the posterior probability alternatively in the first and second variable. Starting from current estimates $\mathbf{d}^{(t)}$ and $\mathbf{n}^{(t)}$ at iteration t, we consider the following updating,

$$\mathbf{d}^{(t+1)} = \arg\max_{\mathbf{d}\in\mathcal{D}} p(\mathbf{d}, \mathbf{n}^{(t)}|\mathbf{I})$$

$$\mathbf{n}^{(t+1)} = \arg\max_{\mathbf{n}\in\mathcal{N}} p(\mathbf{d}^{(t+1)}, \mathbf{n}|\mathbf{I})\,.$$

It follows easily that,

$$p(\mathbf{d}^{(t+1)}, \mathbf{n}^{(t+1)}|\mathbf{I}) \geq p(\mathbf{d}^{(t)}, \mathbf{n}^{(t)}|\mathbf{I})\,.$$

so that the alternation procedure increases the posterior probability at each step. There is no guarantee that such a procedure leads to a global maximum but similar principle has been used in other MRF optimization schemes, such as ICM [1] and satisfactory results are observed in a lot of complex tasks. In addition, using Bayes' theorem, it is easy to show that the alternation above is equivalent to the following one,

$$\mathbf{d}^{(t+1)} = \arg\max_{\mathbf{d}\in\mathcal{D}} p(\mathbf{d}|\mathbf{n}^{(t)}, \mathbf{I}) \tag{6}$$

$$\mathbf{n}^{(t+1)} = \arg\max_{\mathbf{n}\in\mathcal{N}} p(\mathbf{n}|\mathbf{d}^{(t+1)}, \mathbf{I})\,. \tag{7}$$

3.1 Floating Disparity Labels

The method stated above allows us to use directly the conditional distributions defined in the previous section. However, the distribution involved in both (6) and (7) are the MRF models defined in (1) and (4) for which a direct optimization is intractable. We therefore approximate the optimization process using Max-product Belief Propagation (BP). More specifically, at iteration t, the BP procedure in step (6) is initialized with an interpolated disparity map obtained during the plane fitting at the previous iteration (step (7) at iteration $t-1$).

This enables us to follow [15] by considering so-called *floating disparity* labels. Starting from a discrete set of L integer disparity labels $\mathcal{L}_L = \{d_1, \ldots, d_L\}$, we allow them to move to another set of L possibly non-integer labels. The idea is to capture finer geometric features by adapting the initial disparity discretized grid to the image scene. Importantly, this can be done while keeping the discrete pair-wise MRF formulation (3). Considering at iteration t a current continuous

value d at pixel $\mathbf{x}$, we find l such that $d \in [d_l; d_{l+1})$ and then change the disparity label d_l to d. This provides an efficient alternative to the quickly intractable increase of L.

3.2 Overall Optimization Procedure

The resulting alternation procedure is the following: at iteration $t = 0$, the all the normal field values are assumed to be $\{0, 0, 1\}$ and the first step (6) is performed to get a first estimate of the disparity map. Then denoting by $\mathbf{n}^{(t)}$ and $\mathbf{d}^{(t)}$ current estimates of the normal and disparity fields, the two steps below are carried out alternatively,

1) Update the normal field $\mathbf{n}^{(t)}$ into $\mathbf{n}^{(t+1)}$ by:

   (i) segmenting the left image into small regions;

   (ii) computing for each of the obtained regions, the observed normals $\mathbf{p} = \{\mathbf{p}_\mathbf{x}, \mathbf{x} \in \mathcal{S}\}$ from a plane fitting procedure using $\mathbf{d}^{(t)}$ and $\mathbf{I}$ on each of the obtained region; and

   (iii) updating the normal field using Max-product BP for the model defined in (4).

2) Update the disparity field $\mathbf{d}^{(t)}$ into $\mathbf{d}^{(t+1)}$ by:

   (i) computing the first order disparity derivatives using $\mathbf{n}^{(t+1)}$ and

   (ii) updating disparity estimates into $\mathbf{d}^{(t+1)}$ with Max-product BP applied to the conditional disparity model (1).

4 Experimental Results

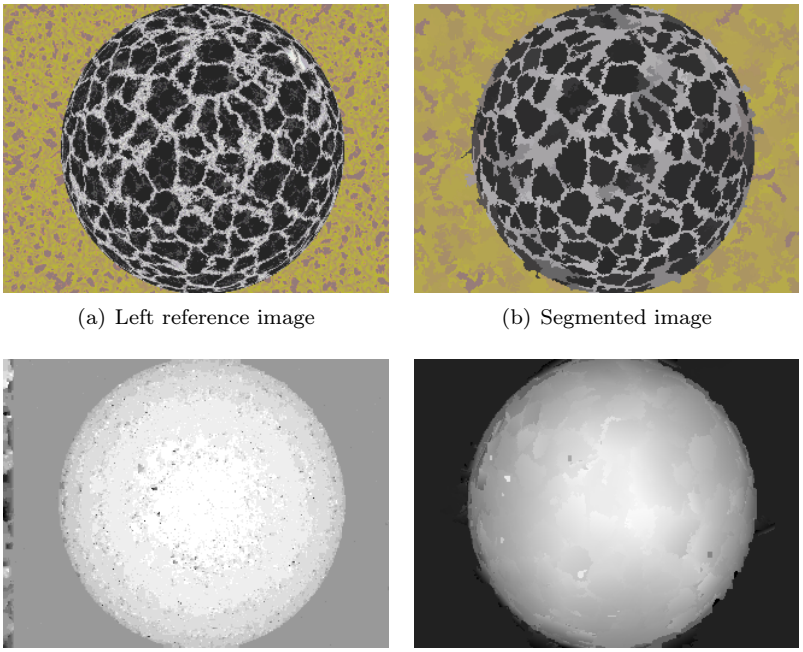
Our approach is illustrated on both grey level and color images. The examples used are Sawtooth (Figure 1), Sphere (Figure 2), Corridor (Figure 3), Head(Figure 4) and images from the Middlebury dataset namely Venus and Teddy (Figures 5–6). The parameter settings are given in the following section.

4.1 Parameter settings

The disparity range and therefore the value of L is fixed to different values depending on each image pair. Our model parameters are fixed to $\lambda = 0.2$, $\epsilon_D = 0.05$, $\sigma_D = 1$, with a window size parameter $w = 5$ pixels for the SAD computation. Parameter T is fixed to the average pixel cost computed over all pixels and disparity labels. For the interaction term, we consider a first order neighborhood where each pixel has 4 neighbors. We perform segmentations using the Mean Shift algorithm [4] for both grey level and color images, with range and distance sigma set to 7 and 5 respectively and a minimum region size of 80 pixels. The plane fitting is performed with a robust approach: we use 1000 iterations of RANSAC [7] with the maximum distance set to 0.5. The alternation procedure is then carried out for a prescribed number of iterations (5 in our case) at 4 different scales, ranging from coarse to fine. A segmentation is carried out once per scale. Each of the BP processes is also performed for 5 iterations.

4.2 Examples

The first example (Figure 1) is the *sawtooth* image[2]. The image size is 380 × 434 pixels with a disparity range of 20 pixels. The underlying scene structure consists of slanted planar surfaces. Figure 1(b) shows the ground truth disparity map. Figure 1(c) shows the result using standard BP – that do not use any normal information, and without plane-fitting post-processing – while Figure 1(d) shows our result using surface constraints.



(a) Left reference image (b) Segmented image

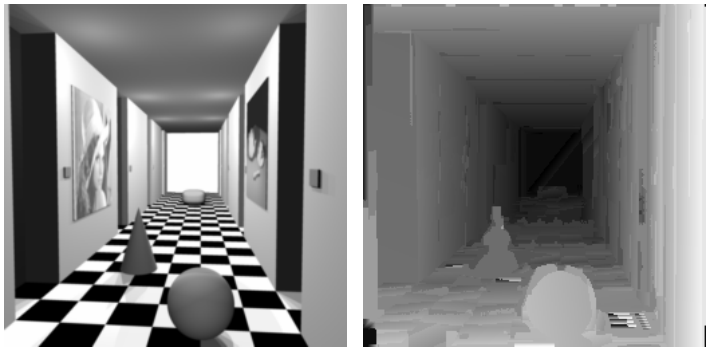
(c) Standard BP disparity map (d) Our approach disparity map

Figure 2: Sphere Image: comparison of our approach (d) with the staircase solution from standard BP (c) for a curved surface.

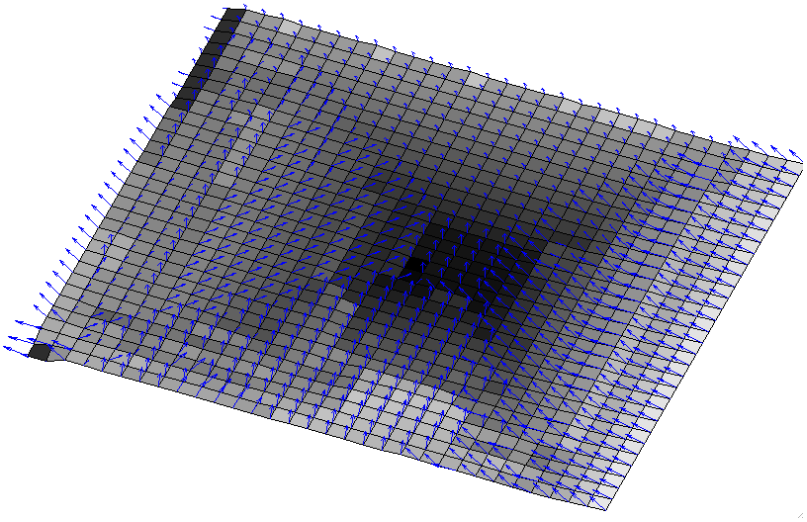
The second example (Figure 2) is a synthetic image of a sphere of size 320×480 with a disparity range of 10 pixels. This example illustrates the performance of our approach on a curved surface. Figure 2(b) shows the segmentation result used in our algorithm. Figure 2(c) shows the result with standard BP while Figure 2(d) shows the result with our approach. Both examples illustrate the bias of standard algorithms towards staircase solutions and show how our approach moves away from the fronto-parallel assumption to obtain a smooth and more accurate disparity map. In particular, Figure 2(d) shows how using the normal information enables us to recover the curved disparity surface.

The third example (Figure 3) is the *corridor* image. The image size is 256 × 256 pixels and the disparity range set to 11 pixels. The estimated disparity and normal maps using our approach are shown respectively in Figures 3(b) and 3(c). These figures indicate that the normals rightly follow the slanted surface

[2]Note that we show only the left image of the stereo pair for all the images used in this paper



(a) Left reference image (b) Disparity map



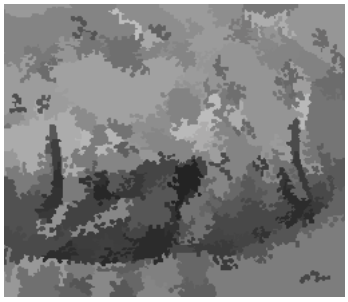
(c) Estimated normals

Figure 3: Corridor Image: consistency between the disparity (b) and normal (c) maps obtained with our approach.

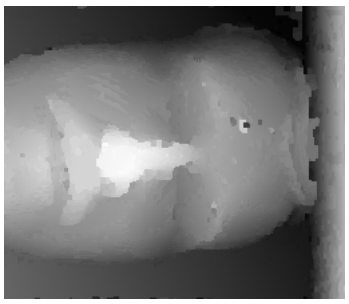
of the corridor disparity map. The disparity result compares favorably with the result obtained in [15] on the same image pair.

Our fourth example (Figure 4) is a face image of size 219×255 pixels with a disparity range of 40 pixels. It is a good example to illustrate how our approach can capture a complex image surface. Figure 4(d) shows the normals obtained by our estimation process. The part near the nose and eye region is highlighted in Figure 4(e). It shows that the surface deformations are satisfyingly captured by the normals but also that our algorithm is limited by the size of the regions obtained in the initial segmentation. If the regions are not small enough, the solution tends to stay close to the plane-fit solution. In this particular case, Figure 4(b) shows that large regions appear during the segmentation process. Conversely, a certain minimum region size is required in order for the RANSAC procedure to provide a good plane representation of the region. It

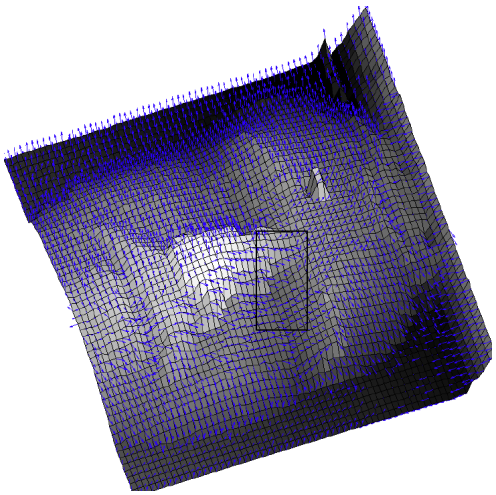
(a) Left reference image



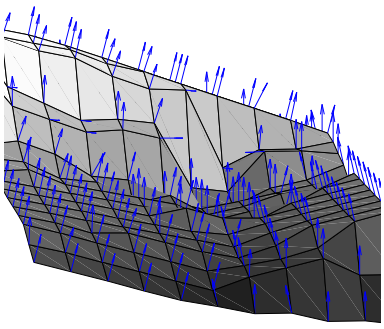
(b) Segmented image



(c) Our approach: disparity map



(d) Our approach: estimated normals



(e) Estimated normals in the highlighted region of Image (d)

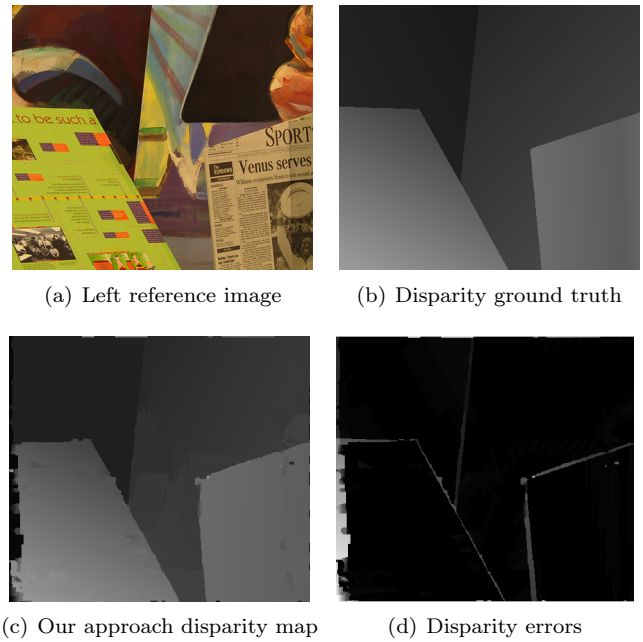
Figure 4: Face Image: a complex disparity surface (a) captured with our approach (c) and (d).

follows that there exists a trade-off between a minimum region size and a good plane representation for that region.

Finally, we illustrate our approach on two image pairs (Figures 5 and 6) from the Middlebury database. Our results reported respectively in Figures 5(c) and 6(c) are satisfying in that smooth disparity surfaces are adequately recovered.

The disparity errors, reported respectively in images 5(d) and 6(d), show that the main source of error is the presence of occluded pixels. This is not surprising as we have not addressed the problem of occlusions in this paper. It appears that when the disparity range is small as in the case of Figure 5 (disparity range 20 pixels) disparity errors are limited (see Figure 5(d)).

However, when the disparity range is large as for Figure 6 (disparity range 60 pixels), this problem becomes predominant (see Figure 6(d)).



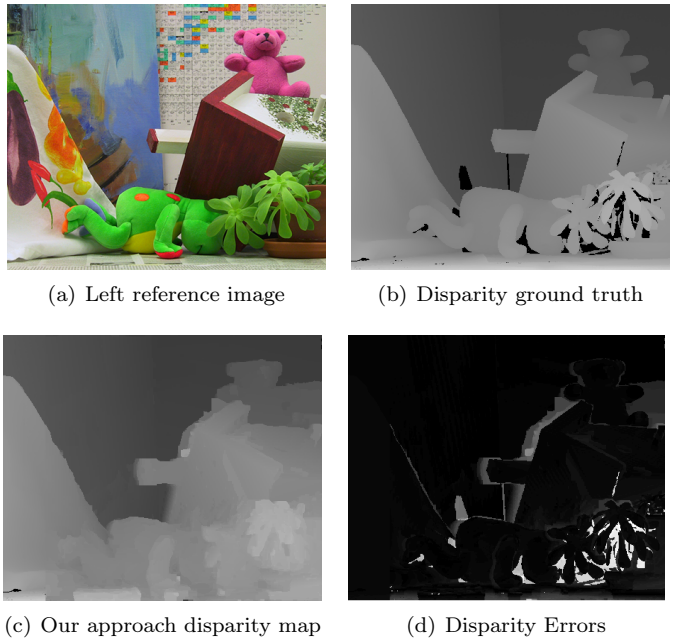
(a) Left reference image (b) Disparity ground truth

(c) Our approach disparity map (d) Disparity errors

Figure 5: Venus Image: disparity errors (d) concentrate in occluded regions (small disparity range).

5 Discussion

We proposed a new joint probabilistic model to efficiently combine disparity and surface geometric information. The main originality was to define such a model through conditional distributions that can model explicitly relationships between disparity and surface normals. As a result, we observed a significant gain in disparity and normal estimation for a number of non-fronto-parallel scenes illustrated in our experiments. However, we also observed that a limitation of our approach was not to address the occluded pixels issue. Future work could then include adding explicit or implicit constraints to handle occlusions. This could be done by performing a mutual consistency check which requires that the disparity values from the left and right disparity maps are consistent as suggested by [25]. Also, the RANSAC plane-fitting could be replaced by the robust plane-fitting approach used in [22]. This should help in finding a proper trade-off for choosing the minimum region size for segmentation.



(a) Left reference image (b) Disparity ground truth

(c) Our approach disparity map (d) Disparity Errors

Figure 6: Teddy Image: disparity errors (d) concentrate in occluded regions (large disparity range).

As for the probabilistic setting itself, we first focused on defining a valid Markovian framework to model cooperations and used a Maximum A Posteriori principle for inference. A natural future direction of research is to investigate the possibility to recast our approach into an EM (Expectation Maximization) framework. This would provide richer modeling alternatives in which iterative estimation of realizations of fields **D** and **N** would be replaced by iterative estimation of full distributions for **D** and **N**.

References

[1] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3), 1986.

[2] S.T. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *ICCV*, 1999.

[3] Y.Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 23(11), November 2001.

[4] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 24(5), May 2002.

[5] F. Devernay and O.D. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. *CVPR*, 94, 1994.

[6] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1), October 2006.

[7] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 1981.

[8] E.B. Gamble, D. Geiger, T.A. Poggio, and D. Weinshall. Integration of vision modules and labeling of surface discontinuities. *IEEE Transactions System, Man and Cybernetics*, 19(6):1576–1581, November 1989.

[9] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 15(12):1217–1232, December 1993.

[10] Andreas Klaus, Mario Sormann, and Konrad F. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *ICPR (3)*, 2006.

[11] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother. Probabilistic fusion of stereo with color and contrast for bilayer segmentation. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 28(9), September 2006.

[12] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother. Probabilistic fusion of stereo with color and contrast for bilayer segmentation. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 28(9):1480–1492, September 2006.

[13] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, 2001.

[14] G. Li and S.W. Zucker. Differential geometric consistency extends stereo to curved surfaces. In *ECCV*, 2006.

[15] G. Li and S.W. Zucker. Surface geometric constraints for stereo in belief propagation. In *CVPR*, 2006.

[16] M.H. Lin and C. Tomasi. Surfaces with occlusions from layered stereo. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 26(8), August 2004.

[17] Ramya Narasimha, Elise Arnaud, Florence Forbes, and Radu P. Horaud. Cooperative disparity and object boundary estimation, oct 2008.

[18] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Internation Journal of Computer Vision*, 47(1-3), April 2002.

[19] C. Strecha, R. Fransens, and L.J. Van Gool. Combined depth and outlier estimation in multi-view stereo. In *CVPR*, 2006.

[20] J. Sun, Y. Li, S.B. Kang, and H.Y. Shum. Symmetric stereo matching for occlusion handling. In *CVPR*, 2005.

[21] J. Sun, N.N. Zheng, and H.Y. Shum. Stereo matching using belief propagation. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 25(7), July 2003.

[22] Z.F. Wang and Z.G. Zheng. A region based stereo matching algorithm using cooperative optimization. In *CVPR*, 2008.

[23] Oliver Woodford, Philip Torr, Ian Reid, and Andrew Fitzgibbon. Global stereo reconstruction under second-order smoothness priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12), 2009.

[24] T.P. Wu, K.L. Tang, C.K. Tang, and T.T. Wong. Dense photometric stereo: A Markov random field approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 2006.

[25] Qyngxiong Yang, Liang Wang, Ruigang Yang, Henrik Stewenius, and David Nister. Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling. In *CVPR*, 2006.

[26] K.J. Yoon and I.S. Kweon. Locally adaptive support-weight approach for visual correspondence search. In *CVPR*, 2005.

[27] C.L. Zitnick and S.B. Kang. Stereo for image-based rendering using image over-segmentation. *International Journal of Computer Vision*, 75(1), October 2007.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399