Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 25 (2007) 623–636
www.elsevier.com/locate/imavis

High-precision stereo disparity estimation using HMMF models

Edgar Arce [a,*,1], J.L. Marroquin [b,2]

[a] Facultad de Ciencias, Av. Salvador Nava Mtz S/N, Zona Universitaria, C.P. 78290, San Luis Potosi, S.L.P., Mexico
[b] Center for Research in Mathematics, Callejon Jalisco s/n, Mineral de la Valenciana, C.P. 36240, Guanajuato, Gto., Mexico

Received 20 December 2004; received in revised form 23 February 2006; accepted 16 May 2006

Abstract

In this paper, stereo disparity reconstruction is formulated as a parametric segmentation problem in a Bayesian framework: the goal is to partition the reference image into a set of non-overlapping regions, inside each one of which a specific disparity model (which consists of two coupled membranes) is adjusted. The problem of simultaneously finding the regions and the parameters of the corresponding models is formulated using a novel probabilistic framework which uses a hidden Markov random measure field model, which allows one to efficiently find the optimal estimators by minimization of a differentiable cost function. This framework also allows for the explicit modeling of occlusions, consistency constraints and correspondence of disparity and intensity discontinuities. It is shown experimentally that this method produces competitive results, with respect to state-of-the-art methods, for discretized (integer) disparities and significantly better results for high-precision real-valued disparities.

Keywords: Stereo correspondence; Disparity map; Image segmentation; Markov random fields; Bayesian method; Doubly stochastic prior model; Parametric disparity model; Subpixel disparity values; Occluded regions

1. Introduction

Stereo correspondence is one of the most investigated topics in computer vision and many methods have been proposed to solve this problem. This paper emphasizes two-frame stereo matching, composed by fixed monochromatic cameras and parallel optical axis, so that the epipolar lines are horizontal; if the image planes are not coplanar, the epipolar lines will not be horizontal; however, it is possible to wrap the images in a preprocessing step, so that the epipolar lines become horizontal. Therefore, this model may be considered a general one.

In spite of two decades of research in this area, in which many algorithms have been proposed, the attainable results tend to show some defects, becoming important as more applications use stereo-vision as a relevant part of their tasks. Among these applications we can mention highly realistic 3D models to visualize and simulate events like flight simulators, film industry or product presentations; the range of applications may be extended to robot vision, passive range finding, topography mapping from aerial photographs, architecture visualizations, virtual reality or surface reconstruction in surgical fields, etc. All these examples require piecewise smooth disparity maps with very high precision and good edge definition, as well as automatic occluded region detection.

To estimate a disparity map, the correspondence problem must be solved. In order to make this task possible, two basic assumptions are made:

- Most of the points in the scene are captured in both images.
- Corresponding regions in the images are similar.

Although these assumptions are true and the search space is reduced by searching only in horizontal epipolar

* Corresponding author.
E-mail address: arce@fciencias.uaslp.mx (E. Arce).
[1] The author was sponsored by UASLP and PROMEP under Grant 103.5/04/1387.
[2] The author was sponsored in part by Conacyt, Mexico, under Grant 42523.

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.05.006

lines [1], this is an ill-posed problem due to the inherent difficulties of the process; among the principal ones we can mention: *occlusions*, where objects that are closer to the cameras occlude some parts of farther objects, thus some points in the scene are visible only to one camera; *noise* caused by some artifacts generating differences between the images; *homogeneous regions*, where there are no significant intensity variations, so that every point may be matched with every other point inside these regions; or *photometric variations* due principally to foreshortening, depth discontinuities, lens blur, and image sampling. All these problems imply that stereo algorithms should provide disparity values for regions which are not found by a matching process and, ideality, they should also locate occluded regions.

Different approaches have been proposed to solve the correspondence problem; an exhaustive survey of the literature is beyond the scope of this paper, but some books [4–7] and review papers [34] cover the basics of this subject.

There are two principal approaches to solve the correspondence problem; the first one is based on local optimization methods. These methods search the smallest value of some correspondence metric for a pixel in the reference image with respect to another pixel in the matching image. Among these, there are window-based methods that estimate correspondences inside a region (window), using as a measure cross-correlation [10,11], squared intensity difference (SD) [8], or absolute intensity difference (AD) [9]. There are also methods that use more than one fixed window in each position [12], or variable windows [13,14], getting better disparity estimates inside homogeneous intensity regions.

The second approach is based on global optimization methods. In this case, the goal is to minimize a function which depends on a disparity field between a reference image I_R and a matching image I_T. In addition to a correspondence term, a spatial coherence disparity term is used to regularize that solution, getting energy functions of the form:

$$E(d) = \sum_r \rho_1(I_R(r), I_T(d(r))) + \sum_{\langle r,s\rangle} \rho_2(d(r) - d(s)) \qquad (1)$$

where $d(r) \in \mathbb{R}$ is the value of the disparity field at each site r; ρ_1 is a function that measures the intensity difference between the images I_R, I_T for a given d, and ρ_2 is some monotonically increasing function that measures the disparity difference between nearest neighbor pairs. To minimize this function many methods have been proposed; regularization-based methods [15] use a quadratic function ρ_2 that makes d smooth, but may yield poor results at disparity discontinuities; to avoid this problem robust functions ρ have been used as in [16,17]. After the seminal paper of Geman and Geman [18] in which a Bayesian interpretation of this kind of energy functions based on Markov random fields (MRF) was made, many algorithms have been proposed using different techniques for the computation of the optimal estimator, such as: simulated annealing [19,20], highest confidence first [21], belief propagation [22], mean-field [23], and recently methods based on max-flow and graph-cuts [24,25,42]; models based on MRF and psychovisual cue [2]. Finally, dynamic programming has been used to minimize similar cost functions that allow the enforcement of constraints involving occlusions [26,28,29], continuity and monotonicity [12].

In this paper, we present a new disparity estimation method, based on a recently proposed Bayesian formulation of a parametric image segmentation problem which uses a hidden Markov measure field (HMMF) model [31]. We will show that this method yields a dense high-precision disparity map at subpixel level inside regions with small disparity variation, localizing at the same time disparity discontinuities and occluded regions.

Our approach is related to the one presented in [42], in the sense that disparity estimation is formulated as a parametric image segmentation problem; however, in [42] a minimization framework based on a two-step (MAP-MAP) algorithm is used, which makes the process slow and sensitive to noise and initialization. In our case, we propose a different model (a "dual membrane"), which permits the recovery of the global shape of the disparity field, as well as small high-frequency variations, getting high-precision disparity maps. Another important difference is that in our approach it is not necessary to use two-step procedures, such as Expectation-Maximization or the MAP-MAP algorithm, to compute the segmentation and the parameters of the models; instead, one directly minimizes a differentiable function obtained from the HMMF model, making our method more robust with respect to the selection of the starting points, less vulnerable to noise and computationally more efficient. In the proposed energy function we incorporate constraints about the localization of disparity discontinuities, using intensity edges [40,34], and occluded regions, characterized by the inconsistency between disparity maps using successively, as reference image, the right and left components of a stereo pair.

The plan of our presentation is as follows: in Section 2, we present the Bayesian formulation of disparity estimation as a parametric segmentation problem using HMMF prior models. In Section 3, we describe the double-membrane parametric model. In Section 4, we propose some modifications to the model to incorporate edge information to improve the localization of disparity discontinuities. In Section 5, we show some results and comparisons with other recent algorithms. Finally, in Section 6 we discuss the results obtained in this work and present some conclusions.

2. Hidden Markov measure field models for stereo

In this work, we propose to use a parametric segmentation approach to estimate disparity; the reference image I_R is divided into M regions $\{R_k, k = 1, \ldots, M\}$. Inside each

region, the disparity $d(r)$, at each position r, is given by the parametric model $\Phi(r,\theta_k)$, so that

$$I_R(r) = I_T(r \pm \Phi(r,\theta_k)) \quad \text{for } r \in R_k \tag{2}$$

where I_T is the matching image of a stereo pair and θ_k is the parameter vector that corresponds to region R_k (in what follows, θ will denote the set of all parameter vectors θ_k, $k = 1,\ldots,M$).

The problem with this approach is that we have to estimate, at the same time, the regions and the parameter values of the corresponding disparity models. To solve this complex problem, we use a new probabilistic formulation based on HMMF models [31], firmly rooted in Bayesian estimation theory, which allows one to introduce constraints about the form and size of each region R_k. To describe this model for estimating disparity, we include the following definitions: let L be the pixel lattice with N sites, where the stereo pair I_R, I_T is observed, with $L = \cup_{k=1}^{M} R_k$; $R_i \cap R_j = \emptyset$, $i \neq j$. Associated with I_R there is a label field f, indicating to which region R_k does every pixel in L belong to: $f(r) = k$ iff $r \in R_k$. The intensity of each pixel $r \in L$ of I_R is given by

$$I_R(r) = I_T(r \pm \Phi(r,\theta_{f(r)})) + n(r) \tag{3}$$

where $n(r)$ is a white noise process with known distribution P_n (e.g., $n(r)$ are independent, zero mean, identically distributed random variables) and the sign depends on the reference image; if it is the left image, the minus sign is used, otherwise the plus sign is applied.

In this model, Fig. 1, the label field f is generated by a two-step stochastic process; in the first step, a Markov random vector field p is generated, where each $p(r)$ satisfies the following constraints:

$$\sum_{k=1}^{M} p_k(r) = 1, \quad p_k(r) \geqslant 0, \ \ k = 1,\ldots,M \tag{4}$$

so that each $p(r)$ can be interpreted as a discrete probability distribution on $\{1,2,\ldots,M\}$. In the second step, the label field f is generated in such a way that each $f(r)$ is an *independent* sample of the corresponding distribution $p(r)$. The important point about this model is that it allows for a formulation of the problem in which instead of trying to estimate the discrete field f directly – which would entail solving a difficult combinatorial optimization problem – one first estimates the p field (which is real-valued, and hence may be done using gradient-based techniques) and then estimates the f field on a pixel-by-pixel basis. We now explain how this is done.

Using the model of Fig. 1, one can compute the posterior probability of p, θ using Bayes rule:

$$P(p,\theta|I_R,I_T) = \frac{1}{Z} P(I_R,I_T \mid p,\theta) P_p(p) P_\theta(\theta) \tag{5}$$

where Z is a normalization constant. The conditional distribution $P(I_R,I_T|p,\theta)$ is obtained as:

$$P(I_R,I_T|p,\theta) = \prod_{r\in L} P(I_R(r),I_T(r)|p,\theta) \tag{6}$$

The individual conditional distributions $P(I_R(r),I_T(r)|p,\theta)$ may be obtained by first computing the joint conditional probability as:

$$P(I_R(r),I_T(r),f(r)|p,\theta) = P(I_R(r),I_T(r)|f(r),p,\theta)P(f(r)|p,\theta) \tag{7}$$

and then marginalizing over $f(r)$:

$$P(I_R(r),I_T(r)|p,\theta) = \sum_{k=1}^{M} P(I_R(r),I_T(r)|f(r)=k,p,\theta) \times P(f(r)=k|p,\theta) \tag{8}$$

Defining the likelihood $v_k(r,\theta)$ as:

$$v_k(r,\theta) = P(I_R(r),I_T(r)|f(r)=k,\theta) = P_n(I_R(r) - I_T(r \pm \Phi(r,\theta_k))) \tag{9}$$

where P_n is the noise distribution (assumed known), and using the fact that the first term of the summation in (8) is independent of p given $f(r) = k$ and θ, so that

$$P(I_R(r),I_T(r)|f(r)=k,p,\theta) = P(I_R(r),I_T(r)|f(r)=k,\theta) = v_k(r,\theta) \tag{10}$$

and also considering that $P(f(r)=k|p,\theta) = p_k(r)$ one obtains that

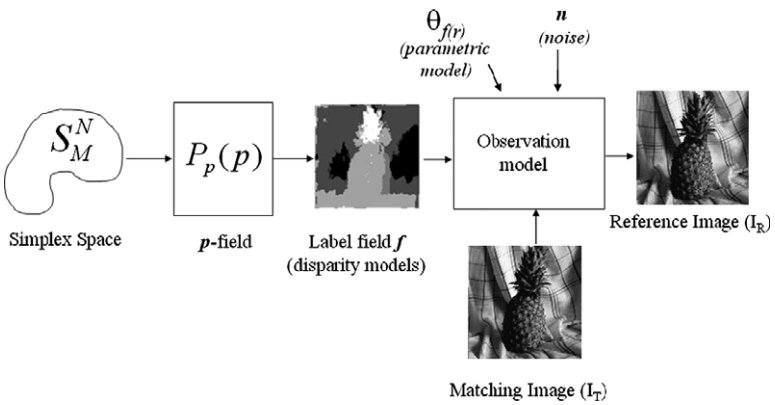

Fig. 1. HMMF model for disparity (see text).

$$P(I_R(r), I_T(r)|p,\theta) = \sum_{k=1}^{M} v_k(r,\theta)p_k(r) = v(r,\theta)\cdot p(r) \quad (11)$$

Note that in this expression the discrete variables $\{f(r), r \in L\}$ do not appear, since we have marginalized the conditional distribution (7) over them. This is the key step that allows one to have a differentiable cost function that depends only on the continuous variables p and θ. Once these are estimated, the f variables may be easily obtained, as shown below.

Usually, P_n is assumed to be Gaussian; however, in the stereo problem, there may be large differences of intensity between corresponding points in the two images, e.g., due to specularities, which may produce a large number of outliers; for this reason, a better model is a distribution with heavier tails. At the same time, it is important that small intensity errors be penalized in an adequate way. A function that satisfies both requirements is the following one:

$$v_k(r,\theta) = \frac{\alpha}{1+\alpha|\xi_k(r)|} \quad (12)$$

with $\xi_k(r) = (I_R - I_T(r \pm \Phi(r,\theta_k)))$, and where α is a parameter that depends on the noise standard deviation.

From (5) and (11) we get that,

$$P(p,\theta|I_R,I_T) = \frac{1}{Z}\exp[-U(p,\theta)] \quad (13)$$

with

$$U(p,\theta) = -\sum_r \log(v(r,\theta)\cdot p(r)) + \sum_c V_C(p) - \log P_\theta(\theta) \quad (14)$$

where $V_C(p)$ is a potential function which depends on the values of the p field at the sites that belong to the clique C (see [31] for details) and where $P_\theta(\theta)$ are prior constraints over θ.

In our model we use a first-order neighborhood, with cliques of size 2, so that each clique corresponds to a horizontal or vertical pair of adjacent pixels. For the potentials V_C, it is important to use functions that do not penalize too much the discontinuities in the p field – which correspond to model changes – since otherwise the minimization process might prefer to bend a spline model (see below) instead of producing a correct discontinuous solution. Thus, instead of the quadratic potentials that are normally used, we use a vectorized Huber potential [32] given by:

$$V_C(p) = \overline{h}(\Delta p) = \sum_{k=1}^{M} h(\Delta p_k) \quad (15)$$

where

$$h(\Delta p_k) = \begin{cases} \Delta p_k^2 & \text{if } |\Delta p_k| \leqslant 0.5 \\ |\Delta p_k| - 0.25 & \text{if } |\Delta p_k| > 0.5 \end{cases} \quad (16)$$

where $\Delta p = p(r) - p(s)$, where r and s are neighboring sites in L.

The choice of the right or left image as the reference image is somewhat arbitrary. In fact, it is better to consider that one has two p fields p^{R}, p^{L}, which are obtained when the right (respectively, the left) image is taken as the reference image. These fields may be used to implement a consistency constraint [42,3] that improves the robustness of the model. This constraint may be expressed as:

$$p_k^{\mathrm{L}}(r) \approx p_k^{\mathrm{R}}(r - \Phi(r,\theta_k)) \quad (17)$$

At the same time, this constraint allows one to introduce an explicit occlusion model; to do this, we consider R_0 (i.e., model 0) to correspond to occluded areas, and consider that the consistency constraint (17) should have a weight proportional to one minus the probability of occlusion $p_0(r)$. Thus, one finally obtains the following posterior energy function that depends on the two fields p^{L},p^{R} and the parameter vectors θ:

$$\begin{aligned} U(p^{\mathrm{L}},p^{\mathrm{R}},\theta) = & -\sum_{r\in L}\log(p^L(r)\cdot v^{\mathrm{L}}(r,\theta)) \\ & -\sum_{r\in L}\log(p^R(r)\cdot v^{\mathrm{R}}(r,\theta)) \\ & +\lambda_1\sum_{\langle r,s\rangle}\overline{h}(p^{\mathrm{L}}(r)-p^{\mathrm{L}}(s)) \\ & +\lambda_1\sum_{\langle r,s\rangle}\overline{h}(p^{\mathrm{R}}(r)-p^{\mathrm{R}}(s)) \\ & +\lambda_2\sum_{r\in L}\sum_{k>0}\left(p_k^{\mathrm{L}}(r)-p_k^{\mathrm{R}}(r-\Phi(r,\theta_k))\right)^2 \\ & \times\left(1-p_0^{\mathrm{L}}(r)\right) \\ & +\lambda_2\sum_{r\in L}\sum_{k>0}\left(p_k^{\mathrm{R}}(r)-p_k^{\mathrm{L}}(r+\Phi(r,\theta_k))\right)^2 \\ & \times\left(1-p_0^{\mathrm{R}}(r)\right) \end{aligned} \quad (18)$$

where v^{L} is the likelihood obtained taking as reference the left image I^{L}, v^{R} the likelihood taking as reference the right image I^{R} and λ_1 is a parameter that controls spatial coherence of the p-fields. The last two expressions in (18) are the inconsistency terms; these reflect the fact that the p-fields should be coherent in all the regions except in the occluded regions which are denoted by the label $f(r) = 0$, and where $v_0^{\mathrm{L}}(r,\theta)$, $v_0^{\mathrm{R}}(r,\theta)$ are assigned constant values. These terms are controlled by λ_2 and weighted by the evidence that any pixel r does not belong to the occluded regions given by $p(r)_0^{\mathrm{L}}$ or $p(r)_0^{\mathrm{R}}$.

To obtain optimal estimators for $f^{\mathrm{L}*}$, $f^{\mathrm{R}*}$, the label fields, we use the following two-step procedure:

(1) Find the MAP estimators $p^{\mathrm{L}*}$, $p^{\mathrm{R}*}$, θ^* for p^{L}, p^{R}, θ:

$$p^{\mathrm{L}*}p^{\mathrm{R}*}\theta^* = \arg\max_{p^{\mathrm{L}},p^{\mathrm{R}},\theta} P(p^{\mathrm{L}},p^{\mathrm{R}},\theta|I^{\mathrm{R}},I^{\mathrm{L}}) \quad (19)$$

(2) Find $f^{\mathrm{L}*}$, $f^{\mathrm{R}*}$ as the maximizers of $P(f^{\mathrm{L}},f^{\mathrm{R}}|p^{\mathrm{L}}=p^{\mathrm{L}*},p^{\mathrm{R}}=p^{\mathrm{R}*},\theta=\theta^*,I^{\mathrm{R}},I^{\mathrm{L}})$

The first step is equivalent to the minimization of $U(p^{\mathrm{L}},p^{\mathrm{R}},\theta)$ subject to the constraints (4) for $p^{\mathrm{L}}(r)$, $p^{\mathrm{R}}(r)$ for all $r \in L$; while the second step consists simply of finding the mode for each discrete measure $p^{\mathrm{L}*}(r)$ and $p^{\mathrm{R}*}(r)$ in a decoupled way:

$$f^{L*}(r) = \arg\max_k p_k^L(r)$$
$$f^{R*}(r) = \arg\max_k p_k^R(r)$$

The computational burden, thus, lies on the first step, but since (18) is differentiable this minimization may be carried out very efficiently using a constrained minimization algorithm called "Gradient Projection Newtonian Descent" [31].

An important implementation detail is that the partial derivatives of $v_k(r,\theta)$ with respect to θ must be computed with high accuracy. Since they involve the evaluation of the gradient of I_T at non-integer locations, the best way is to use spline interpolation for I_T, so that the derivatives may be evaluated analytically.

3. Parametric model

Many disparity estimation algorithms, as in [9,24,25], use constant models (which correspond to fronto-parallel planes) for computing disparity maps. Moreover, most of them use discretized disparities usually with a separation of one pixel. In many real stereo pairs, particularly when they portray natural scenes, one has, in each region, not a constant – or even a smooth – disparity, but rather a slowly varying envelope with high-frequency and low-amplitude variations. This situation cannot be modeled adequately by a piecewise constant integer-valued disparity.

In this work, we propose a dual membrane model (DMM) to estimate disparities. It is defined by the sum of two models:

$$\Phi(r,\theta_k) = \Phi_1\left(r,\theta_k^{(1)}\right) + \Phi_2\left(r,\theta_k^{(2)}\right) \tag{20}$$

The first model corresponds to a spline model, defined as a linear combination of basis functions:

$$\Phi_1\left(r,\theta_k^{(1)}\right) = \sum_{j=1}^{J} \theta_{kj}^{(1)} N_j(r) \tag{21}$$

where $\{N_j,\ j=1,\ldots,J\}$ are B-spline functions [30], translated to a jth node of a coarse sub-grid superimposed on the reference image I_R, called spline-grid:

$$N_j(x,y) = B^2\left(\frac{x-x_j}{\triangle}\right) B^2\left(\frac{y-y_j}{\triangle}\right) \tag{22}$$

where (x_j,y_j) are the jth coordinates (in pixels) of a node in this sub-grid, $\triangle$ is a scale factor indicating the distance between nodes in this grid, and B^2 is given by

$$B^2(x) = \begin{cases} \frac{1}{2}(1.5-2x^2) & \text{if } |x| \in [0,0.5] \\ \frac{1}{2}(x^2-3|x|+2.25) & \text{if } |x| \in [0.5,1.5] \\ 0 & \text{if } |x| > 1.5 \end{cases} \tag{23}$$

The coefficients $\theta_{kj}^{(1)}$ and the B-spline functions describe a surface whose rigidity may be controlled by imposing a Gibbsian prior on $\theta^{(1)}$, of the form:

$$P_\theta\left(\theta^{(1)}\right) = \frac{1}{Z^{\theta^{(1)}}} \exp\left[-\eta_1 \sum_{k=1}^{M} \sum_{\langle u,v\rangle} \left(\theta_{ku}^{(1)} - \theta_{kv}^{(1)}\right)^2\right] \tag{24}$$

where the second sum is taken over the nearest neighbor pair of nodes $\langle u,v\rangle$ in the spline-grid. This model allows smooth variations inside each region R_k controlled by η_1. One advantage of this model is that it extrapolates disparity values as constants outside their support regions, and hence is less prone to produce spurious interactions with other regions.

The second model in (20) is a classical pixel-to-pixel membrane given by,

$$\Phi_2\left(r,\theta_k^{(2)}\right) = \theta_k^{(2)}(r) \tag{25}$$

i.e., $\theta_k^{(2)}$ denotes a scalar field. As in Φ_1, a prior was imposed on $\theta^{(2)}$ of the form:

$$P_\theta\left(\theta^{(2)}\right) = \frac{1}{Z^{\theta^{(2)}}} \exp\left[-\eta_2 \sum_{k=1}^{M} \sum_{\langle r,s\rangle} \left(\theta_k^{(2)}(r) - \theta_k^{(2)}(s)\right)^2\right] \tag{26}$$

where η_2 is a parameter that controls the rigidity of the pixel-to-pixel membrane and $\langle r,s\rangle$ are the nearest neighbor pairs of pixels.

To verify the DMM performance, two experiments were made using the stereo pair shown in Fig. 4. The first one consists in estimating the disparity field by means of a pixel-to-pixel membrane model only. The plot in Fig. 2(a) and (b) shows the profiles (row 128) of the disparity maps obtained by setting η_2 in (26) to 1 and 20, respectively. These plots show that if one sets η_2 to a large value, in order to control intensity variations in the images due to noise, mostly flat disparity regions with small variations are obtained. In contrast, if a small value is set, noise appears on the disparity surface. In the second experiment, only the spline membrane was applied (yielding a model similar to the one in [42]); Fig. 2(c) shows how this model estimates better global smooth disparity variations, but loses the disparity texture in the surface of the pineapple. Finally, in Fig. 2(d) the behavior of the DMM is shown, and one can appreciate that the correct smooth behavior, adequate noise elimination and appropriate modeling of the high-frequency variations are simultaneously achieved.

Fig. 3 illustrates the behavior of this model for two stereo pairs containing different types of surfaces; in panels (a-2) and (a-3), we show the *f-field* (segmentation; occlusion regions indicated in black) and the 3D-reconstruction using the obtained disparity map for the case where the objects in the scene are composed by planar (not fronto-parallel) surfaces. Panel (b) shows the results in the case when the objects are composed by curved surfaces; panel (b-3) is the 3D-reconstruction (obtained with a simple pinhole camera model).

Fig. 4 illustrates the case where the objects are more complex; the parametric disparity surfaces obtained Φ_1, Φ_2 are shown in panels (c) and (d), after finding the optimal values for $\theta^{(1)}$, $\theta^{(2)}$ for the region R_k that corresponds to the pineapple. Panel (e) shows the complete disparity map;

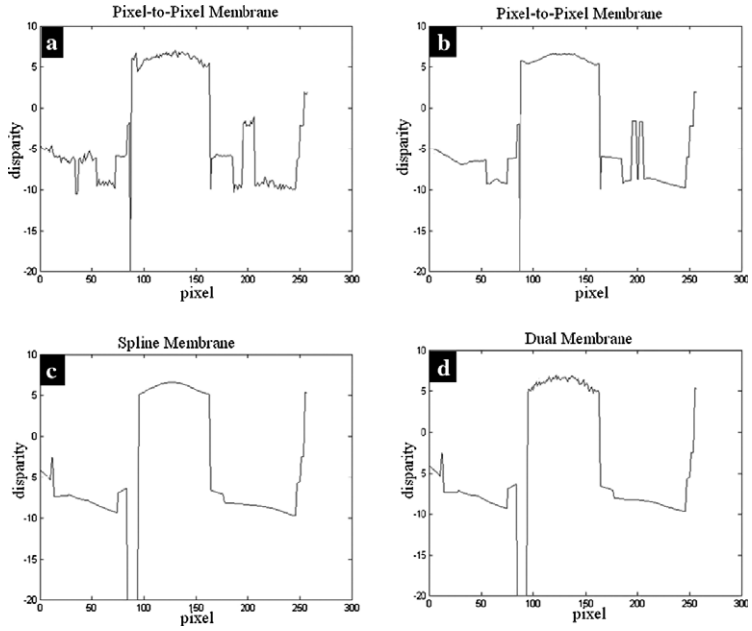

Fig. 2. (a) Pixel-to-pixel disparity map with $\eta_2 = 1$, (b) pixel-to-pixel disparity map with $\eta_2 = 20$, (c) spline membrane, and (d) DMM disparity map.

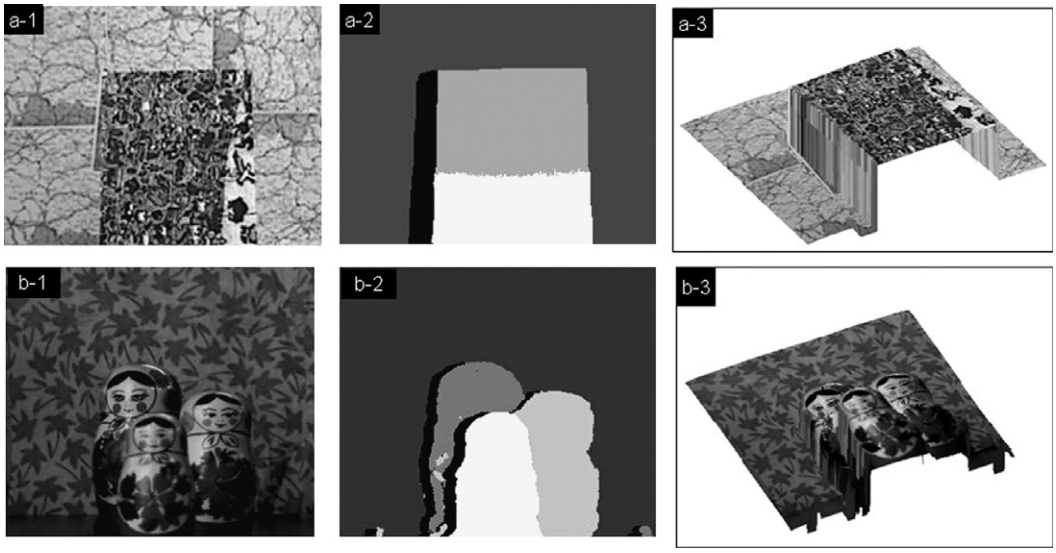

Fig. 3. (a-1) Reference image, (a-2) segmentation, and (a-3) 3D-reconstruction (planar surfaces). (b-1) Reference image, (b-2) segmentation, and (b-3) 3D-reconstruction (curved surfaces).

dark regions correspond to occluded areas. Panel (f) shows the 3D-reconstruction.

One important question is the criterion for selecting the number of models M. Fortunately, the precise value for this parameter is not critical, provided that there are sufficient models to account for all the smooth regions in the image. If one specifies more models than the ones that are strictly needed, the computational complexity will increase, but the final result will not be affected, because the extra models will either be automatically eliminated (i.e., will get consistently low p values) or a smooth region will be described by more than one model (see, for example, Fig. 3(a)); in this case, since the dual membrane models are flexible, they will automatically bend, so that no spurious discontinuities appear at the interfaces. In the experiments reported here, in order to keep a reasonable computational complexity, an approximate value for M was chosen manually in each case, based on a rough estimation of the number of smooth regions present in each image.

4. Edge information

An important issue in stereo disparity estimation is the precise localization of depth discontinuities. It is difficult to do this using only the intensity differences between the two images, since, in many cases, in particular when the intensity

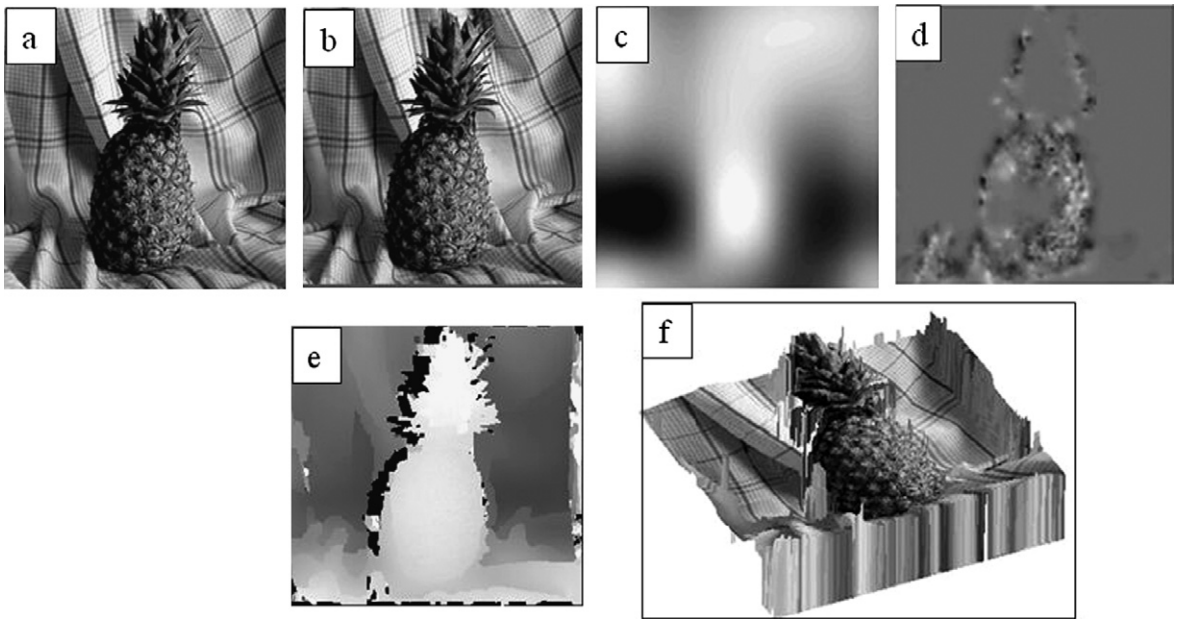

Fig. 4. (a, b) Stereo pair, (c) Φ_1, (d) Φ_2, (e) disparity map, (f) 3D-reconstruction.

of the occluded regions is relatively homogeneous, the disparity may be ambiguous. As in [40], one may tackle this problem by incorporating the prior constraint that the depth discontinuities coincide with significant intensity edges in most places. In the context of our model, this may be done by including a term that biases the p distributions toward high entropy configurations (i.e., distributions that are close to uniform) whenever an intensity edge is present, so that a change in the locally dominant model (the one that corresponds to the mode of p) is facilitated. In particular, we add the following terms to the energy function:

$$\sum_{r\in L}\sum_{k=1}^{M}[(\mu_1+\mu_2)B^{\mathrm{R}}(r)-\mu_2]p_k^{\mathrm{R}}(r)\log p_k^{\mathrm{R}}(r) + \sum_{r\in L}\sum_{k=1}^{M}[(\mu_1+\mu_2)B^{\mathrm{L}}(r)-\mu_2]p_k^{\mathrm{L}}(r)\log p_k^{\mathrm{L}}(r) \quad (27)$$

where $B^{\mathrm{L}}(r)=1$ if there is an intensity edge at pixel r in the left image (respectively, right), and $B^{\mathrm{L}}(r)=0$ otherwise (we compute $B^{\mathrm{L}}(r)$ and $B^{\mathrm{R}}(r)$ using a Canny edge detector [27]). This term works as follows: if $B^{\mathrm{L}}(r)=1$, it will bias the distribution $p^{\mathrm{L}}(r)$ towards high entropy configurations (i.e., to values closer to the uniform distribution $p_k^{\mathrm{L}}(r)=1/M$ for all k), thus facilitating the label changes at those sites. If $B^{\mathrm{L}}(r)=0$, on the other hand, this term will bias $p^{\mathrm{L}}(r)$ towards low entropy (peaked) configurations. μ_1 and μ_2 are positive parameters that control the influence of these terms.

5. Results and comparisons

5.1. Disparity precision

An important contribution of our method is the high disparity precision it produces, to subpixel level. To validate this fact, we made two experiments; one with a synthetic image, where a quantitative performance evaluation is possible, and the other with a stereo pair corresponding to a real scene. In both cases, we compare the method proposed in this paper (HMMF) against two algorithms. One that estimates disparity to subpixel level by computing correlation or sum of squared difference (SSD) using an adaptive window [9] (AW-SSD); the other is one of the best algorithms reported in [34] which minimizes a global function using the graph-cut method (GC) [24,26], using a discrete disparity space.

The first experiment corresponds to a synthetic stereo pair with a disparity pattern formed by a ramp whose average disparity gradient, ADG (computed only on the region where the disparity ramp is localized), is equal to 0.1243. Gaussian noise ($\sigma=1.5$) was added to the right image.

Fig. 5 shows the results obtained using the best set of parameters for each method. One can see on panel (a) that, although AW-SSD computes high-precision continuous disparity values, it is very vulnerable to noise, mostly in the constant disparity region; in (b) one can see that GC strongly discretizes the disparity, increasing in this way the RMS error (see also Fig. 7); finally, HMMF, panel (c), obtains a better approximation everywhere. The overall processing time required to process the 256×256 synthetic stereo pair by each one of these algorithms is shown in Table 1. The algorithm was implemented in C++ under Windows-XP (TM). One can see that AW-SSD is very fast, but the computed disparity map is very noisy. On the other hand, GC takes 12 times more than HMMF for the disparity discretization used.

In Fig. 6 we present a quantitative evaluation of the performance of the three algorithms analyzed, for different values of the corresponding regularization parameter (noise variance σ for AW-SSD, λ for GC and λ_1 for

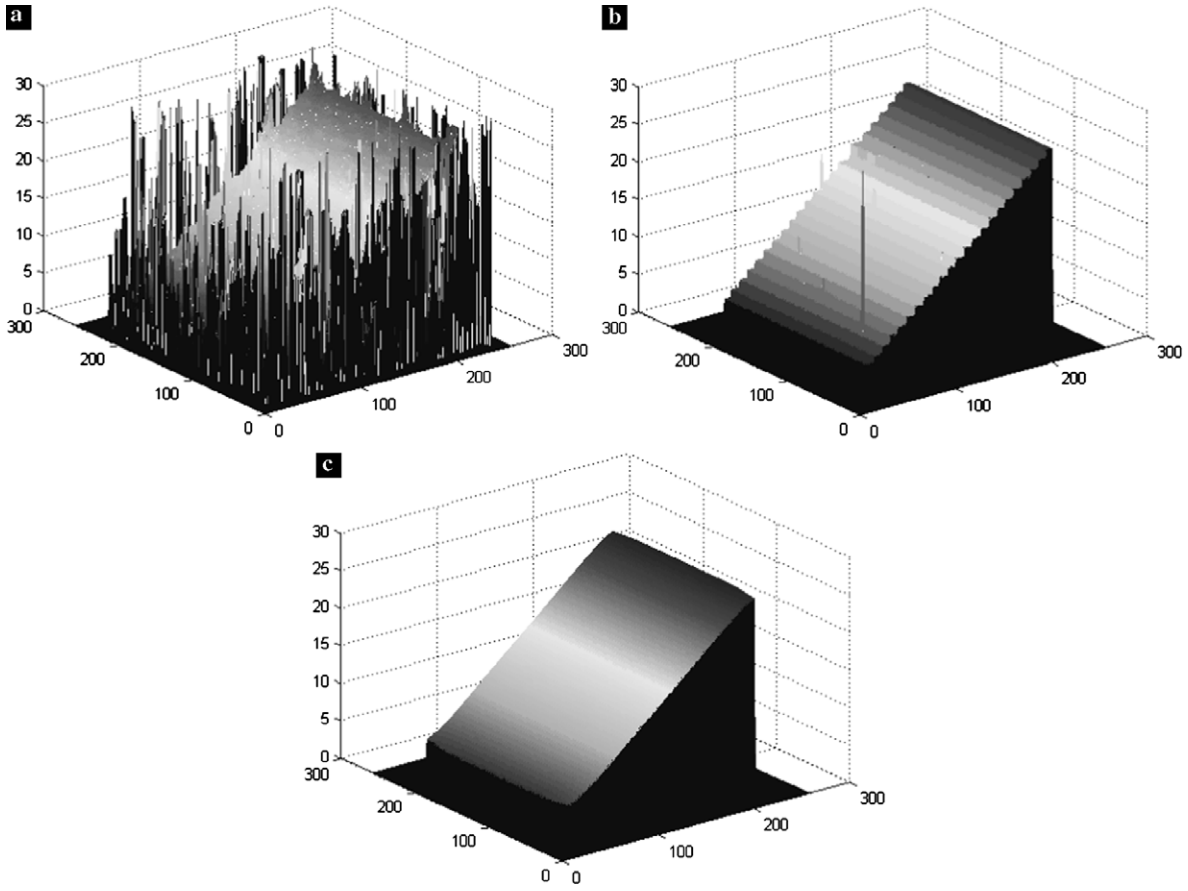

Fig. 5. Disparities maps obtained by: (a) AW-SSD; (b) GC; (c) HMMF (see also Fig. 7).

HMMF). This evaluation is made with respect to the computed ADG in the ramp of the test image (which measures the subpixel precision produced by each algorithm), and with respect to the average root mean squared (RMS) error [34,35], between the computed disparity $d_C(x,y)$ and the ground-truth disparity $d_T(x,y)$:

$$R = \left(\frac{1}{N} \sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2 \right)^{1/2} \tag{28}$$

where N is the total number of pixels.

In the case of AW-SSD, σ was varied from 1.1 to 2.0, and in each one of these values the size of an adaptive window was varied from three pixels minimum to 20 pixels maximum, the results are shown in plot (a) of Fig. 6; plot (a-1) is the ADG and plot (a-2) is the RMS error. One can see that although the ADG values are very small (subpixel level), the RMS errors are large, compared with the results of GC and HMMF.

Table 1
Processing time

	Time
AW-SSD	6.25 min
GC	3.75 h
HMMF	17.57 min

For GC, in theory, one may improve the precision by using a finer (subpixel) discretization of the disparity at the expense of increasing the computational complexity; in this experiment, the precision of the disparity discretization was set to 0.1 pixels; however, in order to reduce the noise influence in the disparity estimation, one must increase the value of the regularization parameter λ, causing ADG values to be higher than the true value (plot (b-1)), promoting piecewise constant solutions. This phenomenon limits the performance of GC in terms of RMS error (plot (b-2)). This effect is visible in more detail in Fig. 7: for high values of λ, the GC algorithm is resistant to noise, but the computed disparity is heavily discretized, regardless of the value of the discretization precision (Figs. 7(a) and (b)), whereas for low values of λ the algorithm is very vulnerable to noise (panel (c)).

The results obtained by HMMF are shown in plot (c-1) and (c-2) of Fig. 6, and in panel (d) of Fig. 7; one can see that the ADG and RMS values remain practically constant for all the range of the regularization parameter (λ_1 in (18)), being possible in this case to reduce the RMS value up to 0.1286 (more than two times lower than GC).

In the next experiment, we test qualitatively the performance of these algorithms with a real stereo pair taken in normal conditions and containing more complex objects. This stereo pair contains a pineapple, see Fig. 4. One

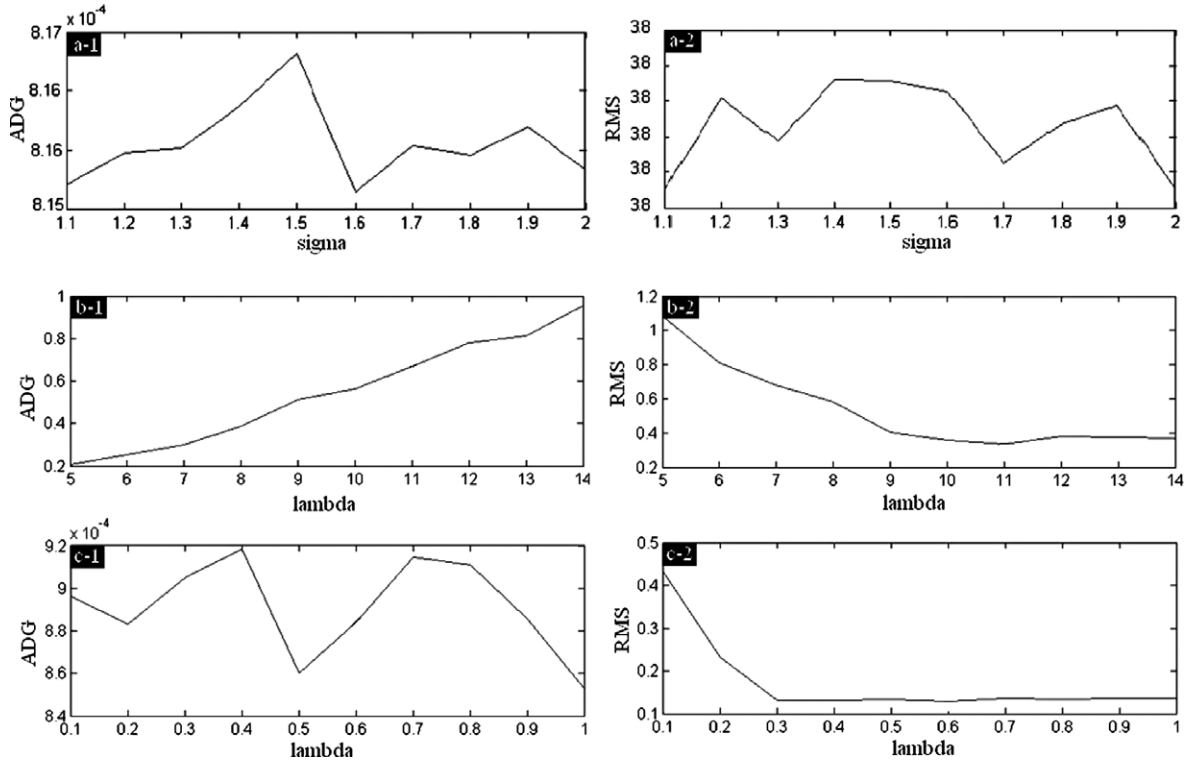

Fig. 6. ADG and RMS errors obtained by (a) AW-SSD, (b) GC, and (c) HMMF.

can see that the pineapple's crest and body are composed by irregular surfaces, making it specially difficult to estimate the disparity map and occluded regions. The disparity maps estimated by AW-SSD, GC, and HMMF are shown in Fig. 8. In Fig. 9, we plot the corresponding disparity profiles of row 128. One can see in the graphs how HMMF, using the dual membrane parametric model, estimates with high precision the complex disparity surface of the pineapple's body and the smooth surface of the tablecloth (the black regions correspond to occluded regions). In the case of GC, the disparity is strongly discretized, mainly on the pineapple's body. It is not possible to increase the precision of GC, since it would be necessary to decrease the value of its regularization parameter, producing noisy disparity maps. Finally, the profile obtained by AW-SSD shows that in order to estimate appropriately disparities on the pineapple's body, one has to reduce the noise parameter, causing noisy disparity values on the surface of the tablecloth.

5.2. Standard benchmarks

We also evaluated the performance of the proposed algorithm, using a set of stereo images available in the web site: www.middlebury.edu/stereo. These stereo pairs have available disparity ground truth, including their occluded regions, being possible to make quantitative comparisons. We use the metric described in (28) and the Percentage of Bad Matching Pixels metric (BMP) [34,35],

$$B = \frac{1}{N}\sum_{(x,y)}(|d_C(x,y) - d_T(x,y)| > \delta_d) \tag{29}$$

to the algorithm evaluations, where δ_d is a disparity error tolerance; for the experiments in this paper we use $\delta_d = 1.0$.

We tested four stereo pairs; three of them, "Venus", "Sawtooth", and "Tsukuba" are color images that were transformed to monochromatic images for this comparison, since there are many different transformations that may be applied to the RGB components, affecting the results obtained by the algorithms; the third one is a monochromatic image, "Map".

It is important to note that the available ground-truth disparity for these stereo pairs is discretized to integer disparity values in all cases; therefore, the advantages of HMMF, in the sense of being able to compute high-precision disparities, are not apparent in this comparison, and, moreover, it becomes a liability, since the comparison is bound to favor (incorrectly) algorithms (such as GC) which estimate only integer disparities.

The comparisons were made with respect to some recent algorithms reported in [34]: algorithms that use as matching cost absolute differences (AD) and squared differences (SD) followed by a winner-take-all optimization; global algorithms as dynamic programming (DP) [28], scanline optimization (SO) [36], graph cuts (GC) [24,25], simulated annealing (SA), and Layered Stereo (LS) [42]. We tested also some variants of these algorithms that use as matching cost the Birchfield and Tomasis's measure [37]; aggregation

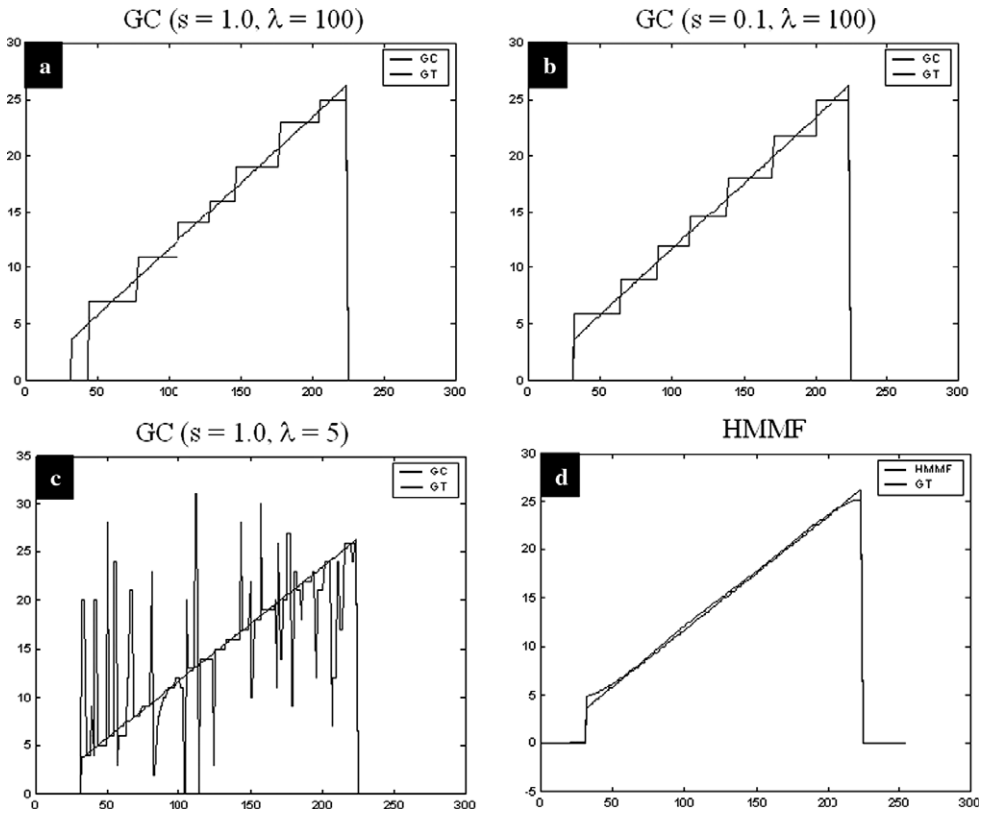

Fig. 7. Center profile of the disparity maps obtained by GC and HMMF: (a) GC: *disparity discretization* $s = 1.0$, $\lambda = 100$; (b) GC: *disparity discretizations* $= 0.1$, $\lambda = 100$; (c) GC: *disparity discretization* $s = 0.1$, $\lambda = 5$; (d) HMMF.

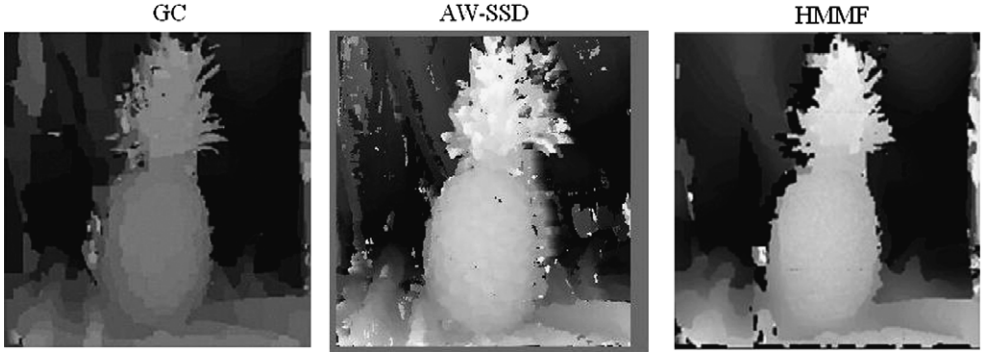

Fig. 8. Disparity maps obtained by AW-SSD, GC, and HMMF.

methods: using different window sizes, minifilters [38], binomial filters or membrane diffusion [39].

The results obtained in these experiments by the algorithms and by the method proposed in this paper (HMMF) are plotted in Fig. 10; we show only the best results from each variant of the algorithms. The average for the two metrics RMS and BMP for the four stereo pairs are shown in Fig. 11. In this set of experiments, we include two versions of the HMMF algorithm: one using the DMM (HHMF-dmm), and the other one using as a parametric model horizontal planes $\Phi(r, \theta_k) = \theta_k, \theta_k \in \Re$ (HMMF-hp). The purpose of including these different versions was to verify the behavior of HMMF using a similar approach to the one used by the other algorithms (piecewise constant disparity models).

The best set of parameter values for the HMMF-dmm algorithm are shown in Table 2; in the case of the HMMF-hp, the same values were used, except for the membrane parameters, which are not used. Table 3 shows numerical values only for the RMS error, since the approach here proposed emphasizes in the precision of the disparity map values. One can see that the performance of HMMF (in both versions HMMF-dmm and HMMF-hp) is highly competitive; the difference between GC and HMMF is minimal in the RMS metric, and it is also minimal between LS and HMMF in the BMP metric, in spite

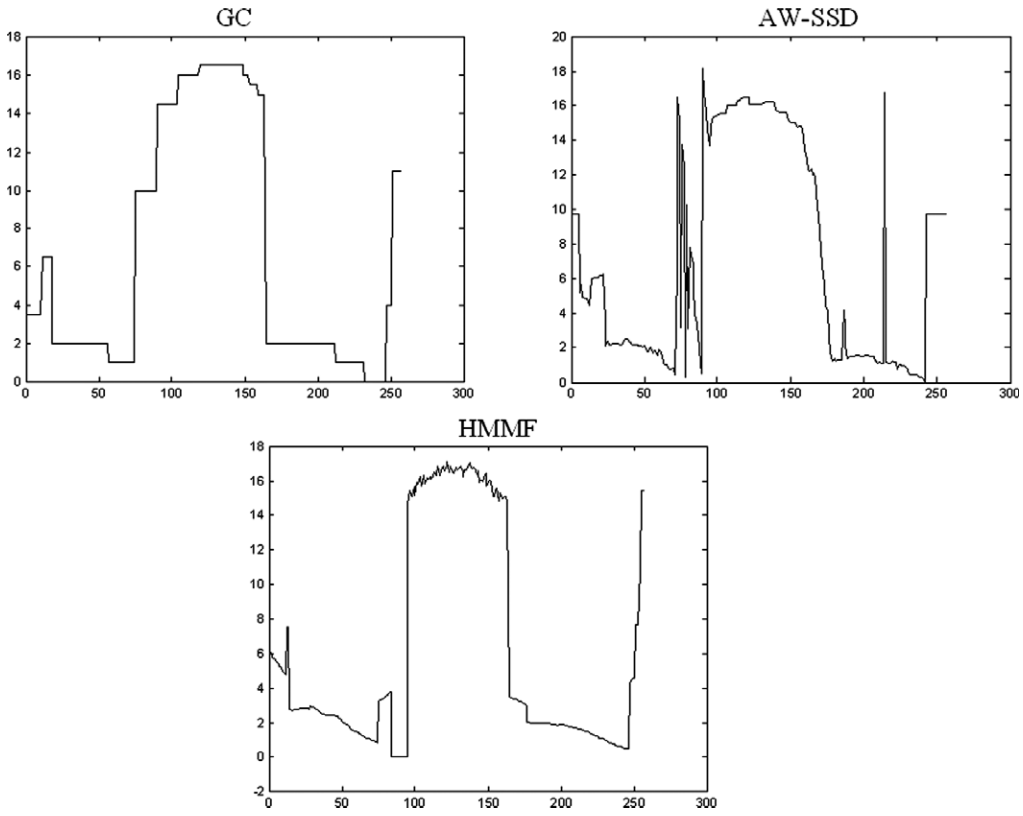

Fig. 9. (Top) Profile disparity maps obtained by GC and AW-SSD; (bottom) disparity profiles of HMMF.

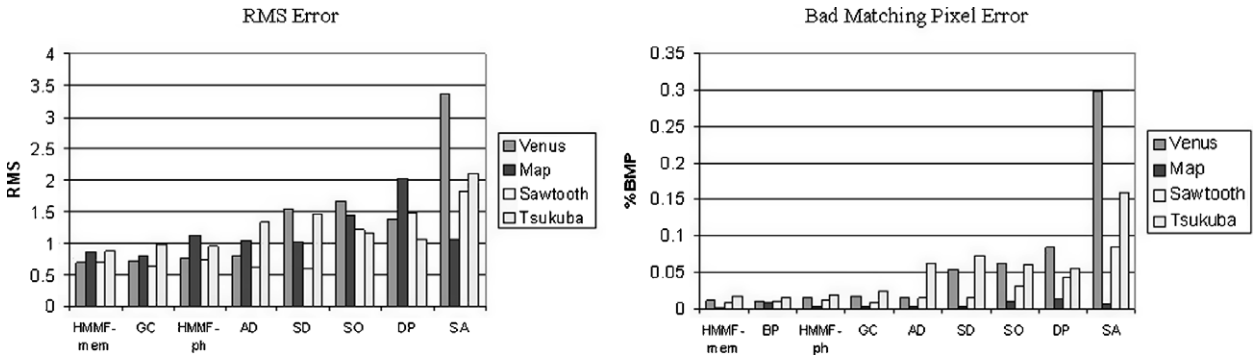

Fig. 10. RMS error and BMP error for "Venus", "Sawtooth", "Tsukuba" and "MAP" stereo pairs.

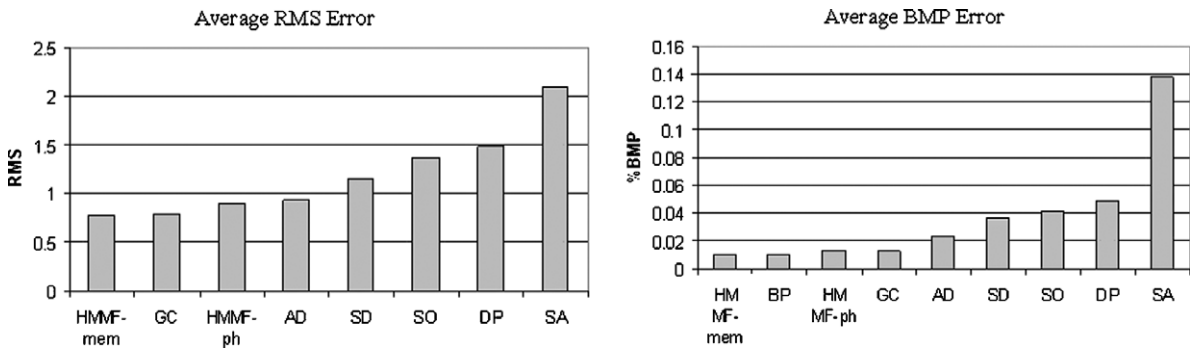

Fig. 11. Average error for RMS and BMP.

Table 2
Main parameter values for the HMMF algorithm

	Parameter of	Venus	Map	Sawtooth	Tsukuba
η_1	Spline membrane	50	50	50	100
η_2	Pixel–pixel membrane	50	50	50	100
α	Likelihood	0.2	0.2	0.2	0.2
τ	Huber potential	0.5	0.5	0.5	0.5
μ_1	Disparity edge	0.8	0.0	0.8	1.0
μ_2	Disparity non-edges	0.02	0.0	0.02	0.01
λ_1	Spatial coherence	0.2	1.5	0.5	0.3
λ_2	Map consistency	0.1	0.1	0.1	0.1
v_0	Occlusion likelihood	0.4	0.4	0.4	0.01
Regions		7	4	7	9

Table 3
RMS error using the best parameters for each image, and the minimum (best) value in each column is shown in bold

	Venus	Map	Sawtooth	Tsukuba
HMMF-mem	**0.6835**	0.8549	0.6932	**0.8706**
GC	0.719704	**0.796838**	**0.645627**	0.984643
HMMF-ph	0.7578	1.1311	0.7474	0.95195
AD	0.784794	1.034352	0.609795	1.333782
SD	1.52871	1.023509	0.592214	1.460442
SO	1.671218	1.444221	1.211812	1.149207
DP	1.380688	2.009353	1.479986	1.07093
SA	3.378821	1.058696	1.826139	2.109155

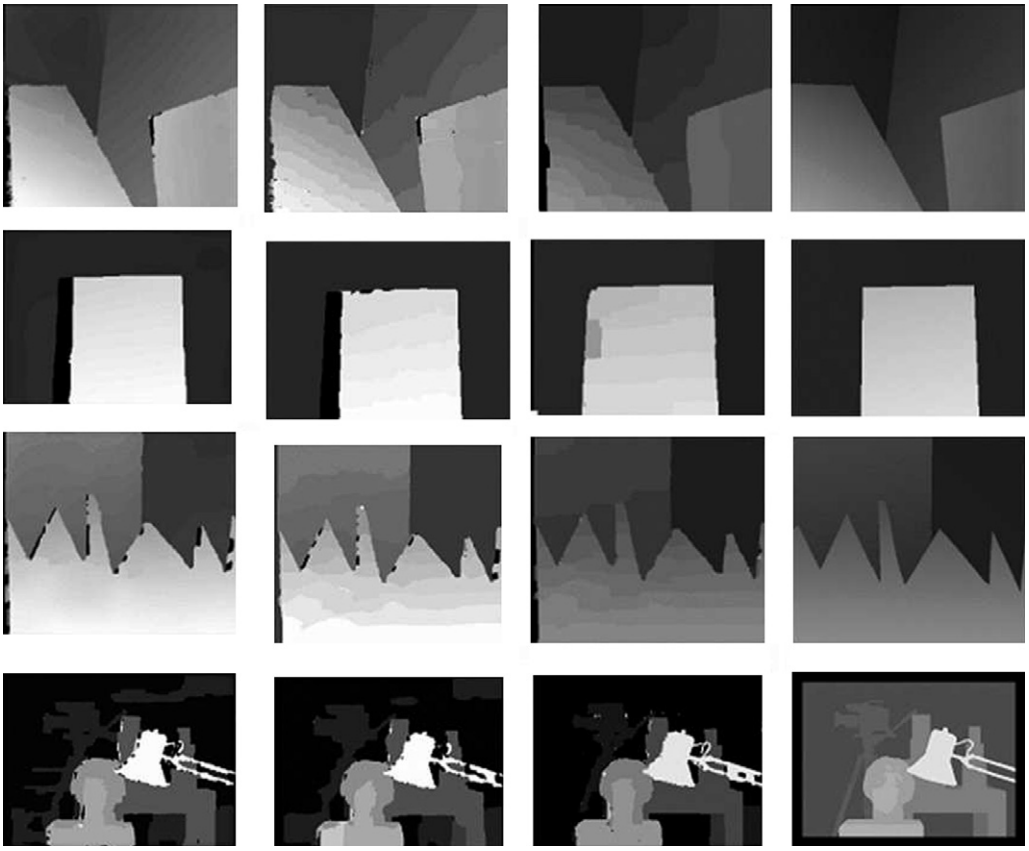

Fig. 12. Results obtained for the stereo pairs: "Venus", "MAP", "Sawtooth", and "Tsukuba"; first column by HMMF-mmd, second column by HMMF-hp, third column by GC, and last column ground-truth disparity maps.

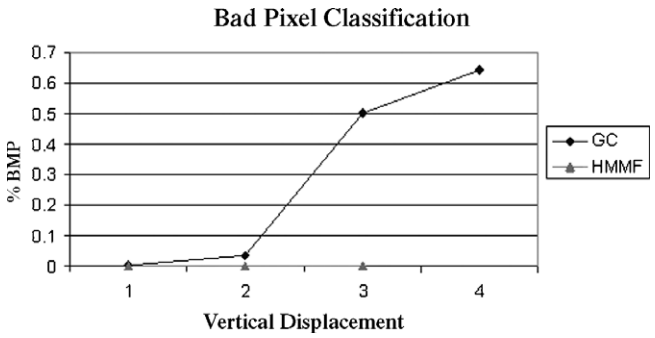

Fig. 13. BPC error for HMMF and GC when vertical displacements occur.

of the provided discretized ground truth. The disparity maps obtained by the two algorithms HMMF (HMMF-dmm, HMMF-hp) and GC for these stereo pairs are shown in Fig. 12, together with the corresponding ground truth.

An important aspect about the real stereo pairs used in these experiments is that they were taken in very controlled conditions (see [34] for details). In practical situations, it is possible that small vertical displacements occur; therefore algorithms must be, in some way, robust with respect to these displacements. In our approach, this robustness is easily incorporated, simply by making the models $\Phi(r,\theta_k)$ vector-valued, so that vertical disparities are automatically corrected. The relative robustness of GC and our approach is illustrated in Fig. 13, which measures the increase in BMP error when an artificial vertical disparity is introduced in the "MAP" stereo pair. As one can see, the performance of our algorithm is practically insensitive to these vertical displacements.

6. Conclusions

In this paper, we have proposed a new algorithm for disparity estimation. It is based on a new Bayesian formulation of the image segmentation model [31] that uses a dual membrane to model disparities. It allows one to estimate with very high precision (subpixel level) disparity maps, since the DMM can model disparities that vary smoothly, and at the same time, small high-frequency disparity variations in the same region. The proposed algorithm is very robust with respect to vertical misalignments, allowing one to process real stereo pairs which were taken in realistic conditions, with excellent results. Another important contribution is the estimation, using the same energy function (18), two p-fields (one time taking as reference image the left component and the other the right one), allowing one, in this way, to check consistencies in both disparity maps, increasing the precision of the computed disparity map, and also, estimating explicitly the location of occluded regions and enforcing the correspondence between intensity and disparity discontinuities.

We have shown that the algorithm presented here represents a significant improvement over other state-of-the-art schemes in terms of subpixel precision of the disparity estimates, a feature that may be relevant in applications such as optical metrology, 3D digitalization, aerial photogrammetry, etc. The presented method is also robust with respect to noise and vertical misalignments, while maintaining a competitive performance on standard benchmarks with discretized ground-truth disparities.

Acknowledgements

The authors thank the anonymous reviewers for their comments, which helped to improve the quality of this paper. J.L. Marroquin was supported in part by Conacyt Grant No. 46270.

References

[1] Emanuele Trucco, Alessandro Verro, Introductory Techniques for 3D Computer Vision, Prentice-Hall, New Jersey, 1998.
[2] Lorenzo J. Tardon, Javier Portillo, Carlos Alberola-Lopez, A novel Markovian formulation of the correspondence problem in stereo vision, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 24 (3) (2004) 428–436.
[3] S. Gutierrez, J.L. Marroqun, Robust approach for disparity estimation in stereo vision, Image and Vision Computing (22) (2004) 183–195.
[4] B. Julez, Foundation of Cyclopean Perception, The University of Chicago Press, Chicago and London, 1971.
[5] W.E.L. Grimson, From Images to Surfaces, MIT Press, Cambridge, MA, 1978.
[6] B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, 1986.
[7] O. Fuageras, Three-dimensional Computer Vision, MIT Press, Cambridge, MA, 1993.
[8] P. Anandan, A computational framework and an algorithm for the measurement of visual motion, International Journal of Computer Vision 2 (3) (1989) 283–310.
[9] T. Kanade, M. Okutomi, A stereo matching algorithm with adaptive window: theory and experiment, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (9) (1994) 920–932.
[10] R.C. Bolles, H.H. Baker, M.J. Hannash, The JISCT stereo evaluation, in: DARPA Image Understanding Workshop, 1993, pp. 263–274.
[11] D.N. Bhat, S.K. Nayar, Ordinal measures for visual correspondence, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'96), 1996, pp. 351–357.
[12] D. Geiger, B. Ladendorf, A. Yuille, Occlusions and binocular stereo, International Journal of Computer Vision 14 (3) (1995) 211–226.
[13] M. Okutomi, T. Kanade, A stereo matching algorithm with an adaptive window, IEEE TPAMI 16 (9) (1996) 920–932.
[14] R.D. Arnold, Automated stereo perception, Tech. Report AIM351, Artificial Intelligence Laboratory, Stanford University, 1983.
[15] T. Poggio, V. Torre, C. Koch, Computational vision and regularization theory, Nature 317 (6035) (1985) 314–319.
[16] M.J. Black, A. Rangarajan, On the unification of line processes, outlier rejection, and robust statistics with application in early vision, International Journal of Computer Vision 19 (1) (1996) 57–91.
[17] Sang Hwa Lee, Yasuaki Kanatsugu, JongIl Park, A hierarchical method of mapbased stochastic diffusion and disparity estimation, IEEE ICIP 2 (2002) 541–544.
[18] S. Geman, D. Geman, Stochastic relaxation, gibbs distribution, and the bayesian restoration of images, IEEE TPAMI 13 (5) (1984) 721–741.
[19] J. Marroquin, S. Mitter, T. Poggio, Probabilistic solution of ill posed problems in computer vision, Journal of the American Statistical Association 82 (397) (1987) 76–89.
[20] S.T. Barnard, Stochastic stereo matching over scale, International Journal of Computer Vision 3 (1) (1989) 17–32.

[21] P.B. Chou, C.M. Brown, The theory and practice of bayesian image labeling, International Journal of Computer Vision 4 (3) (1990) 185–210.
[22] Jian Sun, NanNing Zheng, HeungYeung Shum, Stereo matching using belief propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (7) (2003) 787–800.
[23] D. Geiger, F. Girosi, Parallel and deterministic algorithms for MRF's: surface reconstruction, IEEE TPAMI 13 (5) (1991) 401–412.
[24] V. Kolmogorov, R. Zabih, Computing visual correspondence with occlusions using graph cuts, in: ICCV, vol. 2, No. 2001, 2001, pp. 508–515.
[25] S. Birchfield, C. Tomasi, Multiway cut for stereo and motion with slanted surfaces, in: ICCV'99, 1999.
[26] S. Birchfield, C. Tomasi, Depth discontinuities by pixeltopixel stereo, International Journal of Computer Vision 35 (3) (1999) 269–293.
[27] Bernd Jahne, Digital Image Processing, Springer-Verlag, Berlin, Heidelberg, 2002.
[28] Y. Ohta, T. Kanade, Stereo by intra and inter scanline search using dynamic programming, IEEE TPAMI 7 (2) (1985) 139–154.
[29] P.N. Belhumeur, D. Mumford, A bayesian treatment of stereo correspondence problem using half occluded regions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 1992 (CVPR'92), 1992, pp. 506–512.
[30] I.J. Schoenberg, Cardinal Spline Interpolation, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1973.
[31] J.L. Marroquin, Edgar Arce, S. Botello, Hidden Markov measure field models for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (11) (2003) 1380–1387.
[32] Stan Z. Li, Markov Random Field Modeling in Image Analysis, Springer-Verlag, Tokyo, Berlin, Heidelberg, New York, 2002.
[34] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense twoframe stereo correspondence algorithms, International Journal of Computer Vision 47 (1/2/3) (2002) 7–42.
[35] D. Scharstein, R. Szeliski, High-accuracy stereo depth maps using structured light, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), vol. 1, 2003, pp. 195–202.
[36] A.F. Bobick, S.S. Intille, Large oclusion stereo, International Journal of Computer Vision 33 (3) (1999) 181–200.
[37] S. Birchfield, C. Tomasi, Depth discontinuities by pixeltopixel stereo, in: ICCV, 1998, pp. 1073–1080.
[38] H. Tao, H. Sawhney, R. Kumar, A global matching framework for estereo computation, in: ICCV I, 2001, pp. 532–539.
[39] D. Scharstein, R. Szeliski, Stereo matching with nonlinear diffusion, International Journal of Computer Vision 28 (2) (1998) 155–174.
[40] E. Gamble, T. Poggio, Visual integration and detection of discontinuities: the key role of intensity edges, A.I. Memo 970, 1987, Ai. Lab, MIT Press, Cambridge, MA.
[42] Michael H. Lin, Carlo Tomasi, Surfaces with occlusions from layered stereo, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), vol. 1, pp. I-710–I-717.