Stereo Matching with Mumford-Shah Regularization and Occlusion Handling

Rami Ben-Ari, *Member*, *IEEE*, and Nir Sochen, *Member*, *IEEE*

Abstract—This paper addresses the problem of correspondence establishment in binocular stereo vision. We suggest a novel spatially continuous approach for stereo matching based on the variational framework. The proposed method suggests a unique regularization term based on Mumford-Shah functional for discontinuity preserving, combined with a new energy functional for occlusion handling. The evaluation process is based on concurrent minimization of two coupled energy functionals, one for domain segmentation (occluded versus visible) and the other for disparity evaluation. In addition to a dense disparity map, our method also provides an estimation for the half-occlusion domain and a discontinuity function allocating the disparity/depth boundaries. Two new constraints are introduced improving the revealed discontinuity map. The experimental tests include a wide range of real data sets from the Middlebury stereo database. The results demonstrate the capability of our method in calculating an accurate disparity function with sharp discontinuities and occlusion map recovery. Significant improvements are shown compared to a recently published variational stereo approach. A comparison on the Middlebury stereo benchmark with subpixel accuracies shows that our method is currently among the top-ranked stereo matching algorithms.

Index Terms—Stereo matching, Mumford-Shah functional, variational stereo vision, occlusion handling, Total Variation.

✦

1 INTRODUCTION

IN binocular stereo vision, we have a pair of images captured by two cameras in different locations. The central task is the establishment of *correspondence* (also called stereo matching). The solution to this problem is expressed by the *disparity* map that reveals the difference in location of corresponding pixels, along epipolar lines. For the sake of simplicity, we consider two rectified images or alternatively assume the two images are captured by identical coplanar cameras.

Over the years, numerous algorithms have been proposed for stereo vision. For review on various stereo techniques and recent advances in computational stereo the reader is referred to [9]. Many stereo techniques use features to establish correspondence [45], [38]. While their advantage is to gain accurate results on specific cues, the main drawback of these methods is the sparsity of the recovered disparity/depth map. Dense disparity maps provide more detailed information and thus there is much interest in these methods recently.

In general, dense matching algorithms can be classified into local and global methods. Local approaches utilize the color or intensity values within a finite window to determine the disparity in each pixel [20], [16], [50]. Alternatively, global approaches incorporate explicit smoothness assumptions and determine dense disparity

by energy minimization techniques, such as graph cuts [24], belief propagation [22], [42], dynamic programming [26], or variational framework [40], [36], [27], [2], [21], [41]. The variational framework offers the advantage of providing mathematically sound techniques with clear modeling.

Correspondence establishment by variational methods has attracted much interest in recent years. This framework currently suggests the most accurate results in the optical flow field [10], [4], [43], where correspondences in 2D space are established (in opposed to 1D search in stereo). The main reason behind this success is the spatially continuous formalization and the consequent real (noninteger) estimates. The continuous approach enables the variational methods to have an inherent subpixel evaluations obtained without refinement of the search region (causing a computational overhead) or a postprocessing process.

Stereo research has recently entered a new era, with public access to real data sets having accurate ground truth disparities [37], [34]. The Middlebury stereo benchmark [34] is now a well-known ranking method for various stereo matching algorithms. In the first version of the Middlebury ranking table, variational methods were rarely present, apparently due to the discrete error measures, discarding subpixel accuracy. Recently, with adding the option for subpixel error measures (although still discrete), new variational methods are participating and presenting competitive performances [35], [30]. In this respect, our proposed algorithm also promotes the ranking of the variational approaches in the Middlebury benchmark.

The first ingredient of the objective functional is the data-fidelity term. This term expresses a relation between one or more attributes in the images. While, commonly, gray-level constancy is considered, in this paper we suggest a data term that incorporates image gradients and also deals with color images.

● *R. Ben-Ari is with Orbotech Ltd., Yavneh 81101, Israel.*
  E-mail: rami-ba@Orbotech.com, benari.ram@gmail.com.
● *N. Sochen is with the Department of Applied Mathematics, School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel. E-mail: sochen@post.ac.il.*

The inherent ill-posedness of the minimization problem is solved by adding a regularization (also referred as smoothness) term, reflecting some a priori preference on the solution. Seeking piecewise smooth modeling of the disparity map with abrupt crossing in between, we propose a novel regularizer based on the Mumford-Shah (MS) functional [31] in its Γ-convergence approximation [3], [12]. Similarly to the MS approach in image segmentation, a discontinuity map is evaluated in our method revealing the discontinuities in the disparity map. In this paper, we view the discontinuity map as a feature that may be enforced to follow new additional constraints in order to obtain a more accurate disparity/depth boundary allocations .

Recently published methods in variational stereo vision, e.g., [2], [25], [27], [35], [36], [41], disregard the half-occlusions (also briefly referred as occlusions) in the stereo problem. These regions consist of points projected from the scene that are visible in one view but not in the other. Our study shows that, without detecting the half-occlusions and discarding them form the objective functional, poor results are obtained. We therefore suggest a novel domain segmentation for extraction of the half-occlusions. Labeling the occlusion domain is also useful in indicating the inestimable depth points (by stereopsis).

We evaluate the performance of our method on data sets with widely different characteristics, available in the Middlebury page [34]. Experimental results demonstrate the superior performance of our method in comparison to a recently published variational method of Slesareva et al. [41], as well as high subpixel accuracies comparing state of the art stereo techniques in the Middlebury benchmark.

This paper extends work previously published at a conference [7]. Substantial differences are, among other things:

- A thorough introduction and discussion on related work is presented here, including an introduction to the Mumford-Shah framework.
- The discontinuity function is differently calculated, yielding improved results.
- An extensive experimental evaluation is shown, including tests on the new Middlebury data set and ranking of our method in the Middlebury benchmark [34]. To this end, an automatic parameter setting is suggested in this paper.
- The proposed method for Middlebury performance assessment discards the additional constraints imposed on the discontinuity function in [7], presenting a simpler method with higher accuracies. The additional constraints are optionally added to yield a new model for improved allocation of the discontinuities.
- Comparison between L^2 and L^1 regularizer in the Mumford-Shah functional is newly presented here.

The paper is organized as follows: In Section 2, we survey the related work. Section 3 includes an introduction to Mumford-Shah framework. In Section 4, we present our baseline method relying on the previously published stereo matching approach [41]. Next, a new inhomogeneous image-based approach is introduced using the MS framework. This method will then be used for initialization of our suggested stereo matching algorithm. In Section 6, we

introduce our novel variational model concurrently handling half-occlusions and discontinuities. Section 7 then deals with the minimization strategy of the defined energy functionals. The final algorithm is described stage by stage in Section 9. Section 10 presents extensive experimental results including quantitative and qualitative comparison to the baseline approach and the Middlebury benchmark methods. A discussion on the error measures and performance assessment is presented in Section 11. Finally, the paper is concluded in Section 12 by a short summary.

2 RELATED WORK

In this section, we survey various regularization methods in variational framework as well as different occlusion handling approaches.

The first proposition for the regularization can be acknowledged as the Tikhonov and Arsenin [44] L^2 smoothness of the gradients. The well-known consequent of the L^2 term is the *homogenous* diffusion resulting from the blur of the discontinuities in the revealed disparity map. In order to cope with this problem, two classes of regularization terms were suggested using different strategies, the *image-driven* and the *disparity-driven* (also referred to as data-driven) methods. The image-driven terms, pioneered by Mansouri et al. in [27] and further used in [2], exploit the intensities in the image to determine a criterion for spatial changes in the diffusion. The suggested regularizer in [27], [2] is anisotropic in the resulting diffusion-reaction process and the diffusion at each point is steered according to the local intensity gradient. Kim and Sohn [21] recently presented a model with a different and inhomogeneous image-driven term. All of the image-driven methods are based on the assumption that an intensity edge indicates a disparity (depth) boundary. Obviously, this hypothesis is not necessarily valid in textured areas and, therefore, the emerged disparity maps often suffer from a noisy pattern in textured regions.

The disparity-driven terms seem to be a more natural choice since the diffusion follows the disparity pattern. However, the disparity map is a priori unknown and part of the evaluation process. Various disparity-driven regularizers tackled this problem, e.g., [25], [36], [40], [41]. A preliminary suggestion for a disparity-driven term goes back to 1993, where Shah [40] proposed a nonlinear diffusion model based on the Mumford-Shah framework in segmentation [31]. Later on, Robert and Deriche [36] proposed a framework for several disparity-driven regularizers.

A well-known regularization type in this class is the Total Variation (TV) resulting from an *inhomogeneous* diffusion process [47]. Although this term is *isotropic*, it has gained large popularity due to its simple form and implementation and often adequate results in preserving the discontinuities. Employment of the TV method in stereo applications can be found in [25], [30], [35], [41]. Inspired by the accurate optical flow evaluation of Brox et al. [10], Slesareva et al. [41] recently introduced a new variational method for stereo, including the TV regularizer. Using a robust data term and an efficient numerical seding to the previous work of Alvarez et al. [2].

The TV regularizer is considered a discontinuity preserving term due to the monotonically decreasing diffusivity [47] with respect to growth of data gradients. However, the diffusivity often is not dropped as abruptly as desired,

resulting in relatively wide disparity edges. Our approach suggests embedding the TV regularizer in the MS functional, resulting in significantly sharper discontinuities.

Several kinds of constraints are typically used for occlusion detection, e.g., ordering, monotonicity, uniqueness, and stereo match consistency [9]. The ordering constraint requires the conservation of the relative order between points along a pair of scan-lines [30]. The monotonicity term [40], [18] is a variant of the ordering constraint enforcing the correspondence mapping function to be monotonically increasing. It is well known that these constraints are violated in scenes with *narrow occlusions* [13], i.e., with foregrounds including thin objects or narrow holes (e.g., an object behind a fence).

The uniqueness constraint used in [28], [24], [30] enforces a one-to-one mapping between pixels in the stereo pair. However, this requirement is appropriate for integer valued methods and is not satisfied in scenes containing horizontally slanted surfaces [42].

Stereo match *consistency* is widely used to detect half-occlusions [8], [26], [48]. In this approach two disparity maps are computed, one based on the correspondence from the left image to the right and the other based on the counterpart matching. Inconsistent disparities are then declared as half-occlusions. The strategy adopted in this paper is based on the robust approach of stereo match consistency [9]. While common methods use this strategy as a postprocessing stage for occlusion detection, in our approach this procedure is fused into the disparity evaluation process. The occlusion map is then built gradually, yielding accurate disparity and occlusion maps.

Although a variational approach for stereo vision including Mumford-Shah type regularization with occlusion handling was presented by Shah in [40]; the method suggested in this paper does not have much in common with [40]. The main differences are:

1. Instead of the quadratic data term in [40], the modified L^1 norm is employed here.
2. We employ a Mumford-Shah smoothing term incorporating the TV regularizer in the sense of [39] instead of the original type based on the L^2 term in [40].
3. The half-occlusion detection in [40] is based on the monotonicity assumption, a variant of the ordering constraint, while in our method, robust stereo match consistency is employed, including a novel energy functional for occlusion detection based on domain segmentation and level sets.
4. Instead of the standard gradient descend approach for minimization [40], highly sensitive to local minimas, the efficient fixed point iteration scheme is used here.
5. The method in [40] approximates the scene to fronto parallel planes, therefore imposing a strong limitation for its application. In our method no assumptions are made regarding the structure of the scene.

3 MUMFORD-SHAH FRAMEWORK

The Mumford-Shah functional, originally proposed in [31], deals with the denoising and segmentation of gray-level images. The following segmentation functional which was introduced by Mumford and Shah [31], models an image as a set of piecewise smooth segments separated by well-behaved contours: Given an observed image $I_0 : \Omega \to \mathbb{R}$, with Ω denoting the image domain, find a set of smooth curves C_i and an optimal piecewise smooth approximation I of I_0 such that I varies smoothly within $\Omega \backslash C_i$ and discontinuously across the set of the edges C_i:

$$E(I, C_i) = \int_{\Omega} (I - I_0)^2 dA + \beta_i \int_{\Omega \backslash C_i} \|\nabla I\|^2 dA + \alpha_i \int_{C_i} d\sigma, \tag{1}$$

where $\nabla := (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^T$. The last term in (1) is a line integral penalizing the length of C_i. The coefficients $\alpha_i, \beta_i > 0$ are fixed parameters presenting the relative weights for penalization of the smoothness and the edge length, respectively. The index i here relates the parameters to the image cue. In the homogenous regions, therefore, *linear* diffusion is imposed and the smoothing is inhibited across the discontinuities. The primary difficulty in the minimization process of (1) is the presence of the unknown discontinuity set C_i in the integration domain. Due to the irregularity of this functional, classical calculus of variation methods are not applicable, and approximation approaches have to be used. In this respect, De Giorgio introduced an approximation of the MS segmentation functional (1) by regular functionals in the framework of Γ-convergence [12].

The main idea of Γ-convergence is to approximate the irregular functional $E(I, C_i)$ by a sequence of regular functionals $E_{\epsilon_i}(I)$ such that

$$\lim_{\epsilon_i \to 0} E_{\epsilon_i}(I) = E(I, C_i)$$

and the minimizers of E_{ϵ_i} approximate the minimizer of E.

Ambrosio and Tortorelli [3] used the Γ-convergence framework for minimization of the Mumford-Shah functional, representing the discontinuity set by a characteristic function $(1 - \chi^{C_i})$. Then, they introduced an approximation for χ^{C_i} by a smooth auxiliary function $v_i(x, y)$, where $v_i \approx 0$ if $(x, y) \in C_i$ and $v_i \approx 1$ otherwise. Finally, the solution can be approximated by the minimizer of the following functional:

$$E^{MS}(I, v_i) = \int_{\Omega} (I - I_0)^2 dA + \beta_i \int_{\Omega} \|\nabla I\|^2 v_i^2 dA + \alpha_i \int_{\Omega} \epsilon_i \|\nabla v_i\|^2 + \frac{(v_i - 1)^2}{4\epsilon_i} dA. \tag{2}$$

The first term denotes the data-fidelity term. The second term favors piecewise smooth solution and corresponds to the term $\int_{\Omega \backslash C_i} \|\nabla I\|^2$. The third term demands smoothness of the discontinuity set and is related to the line integral $\int_{C_i} d\sigma$, in (1).

The Total Variation Mumford-Shah functional was suggested by Shah [39]:

$$E^{MSTV}(I, v_i) = \int_{\Omega} (I - I_0)^2 dA + \beta_i \int_{\Omega} \|\nabla I\| v_i^2 dA + \alpha_i \int_{\Omega} \epsilon_i \|\nabla v_i\|^2 + \frac{(v_i - 1)^2}{4\epsilon_i} dA, \tag{3}$$

where the L^2 norm $\|\nabla I\|^2$ in (2) was replaced by the TV regularizer (L^1 norm) $\|\nabla I\|$. The Γ-convergence of this functional was proven by Alicandro et al. [1]. The latter

Mumford-Shah functional with the TV regularizer (MSTV) is more tolerant to outliers than the seminal MS stabilizer (2) and drives the recovered image closer to the piecewise constant limit [39]. In the case of high noise density, the cartoon-like restored image is much cleaner than the recovered image using the original method [6].

4 THE BASELINE APPROACH

Consider a normalized image pair $I_j : \Omega_j \rightarrow [0, 1]$ with $j \in \{l, r\}$ standing for the left and right images, respectively. For every point $\mathbf{x} = (x, y) \in \Omega_l$ we describe the location of the corresponding point in the domain of the right image by $\mathbf{g}_l = (x - d_l(\mathbf{x}), y) \in \Omega_r$, where $d_l : \Omega_l \rightarrow \mathbb{R}$ denotes the *left* disparity map.

A common assumption relating the corresponding pixels is the brightness constancy. In practice, the use of other attributes, such as intensity gradients, helps in establishing the right correspondences by reducing ambiguity [22], [41]. Recently, Slesareva et al. [41] suggested a robust data term based on the modified L^1 norm for constancy in both the brightness and the brightness gradients:

$$e_d(s_d^2) = \sqrt{s_d^2 + \eta^2}, \qquad (4)$$

where

$$s_d^2 = \|I_r(\mathbf{g}_l) - I_l(\mathbf{x})\|^2 + \lambda \|\nabla I_r(\mathbf{g}_l) - \nabla I_l(\mathbf{x})\|^2. \qquad (5)$$

The parameter η is a small positive constant commonly viewed as a stabilizing factor (since the pure L^1 norm is not derivable at zero). The *modified* L^1 norm shares the robustness to outliers of the L^1 norm, but prevents the resulting PDE from being singular at zero. Note that when $s_d^2 \ll \eta^2$, the norm in (4) tends to L^2 since, by Taylor expansion rule:

$$\sqrt{s_d^2 + \eta^2} = \eta \sqrt{1 + \left(\frac{s_d}{\eta}\right)^2} \approx \eta \left(1 + \frac{s_d^2}{2\eta^2}\right), \qquad (6)$$

which has the quadratic form. Hence, the parameter η interpolates between the L^1 and L^2 norms, converging to L^1 for $\eta \rightarrow 0$. We set this constant by the small positive value $\eta = 0.001$ in order to be close to the L^1 norm.

While the data term in (5) handles gray-level values, we elaborate this term to also accommodate color images. Given a pair of images in RGB form, the generalization to multichannel color space is conducted by extension of the norms in (5) as Frobenius norms:

$$\mathbf{I}_j := (R_j, G_j, B_j)^T : \Omega_j \rightarrow [0, 1]^3, j \in \{l, r\} \qquad (7)$$

and

$$\nabla \mathbf{I}_j := (\nabla R_j, \nabla G_j, \nabla B_j)^T. \qquad (8)$$

In our baseline approach we set the smoothness prior similar to [41] as the (modified) TV regularizer:

$$e_s = \beta \sqrt{\|\nabla d_l\|^2 + \eta^2} \qquad (9)$$

where β is a weight parameter.

Considering the data and smoothness terms in (4), (9), the following energy functional is obtained for disparity evaluation:

$$E(d_l) = \int_{\Omega_l} \sqrt{\|I_r(\mathbf{g}_l) - I_l(\mathbf{x})\|^2 + \lambda \|\nabla I_r(\mathbf{g}_l) - \nabla I_l(\mathbf{x})\|^2 + \eta^2}$$
$$+ \beta \sqrt{\|\nabla d_l\|^2 + \eta^2}. \qquad (10)$$

Note that this functional is the same as proposed by Slesareva et al. in [41], except for the extension of the data term to color space.

5 AN INHOMOGENEOUS IMAGE-BASED APPROACH

Several methods have been recently proposed based on (over)segmentation of the stereo images [22], [42], [15]. These approaches are based on two main assumptions: 1) The disparity varies smoothly on each image segment and 2) the disparity discontinuities coincide with the segment *appearance* boundaries. Assuming that the object's appearance boundaries in the image C_i are characterized by high intensity gradients, according to the latter assumption, the disparity set of edges C_d should satisfy $C_d \subseteq C_i$ (note that C_i also includes texture gradients). Following this paradigm, we introduce the following image-based approach for correspondence establishment. An image edge map is first obtained by minimization of the energy functional (2) with respect to the discontinuity function v_i, setting $\beta_i = 1$ (w.l.o.g). According to Euler-Lagrange equations, v_i should satisfy:

$$\frac{\delta E_{\epsilon_i}}{\delta v_i} = \|\nabla I\| v_i + \alpha_i \frac{(v_i - 1)}{4\epsilon_i} - \alpha_i \epsilon_i \nabla^2 v_i = 0, \qquad (11)$$

where ∇^2 stands for the euclidean Laplacian operator. We then use $v_i \in [0, 1]$ as an inhomogeneous diffusion factor in the following energy functional for disparity estimation d_{l_0}:

$$E(d_{l_0}) = \int_{\Omega_l} \sqrt{\|I_r(x - d_{l_0}, y) - I_l(x, y)\|^2 + \eta^2}$$
$$+ \widehat{\beta} v_i^2 \sqrt{\|\nabla d_{l_0}\|^2 + \eta^2} dA, \qquad (12)$$

where $\widehat{\beta} > 0$ is a weight parameter. The above scheme involves an inhomogeneous diffusion process where the regularization rate is decreased at high intensity gradients. Moreover, this scheme presents a unique diffusivity term:

$$\frac{\delta E}{\delta d_{l_0}} = e_d' \frac{\partial s_d^2}{\partial d_{l_0}} - \widehat{\beta} \mathrm{div} \left(\underbrace{\frac{v_i^2}{\sqrt{\|\nabla d_{l_0}\|^2 + \eta^2}}}_{\text{Diffusivity}} \nabla d_{l_0} \right) = 0. \qquad (13)$$

The diffusivity approaches its minimum value for $v_i \rightarrow 0$ (where an intensity edge exists) and $\|\nabla d_{l_0}\| \rightarrow \infty$ (i.e., at a discontinuity in the disparity). So, the diffusion is highly decreased at locations where there is a coincidence between disparity discontinuity and image edge.

The results from minimization of (12) for four Middlebury data sets are shown in Fig. 1. The obtained disparity maps
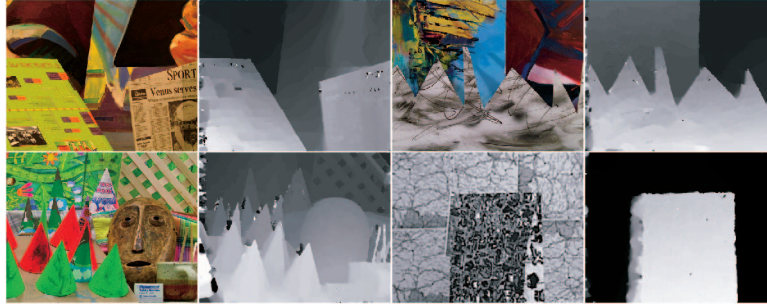
Fig. 1. The image-based approach results. *First row*: Venus and Sawtooth. *Second row*: Cones and Map. Left: The reference image. Right: Evaluated disparity map.

suffer from artifacts appearing mainly in the highly textured regions. However, the disparity boundaries are well localized thanks to the unique diffusivity term (see (13)).

In the following section, we will describe a new model based on the MS functional. Since, in such a method, the initial allocation of the disparity boundaries affects the final results, we use the inhomogeneous image-based approach as an initialization stage for our suggested algorithm.

6 THE PROPOSED MODEL

In this section, we evaluate the disparity map, half-occlusions, and the discontinuity function separately for each view. For the sake of abbreviation, we present the scheme for the left image. The formulation for the counterpart right image is obtained similarly.

6.1 Detecting the Occlusion Domain

We use the superscripts **v** and **o** in our notation to indicate attribution to the *visible* (nonoccluded) and *half-occluded* domains, respectively.

The image domain Ω_l can be divided into two *disjoint* sets:

$$\Omega_l = \Omega_l^{\mathbf{v}} \cup \Omega_l^{\mathbf{o}},$$

where $\Omega_l^{\mathbf{v}}$ denotes the visible and $\Omega_l^{\mathbf{o}}$ the occluded domain. Our visibility constraint requires that a pair of corresponding points should be mapped to each other by the associated disparities (homography) or, formally,

$$d_l(\mathbf{x}) = -d_r(\mathbf{g}_l), \tag{14}$$

where $d_r(\mathbf{g}_l)$ presents the disparity at the point $\mathbf{g}_l$ in the right image. We define the stereo match *inconsistency* measure by:

$$u_l(\mathbf{x}) := d_l(\mathbf{x}) + d_r(\mathbf{g}_l), \quad \forall \mathbf{x} \in \Omega_l, \tag{15}$$

where $u_l \approx 0$ indicates consistent points and $u_l \gg 0$ presents inconsistency. Frequently, (15) is used as a *pointwise* score for extracting the occluded points [8], [26], [48]. However, we model the half-occlusions as a *region* that should be segmented out of the whole domain. Since the disparity range has an impact on the score in (15), we give every inconsistency measure a penalty by a truncated L^1 function:

$$e_{occl}^l = -\ln\left(\epsilon_{occl} + (1 - \epsilon_{occl})e^{-|u_l|}\right). \tag{16}$$

The function e_{occl}^l is the new occlusion cost function and $\epsilon_{occl} > 0$ is a constant. The truncation yields a saturation level controlled by ϵ_{occl}. Typical plots of e_{occl}^l are depicted in Fig. 2.

Inspired by the Chan and Vese binary segmentation approach in images [11], [46], we extract the half-occlusion domain form e_{occl}^l space. We further make use of the powerful Osher and Sethian level set method [32], allowing complex domain topologies to be recovered (e.g., several disjoint regions). To this end, a level set function $\phi_l : \Omega_l \to \mathbb{R}$ is defined. The *Heaviside* function of ϕ_l,

$$H(\phi_l) = \begin{cases} 1, & \phi_l \geq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{17}$$

is then used to indicate the visible and occluded regions in the image domain, corresponding to the nonnegative and negative levels of ϕ_l, respectively. Practically, a smooth approximation of the Heaviside function H_ϵ, rather than a step function is used [11]:

$$
\begin{aligned}
H_\epsilon(\phi) &= \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{\phi}{\epsilon}\right)\right) \\
\hat{\delta}_\epsilon(\phi) &= \frac{dH_\epsilon(\phi)}{d\phi} = \frac{1}{\pi}\frac{\epsilon}{\epsilon^2 + \phi^2},
\end{aligned}
\tag{18}
$$

where $\hat{\delta}_\epsilon$ is an approximation for the Dirac delta function. In our application we set $\epsilon = 0.1$.

The basic idea is to look for a particular partition of the domain into two regions, one representing the visible and the other the half-occluded set. Differently from the minimal variance requirement in [11], we define our partition model using a threshold level t:
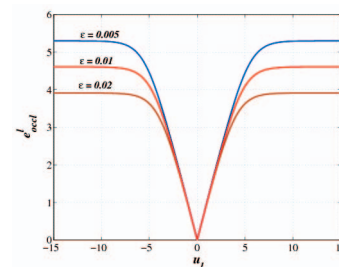

Fig. 2. The truncated L^1 function for $\epsilon_{occl} = 0.005, 0.01, 0.02$.

$$E_{occl}^l(\phi_l) = \int_{\Omega_l} \left(e_{occl}^l * G_\sigma\right) H(\phi_l) + t(1 - H(\phi_l))$$
$$+ \nu \|\nabla H(\phi_l)\| dA, \tag{19}$$

where G_σ is a Gaussian with standard deviation of $\sigma = 0.1$ and $*$ denotes the convolution operator. Here, pixels having penalty cost over the threshold t will be favorably assigned as occluded. Demand for smooth boundaries between the regions is achieved by penalizing the length of the separating contour, imposed by the term $\|\nabla H(\phi_l)\|$ and the corresponding weight ν [14]. The low pass filter in (19) enforces the smoothness of the half-occlusion domain and is helpful to remove some isolated points or small holes from the half-occlusion region.

6.2 Mumford-Shah Functional for Stereo Matching

We define $d_l^{\mathbf{v}}, d_l^{\mathbf{o}} : \Omega_l \to \mathbb{R}$ as two disparity maps, one for the visible (with superscript $\mathbf{v}$) and the other for the half-occlusion (with superscript $\mathbf{o}$) domain. For a given half-occlusion map, the disparity on Ω_l is then calculated by combining the two contributions:

$$d_l = d_l^{\mathbf{v}} H(\phi_l) + d_l^{\mathbf{o}}(1 - H(\phi_l)). \tag{20}$$

Reflecting the preference for piecewise smooth disparity maps in the visible domain, elements from the Mumford-Shah segmentation functional [31] are used for regularization. In this stabilizer (see (1)), the assigned energy functional depends on the set of discontinuities in the disparity map C_d:

$$E_L\left(d_l^{\mathbf{v}}, C_d\right) = \int_{\Omega_l^{\mathbf{v}}} e_d + \int_{\Omega_l^{\mathbf{v}} \backslash C_d} e_s + \alpha \int_{C_d} d\sigma, \tag{21}$$

where $\alpha > 0$ is a fixed parameter presenting the weight for penalization of the total discontinuity length. Note that the data-fidelity and the smoothness constraints are solely demanded on the visible domain since, in the half-occlusions, matching points do not exist. The data similarity measure e_d is the same as the baseline method shown in Section 4. As for the smoothness prior, the TV regularizer in the baseline approach is embedded here in the MS framework. The choice of the TV (L^1) prior instead of the standard L^2 norm has an important effect. While the higher robustness of the L^1 norm in the MS stabilizer was previously reported in [39] in the context of segmentation and denoising, the benefit of the TV regularizer is even more important here. In the stereo matching problem, there is a need for recovery of the discontinuities concurrently with the disparity formation as opposed to the image segmentation, where the objects edges are apparent a priori, characterized by high intensity gradients. The influence of changing the classical L^2 norm with the L^1 is also demonstrated in Section 10.3 by experimental results.

Following the Ambrosio and Tortorelli Γ-convergence framework [3], the discontinuity set C_d can be approximated by a regular function v_l, where $v_l(\mathbf{x}) \approx 0$ if $\mathbf{x} \in C_d$ and $v_l(\mathbf{x}) \approx 1$ otherwise. Similarly, we introduce the following approximation for the functional in (21):

$$E_L\left(d_l^{\mathbf{v}}, v_l\right) \cong \int_{\Omega_l^{\mathbf{v}}} e_d + e_s v_l^2 + \widehat{\alpha}(v_l - 1)^2 dA$$
$$+ \widehat{\epsilon} \int_{\Omega_l} \|\nabla v_l\|^2 dA, \tag{22}$$

where $\widehat{\alpha}, \widehat{\epsilon}$ are related to the standard MS parameters in (1) and can be viewed as new weight scalars,

$$\widehat{\alpha} = \frac{\alpha}{4\epsilon}, \quad \widehat{\epsilon} = \alpha\epsilon.$$

The discontinuity set is now approximated by the function $v_l(\mathbf{x})$, where $v_l \approx 1$ in the homogenous regions and $v_l \approx 0$ at a discontinuity. Note that the integration for the left term in (22) is just over the visible domain $\Omega_l^{\mathbf{v}} \subset \Omega_l$, which is initially unknown. We alleviate this irregularity using the level set function ϕ_l and express the functional in (29) by a single integral on the whole image domain Ω_l:

$$E_L\left(d_l^{\mathbf{v}}, v_l\right) = \int_{\Omega_l} \left[e_d + e_s v_l^2 + \widehat{\alpha}(v_l - 1)^2\right] H(\phi_l)$$
$$+ \widehat{\epsilon} \|\nabla v_l\|^2 dA. \tag{23}$$

We alos require the smoothness of the discontinuity function v_l in the occluded domain since, w.l.o.g, a constant value for v_l is expected in this region.

The minimization process also requires assignment of values to the disparities in the occluded region ($d_l^{\mathbf{o}}$ in (20)). To this end, a soft constraint is set, allowing the diffusion from the visible domain into the occluded region but not vice versa. This procedure causes the regions that are falsely labeled as occluded at the preliminary stages of minimization to later become part of the visible set. Prohibiting the diffusion from the occluded region to the visible domain prevents the correct disparities in the visible domain from being deviated from by the unreliable values of the occluded region. These demands are satisfied by driving $d_l^{\mathbf{o}}$ toward the solution of:

$$\operatorname{div}\left(e_s'(\nabla d_l^{\mathbf{o}}) v_l^2 \nabla d_l^{\mathbf{o}}\right) = 0, \tag{24}$$

which presents a standard diffusion process. The desired unique diffusion process is obtained by the following two steps: First, at each iteration, the computation is initialized by $d_l^{\mathbf{o}} = d_l^{\mathbf{v}}$. This periodic initialization causes the diffusion from the visible to the occlusion domain to take place. Then, d_l is updated by (20), preventing the diffusion in the opposite direction.

7 MINIMIZATION

According to calculus of variations, minimizers of (19) and (23) must fulfill the Euler-Lagrange (EL) equations. The associated PDE for (19) is:

$$\frac{\delta E_{occl}^l}{\delta \phi_l} = \left(e_{occl}^l - t - \nu \operatorname{div}\left(\frac{\nabla \phi_l}{\|\nabla \phi_l\|}\right)\right) \widehat{\delta}(\phi_l) = 0, \tag{25}$$

and the corresponding EL equations for (23) are given by:

$$\frac{\delta E_L}{\delta d_l^{\mathbf{v}}} = \left(e_d' \frac{\partial s_d^2}{\partial d_l^{\mathbf{v}}} - 2 \operatorname{div}(e_s' v_l^2 \nabla d_l^{\mathbf{v}})\right) H(\phi_l) = 0, \tag{26}$$

Fig. 3. Additional constraints. Left: Sawtooth image set with half-occlusions labeled by yellow. A disparity edge at the adjoining occlusion-visible domain in left (red) is mapped to an edge inside the visible domain. Right: Left and right occlusion maps. An occluded point in the right image domain "O" is mapped to a point in the visible domain of the counterpart image "V" and vice versa.

$$\frac{\delta E_L}{\delta v_l} = (e_s v_l + \widehat{\alpha}(v_l - 1))H(\phi_l) - 2\widehat{\epsilon}\,\mathrm{div}(\nabla v_l) = 0, \quad (27)$$

where, in (26):

$$\frac{\partial s_d^2}{\partial d_l^v} = -2\sum_{i=1}^{3} I_{rx}^i(\mathbf{g}_l)\big(I_r^i(\mathbf{g}_l) - I_l^i(\mathbf{x})\big)$$

$$- 2\lambda \sum_{i=1}^{3} \big(I_{rx}^i(\mathbf{g}_l) - I_{lx}^i(\mathbf{x})\big)I_{rxx}^i(\mathbf{g}_l)$$

$$+ \big(I_{ry}^i(\mathbf{g}_l) - I_{ly}^i(\mathbf{x})\big)I_{rxy}^i(\mathbf{g}_l).$$

The subscripts x, y indicate partial derivatives.

We apply an alternate minimization (AM) approach to solve the coupled system of PDEs in (25)-(27). Therefore, at each iteration step we minimize with respect to one function keeping the others fixed. The EL equation (27) presents a linear PDE with respect to v_l. Discretization of (27) leads to a sparse linear system of equations that can be efficiently solved with iterative solvers, such as Jacobi, Gauss-Seidel, successive over-relaxation (SOR), or even multigrid methods. On the contrary, the PDEs in (25) and (26) are nonlinear. We therefore employ a semi-implicit gradient descend method for obtaining the solution of (25). For the solution of (26), nested fixed point iterations similar to [10], [41] are used, yielding a sparse linear system of equations that are then solved by the SOR scheme.

All spatial derivatives are approximated by central differences in discretization and the image warping is performed by linear interpolation. The PDEs (25)-(27) are subjected to the Neuman boundary condition stating that, for the variable function f, the condition $\frac{\partial f}{\partial n} = 0$ is imposed, where n is a unit vector normal to the image boundary $\partial\Omega_l$.

8 ADDITIONAL CONSTRAINTS FOR IMPROVED RECOVERY OF THE DISCONTINUITY MAP

The discontinuities in the disparity map present an important feature for a realistic rendering and are often under particular interest. There are two causes for these discontinuities: 1) jumps in the depth and 2) transitions between visible-occluded domains. The discontinuities due to domain transitions are frequently handled in a post-processing stage by filling of the occluded area by the smoothness requirement since, in this region, the disparities are inestimable by stereopsis.

Experimental tests show that recovery of the discontinuities by (27) yields unsatisfactory results for the adjoining visible-occlusion boundaries. To this end, two constraints based on the stereo configuration may be added to our model in (23).

Occlusions often cause the blurring of the discontinuities in the recovered disparity map at the adjoining visible-occluded boundaries. Due to stereo geometry, these discontinuities in one image are mapped *into* the visible domain in the counterpart image, and therefore, are estimated reliably there. Suppose $T_r : \Omega \subseteq \Omega_r \to \widetilde{\Omega}_r \subset \Omega_l$ is a coordinate transformation from the right image to the left. Let $\widetilde{v}_r : \widetilde{\Omega}_r \to \mathbb{R}$ denote a discontinuity function given by:

$$\widetilde{v}_r(x, y) := v_r(x - d_r, y). \quad (28)$$

The mapping constraint for the discontinuities at the adjoining visible-occlusion domain is formally expressed by: $v_l \approx \widetilde{v}_r$ in $\Omega_l^o \cap \widetilde{\Omega}_r^v$.

The next constraint states that a half-occluded point in the right image cannot be mapped to a half-occluded point in the left view and vice versa. Otherwise, such a point will be occluded in both views, yielding a contradiction to the definition of the *half*-occluded point. This condition is also valid for points in the discontinuity function v_l. We therefore impose $v_l \approx 1$ on $\Omega_l^o \cap \widetilde{\Omega}_r^o$, suppressing the localization of discontinuities in this nonlegitimum domain. The two proposed constraints are illustrated in Fig. 3. Enforcing the two new constraints yields the following energy functional for d_l^v, v_l:

$$E_L^{AC}\big(d_l^v, v_l\big) = E_L^v + \int_{\Omega_l^o \cap \widetilde{\Omega}_r^v} (v_l - \widetilde{v}_r)^2 dA$$

$$+ \int_{\Omega_l^o \cap \widetilde{\Omega}_r^o} (v_l - 1)^2 dA. \quad (29)$$

Note that the new model in (29) introduces a novel idea, where the discontinuity function is not just considered as an auxiliary function but introduces an attribute that has to obey certain constraints, emerged from the stereo configuration.

The irregular energy functional in (29) is replaced by the following regular functional using the level set function ϕ_l:

$$E_L^{AC}\big(d_l^v, v_l\big) = E_L + \int_{\Omega_l} (v_l - \widetilde{v}_r)^2(1 - H(\phi_l))H(\widetilde{\phi}_r)$$

$$+ (v_l - 1)^2(1 - H(\phi_l))(1 - H(\widetilde{\phi}_r))dA, \quad (30)$$

where $\widetilde{\phi}_r := \phi_r(x - d_r, y)$. The new Euler-Lagrange equation for v_l now includes additional linear terms:

$$\frac{\delta E_L^{AC}}{\delta v_l} = \frac{\delta E_L}{\delta v_l} + (v_l - \widetilde{v}_r)(1 - H(\phi_l))H(\widetilde{\phi}_r)$$

$$+ (v_l - 1)(1 - H(\phi_l))(1 - H(\widetilde{\phi}_r)) = 0. \quad (31)$$

In the minimization process the function $v_l(\mathbf{x})$ is initialized by the solution of (27) and the intermediate values for functions $\widetilde{v}_r, \widetilde{\phi}_r$ are calculated by linear interpolation.

9   ALGORITHM

Solving the system of PDE's (25)-(27) (or, alternatively, (25), (26), and (31)) in alternation leads to the following iterative algorithm:

1. Initialize the left and right disparity maps using the inhomogeneous image-based method in Section 5, assuring rather accurate localization of the discontinuities.
2. Perform *domain* segmentation for the left and right images using (25) and its counterpart.
3. Compute the discontinuity maps v_l and v_r using (27) (or, alternatively, (31)) and the corresponding counterpart.
4. Solve (26) and its counterpart for the unknown d_l^γ and d_r^γ using the current values of v_l, v_r and ϕ_l, ϕ_r.
5. Fill the disparity gaps in the half-occluded domain by (24).
6. Update the values for d_l and d_r by (20) and its counterpart.
7. Repeat steps 2-6 until convergence.

10   EXPERIMENTAL RESULTS

We evaluate the proposed algorithm on real images from the Middlebury site [34]. At first, the estimated disparity maps are compared to our baseline approach presenting an implementation of a previously published variational method [41], with a data term generalized to accommodate color images. Next, the method is tested on the new Middlebury data set in order to obtain a global rank. In the following section, our nominal approach with the combined MS and TV stabilizer (see Section 8) is evaluated against two alternatives, one with MS regularizer including an L^2 smoother and the second with the model including additional constraints on the discontinuity function.

10.1   Automatic Parameter Setting

The proposed method includes parameters from three main categories. 1) Constants related to image segmentation: α_i, ϵ_i. 2) Parameters associated with the magnitude of the disparity gradients: $\widehat{\alpha}, \widehat{\epsilon}$. 3) The data internal weight λ and the data-smoothness weights $\beta, \widehat{\beta}$. In this section, we present a fully automatic parameter setting for our method.

We start by setting $\epsilon_i = 0.1$ and the internal data weight $\lambda = 2$ for the color sets and $\lambda = 1$ for gray-level images. The parameter $\widehat{\beta} = 0.2$ is chosen for color images and $\widehat{\beta} = 0.1$ for gray-level sets.

Next, we perform the following parameter adjustment:

$$\beta = \frac{2}{\text{Disparity Range}},$$
$$\alpha_i = \frac{0.002}{\text{mean } \|\nabla I\|}, \tag{32}$$
$$\widehat{\alpha} = 0.9375\beta, \quad \widehat{\epsilon} = 0.25\beta,$$

where the factors in these relations are empirically found. Not having a priori an estimation for the magnitude of the disparity gradients, we relate β to the disparity range. The parameter α_i depends on the image's contrast and therefore is set according to the intensity/color gradients.

For our half-occlusion model, we assign $\epsilon_{occl} = 0.01$ in (16), thus setting the saturation level to ~5 pixels. In order to assure gradual formation of the occlusion domain and avoid false visible/occlusion labeling, the stereo consistency threshold t is reduced continuously by an exponentially decay function from the initial value $t = 5$, where all of the domain is determined as visible, to the final value of $t = 1$. The weight for the segmenting contour length in (19) is set to $\nu = 0.01$. All of the results obtained from our method in this paper are based on the above automatic parameter setting.

10.2   Evaluation

The experimental tests are conducted on data from the old and new Middlebury sets. The tests include the Tsukuba and the Map sets characterized by fronto parallel scenes, the Venus and Sawtooth image sets consisting of slanted planar objects and the Cones and Teddy pairs presenting non-planar scenes with curved surfaces. The disparity ranges are 19 pixels for Venus and Sawtooth, 15 pixels for Tsukuba, 25 pixels for Map, and 59 pixels for Teddy and Cones sets.

Having the half-occlusion field evaluated, we use it for occlusion filling. We employ the simple occlusion filling heuristic in [24], where the occluded pixels are filled by the nearest background disparity values on the same scanline.

We also asses the performances quantitatively. Our error measure for comparison to the baseline is the *continuous* measure, Average Absolute Disparity Error (AADE) used in [2], [41]:

$$AADE = \frac{1}{N} \sum_{i=1}^{N} \left| d_i^{truth} - d_i^{estimated} \right|, \tag{33}$$

where N is the number of pixels in the domain of interest. Often the estimation of the discontinuities in the disparity map are of particular concern. We follow the definitions in [37] and use the *masks* from the Middlebury page [34] for accuracy assessment at visible pixels (ignoring a border of 10 pixels) and regions near disparity discontinuities.

10.2.1   Comparison to Previous Work in Variational Framework

We now show an experimental comparison to our baseline model (10) based on the previous variational method of Slesareva et al. in [41] . This model involves two parameters, β, the relative data-smoothness weight, and σ, the standard deviation of the Gaussian kernel in the image presmoothing. In order to present the best possible performances from this method, the parameters were tuned individually for each image set. Fig. 4 depicts the recovered disparity maps along with the corresponding error distribution obtained from the baseline method in comparison to our approach. Clearly, our piecewise smooth model yields improved results even in the highly textured Map set where one might expect a failure from an MS-based method involving segmentation principals. In all of the test cases, the results from our approach present sharper disparity boundaries with notably lower errors rates. In particular, the proposed method succeeds in recovering thin objects like the narrow sticks at the right part of the Cones image, where the baseline approach fails. Preserving the discontinuities at the boundaries of occluded regions is often notorious. However, our method also deals with this task adequately.
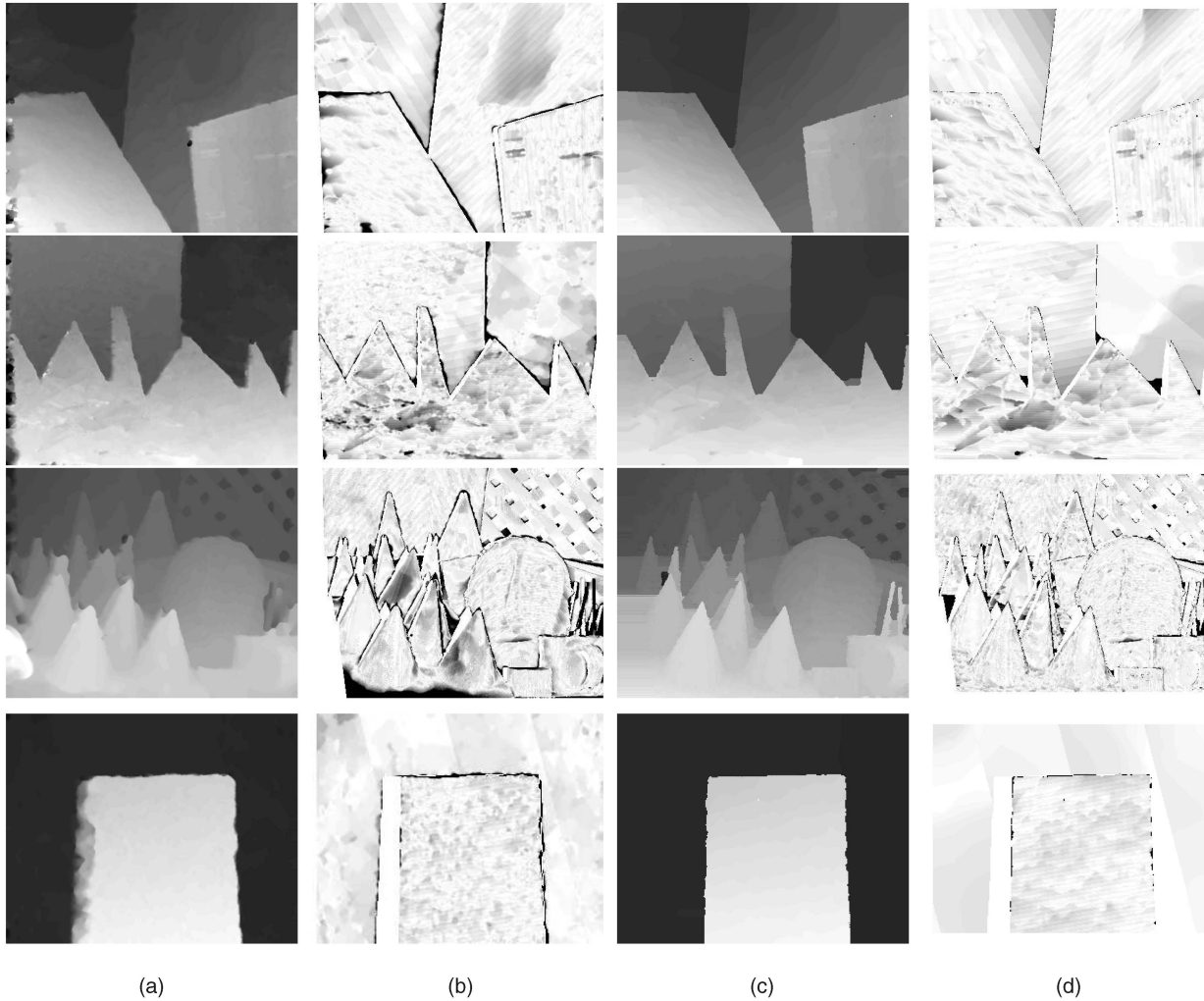
Fig. 4. Disparity maps. *First row*: Venus, *Second row*: Sawtooth, *Third row*: Cones, *Fourth row*: Map. (a) Disparity: Baseline method, (b) Error map: Baseline method, (c) Disparity: Our method, (d) Error map: Our method. The error maps correspond to absolute disparity errors in the nonoccluded domain. The gray level range is $[0, 2]$, where darker points present higher errors.

The AADE error measures for the baseline and the inhomogeneous image-based methods are summarized in Table 1 against the suggested approach. The errors are calculated in two regions, the nonoccluded domain (nonoccl) and the discontinuity set (disc). The results show a significant decrease in error figures by our method in both, the nonoccluded (visible and the discontinuity regions). Note that the same data term is used in the baseline and the

proposed method. The improvement, therefore, is due to our novel approach for regularization and occlusion handling.

Furthermore, Fig. 5 shows a comparison of recovered occlusion maps with ground truth results. The estimated half-occlusion field is realistic and close to ground truth, even in the challenging case of the Cones set.

One should note that errors in evaluation of the occlusion map may yield degradation of the results when

TABLE 1
Performance Comparison of Stereo Algorithms Using AADE Measure

	Venus		Sawtooth		Cones		Map	
	nonoccl	disc	nonoccl	disc	nonoccl	disc	nonoccl	disc
Baseline method	0.221	1.066	0.304	1.010	0.587	1.266	0.228	1.396
Inhom. Image Based Approach	0.353	0.420	0.264	0.617	0.488	0.975	0.250	1.073
Our method	0.135	0.291	0.227	0.579	0.387	0.850	0.198[2]	1.149[2]
Improvement[1]	39%	73%	25%	43%	34%	33%	13%	18%

Notes: [1] *Improvement for our method w.r.t the baseline.* [2] *The corresponding figures* **without** *occlusion filling are:* $AADE(nonoccl) = 0.170$, $AADE(disc) = 0.755$.

Fig. 5. The half-occlusion maps. Left: Ground truth. Right: Our evaluation.

occlusion filling is conducted, as in the case of the Map set (see Table 1). Occlusion labeling errors in the Map set are caused by the high texture pattern and appear at the right and (probably) upper side of the foreground. The existence of these errors even in the ground truth data is evidence of the difficulty of accurate occlusion estimation in this case.

10.2.2 Middlebury Rank

In this section, we provide a full evaluation of our continuous algorithm on the new Middlebury stereo database [34], using the parameter setting in Section 10.1. Fig. 6 shows the estimated disparity maps by our method. The results demonstrate the ability of our algorithm to compute smooth disparity maps with accurate discontinuities in all the data sets. Note that the discrepancies in the upper right part of the Teddy set are due to the untextured area in this region. The obtained disparity maps were submitted to the Middlebury Web site in February 2009.

Table 2 summarizes the performances according to Middlebury measures, for error threshold of $T = 0.5$. Here, in addition to the nonoccluded and the discontinuity regions, errors are also computed on the whole image domain (all) using the simple occlusion filling of [24] (constant plane extension). Our method is currently ranked number 4 out of 55 stereo algorithms (as for February 2009). Note that our algorithm performs well on all of the data sets rather than having a peak performance on specific data. Prior knowledge about the structure of the scene may improve the results (e.g., assuming planar surfaces in the occluded domain). Such information was used in a postprecessing process in [19], decreasing the outliers in the final disparity maps. However, one should consider that such assumptions may become invalid in other scenarios.

The computation time for the 375×450 Cones color set is ~ 30 minutes (including the initialization) for our implementation in MATLAB 7.4 run on a 1.86 GHz PC with 512 MB RAM. The computation time can be reduced by decreasing the number of iterations, compromising the accuracy. Furthermore, since PDE-based approaches are generally suitable for parallel computing on graphics cards [35], [23], employing GPU (Graphics Processing Unit) processing techniques is expected to cut computation time significantly (as demonstrated in [35]).

10.3 L^2 versus L^1 Regularizer and Applying the Additional Constraints

In this section, we first present a comparison between a traditional MS model with L^2 stabilizer, i.e., $e_s = \|\nabla d_l\|^2$,

against the Total Variation MS, including an L^1 regularizer, i.e.,

$$e_s = \sqrt{\|\nabla d_l\|^2 + \eta^2}.$$

Comparative tests conducted on the Venus and Sawtooth sets are shown in Fig. 7. The evaluated disparity maps reveal that discontinuities are better preserved by the L^1 choice, as expected.

An additional advantage of our method is in evaluation of a discontinuity function for the disparity map. This function is continuously distributed in the range $[0, 1]$, indicating the disparity boundaries (where $v_l \approx 0$) and presenting a measure for the intensity of a discontinuity. Recovery of v_l can be improved, particularly at the regions adjoining the visible-occlusion domain, by additional constraints (see Section 8). Fig. 8 shows the impact of the additional constraints on the recovered discontinuity function. The accurate allocation of the boundaries is demonstrated in Fig. 8 by plotting the set of points corresponding to $v_l \leq 0.6$ on the ground truth disparity maps. The results show that while the boundaries in the visible region are accurately indicated in both methods, the boundaries at the adjoining visible-occlusion domains are well recovered by the model with additional constraints. These constraints further suggest higher fidelity for all the discontinuities resulting improved results in noisy images.

Table 3 lists the error measures for the three discussed alternatives. Comparison between the first and second rows in Table 3 reinforces the qualitative observations in favor of the L^1 regularizer. The figures also show a slight impact on the accuracy when the additional constraints are used.

11 DISCUSSION

When dealing with assessment of a continuous disparity map one should examine the accuracies of the ground truth data. In the case of the Middlebury stereo sets, the ground truth precision for the Venus is $\frac{1}{8}$ pixels, for Cones and Teddy $\frac{1}{4}$ pixels, while in the Tsukuba case, the accuracy is only 1 pixel. These rates, especially in the Tsukuba set, raise a limitation for evaluating accuracy measures of continuous methods. Due to the importance of subpixel disparity evaluation, we employed the AADE metric (Average Absolute Disparity Error) for comparison to a previous variational method. This type of metric is also used in the Middlebury optical flow page [33] (average error angle and average end point error), where high subpixel precisions are required. However, the AADE is sensitive to the outliers, and
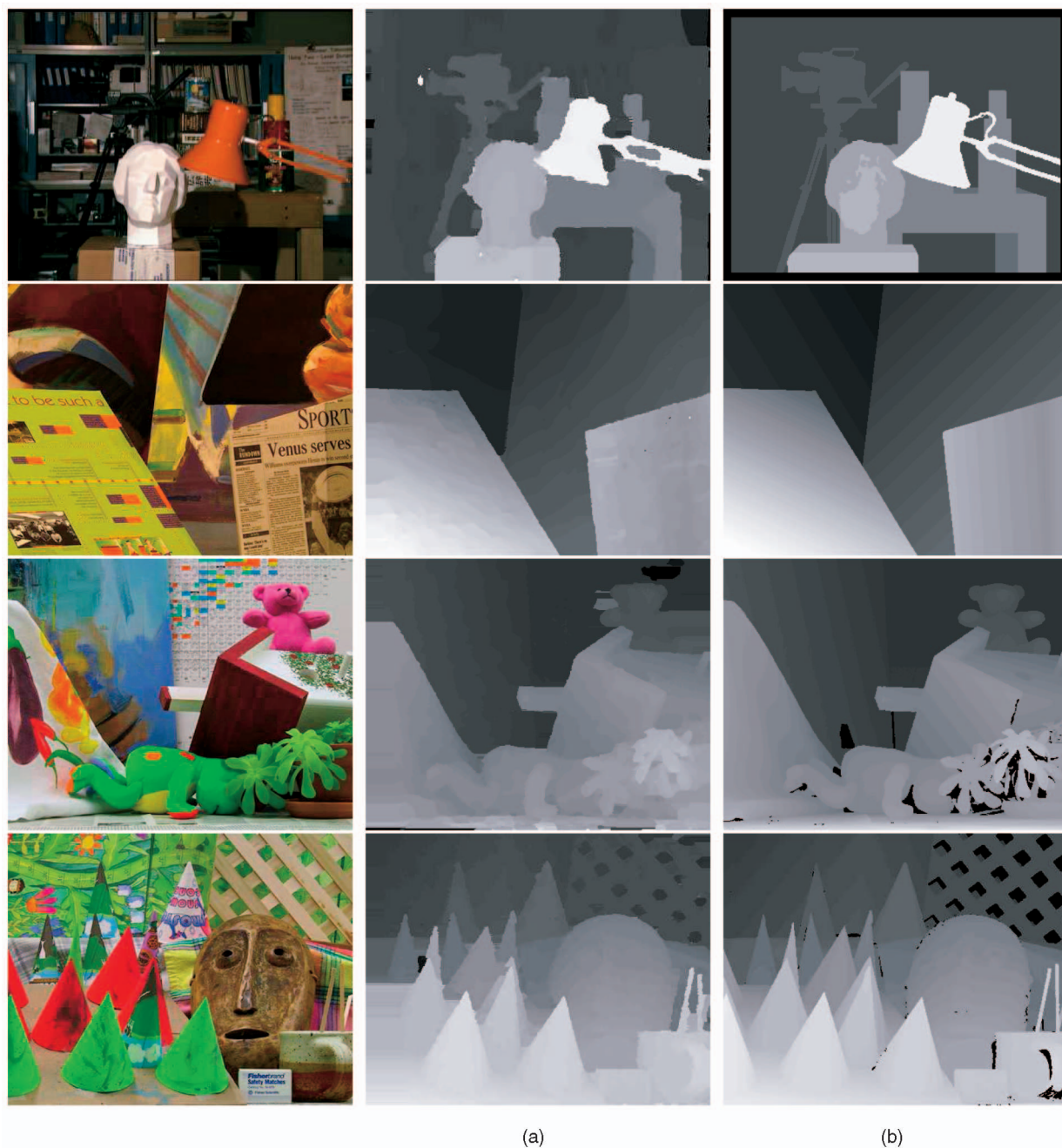
Fig. 6. Disparity maps. *First row*: Tsukuba, *Second row*: Venus, *Third row*: Teddy, *Fourth row*: Cones. (a) Our result. (b) Ground truth.

the average error may be significantly drifted due to a set of mismatched points obtaining the bounding values of the disparity range. While the common statistics metric, with thresholding the errors, copes with this problem, it can also create a wrong impression. For instance, if all of the pixels had an error of $T - \varepsilon$ pixels, with T being the threshold and ε a small number, the standard error statistics will obviously present a zero error. It seems that a better measure would be a combination of the mentioned factors, i.e., calculating the AADE only for the inliers. We performed this analysis for our comparison with [40]. Although a noticeable change in the AADE figures can be obtained (depending on the definition of the outlier bound) the relative improvement stays rather the same.

Using $T = 1$ in the Middlebury benchmark, our ranking is 34.6 out of 60 (as of February 2009). In order to get an overview of this change in our grading, we present in Fig. 9 the error distribution for various thresholds in comparison to two other methods, the "C-SemiGlob" [19] ranked one step higher (see Table 2) for $T = 0.5$ benchmark and the "FastBilateral" [29] ranked 30 steps lower for $T = 0.5$ but nine steps higher for $T = 1$. The comparison is made for two typical data sets, Venus and Cones. Comparing to the higher $T = 1$ ranked "FastBilateral" method, one observes how the significant improvement in our results for subpixel accuracy are reduced to moderate and negligible differences for $T = 1, 2$. Eventually, our method obtains 98.5 percent and 93.8 percent of pixels below 0.5 pixel disparity error in the Venus and Cones sets, respectively. The results in Fig. 9 also

TABLE 2
Table from Middlebury Stereo Evaluation [34] Using the Minimum Available Disparity Error Threshold $T = 0.5$ Pixel

Algorithm	S	Tsukuba			Venus			Teddy			Cones			AB
		nonoc	all	disc	nonoc	all	disc	nonoc	all	disc	nonoc	all	disc	
SubPixDoubleBP [49]	5.9	8.78	9.45	14.9	0.72	1.12	5.24	10.1	16.4	**21.3**	8.49	14.7	16.5	10.7
Undr+OverSeg [5]	7.5	**4.84**	**5.22**	9.39	2.00	2.15	5.44	12.9	17.9	27.6	9.16	14.6	17.6	10.7
C-SemiGlob [19]	8.8	13.9	14.7	18.9	3.30	3.82	10.9	**9.82**	17.4	22.8	5.37	11.7	12.8	12.1
Our method	9.2	7.18	8.56	20.1	1.46	2.12	7.87	12.9	19.4	27.5	6.22	12.6	15.8	11.8
ImproveSubPix [17]	9.3	8.96	9.66	16.2	4.62	5.41	16.9	11.0	17.8	24.1	**4.90**	10.8	**12.2**	11.9
Currently 50 more entries ...														

S = Score measured as average rank, AB = Average Percent of erroneous pixels. The scores with bold font present the best performance in the corresponding category.
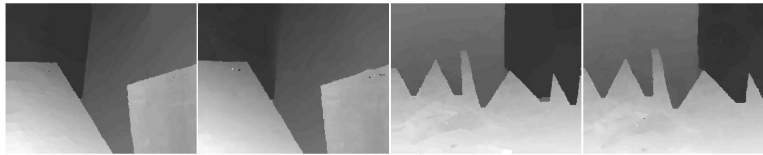

Fig. 7. Comparison between L^1 and L^2 regularizers in the MS framework. Left: L^1. Right: L^2.
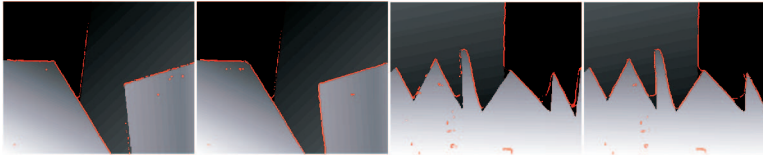

Fig. 8. Influence of the additional constraints on the discontinuity function. Left: Without the additional constraints. Right: With the additional constraints. The red dots indicate points where $v_l \leq 0.6$.

show that the minor factor of points having errors beyond 0.5-0.75 pixel are outliers mainly associated with errors over two pixels. It seems, therefore, that by an appropriate postprocessing (e.g., as performed in the "C-SemiGlob" method [19]), the factor of outlier points can be reduced, yielding an upswing in our method's rank for $T = 1$ benchmark table.

12 SUMMARY

In this paper, we have presented a variational approach for correspondence establishment in binocular stereo vision. Our algorithm is defined in a spatially continuous setting, providing an inherent subpixel evaluations.

The model involves two coupled energy functionals, one for the half-occlusions and the other for the disparity and the discontinuity map. The emerged coupled PDEs are solved by alternate minimization.

The data term in our objective function is based on the robust L^1 norm for brightness and brightness gradient constancy. However, for half-occluded points that are visible on only one image, this assumption is no longer valid. These points yield deviating disparity values (outliers) that influence the results in the neighboring visible area through the diffusion process. To this end, the occlusions are detected via separate but concurrent minimization process and then discarded from the disparity cost function.

The discontinuities are preserved in our model by using a Mumford-Shah framework with embedded TV regularizer.

TABLE 3
AADE in the Nonoccluded (Visible) and
the Discontinuity Regions

	Venus		Sawtooth	
	nonoccl	disc	nonoccl	disc
L^1 **without** additional constraints	0.135	0.420	0.227	0.579
L^2	0.156	0.453	0.234	0.629
L^1 **with** additional constraints	0.135	0.299	0.230	0.608

The standard MS versus Total Variation MS and the impact of the additional constraints on the performances.
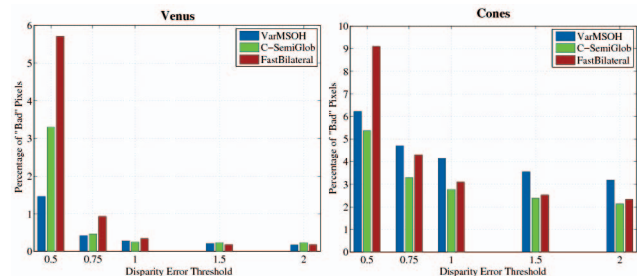

Fig. 9. Error histogram for Venus and Cones sets in the nonoccluded region.

The choice of the TV term in the MS regularizer instead of the standard L^2 stabilizer was justified empirically. Similar observations have been obtained in image segmentation [39] and deblurring [6].

In addition to an accurate subpixel dense disparity map, our method also recovers the half-occlusion domain and the discontinuity field v_l for the disparity map. The function v_l is often viewed as an auxiliary function in the MS framework. Nevertheless, we handle v_l as a feature that can be enforced to obey additional constraints raised from the stereo geometry. It has been shown that the proposed constraints improve the recovered discontinuity function specially at the adjoining visible-occluded domains.

Experimental tests show that our approach can compete with the currently best stereo methods, showing highly accurate subpixel evaluations *without/before* any pre/post-processing procedure. The wide range of scenarios tested here indicates the robustness of our method and the validity of our automatic parameter setting.

REFERENCES

[1] R. Alicandro, A. Braides, and J. Shah, "Free-Discontinuity Problems via Functionals Involving l^1-Norm of the Gradient and Their Approximation," *Interface and Free Boundaries,* vol. 1, pp. 17-37, 1999.

[2] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert, "Dense Disparity Map Estimation Respecting Image Discontinuities: A PDE and Scale-Space Based Approach," *J. Visual Comm. and Image Representation,* vol. 13, pp. 3-21, 2002.

[3] L. Ambrosio and V.M. Tortorelli, "Approximation of Functionals Depending on Jumps by Elliptic Functionals via Γ-Convergence," *Comm. Pure and Applied Math.,* vol. 43, pp. 999-1036, 1990.

[4] T. Amiaz and N. Kiryati, "Dense Discontinuous Optical Flow via Contour Based Segmentation," *Proc. IEEE Int'l Conf. Image Processing,* vol. 3, pp. 1264-1267, 2005.

[5] Anonymous, "Stereo Matching Based on Under and Over-Segmentation with Occlusion Handling," http://vision.middlebury.edu/stereo/eval/, 2010.

[6] L. Bar, A. Brook, N. Sochen, and N. Kiryati, "Deblurring of Color Images Corrupted by Impulsive Noise," *IEEE Trans. Image Processing,* vol. 16, no. 4, pp. 1101-1111, Apr. 2007.

[7] R. Ben-Ari and N. Sochen, "Variational Stereo Vision with Sharp Discontinuities and Occlusion Handling," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 1-7, 2007.

[8] M. Bleyer and M. Gelautz, "A Layered Stereo Algorithm Using Image Segmentation and Global Visibility Constraints," *Proc. IEEE Int'l Conf. Image Processing,* vol. 5, pp. 2997-3000, 2004.

[9] M.Z. Brown, D. Burschka, and G.D. Hager, "Advances in Computational Stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 8, pp. 993-1008, Aug. 2003.

[10] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," *Proc. Eighth European Conf. Computer Vision,* pp. 25-36, 2004.

[11] T. Chan and L. Vese, "Active Contours without Edges," *IEEE Trans. Image Processing,* vol. 10, no. 2, pp. 266-277, Feb. 2001.

[12] G. Dal Maso, "An Introduction to Γ-Convergence," *Progress in Nonlinear Differential Equations and Their Applications,* Birkhauser, 1993.

[13] U.R. Dhond and J.K. Aggarwal, "Stereo Matching in Presence of Narrow Occluding Objects Using Dynamic Disparity Search," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 7, pp. 719-724, July 1995.

[14] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions.* CRC Press, 1992.

[15] L. Di Stefano, F. Tombari, S. Mattoccia, "Segmentation-Based Adaptive Support for Accurate Stereo Correspondence," *Proc. IEEE Pacific-Rim Symp. Image and Video Technology,* 2007.

[16] A. Fusiello, V. Roberto, and E. Trucco, "Efficient Stereo with Multiple Windowing," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 858-863, 1997.

[17] S. Gehrig and U. Franke, "Improving Sub-Pixel Accuracy for Long Range Stereo," *Proc. IEEE Int'l Conf. Computer Vision VRML Workshop,* 2007.

[18] D. Geiger, B. Landendorf, and A.L. Yuille, "Occlusions and Binocular Stereo," *Int'l J. Computer Vision,* vol. 14, no. 3, pp. 211-226, 1995.

[19] H. Hirschmüller, "Stereo Processing by Semi-Global Matching and Mutual Information," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 2, pp. 328-341, Feb. 2008.

[20] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiments," *IEEE Trans. Pattern Recognition and Machine Intelligence,* vol. 16, no. 9, pp. 920-932, Sept. 1994.

[21] H. Kim and K. Sohn, "Hierarchical Disparity Estimation with Energy Based Regularization," *Proc. 10th IEEE Int'l Conf. Image Processing,* vol. 1, pp. 373-376, 2003.

[22] A. Klaus, M. Sormann, and K. Karner, "Segment-Based Stereo Matching Using Belief Propagation and Self-Adapting Dissimilarity Measure," *Proc. Int'l Conf. Pattern Recognition,* vol. 3, pp. 15-18, 2006.

[23] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers, "An Experimental Comparison of Discrete and Continuous Shape Optimization Methods," *Proc. European Conf. Computer Vision,* 2008.

[24] K. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions Using Graph Cuts," *Proc. IEEE Int'l Conf. Computer Vision,* vol. 2, pp. 508-515, 2001.

[25] A. Kumar, C.V.S. Haker, S. Zucker, and A. Tannenbaum, "Stereo Disparity and l^1 Minimization," *Proc. 36th IEEE Conf. Decision and Control,* vol. 2, pp. 1125-1129, 1997.

[26] C. Lei, J. Selzer, and Y.H. Yang, "Region Tree Based Stereo Using Dynamic Programming Optimization," *Proc. IEEE Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 2378-2385, 2006.

[27] A.R. Mansouri, A. Mitiche, and J. Konard, "Selective Image Diffusion: Application to Disparity Estimation," *Proc. IEEE Int'l Conf. Image Processing,* vol. 3, pp. 114-118, 1998.

[28] D. Marr and T.A. Poggio, "Cooperative Computation of Stereo Disparity," *Science,* vol. 194, no. 4262, pp. 283-287, 1976.

[29] S. Mattoccia, S. Giardino, and A. Gambini, "Accurate and Efficient Cost Aggregation Strategy for Stereo Correspondence Based on Approximated Joint Bilateral Filtering," *Proc. Asian Conf. Computer Vision,* 2009.

[30] W. Miled and J.C. Pesquet, "Disparity Map Estimation Using a Total Variation Bound," *Proc. Third Canadian Conf. Computer and Robot Vision.*

[31] D. Mumford and J. Shah, "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Comm. Pure and Applied Math.,* vol. 42, pp. 577-684, 1989.

[32] S. Osher and J.A. Sethian, "Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations," *J. Computational Physics,* vol. 79, pp. 577-684, 1988.

[33] Middlebury Optical Flow Benchmark Page. http://vision.middlebury.edu/flow/, 2010.

[34] Middlebury Stereo Benchmark Page. http://vision.middlebury.edu/stereo/, 2010.

[35] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, "A Convex Formulation of Continuous Multi-Label Problems," *Proc. European Conf. Computer Vision,* 2008.

[36] L. Robert and R. Deriche, "Dense Depth Map Reconstruction: A Minimization and Regularization Approach Which Preserves Discontinuities," *Proc. European Conf. Computer Vision,* pp. 439-451, 1996.

[37] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision,* vol. 47, no. 1, pp. 7-42, 2002.

[38] C. Schmid and A. Zisserman, "The Geometry and Matching of Curves in Multiple Views," *Proc. European Conf. Computer Vision,* pp. 104-118, 1998.

[39] J. Shah, "Segmentation By Nonlinear Diffusion," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 202-207, 1991.

[40] J. Shah, "A Nonlinear Diffusion Model for Discontinuous Disparity and Half-Occlusions in Stereo," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 34-40, 1993.

[41] N. Slesareva, A. Bruhn, and J. Weickert, "Optical Flow Goes
      Stereo: A Variational Method for Estimating Discontinuity-
      Preserving Dense Disparity Maps," *Proc. 27th DAGM Symp.
      Pattern Recognition,* W. Kropatsch, R. Sablatnig, and A. Hanbury,
      eds., pp. 33-40, 2005.
[42] J. Sun, Y. Li, S.B. Kang, and H.-Y. Shum, "Symmetric Stereo
      Matching for Occlusion Handling," *Proc. IEEE Int'l Conf. Computer
      Vision and Pattern Recognition,* pp. 399-406, 2005.
[43] R. Kimmel, T. Nir, and A.M. Bruckstein, "Over-Parametrized
      Variational Optical Flow," *Int'l J. Computer Vision,* vol. 76, no. 2,
      pp. 205-216, 2008.
[44] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems.*
      Winston and Sons, 1977.
[45] V. Venkateswar and R. Chellapa, "Hierarchical Stereo and Motion
      Correspondence Using Feature Grouping," *Int'l J. Computer Vision,*
      vol. 15, pp. 245-269, 1995.
[46] L.A. Vese and T.F. Chan, "A Multiphase Level Set Framework for
      Image Segmentation Using the Mumford and Shah Model," *Int'l J.
      Computer Vision,* vol. 50, no. 3, pp. 271-293, 2002.
[47] J. Weickert, *Anisotropic Diffusion in Image Processing.* Teubner
      Stuttgart, 1998.
[48] Q. Yáng, R.Y.L. Wang, H. Stewénius, and D. Nistér, "Stereo
      Matching with Color-Weighted Correlation, Hierarchial Belief
      Propagation and Occlusion Handling," *Proc. IEEE Conf. Computer
      Vision and Pattern Recognition,* vol. 2, pp. 2347-2354, 2006.
[49] Q. Yáng, R. Yang, J. Davis, and D. Nistér, "Spatial-Depth Super
      Resolution for Range Images," *Proc. IEEE Conf. Computer Vision
      and Pattern Recognition,* 2007.
[50] K. Yoon and I. Kweon, "Locally Adaptive Support-Weight
      Approach for Visual Correspondence Search," *Proc. IEEE Conf.
      Computer Vision and Pattern Recognition,* vol. 2, pp. 924-931, 2005.

Rami Ben-Ari received the BSC and MSc degrees from the Israel Institute of Technology-Technion in 1990 and 1993, respectively, and the PhD degree in applied mathematics from Tel-Aviv University in 2008. During 2009, he was an associate researcher at the Israel Institute of Technology-Technion and Ben-Gurion University in Israel. Currently, he is a senior algorithm researcher at Orbotech Ltd., the world leader in hi-tech inspection and imaging, developing new vision technologies for future AOI machines. His research interests include variational methods in computer vision, 3D shape estimation, optical flow, statistical filtering, and visual tracking. He is a member of the IEEE and the IEEE Computer Society.

Nir Sochen received the BSc degree in physics and the MSc degree in theoretical physics from the University of Tel-Aviv in 1986 and 1988, respectively. He received the PhD degree in theoretical physics in 1992 from the Université de Paris-Sud, France. He continued with a one year research at the Ecole Normale Superieure in Paris on a Haute Etude Scientifique fellowship, and a three-year US National Science Foundation fellowship in the Physics Department of the University of California, Berkeley. It is at Berkeley that his interests shifted from quantum field theories and integrable models related to high-energy physics and string theory to computer vision and image processing. He spent one year in the Physics Department at the University of Tel-Aviv and two years on the Faculty of Electrical Engineering in the Technion-Israel Institute of Technology. Currently, he is an associate professor in the Department of Applied Mathematics, University of Tel-Aviv. He is also a member of the Ollendorf Center at the Technion. His research interests include the applications of differential geometry and statistical physics in image processing and computational vision. He is a member of the IEEE and the IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.