Variational Stereo Vision with Sharp Discontinuities and Occlusion Handling

Rami Ben-Ari  
Nir Sochen  
Tel Aviv University - Department of Applied Mathematics  
Ramat-Aviv, Tel-Aviv 69978, Israel  
{ramiben,sochen}@post.tau.ac.il

Abstract

This paper addresses the problem of correspondence establishment in binocular stereo vision. We suggest a novel variational approach that considers both the discontinuities and occlusions. It deals with color images as well as gray levels. The proposed method divides the image domain into the visible and occluded regions where each region is handled differently. The depth discontinuities in the visible domain are preserved by use of the total variation term in conjunction with the Mumford-Shah framework. In addition to the dense disparity and the occlusion maps, our method also provides a discontinuity function revealing the location of the boundaries in the disparity map. We evaluate our method on data sets from Middlebury site showing superior performance in comparison to the state of the art variational technique.

1. Introduction

We consider the binocular stereo vision where a pair of images are captured by two cameras in different locations. The central task in stereo vision is the establishment of correspondence. The solution of this problem is expressed by the disparity, that reveals the difference in location of corresponding pixels, along epipolar lines. In this paper, we consider two rectified images or alternatively assume that the images are captured by identical coplanar cameras.

For review over various stereo techniques and recent advances in computational stereo the reader is referred to [4]. 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 at each pixel [8, 24]. Global approaches incorporate explicit smoothness assumptions and determine the disparity map by energy minimization techniques, such as graph cuts [11], belief propagation [10], dynamic programming [13] or variational framework [1, 9, 14, 17, 21, 22].

Variational methods in stereo vision has attracted much interest in recent years. In contrary to the common techniques which produce integer valued disparity map (e.g. [10, 11, 13]), sub-pixel evaluation is an inherent part of the variational methods. Moreover, the variational framework offers the advantage of well based mathematical formalism. The objective functional consists of a data-fidelity term and a smoothing prior (also called the regularization term). Commonly gray value images are considered in the data term [1, 14, 17, 22]. The method in this paper accommodates a data term for color images as well as gray levels.

During the years, various regularization terms were suggested considering the discontinuities in the disparity map. These terms can generally be divided into two classes. Image driven regularizers were presented in [1, 14], proposing an anisotropic method that smoothes along the intensity gradients but not across them. Kim and Sohn [9] presented a similar approach using an inhomogeneous term. The image driven regularizers are based on the assumption that an intensity edge in an image indicates a depth discontinuity in the scene. Obviously, this hypothesis is violated in textured images.

The second class of regularizers include the disparity driven terms. Suggestion for a discontinuity preserving smoothness constraint goes back to [21] where a nonlinear diffusion model based on Mumford-Shah framework was presented. Robert et al. [17] proposed a framework for several other disparity driven regularizers. A well known regularization type in this class is the total variation (TV) [12]. Recently, Slesareva et al. [22] used this regularizer in a method based on [5] presenting the state of the results in the class of variational methods. However, TV regularizer does not handle discontinuities adequately. It “penalizes” each discontinuity proportionally to the height of the “jump”. An ideal regularizer should not penalize large jumps (probable discontinuities), definitely not more than small ones (probable noise). As a consequence, in methods based on the TV regularizer, disparity/depth discontinuities are not sharp. Seeking for piecewise smooth modeling of the disparity map we propose a novel approach based on the Mumford-Shah (MS) framework [15].

Recently published methods in variational stereo vision
[1, 12, 14, 17, 22] do not handle the half-occlusions (also briefly referred as occlusions) in the stereo problem. Since the disparity at half-occluded points is not defined, these methods often yield over-smoothing and artifacts at the occluded regions, the adjoining visible-occlusion discontinuities and also the nearby visible domain. In the method proposed here the occlusion map is evaluated and utilized for improving disparity estimations. Production of the occlusion map is also useful in labeling these regions where the disparities are unreliable and should be estimated differently.

Three kinds of constraints are typically used for occlusion detection, ordering, uniqueness and stereo match consistency [4]. The ordering constraint is important in scenes containing narrow occlusions (e.g., an object behind a fence). The uniqueness requirement appropriate for an integer valued method is not satisfied in scenes containing horizontally slanted surfaces [23]. The strategy adopted in this paper is based on the latter robust approach of stereo match consistency [4].

Although a variational approach for stereo vision including MS type regularization and occlusion handling was presented by Shah in [21], the method suggested here has not much in common with this work. The main differences are: 1) Instead of the quadratic data term in [21] the robust term is employed here. 2) We employ a MS smoothing term incorporating the TV regularizer in the sense of [20] instead of the original type based on the Tikhonov term, in [21]. 3) The occlusion detection in [21] is based on the monotonicty assumption, a variant of the ordering constraint while in our method the robust stereo match consistency is employed. 4) The method in [21] 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.

Stereo research has recently entered a new era with public access to real data sets with accurate ground truth disparities [18, 19]. We evaluate the performance of our method on three color sets available in [19]. Experimental results demonstrate the superior performance of our method in comparison to a color version of the state of the art variational method [22].

Often object boundaries in the scene are indicated by disparity discontinuities. In addition to the improved disparity map and a reliable occlusion field we also provide a discontinuity function presenting the locations of the discontinuities in the depth/disparity map.

The paper is organized as follows: In Section 2 we introduce the variational model considering both occlusions and discontinuities. Section 3 then deals with the minimization of the defined energy functionals. The initialization method is presented in section 4 followed by the description of the suggested algorithm in section 5. Section 6 presents experimental results and comparison to the state of the art variational method. Finally, the paper is concluded in section 7 by a brief summary and discussion.

2. The Model

In this paper we evaluate the disparity, the occlusion and the discontinuity maps for both views. For sake of abbreviation we present the model just for the left image. The model for the counterpart, right image, is obtained similarly.

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 \( x = (x, y) \in \Omega_l \) we describe the location of the corresponding point in the domain of the right image by \( g_j = (x - d_l(x), y) \in \Omega_r \) where \( d_l : \Omega_l \rightarrow \mathbb{R} \) denotes the left disparity.

2.1. Occlusion Domain

We use the superscripts \( v \) and \( o \) in our notation to indicate attribution to the visible (non-occluded) or occluded domain respectively.

The image domain in a stereo pair can be divided into two disjoint sets \( \Omega_l = \Omega^v_l \cup \Omega^o_l \)

where \( \Omega^v_l \) denotes the visible and \( \Omega^o_l \) the occluded domain in \( \Omega_l \). Under the assumption that a visible point satisfies stereo match consistency, we define an \textit{inconsistency} measure for every \( x \in \Omega_l \) by

\[
\epsilon_{occl}^l = -\ln \left( \epsilon_{occl} + (1 - \epsilon_{occl}) e^{-|u_l|} \right)
\]

where \( u_l(x) := |d_l(x)| - |d_r(g_l)| \), and \( d_r(g_l) \) denotes the disparity at the point \( g_l \) in the right image. The cost function \( \epsilon^l_{occl} \) presents a truncated \( L^1 \) penalization with \( \epsilon_{occl} \) controlling the saturation level. In Our method we assign \( \epsilon_{occl} = 0.01 \) thus setting the saturation level to \( \sim 4.6 \).

The occlusion domain is obtained by a two phase segmentation of the field \( \epsilon_{occl}^l \). We use the level set framework for segmentation [16]. To this end, a level set function \( \phi_l : \Omega_l \rightarrow \mathbb{R} \) is defined. Similarly to [6], the \textit{Heaviside} function of \( \phi_l \)

\[
H(\phi_l) = \begin{cases} 1 & \phi_l \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]

is used to indicate the visible and occluded regions in the image corresponding to the non-negative and negative levels in \( \phi_l \) respectively. The domain segmentation is obtained by minimization of

\[
E^l_{occl}(\phi_l) = \int_{\Omega_l} (\epsilon^l_{occl} * G_\sigma) H(\phi_l) + t(1 - H(\phi_l)) + \nu |\nabla H(\phi_l)| \, dA.
\]


where $t$ is a threshold and $\nabla := \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T$. The symbol $\ast$ denotes the standard convolution operator and $G_\sigma$ is a Gaussian with standard deviation of $\sigma = 0.1$. This low pass filter enforces the smoothness of the occlusion domain and is helpful in removing some isolate points or small holes from the occlusion region. The term $|\nabla H(\phi_l)|$ denotes the length of the contour [7] separating between the occluded and the visible domain and $\nu$ is a real positive weight parameter. Practically, a smooth approximation of the Heaviside function $H_\epsilon$, rather than a step function is used [6]

$$H_\epsilon(\phi) = \frac{1}{2} \left( 1 + \frac{2}{\pi} \arctan \left( \frac{\phi}{\epsilon} \right) \right)$$

$$\delta_\epsilon(\phi) = \frac{dH_\epsilon(\phi)}{d\phi} = \frac{1}{\pi \epsilon^2 + \phi^2}$$

In our application we set $\epsilon = 0.1$ and $\nu = 0.01$.

### 2.2. Mumford-Shah Formulation

In order to handle differently the disparities in the visible and the occluded domains we introduce two disparity functions defined on the whole domain, i.e. $d^v_l, d^o_l : \Omega \to \mathbb{R}$. Given an occlusion map, the disparity on $\Omega_l$ is then calculated by

$$d_l = d^v_l H(\phi_l) + d^o_l (1 - H(\phi_l)).$$

(5)

We seek to model the disparity map in the visible domain by a piece of a piecewise smooth segments separated by simple contours. Using the MS formulation [15] in the Ambrosio and Tortorelli’s $\Gamma$-convergence framework [2], we consider first the pair $d^v_l, v_l$ as the minimizer of the following functional:

$$E_L^v (d^v_l, v_l) = \int_{\Omega_l} e_d + e_av_l^2 + \tilde{\alpha} (v_l - 1)^2 + \tilde{\epsilon} |\nabla v_l|^2 \, dA$$

(6)

where $e_d$ denotes the data term and $e_a$ is the embedded smoothing prior in our MS formulation. The parameters $\tilde{\alpha}, \tilde{\epsilon}$ are related to the standard MS parameters in $\Gamma$-convergence formulation. The smooth function $v_l(\mathbf{x})$ presents a discontinuity map for the disparity function. The discontinuity set is approximated in $\Omega_l$ by $v_l \approx 0$, while the constant disparity regions are indicated by $v_l \approx 1$. Note that the integration is solely over the visible domain.

For the data term in (6) we employ the $L^1$ cost function for constancy in both the brightness and the brightness gradients as suggested in [5, 22]

$$e_d(s^2_d) = \sqrt{s^2_d + 0.001^2}$$

(7)

where

$$s^2_d := \|I_\epsilon(\mathbf{g}) - I_\epsilon(\mathbf{x})\|^2 + \lambda \|\nabla I_\epsilon(\mathbf{g}) - \nabla I_\epsilon(\mathbf{x})\|^2.$$
The corresponding EL equations for (10) are given by
\[
\frac{\delta E_L}{\delta d_l^2} = \left( e_d \frac{\partial s_d^2}{\partial d_l^2} - 2\text{div}\left(e_v v_l^2 \nabla d_l^0\right)\right) H(\phi_l) = 0, \quad (12)
\]
\[
\frac{\delta E_L}{\delta v_l} = (e_v v_l + \tilde{\alpha} (v_l - 1)) H(\phi_l)
+ (v_l - \tilde{v}_r) (1 - H(\phi_l)) \tilde{H} (\tilde{\phi}_r)
+ (v_l - 1) (1 - H(\phi_l)) \left(1 - H(\tilde{\phi}_r)\right)
- \tilde{\alpha} \text{div}(\nabla v_l) = 0, \quad (13)
\]
where
\[
\frac{\partial s_d^2}{\partial d_l^2} = -2 \sum_i I^i_x (g_l) (I^i_y (g_l) - I^i (x))
- 2\lambda \sum_i (I^i_x (g_l) - I^i_x (x)) I^i_{xxx} (g_l)
+ (I^i_{xy} (g_l) - I^i_{y} (x)) \ell^i_{xy} (g_l)
\]
in (12). The subscripts x, y indicate partial derivatives and i presents the color channel in RGB space. Minimization with respect to both \(d_l^0\) and \(v_l\) is carried out subjected to Neumann boundary condition.

In order to fill the occlusion gaps we enforce smoothness in these regions by
\[
\text{div} \left(e_v' \left(\|\nabla d_l^0\|\right) v_l^2 \nabla d_l^0\right) = 0 \quad (14)
\]
with initialization \(d_l^0 = d_l^f\). Note that by performing (14) and then (5), diffusion from the visible domain into the occluded region is allowed but not vice versa. This procedure causes the regions that are falsely labeled as occluded to be transferred to the visible domain.

We apply alternate minimization (AM) to solve the coupled system of PDEs in (11-13). Therefore, in each step of the iterative procedure we minimize with respect to one function keeping the others fixed.

The EL equation (13) presents a linear PDE with respect to \(v_l\), for a given \(d_l^0\) and \(\phi_l\), and can be solved directly by the standard Jacobi, Gauss-Seidel or SOR iterative methods. On the contrary, (11) and (12) are non-linear PDEs. We employ the gradient descend method for obtaining the solution of (11). For obtaining the solution of (12) a nested fixed point iterations similar to [5, 22] is used, reducing the chance for being trapped in an irrelevant local minima. All spatial derivatives are approximated by central differences. The image warping is performed by linear interpolations. Since floating point disparities are considered here, the functions \(\tilde{v}_r, \tilde{\phi}_r\) are obtained on a non-uniformly spaced coordinates. The corresponding values on a uniform grid points are calculated by a surface fitting based on linear interpolation.

4. Initialization

Object boundaries are typically indicated in the image by intensity edges. Thus we often expect a discontinuity in the disparity (or equivalently in the depth) to coincide with an intensity edge. In order to start with a disparity map having a reliable approximation of the boundaries, we first calculate the image edge map in terms of the discontinuity function in MS framework. To this end, we consider minimization of the following energy functional with respect to the image discontinuity function \(v_l\) [3]:
\[
E_I (I_l, v_l) = \int_{\Omega_l} \|I_l - I_0\|^2 + \|\nabla I_l\| v_l^2
+ \alpha_l \left(\epsilon_i \|\nabla v_l\|^2 + \frac{(v_l - 1)^2}{4\epsilon_i}\right) \, dA
\]
where \(I_0\) is a given noisy image and \(\alpha_l, \epsilon_i\) are constants [2]. The associated EL is then given by:
\[
\frac{\delta E_I}{\delta v_l} = \|\nabla I_l\| v_l + \alpha_l \left(\epsilon_i \|\nabla v_l\|^2 + \frac{(v_l - 1)^2}{4\epsilon_i}\right) \, dA
\]
We use the resulting \(v_l\) as a variable regularization weight in our stereo formulation, enforcing alignment between disparity boundaries and image edges. The initial disparity map is then determined as the minimizer of
\[
E(d_l^0) = \int_{\Omega_l} e_d (d_l^0) + \beta (\sqrt{\|\nabla d_l^0\|^2 + 0.001^2}) \, dA
\]
where \(\lambda = 0\) is chosen in (8) for simplicity. The energy functional in (17) is minimized similar to (10). We start the initialization with a simple \(4 \times 4\) SAD (Sum of Absolute Differences) window correlation method.

5. Algorithm

Solving the system of PDE’s (11-13) in alternation leads to the following iterative algorithm:
1. Initialization: (see section 4).
2. Perform domain segmentation for the left and right image using Eq. (11) and its counterpart.
3. Compute the discontinuity maps \(v_l\) and \(v_r\) using Eq. (13) and the corresponding counterpart.
4. Solve (12) and its counterpart for the unknown \(d_l^0\) and \(d_r^e\) using the current values of \(v_l, v_r\) and \(\phi_l, \phi_r\).
5. Fill the disparity gaps in the occluded domain by Eq. (14).
6. Update the values for \(d_l\) and \(d_r\) by Eq. (5) and its counterpart.
7. Repeat steps 2-6 until convergence.
6. Evaluation

The results for three color data sets from [19] are shown, the Venus, Sawtooth and Cones.

6.1. Parameter Setting

The proposed method includes parameter setting from three main categories. 1) Parameters related to image segmentation, \( \alpha_i, \epsilon_i \). 2) Constants regarding the disparity gradients at the discontinuities, \( \tilde{\alpha}, \tilde{\epsilon} \). 3) The data internal weight \( \lambda \) and the data-smoothness weights \( \beta, \tilde{\beta} \). In order to reduce the number of control parameters we first set \( \epsilon_i = 0.1 \) and perform the following automatic adjustment for \( \tilde{\alpha} \) and \( \tilde{\epsilon} \):

\[
\tilde{\alpha} = 0.75\beta, \quad \tilde{\epsilon} = 0.2\beta \tag{18}
\]

where the factors in the linear relations are empirically found. We further fix the internal data weight \( \lambda = 2 \) (also in the comparing method of [22]) and set \( \beta = 0.2 \), for all the test cases. Next we change the stereo consistency threshold \( t \) by an exponentially decay function from the initial value 5 (where all the domain is determined as visible) to the final value of 1. Consequently, the occlusion region is grown gradually and false labeling of the visible-occlusion boundaries as occluded is avoided. The images are smoothed as in [22] with \( \sigma = 0.8 \) Gaussian. After these adjustments only one extra tuning parameter \( \epsilon_i \) is remained in comparison to [22], which is set according to the image’s contrast. The values of the tuning parameters for our test-bed case are listed in Table 1. The tuning parameters were optimized empirically for both of the comparing methods achieving minimum disparity error in the visible domain.

6.2. Experimental Results

Having the occlusion field evaluated it is used here in the visualization of our results. We employ the simple occlusion filling heuristic in [11] where the occluded pixels are assigned by the values from the nearest background disparity, lying on the same scan-line (at the image borders the direction of the filling is reversed). Figure 1 shows the recovered disparity maps comparing to the color version of the method [22]. Obviously, our piecewise smooth model yields improved results. In our method the disparity boundaries are sharp and located rather accurately. Generally, restoration of the discontinuities lying on the boundaries of occluded regions is harsh. However, the proposed method deals also with these discontinuities adequately. Evidently, the over-smoothing and artifacts yielded in the comparing method are significantly decreased in our approach.

We also assess the performance of our method quantitatively. Our error measures are: the percentage of “bad” matchings (where the absolute disparity error is greater than 1 as in [18]) and the average absolute disparity error (AADE) used in [1, 22]. Since in the occluded region the disparities are inestimable, we exclude the these pixels from our statistics. For each image pair, the error measures in the visible domain and the discontinuity regions are calculated based on the masks provided in [19]. In all the measures a border of 10 pixels is ignored. The results summarized in Table 2 clearly show the improvement in the accuracy by our model. Note that in both of the compared methods the same term is used for data fidelity.

Another advantage of our method is evaluation of the occlusion field and production of a discontinuity map. Figure 2 shows the evaluated occlusion map with the estimated discontinuities. The received occlusion field is realistic and close to ground truth even in the challenging case of the Cones set. Note the high accuracy in localization of the discontinuities in the visible domain.

7. Summary and Discussion

The occlusions are an important part of the stereo problem and can often occupy relatively large portions of the image domain. In our novel approach we neutralize the “deceiving” data-fidelity constraint in the occluded region since there are no matches for the pixels in this area. Moreover, the proposed method employs a Mumford-Shah regularization term for accurate recovery of the disparity/depth boundaries. Additional constraints on the discontinuity map ensure rejection of outliers and rather accurate restoration of the disparity boundaries between the occluded and the visible regions.

Although there are a few parameters involved in our method, most of them are fixed once and kept unchanged for all the test cases. empirically examining the results shows low sensitivity to the parameters. Furthermore, the highly different characteristics of the data sets show the robustness of our fix setting.

Experimental results show qualitative and quantitative improvements in comparison to the state of the art variational method [22], extended here to color space. Our method provides in addition to a dense disparity map also reliable estimations for the occlusion field and the location of the disparity discontinuities.
Figure 1. Disparity maps. Top: Venus, Middle: Sawtooth, Bottom: Cones. The black areas in the Cones ground truth map indicate unknown disparities.

<table>
<thead>
<tr>
<th>Visible region</th>
<th>Venus</th>
<th>Sawtooth</th>
<th>Cones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Version of [22]</td>
<td>bad</td>
<td>2.77</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td>AADE</td>
<td>3.55</td>
<td>0.304</td>
</tr>
<tr>
<td>Our method</td>
<td>0.76</td>
<td>0.176</td>
<td>1.09</td>
</tr>
<tr>
<td>Improvement</td>
<td>72%</td>
<td>20%</td>
<td>69%</td>
</tr>
<tr>
<td>Discontinuity region</td>
<td>bad</td>
<td>29.0</td>
<td>1.066</td>
</tr>
<tr>
<td>Color Version of [22]</td>
<td>bad</td>
<td>19.5</td>
<td>1.010</td>
</tr>
<tr>
<td>Our method</td>
<td>10.5</td>
<td>0.480</td>
<td>8.32</td>
</tr>
<tr>
<td>Improvement</td>
<td>64%</td>
<td>55%</td>
<td>35%</td>
</tr>
</tbody>
</table>

Table 2. The disparity errors in the visible (non-occluded) and the discontinuity regions.

References


Figure 2. Occlusion and discontinuity maps. Left: Ground truth occlusion map. Middle: Our evaluation. Right: The evaluated discontinuities mapped on the ground truth disparity. The red dots indicate points where $v_{l} \leq 0.5$.


