Symmetric Stereo Matching for Occlusion Handling.

3 authors, including:

Yin Li
Sun Yat-Sen University
103 PUBLICATIONS   **2,415** CITATIONS

SEE PROFILE

Sing Bing Kang
Microsoft
225 PUBLICATIONS   **11,094** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Real-time rendering of realistic rain View project

Project Al matrix composites View project

Symmetric Stereo Matching for Occlusion Handling

Jian Sun[1]   Yin Li[1]   Sing Bing Kang[2]   Heung-Yeung Shum[1]

[1]Microsoft Research Asia   [2]Microsoft Research
Beijing, P.R. China   Redmond, WA, USA

Abstract

In this paper, we propose a symmetric stereo model to handle occlusion in dense two-frame stereo. Our occlusion reasoning is directly based on the visibility constraint that is more general than both ordering and uniqueness constraints used in previous work. The visibility constraint requires occlusion in one image and disparity in the other to be consistent. We embed the visibility constraint within an energy minimization framework, resulting in a symmetric stereo model that treats left and right images equally. An iterative optimization algorithm is used to approximate the minimum of the energy using belief propagation. Our stereo model can also incorporate segmentation as a soft constraint. Experimental results on the Middlebury stereo images show that our algorithm is state-of-the-art.

1 Introduction

Occlusion is one of the biggest challenges in stereo. For two-frame stereo, occluded pixels are only visible in one image. Accurate depth and occlusion information is important to applications in vision and robotics, such as tracking, recognition, and path planning. In certain applications such as 3D modeling and view interpolation, a stereo algorithm should not only estimate accurate depths at visible areas, but it should also provide reasonable guesses of depths at occluded areas.

1.1 Previous work

For a comprehensive discussion on dense two-frame stereo matching, we refer the reader to the survey by Scharstein and Szeliski [17]. In this paper, we review two-frame stereo algorithms that can handle occlusion. Two kinds of hard constraints are typically used: ordering constraint and uniqueness constraint.

The ordering constraint preserves order along scanlines in both input images. The monotonicity constraint [9] is a variant of the ordering constraint which requires neighboring pixels to be matched. On the other hand, the uniqueness constraint [15] enforces a one-to-one mapping between pixels in two images.

Most approaches [1, 6, 11, 3] that exploit the ordering constraint or uniqueness constraint use dynamic programming. Stereo matching is formulated as finding a minimum-cost path in the matrix of all pairwise matching costs between two corresponding scanlines. Dynamic programming can independently yield a global minimum for each scanline in a polynomial time. The "horizontal" and "vertical" discontinuities of the path correspond to the left and right occlusions, respectively.

The simplest method that uses the uniqueness constraint to detect occlusion is that of cross-checking [7]. In cooperative algorithm [23], the uniqueness constraint is enforced within a 3D array of match values in disparity space using an iterative update algorithm. In [12], stereo matching is formulated as finding a subset of assignments that may potentially correspond by graph cuts algorithm. A Potts energy to be minimized is defined on the assignments rather than pixels. The uniqueness constraint is always satisfied using graph cuts. To date, graph cuts consistently produces the best results (disparity and occlusion) for frontal-parallel scenes.

Unfortunately, the ordering and uniqueness constraints have limitations. The ordering constraint is not always true in general; it is violated in scenes that contain thin foreground objects or narrow holes. Figure 2 shows such a scene ("double nail illusion"). While point A is to the left of the thin foreground object in the left image, it is to the right of the same object in the right image. Furthermore, the ordering constraint can only be enforced for each scanline *independently*. Without the smoothness assumption between epipolar lines, the results of dynamic programming based approaches often show a streaking effect.

Using a discrete representation of disparity, the uniqueness constraint is not appropriate for scenes containing horizontally slanted surfaces (Figure 1(a)), which result in correspondence between unequal numbers of pixels. As observed in [16], a horizontally slanted surface will appear more horizontally stretched in one image compared with the other. Figure 1(b) is the output of the graph cuts [12] algorithm (www.cs.cornell.edu/people/vnk/software.html). Neither disparity nor occlusion can be recovered correctly using graph cut with a Potts model.
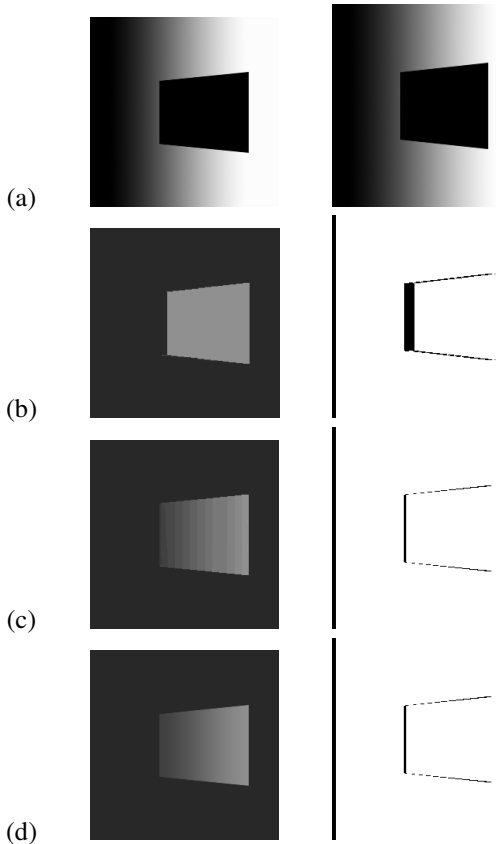
Figure 1: Horizontally slanted object. (a) Left and right images, (b) results using graph cuts [12], (c) our results, and (d) ground truth. In (b-d), the disparity map is shown in the left column while the occlusion map is shown in the right column. Here, graph cuts with the uniqueness constraint yields both incorrect disparity and occlusion.

1.2 Visibility constraint

In order to handle occlusions for more general scenes, we adhere to the basic visibility constraint. All this visibility constraint requires is that *an occluded pixel must have no match on the other image and a non-occluded pixel must have at least one match.*

The visibility constraint is self-evident because it is derived directly from the definition of occlusion. A pixel in the left image will be visible in both images if there is at least one pixel in the right image matching it. Unlike the uniqueness constraint, the visibility constraint permits many-to-one matching. Furthermore, the ordering constraint need not be satisfied. As a result, the visibility constraint is a more flexible but weaker constraint than the ordering and uniqueness constraints. The visibility constraint only enforces consistency between occlusion in one image and disparity in the other.

Applying the visibility constraint is nontrivial because both disparity and occlusion are unknown. In this paper, we propose to enforce the visibility constraint using an algo-

rithm that iteratively performs these two steps: (1) infer the disparity map in one view considering the occlusion map of the other view, and (2) infer the occlusion map in one view from the disparity map of the other view.

Step (1) improves the disparity map by enforcing piecewise smoothness while using the occlusion maps for both views as constraints. Step (2) improves the occlusion map using two observations illustrated in Figure 3.

The first observation is that if pixels on both sides of a discontinuous boundary are visible, their disparities tend to be unambiguous. (As an example, observe the discontinuity between regions C and E in Figure 3(b).) Hence, the occlusion region in the other view can be directly inferred using the visibility constraint (using the same example, region D as shown in Figure 3(c)).

The second observation is that disparities in occlusion regions are usually ambiguous (e.g., the hatched region in Figure 3(b)). This observation can be verified by inspecting Figure 6(a), especially the disparities in the occlusion regions. The key is that these pixels rarely affect the occlusion reasoning on the other view. For example, a pixel in region B (no matter what disparity it ends up with inside the hatched region) will not influence the occlusion reasoning, since regions A and C (where the pixel may land after warping) have already been matched.

We use these observations to enforce the visibility constraint and thus estimate disparity and occlusion regions more accurately. We now describe our symmetric stereo model that explicitly uses the visibility constraint in an energy minimization framework and an iterative algorithm that approximates the minimum.

2 Symmetric Stereo Model

Given a stereo image pair $I = \{I_L, I_R\}$, we want to compute the disparity $\{D_L, D_R\}$ and occlusion $\{O_L, O_R\}$ for left view I_L and right view I_R, respectively. For each pixel s, its disparity d_s is in the range $[d_{min}, d_{max}]$ and its oc-
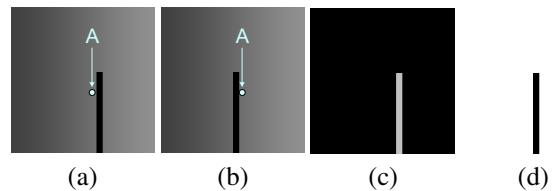

Figure 2: Double nail illusion. (a) and (b) are the left and right input images. Point A is to the left of the thin foreground object in the left image but is to the right of the same foreground object in the right image. Our stereo algorithm yields the disparity map (c) and occlusion map (d) associated with the left image. (The same maps for the right image are not shown here.)
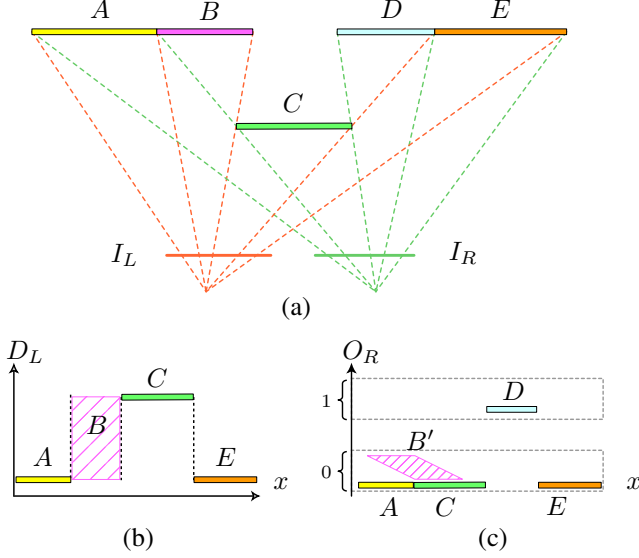
Figure 3: Occlusion reasoning. (a) A stereo image pair $\{I_L, I_R\}$ captures a scene with objects A to E. (b) The disparity map of the left image. The disparity for the occlusion region B is ambiguous (usually within a range specified by the hatched area). (c) The inferred occlusion map of the right image. The depth discontinuity between C and E in the left image causes the occlusion region D in the right image. In addition, region B warps to either A or C, which is already matched. Here, occlusion reasoning is independent of the disparity for region B.

clusion $o_s \in \{0, 1\}$ is a binary variable. In this paper, we denote $\widehat{D}$ and $\widehat{O}$ as disparity and occlusion in a single view, either left or right.

Most local stereo algorithms [17] and global stereo algorithms [4, 19] only compute the disparity and occlusion in the reference view (say, the left image). In global algorithm, stereo matching can be formulated in an energy minimization framework:

$$E(D_L; I) = E_d(D_L; I) + E_s(D_L), \qquad (1)$$

where the data term $E_d(D_L; I)$ measures how well the disparity D_L fits the given stereo pair I, and the smoothness term $E_s(D_L)$ encodes a smoothness assumption on disparity. The occlusion is either considered implicitly or treated as an outlier process. The disparity D_R of the right image is computed independently. Here, consistency between two views is *not* enforced.

The visibility constraint requires occlusion in one image and disparity in the other to be consistent. Hence we embed the visibility constraint within the visibility terms $E_v(O_L, D_R)$ and $E_v(O_R, D_L)$. This allows us to formulate the occlusion reasoning depicted in Figure 3. Now, we formulate the stereo matching problem with two novel visibility terms:

$$E(D_L, D_R, O_L, O_R; I) =$$
$$E_d(D_L, O_L; I) + E_d(D_R, O_R; I)$$
$$+E_s(D_L, O_L) + E_s(D_R, O_R)$$
$$+E_v(O_L, D_R) + E_v(O_R, D_L). \qquad (2)$$

We call it the "symmetric stereo model" because of the symmetric relationship between $\{D_L, O_L\}$ and $\{D_R, O_R\}$ in the formulation. The two input images are treated equally.

2.1 Data term

The data term only encodes the intensity consistency of pixel correspondences for hypothesized disparity and occlusion. For a single view, the data term $E_d(\widehat{D}, \widehat{O}; I)$ is defined as:

$$E_d(\widehat{D}, \widehat{O}; I) = \sum_{s \notin \widehat{O}} \rho_d(F(s, d_s, I)) + \sum_{s \in \widehat{O}} \eta_o, \qquad (3)$$

where $F(s, d_s, I)$ is the matching cost function of pixel s with disparity d_s given observation I. $\rho_d(x)$ is a truncated $L1$ norm function that is robust to noise or outliers:

$$\rho_d(x) = -\ln((1 - e_d)\exp(-|x|/\sigma_d) + e_d),$$

where parameters σ_d and e_d control the shape of robust function. As matching cost, we use the intensity difference between two gray pixels or the Euclidean distance between two color pixels, namely, $F(s, d_s, I) = ||I_L(x_s, y_s) - I_R(x_s + d_s, y_s)||_2$ (for the left view), where (x_s, y_s) is the image coordinates of pixel s. The matching cost is similarly defined for the right view. The cost η_o is the penalty for occlusion labeling. This term is necessary to prevent the whole scene from being labeled as occlusion.

2.2 Smoothness term

We encode piecewise smoothness on disparity $\widehat{D}$ (given occlusion $\widehat{O}$) in the smoothness term $E_s(\widehat{D}, \widehat{O})$. Let $N(s)$ be the neighbors of the pixel s and $C = \{s, t | s < t, t \in N(s)\}$ be the set of all adjacent pixel pairs. The smoothness term on disparity conditioned on the occlusion is

$$E_s(\widehat{D}, \widehat{O}) = \sum_{s, t \in C \setminus B} \rho_p(d_s, d_t), \qquad (4)$$

where $B = \{s, t | o_s \neq o_t, s, t \in C\}$ is the set of discontinuities at the boundaries between occluded and unoccluded pixels. This term only enforces smoothness within occluded and unoccluded regions. Our robust function is

$$\rho_p(d_s, d_t) = \min(\lambda|d_s - d_t|, T), \qquad (5)$$

where λ is the rate of increase in the cost and T controls the limit of the cost. There are two advantages of using this robust function. For one, it preserves discontinuity. The other advantage is that we can apply a very efficient implementation [8] of belief propagation to minimize the energy involving this smoothness term (described in Section 3).

2.3 Visibility term

In order to enforce the visibility consistency constraint, we define the visibility term $E_v(O_L, D_R)$ as

$$E_v(O_L, D_R) = \sum_s \beta_w |o_s - W_L(s; D_R)| + \sum_{s,t \in C} \beta_o |o_s - o_t|,$$

(6)

where $W_L(s; D_R) \in \{0, 1\}$ is a binary map defined on the left image. For each pixel s, its binary value in $W_L(s; D_R)$ indicates whether or not there exists one or more pixels matching s from the right view according to the disparity D_R. The value at pixel s is set to 1 if there is no pixel in the right view corresponding to pixel s. The binary map $W_L(s; D_R)$ can be computed by forward warping all the pixels in right view using disparity D_R. The parameter β_w controls the strength of the visibility constraint. The last term in (6) enforces the smoothness of the occlusion. It is a classical Ising prior that encourages spatial coherence and is helpful to remove some isolate pixels or small holes of the occlusion. The parameter β_o controls the strength of the smoothness. The binary map $W_R(s; D_L)$ and the visibility term $E(O_R, D_L)$ are defined in a similar way.

The combination of (3), (4), and (6) is our basic stereo model. We now describe our iterative optimization algorithm to minimize the energy (2).

3 Iterative Optimization using BP

Minimizing the energy of disparity and occlusion in (2) simultaneously is difficult because exact inference is intractable. Instead, we propose an iterative optimization algorithm to minimize the energy of disparity and occlusion iteratively, using belief propagation (BP).

3.1 Markov network and belief propagation

A Markov network is an undirected graph. For a graph with node $s = 1, ..., N$, the state of each node is denoted by x_s. Each hidden node is connected with an observed node y_s. The Gibbs energy of this pairwise Markov network is given by

$$E(X; Y) = \sum_s \phi_s(x_s, x_y) + \sum_{s,t,s<t} \psi_{s,t}(x_s, x_t), \quad (7)$$

where $\phi_s(\cdot)$ is the local evidence for node s and $\psi_{st}(\cdot)$ is the compatibility function between node s and t. Belief propagation is an approximate algorithm to minimize the Gibbs energy (7). Recent empirical results on stereo matching [19, 21] show that BP often gives a very good approximation.

3.2 Iterative optimization

The optimization process has two steps: 1) estimate occlusion given disparity, and 2) estimate disparity given occlusion.

3.2.1 Estimate occlusion given disparity

Given the current estimated disparity $\{D_L, D_R\}$, the energy (2) can be written as the sum of two functions with respect to O_L and O_R: $E(D_L, D_R, O_L, O_R; I) = E_{O_L} + E_{O_R}$ where

$$E_{O_L} = E_d(D_L, O_L; I) + E_s(D_L, O_L) + E_v(O_L, D_R)$$
$$E_{O_R} = E_d(D_R, O_R; I) + E_s(D_R, O_R) + E_v(O_R, D_L).$$

Then, the occlusion $\{O_L, O_R\}$ is computed as follows:

$$O_L^* = \arg\min_{O_L} E_{O_L}, \quad O_R^* = \arg\min_{O_R} E_{O_R}. \quad (8)$$

Because occlusions are binary variables, $E_d(D_L, O_L; I)$ can be rewritten as:

$$E_d(D_L, O_L; I) = \sum_s ((1 - o_s)\rho_d(F(s, d_s, I)) + o_s \eta_o).$$

(9)

We ignore $E_s(D_L, O_L)$ because we cannot recover the disparity of the pixel in the occluded region. It could be worse to arbitrarily guess disparity in the same view. Furthermore, the difference in disparity between adjacent pixels is weak evidence for occlusion. Finally, by combining (6) and (9), we get:

$$E_{O_L} \approx \sum_s ((1 - o_s)\rho_d(F(s, d_s, I)) + o_s \eta_o) +$$
$$\sum_s \beta_w |o_s - W(s; D_R)| + \sum_{s,t \in C} \beta_o |o_s - o_t|. \quad (10)$$

Note that the first two terms on the right hand side of (10) can be viewed as an evidence $\phi_s^o(\cdot)$ and the last term can be viewed as a compatibility function $\psi_{st}^o(\cdot)$. So, (10) is the Gibbs energy of a Markov network that is defined by evidence $\phi_s^o(\cdot)$ and compatibility function $\psi_{st}^o(\cdot)$. As a result, we can apply the max-product version of belief propagation to approximately minimize (10).

3.2.2 Estimate disparity given occlusion

Given the current estimation of occlusion $\{O_L, O_R\}$, the energy (2) can be rewritten as the sum of two functions respect to D_L and D_R: $E(D_L, D_R, O_L, O_R; I) = E_{D_L} + E_{D_R}$ where

$$E_{D_L} = E_d(D_L, O_L; I) + E_s(D_L, O_L) + E_v(O_R, D_L)$$
$$E_{D_R} = E_d(D_R, O_R; I) + E_s(D_R, O_R) + E_v(O_L, D_R).$$

The disparity $\{D_L, D_R\}$ is estimated by minimizing E_{D_L} and E_{D_R}:

$$D_L^* = \arg\min_{D_L} E_{D_L}, \quad D_R^* = \arg\min_{D_R} E_{D_R} \quad (11)$$

The visibility term $E_v(O_R, D_L)$ encodes the visibility constraint. Actually, the visibility constraint imposes two kinds of constraints on the disparity D_L given the occlusion O_R:

First, for each pixel s in the left image, it should not match the occluded pixels in the right image. In other words, its disparity d_s should be restricted such that $O_R(x_s + d_s, y_s) = 0$. Second, for each non-occluded pixel in the right image, at least one pixel in the left image should match it. The restriction on d_s is a local constraint that is easy to encode. However, the second constraint is a global on disparities of all pixels in the left image, which is implicitly enforced in matching process. Therefore, in this step, we approximate the visibility term $E_v(O_R, D_L)$ by considering only the local constraint as follows:

$$E_v(O_R, D_L) \approx \sum_s \beta_w O_R(x_s + d_s, y_s), \quad (12)$$

where $O_R(x_s + d_s, y_s)$ indicates whether or not the corresponding pixel of s is occluded, given disparity d_s. Hence, for the left view, combining (3), (4), and (12), we get

$$E_{D_L} \approx \sum_{s \notin O} \rho_d(F(s, d_s, I)) + \sum_s \beta_w O(x_s + d_s, y_s)$$
$$+ \sum_{s,t \in C \setminus B} \rho_p(d_s, d_t). \quad (13)$$

Above Equation (13) can also be interpreted as the Gibbs energy of a Markov network with respect to occlusion. The evidence $\phi_s^d(\cdot)$ is the first two terms on the right hand side of (13) with the compatibility function $\psi_{st}^d(\cdot)$ being the last term. Again, we can apply belief propagation .

In summary, our iterative optimization algorithm alternates between the following two steps:

1. Estimate occlusion $\{O_L, O_R\}$ using (8), given current estimation of disparity $\{D_L, D_R\}$, and

2. Estimate disparity $\{D_L, D_R\}$ using (11), given current estimation of occlusion $\{O_L, O_R\}$.

The values of occlusion $\{O_L, O_R\}$ are initially set to zeros (i.e., all pixels are initially visible).

Figure 5 shows the recovered occlusion for stereo pairs in the Tsukuba and Middlebury data sets. The iteration number is typically 2 or 3 in our experiments. Our results in Figure 5 appear to be very close to the ground truth.

4 Segmentation as Soft Constraint

Recently, segmentation-based stereo approaches (e.g., [20, 10, 2, 14, 22]) have demonstrated that the difficulties and ambiguities caused by textureless or occlusion can be handled by using groups of pixels with similar colors. Such approaches assume that there are no large discontinuities within each segment. We call this assumption the "segment constraint" in this paper. We exploit the segment constraint

by incorporating it into our stereo model as a *soft* constraint:

$$E_d(\widehat{D}, \widehat{O}; I) = \sum_{s \notin \widehat{O}} \rho_d(F(s, d_s, I)) + \sum_{s \in \widehat{O}} \eta_o +$$
$$\sum_s \gamma |d_s - (a_s x_s + b_s y_s + c_s)|, \quad (14)$$

where γ controls the strength of segment constraint and $[a_s \ b_s \ c_s]$ are the 3D plane parameters for the segment containing pixel s. The 3D plane parameters for each region are estimated by a robust fitting algorithm [20].

All previous segmentation-based approaches *commit* to an initial segmentation result prior to 3D plane fitting and stereo processing. This is not ideal because segmentation errors typically cannot be undone and are thus propagated to the stereo processing stage. Our segmentation-based approach does not suffer from this problem because the segment constraint is soft. This advantage is evident in Figure 4.

Figure 4(a) shows the mean-shift color segmentation results [5] using default parameters. In Figure 4(b), the gray pixels are reliable pixels selected by robust 3D plane fitting algorithm. These pixels are subsequently used for 3D plane estimation. We ignore regions that are too small ($<$500 vis-

Figure 5: Occlusion maps for the Tsukuba and Middlebury stereo data. The left column: our results. The right column: ground-truth.

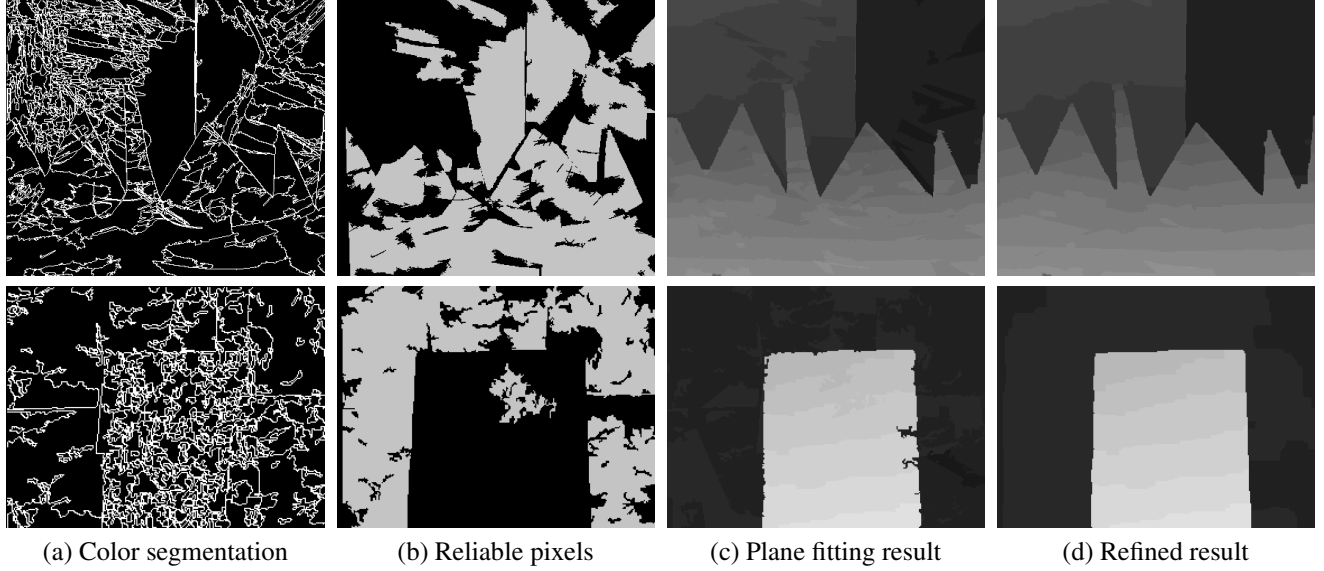|			
(a) Color segmentation	(b) Reliable pixels	(c) Plane fitting result	(d) Refined result

Figure 4: The effect of soft segmentation. The first row is the result for the *Sawtooth* input pair and the second row for *Map* input pair. In (a), the white lines are the boundaries between segments. The grey pixels in (b) are reliable pixels used for 3D plane fitting. The 3D plane fitting results are shown in (c), with refined results using our symmetric stereo model with soft segmentation shown in (d). Note that the refined results are an improvement, especially for the *Map* data.

ible pixels) for reliable 3D plane parameter estimation. Figure 4(c) shows 3D plane fitting results. The disparity of the *Sawtooth* data is significantly improved, especially around the occlusion boundaries. However, the disparity is worse for the *Map* data because the segment constraint is violated. No previous segment-based approaches [20, 10, 2, 14, 22] have been able to handle this image well. By using the segment constraint as a soft constraint, we get better results, as shown in Figure 4(d). The improvement is more dramatic for the *Map* data; here, the errors caused by the violation of the segment constraint are significantly reduced.

5 Parameter Setting

Table 1 lists the values for the symmetric stereo model parameters used in Tsukuba and Middlebury data. We found the most sensitive parameter is the smoothness strength (λ in (5)). To make our approach more practical, we propose a technique to set the value of λ automatically.

λ	T	η_o	σ_d	e_d	β_w	β_o	γ
Automatic	2	2.5	4.0	0.01	4.0	1.4	2.0

Table 1: Parameters for Tsukuba and Middlebury data.

Our approach is based on the following intuition: if the similarity of the data terms (matching distributions) between two neighboring pixels is high, λ should be large to encourage the propagations between them. For a pixel s, its matching distribution is measured by $p_s(d_s) \propto$

$\exp(-F(s, d_s, I))$. The similarity between two neighboring pixels p_s and p_t can be measured with a symmetric version of the Kullback-Leiber (KL) divergence, that is, $KL(p_s||p_t) = \sum_i (p_s^i - p_t^i)(\log(p_s^i) - \log(p_t^i))$.

The average KL divergence $\overline{KL}$ is computed over the whole image. Since we have ground truth depth for the *Tsukuba* data, we use it to estimate the multiplicative constant α in $\lambda = \alpha\overline{KL}$. α is found to be 5.75, which works very well in all our experiments, even for the synthetic data in Figures 1 and 2. Our adaptive λ selection allows our approach to consistently outperform previous approaches for most data sets as shown in Table 2.

6 Experimental Results

All experiments are performed on a 2.8 GHz Pentium 4 PC. The main computation cost is the belief propagation. We adopted the implementation of belief propagation proposed by [8] for efficiency. A data set typically takes about 45 secs to process (e.g., *Tsukuba* data, 3 iterations).

To evaluate the performance of our approach, we follow the methodology proposed by Scharstein and Szeliski in [17]. The quality is measured by the percentages of bad matching (where the absolute disparity error is greater than 1 pixel) in the image ("bad"), in the textureless region ("untex."), and in the discontinuity region ("disc."). Note that all occluded pixels are excluded from the evaluation.

The quantitative comparison in Table 2 and the depth map in Figure 6 demonstrate the high-quality performance of our approaches. Further, Figure 5 shows the visual com-

Algorithm	Tsukuba			Sawtooth			Venus			Map	
	bad	untex.	disc.	bad	untex.	disc.	bad	untex.	disc.	bad	disc.
Symm. BP with Seg.	0.97 ₁	0.28 ₂	5.45 ₁	0.19 ₁	0.00 ₁	2.09 ₁	0.16 ₂	0.02 ₂	2.77 ₄	0.16 ₁	2.20 ₁
Symmetric BP	1.01 _	0.28 _	5.79 _	0.57 _	0.05 _	3.46 _	0.66 _	0.71 _	8.72 _	0.14 _	1.97 _
One way BP	1.42 _	0.64 _	8.01 _	1.18 _	0.38 _	7.71 _	1.21 _	1.57 _	15.22 _	0.15 _	1.85 _
Segm.-based GC [10]	1.23 ₄	0.29 ₃	6.94 ₅	0.30 ₄	0.00 ₁	3.24 ₄	0.08 ₁	0.01 ₁	1.39 ₁	1.49 ₂₀	15.46 ₂₆
Segm.+glob.vis. [2]	1.30 ₆	0.48 ₆	7.50 ₇	0.20 ₂	0.00 ₁	2.30 ₂	0.79 ₅	0.81 ₆	6.37 ₈	1.63 ₂₂	16.07 ₂₇
Layered [14]	1.58 ₈	1.06 ₁₀	8.82 ₉	0.34 ₅	0.00 ₁	3.35 ₅	1.52 ₁₁	2.96 ₂₀	2.62 ₃	0.37 ₁₁	5.24 ₁₁
Belief prop [19]	1.15 ₂	0.42 ₄	6.31 ₂	0.98 ₁₁	0.30 ₁₅	4.83 ₉	1.00 ₇	0.76 ₅	9.13 ₁₄	0.84 ₁₇	5.27 ₁₂
MultiCam GC [13]	1.85 ₁₁	1.94 ₁₆	6.99 ₆	0.62 ₉	0.00 ₁	6.86 ₁₃	1.21 ₉	1.96 ₁₁	5.71 ₇	0.31 ₈	4.34 ₁₀
Region-progress. [22]	1.44 ₇	0.55 ₇	8.18 ₈	0.24 ₃	0.00 ₁	2.64 ₃	0.99 ₆	1.37 ₉	6.40 ₉	1.49 ₂₁	17.11 ₂₈
GC+occl. [12]	1.19 ₃	0.23 ₁	6.71 ₃	0.73 ₁₀	0.11 ₁₀	5.71 ₁₁	1.64 ₁₄	2.75 ₁₈	5.41 ₆	0.61 ₁₄	6.05 ₁₃

Table 2: Comparison of results on Tsukuba and Middlebury data. The subscribe of each number is the rank and the underlined number is the best for each data set. Note that our algorithm "Symmetric BP with segment" consistently outperforms most of other algorithms listed. The complete set of results can be found in http://www.middlebury.edu/stereo/.

(a) One way BP   (b) Symmetric BP   (c) Symmetric BP with segment   (d) Ground Truth
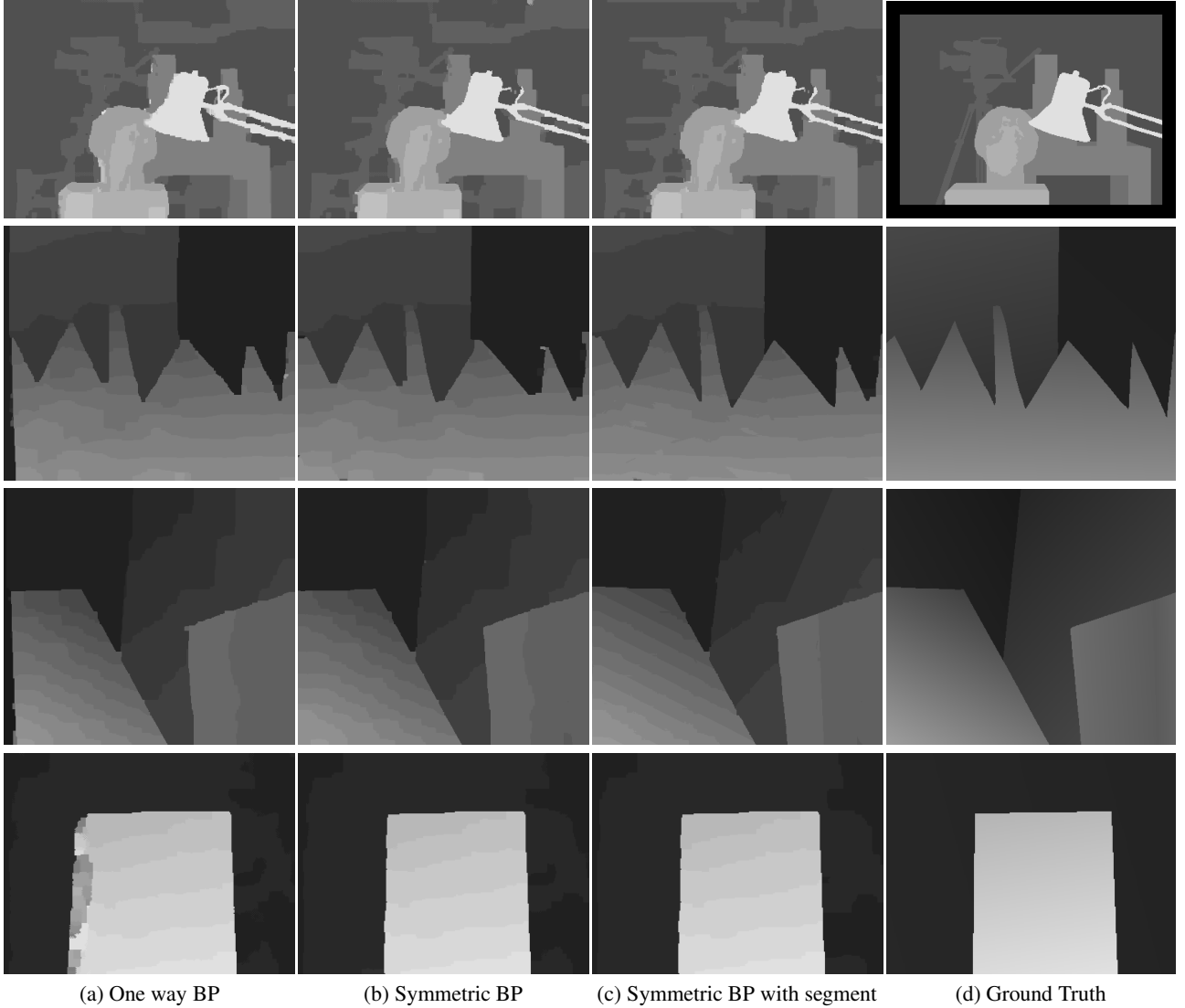
Figure 6: Results for (from top to bottom) *Tsukuba*, *Sawtooth*, *Venus*, and *Map* image pairs.

parison of recovered occlusion maps with ground truth. "One way BP" is the result without visibility term, and "Symmetric BP" is with it. "Symmetric BP with segment" is the result with both visibility and segment constraints. The disparity of occluded pixels are produced by a post-processing method (e.g., for the left image, assigns the disparity next to the right boundary of the occluded region.) For the *Tsukuba* and *Map* data, "Symmetric BP" yields the most significant improvements using the visibility constraint. For the *Sawtooth* and *Venus* data, both "Symmetric BP" and "Symmetric BP with segment" give significant improvements using both visibility and segment constraints. Note that "Symmetric BP with segment" is ranked at or very close to the first place among more than two dozen stereo algorithms.

In order to evaluate the accuracy of our occlusion detection, we also show the incidence of false negatives and positives in Table 3. The false negative rate is measured by the percentage of missed occluded pixels in occluded regions, while the false positive rate is expressed as the percentage of detected occluded pixels in non-occluded regions. Figure 5 shows the visual comparison of our occlusion maps with ground truth.

	Tsukuba	*Sawtooth*	*Venus*	*Map*
False negatives	29.9%	17.0%	25.4%	8.7%
False positives	0.7%	0.2%	0.2%	0.3%

Table 3: Occlusion detection rates.

7 Conclusions

In this paper, we have proposed a symmetric stereo matching model to handle occlusion using the visibility constraint. The visibility constraint is a general constraint that allows us to correctly recover horizontally slanted surfaces and thin foreground objects. Our iterative optimization algorithm, in conjunction with the soft segment constraint and automatic parameter estimation, is validated by state-of-the-art results.

Although our iterative optimization algorithm does not guarantee convergence, we have not observed oscillatory behavior in experiments. More studies are needed to fully understand the behavior of our algorithm and perhaps reformulate it in an Expectation Maximization (EM) framework [18].

References

[1] P. N. Belhumeur. A bayesian-approach to binocular stereopsis. *IJCV*, 19(3):237–260, 1996.

[2] M. Bleyer and M. Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. *ICIP*, pages 2997–3000, 2004.

[3] A. F. Bobick and S. S. Intille. Large occlusion stereo. *IJCV*, 33(3):1–20, 1999.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2001.

[6] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum-likelihood stereo algorithm. *CVIU*, 63(3):542–567, 1996.

[7] G. Egnal and R. Wildes. Detecting binocular halfocclusions: empirical comparisons of five approaches. *PAMI*, 24(8):1127–1133, 2002.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *CVPR*, Vol I:261–268, 2004.

[9] D. Geiger, B. Ladendorf, and A. L. Yuille. Occlusions and binocular stereo. *IJCV*, 14(3):211–226, 1995.

[10] L. Hong and G. Chen. Segment-based stereo matching using graph cuts. *CVPR*, Vol I:74–81, 2004.

[11] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. *ECCV*, pages 425–433, 1998.

[12] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. *ICCV*, pages 508–515, 2001.

[13] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *ECCV*, Vol III:82–96, 2002.

[14] M. Lin. Surfaces with occlusions from layered stereo. *Ph.D. thesis, Stanford University*, 2002.

[15] D. Marr and T.A. Poggio. Cooperative computation of stereo disparity. *Science*, 194(4262):283–287, 1976.

[16] A. S. Ogale and Y. Aloimonos. Stereo correspondence with slanted surface: critical implication of horizontal slant. *CVPR*, Vol I:568–573, 2004.

[17] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.

[18] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. *CVPR*, Vol I:552–559, 2004.

[19] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. *ECCV*, Vol II:510–524, 2002.

[20] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. *ICCV*, Vol I:532–539, 2001.

[21] F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. *ICCV*, Vol II:900–906, 2003.

[22] Y. Wei and L. Quan. Region-based progressive stereo matching. *CVPR*, Vol I:106–113, 2004.

[23] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *PAMI*, 22(7):675–684, 2000.