Stereo Matching Using Belief Propagation

Jian Sun*[1,2], Heung-Yeung Shum[2], and Nan-Ning Zheng[1]

[1] Artificial Intelligence and Robotics Lab, Xi'an Jiaotong University, China
{sj,nnzheng}@aiar.xjtu.edu.cn
[2] Visual Computing Group, Microsoft Research Asia, Beijing
hshum@microsoft.com

Abstract. In this paper, we formulate the stereo matching problem as a Markov network consisting of three coupled Markov random fields (MRF's). These three MRF's model a smooth field for depth/disparity, a line process for depth discontinuity and a binary process for occlusion, respectively. After eliminating the line process and the binary process by introducing two robust functions, we obtain the maximum a posteriori (MAP) estimation in the Markov network by applying a Bayesian belief propagation (BP) algorithm. Furthermore, we extend our basic stereo model to incorporate other visual cues (e.g., image segmentation) that are not modeled in the three MRF's, and again obtain the MAP solution. Experimental results demonstrate that our method outperforms the state-of-art stereo algorithms for most test cases.

1 Introduction

Stereo vision infers scene geometry from two images with different viewpoints. Classical dense two-frame stereo matching computes a dense disparity or depth map from a pair of images under a known camera configuration. In general, the scene is assumed Lambertian or intensity-consistent from different viewpoints, without specularities, reflection, or transparency.

Stereo matching is difficult because of the following reasons.

- Noise: There are always unavoidable light variations, image blurring, and sensor noise in image formation.
- Textureless region: Information from highly textured regions needs to be propagated into textureless regions for stereo matching.
- Depth discontinuity: Information propagation should stop at object boundaries .
- Occlusion: Those occluded pixels in the reference view cannot be matched with the other view.

Therefore, stereo matching is an ill-posed problem with inherent ambiguities. Obviously, some constraints are needed to get a good "guess" of scene structure. Many methods have been proposed to encode various constraints,

* This work was performed while the first author was visiting Microsoft Research Asia

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2351, pp. 510–524, 2002.

e.g, intensity-consistency, local smoothness constraints, generalized order constraints, and uniqueness constraints. It has been shown that these constraints can be modeled well as priors in the Bayesian approach to stereo matching.

In this paper, after reviewing related works in Section 2 on stereo matching, and especially on the Bayesian approaches, we propose in Section 3 a Bayesian stereo matching approach to explicitly model discontinuities, occlusion and the disparity field in the Bayesian framework. In Section 4, Bayesian Belief Propagation is used to infer the stereo matching. The basic stereo model is then extended in Section 5 to integrate multiple cues, e.g, region similarity. The experimental results shown in Section 6 demonstrate that our model is effective and efficient. Finally, we discuss in Section 7 why our stereo matching with belief propagation outperforms the state-of-art stereo algorithms.

2 Related Works

In this section, we review related stereo algorithms and especially those using the Bayesian approach. We refer the reader to a detailed and updated taxonomy of dense, two-frame stereo correspondence algorithms by Scharstein and Szeliski [21]. It also provides a testbed for quantitative evaluation of stereo algorithms.

A stereo algorithm is called a global method if there is a global objective function to be optimized. Otherwise it is called a local method. The central problem of local or window-based stereo matching methods is to determine the optimal support window for each pixel. An ideal support region should be bigger in textureless regions and should be suspended at depth discontinuities. The fixed window is obviously invalid at depth discontinuities. Some improved window-based methods, such as adaptive windows [16], shiftable windows [5] and compact windows [23] try to avoid the windows that span depth discontinuities.

Bayesian methods (e.g., [11,1,5,8,15]) are global methods that model discontinuities and occlusion. Geiger et al. [11] derived an occlusion process and a disparity field from a matching process. Assuming an "order constraint" and "uniqueness constraint", the matching process becomes a "path-finding" problem what the global optimum is obtained by dynamic programming. Belhumeur [1] defined a set of priors from a simple scene to a complex scene. A simplified relationship between disparity and occlusion is used to solve scan line matching by dynamic programming. Unlike Geiger and Belhumeur who enforced a piecewise smooth constraint, Cox et al. [8] and Bobick & Intille [5] did not require the smoothing prior. Assuming corresponding features are normally distributed and a fixed cost for occlusion, Cox also proposed a dynamic programming solution using only the occlusion constraint and ordering constraints. Bobick & Intille incorporated Ground Control Points constraint to reduce the sensitivity to occlusion cost and complexity of Cox's dynamic programming. These methods use dynamic programming and assume that the occlusion cost is the same in each scanline. Ignorance of dependence between scanlines results in the characteristic "streaking" in the disparity maps.

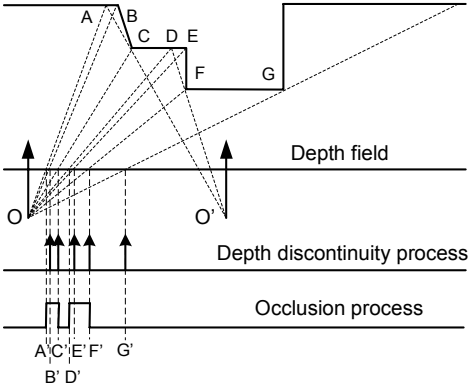


Fig. 1. A scene illustrates the geometry relationship among depth, discontinuities and occlusions. O and O' are optical centers of two cameraes. Discontinuities occur at B',C',E',F' and G'. Occlusion occur in $[A, C]$ and $[D, F]$

In general, Bayesian stereo matching can be formulated as a maximum a posteriori MRF (MAP-MRF) problem. There are several methods to solve the MAP-MRF problem: simulated annealing [12], Mean-Field annealing [10], the Graduated Non-Convexity algorithm(GNC) [4], and Variational approximation [14]. Finding a solution by simulated annealing can often take an unacceptably long time although global optimization is achieved in theory. Mean-Field annealing is a deterministic approximation to simulated annealing by attempting to average over the statistics of the annealing process. It reduces execution time at the expense of solution quality. GNC can only be applied to some special energy functions. Variational approximation converges to a local minimum. Graph Cut (GC) [6] is a fast efficient algorithm to solve a MAP-MRF whose energy function is Potts or Generalized Potts.

3 Basic Stereo Model

In our work, to handle occlusion and depth discontinuity explicitly, we model stereo vision by three coupled MRF's: D is the smooth disparity field of the reference view, L is a spatial line process located on the dual lattice and representing explicitly the presence or absence of depth discontinuities in the reference view, and O is a spatial binary process to indicate occlusion regions in the reference view. Figure 1 illustrates these processes in the 1D case. By using Bayes' rule, the joint posterior probability over D, L and O given a pair of stereo images ($I = (I_L, I_R)$ where I_L is the left and reference image) is:

$$P(D, L, O|I) = \frac{P(I|D, L, O)P(D, L, O)}{P(I)}. \tag{1}$$

Without occlusion, $\{D, L\}$ are coupled MRF's that model a piece-wise smooth surface by two random fields: one represents the variable required to

estimate, the other represents its discontinuities. This model was proposed by [12]. However, the occlusion problem in stereo vision is not included in this kind of model explicitly. In image formation, the piece-wise smooth scene is projected on a pair of stereo images. Some regions are only visible in one image. There is no matching pixel in the other view for each pixel in the occlusion region. We assume that likelihood $P(I|D,O,L)$ is independent of L because the observation is pixel-based, and ignore the statistical dependence between O and $\{D,L\}$:.

$$P(I|D,O,L) = P(I|D,O), \tag{2}$$

$$P(D,O,L) = P(D,L)P(O). \tag{3}$$

The basic stereo model now becomes

$$P(D,O,L|I) = \frac{P(I|D,O)P(D,L)P(O)}{P(I)}. \tag{4}$$

3.1 Likelihood

Assuming observation noises follow an independent identical distribution(i.i.d), we can define the likelihood $P(I|D,O)$ as:

$$P(I|D,O) \propto \prod_{s \notin O} \exp(-F(s,d_s,I)) \tag{5}$$

where $F(s,d_s,I)$ is the matching cost function of pixel s with disparity d_s given observation I. Our likelihood considers the pixels only in non-occluded areas $s \notin O$ because likelihood of the pixels in occluded areas can not be well defined. We use the pixel dissimilarity that is provably insensitive to sampling [2]:

$$F(s,d_s,I) = \min\{\overline{d}(s,s',I)/\sigma_f, \overline{d}(s',s,I)/\sigma_f\}$$

where $\overline{d}(s,s',I) = \min\{\left|I_L(s) - I_R^-(s')\right|, \left|I_L(s) - I_R(s')\right|, \left|I_L(s) - I_R^+(s')\right|\}$, s' is the matching pixel in right view of s with disparity d_s, $I_R^-(s')$ is the linearly interpolated intensity halfway between s' and its neighboring pixel to the left, $I_R^+(s')$ is to the right, $\overline{d}(s',s,I)$ is the symmetric version of $\overline{d}(s,s',I)$ and σ_f is the variance to be estimated.

3.2 Prior

Deriving appropriate priors to encode constraints directly is not only hard but may also result in too many annoying hyper parameters to find the solution easily. The Markov property asserts that the probability of a site in the field depends only on its neighboring sites. By specifying the first order neighborhood $G(s)$ and $N(s) = \{t|t > s, t \in G(s)\}$ of site s, the prior 3 can be expanded as:

$$P(D,L,O) \propto \prod_s \prod_{t \in N(s)} \exp(-\varphi_c(d_s,d_t,l_{s,t})) \prod_s \exp(-\eta_c(o_s)) \tag{6}$$

where $\varphi_c(d_s, d_t, l_{s,t})$ is the joint clique potential function of d_s, d_t and $l_{s,t}$, and $\eta_c(o_s)$ is the clique potential function of o_s. $\varphi_c(d_s, d_t, l_{s,t})$ and $\eta_c(o_s)$ are user-customized functions to force the constraints for stereo matching. $\varphi_c(d_s, d_t, l_{s,t})$ and $\eta_c(o_s)$ also determine the distributions of $\{D, L, O\}$. To enforce spatial interactions between d_s and l_s, we define $\varphi_c(d_s, l_s)$ as follows:

$$\varphi_c(d_s, d_t, l_{s,t}) = \varphi(d_s, d_t)(1 - l_{s,t}) + \gamma(l_{s,t}) \tag{7}$$

where $\varphi(d_s, d_t)$ penalizes the different assignments of neighbor sites when no discontinuity exists between them, and $\gamma(l_{s,t})$ penalizes the occurrence of a discontinuity between site s and t.

Combining (5),(6) and (7), our basic stereo model becomes:

$$\begin{aligned} P(D, O, L|I) \propto & \prod_{s \notin O} \exp(-F(s, d_s, I)) \prod_{s} \exp(-\eta_c(o_s)) \\ & \prod_{s} \prod_{t \in N(s)} \exp(-(\varphi(d_s, d_t)(1 - l_{s,t}) + \gamma(l_{s,t}))). \end{aligned} \tag{8}$$

4 Approximate Inference by Belief Propagation

In the last section, we model stereo matching by three coupled MRFs. After converting MRFs to the corresponding Markov network, the approximate inference algorithm Loopy Belief Propagation can be used to approximate the posterior probability for stereo matching.

4.1 From Line Process to Outlier Process

It is hard not only to specify appropriate forms of $\varphi(d_s, d_t)$, $\gamma(l_{s,t})$ and $\eta_c(o_s)$, but also to do inference in a continuous MRF and two binary MRFs. Fortunately, the unification of line process and robust statistics [3] provides us a way to eliminate the binary random variable from our MAP problem. If we simplify $\eta_c(o_s)$ by ignoring the spatial interaction of occlusion sites[1]

$$\eta_c(o_s) = \eta(o_s) \tag{10}$$

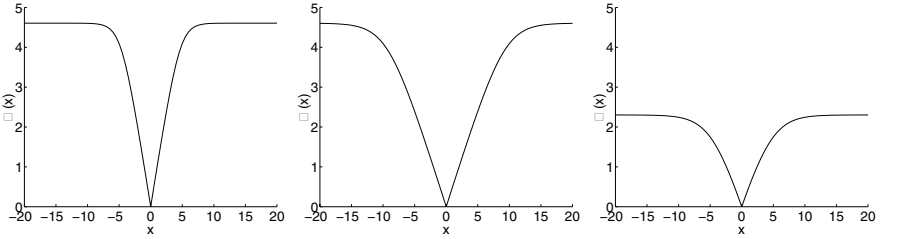
we can rewrite our MAP problem as:

$$\begin{aligned} \max_{D,L,O} P(D, L, O|I) = \max_{D} \Bigg\{ & \max_{O} \prod_{s} \exp(-(F(s, d_s, I)(1 - o_s) + \eta(o_s))) \\ & \max_{L} \prod_{s} \prod_{t \in N(s)} \exp(-(\varphi(d_s, d_t)(1 - l_{s,t}) + \gamma(l_{s,t}))) \Bigg\} \end{aligned} \tag{11}$$

[1] The complete form of $\eta_c(o_s)$ should be:

$$\eta_c(o_s) = \eta(o_s) + \sum_{t \in N(s)} \eta'(o_s, o_t) \tag{9}$$

where $\eta(o_s)$ is a single-site clique potential function that penalizes the occurrence of occlusion, and $\eta'(o_s, o_t)$ is a pair-site cliques potential function that penalizes the different assignments of o_s and o_t.



$e = 0.01, \sigma = 1.0$ $e = 0.01, \sigma = 2.0$ $e = 0.1, \sigma = 2.0$

Fig. 2. Robust function $\rho(x) = -\ln((1-e)\exp(-\frac{|x|}{\sigma}) + e)$ derived from TV model.

Now, we upgrade the binary process l_{st} and o_s to analog process l^a_{st} and o^a_s ("outlier process"[3]) by allowing $0 \le l^a_{st} \le 1$ and $0 \le o^a_s \le 1$. For the first term,

$$\begin{aligned} &\max_O \prod_s \exp(-(F(s, d_s, I)(1 - o^a_s) + \eta(o^a_s))) \\ &= \exp(-\min_O \textstyle\sum_s (F(s, d_s, I)(1 - o^a_s) + \eta(o^a_s))) \end{aligned} \tag{12}$$

where $\min_O \sum_s (F(s, d_s, I)(1 - o^a_s) + \eta(o^a_s))$ is the objective function of a robust estimator. The robust function of this robust estimator is

$$\rho_d(d_s) = \min_{o^a_s}(F(s, d_s, I)(1 - o^a_s) + \eta(o^a_s)) \tag{13}$$

and for the second term, we also have a robust function $\rho_p(d_s, d_t)$:

$$\rho_p(d_s, d_t) = \min_{l^a_{s,t}}(\varphi(d_s, d_t)(1 - l^a_{s,t}) + \gamma(l^a_{s,t})). \tag{14}$$

We get the posterior probability over D defined by two robust functions:

$$P(D|I) \propto \prod_s \exp(-\rho_d(F(s, d_s, I)) \prod_s \prod_{t \in N(s)} \exp(-\rho_p(d_s, d_t)). \tag{15}$$

Thus, we not only eliminate two analog line processes via an outlier process but also model outliers in measurements. We convert the task of modelling the priors of the occlusion process and depth discontinuity process explicitly into defining two robust functions that model occlusion and discontinuity implicitly.

In this paper, our robust functions are derived from the Total Variance(TV) model [18] with the potential function $\rho(x) = |x|$ because of its discontinuity preserving property. We truncate this potential function as our robust function:

$$\rho_d(d_s) = -\ln((1 - e_d)\exp(-\frac{|F(s, d_s, I)|}{\sigma_d}) + e_d)$$

$$\rho_p(d_s, d_t) = -\ln((1 - e_p)\exp(-\frac{|d_s - d_t|}{\sigma_p}) + e_p)$$

Figure 2 shows different shapes of our robust functions. By varying parameters e and σ, we control the shape of the robust function, and therefore the posterior probability.

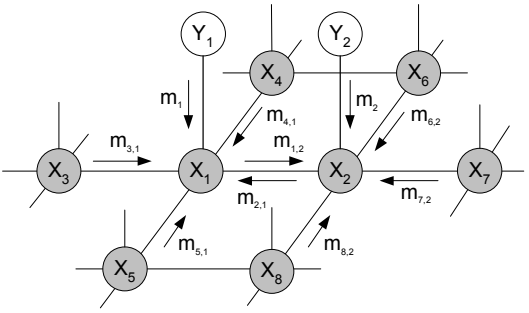


Fig. 3. Local message passing in Markov Network. In "max-product" algorithm, the new message sent from node x_1 to x_2 is: $m_{1,2}^{new} \leftarrow \kappa \max_{x_1} \psi_{12}(x_1, x_2) m_1 m_{3,1} m_{4,1} m_{5,1}$. The belief at node x_1 is computed as: $b_1 \leftarrow \kappa m_1 m_{2,1} m_{3,1} m_{4,1} m_{5,1}$

.

4.2 Belief Propagation

The model that is most similar to our posterior probability (15) is Scharstein & Szeliski's [20]. Unlike Scharstein & Szeliski, where a nonlinear diffusion algorithm is used, we address this MAP problem by Belief Propagation. Belief Propagation is an exact inference method proposed by Pearl[19] in the belief network without loops. Loopy Belief Propagation is just Belief Propagation that ignores the existence of loops in the networks. It has been applied successfully to some vision [9] and communication [24] problems despite the presence of network loops.

The posterior probability (15) over D is exactly a Markov Network in the literature of probabilistic graph models as shown in Figure 3. In the Markov Network, random variable d_s in our stereo model is represented by a hidden node x_s. A "private" observation node y_s is connected to each x_s. Each y_s is a vector where each element is the matching cost given different assignments of node x_s. By denoting $X = \{x_s\}$ and $Y = \{y_s\}$, (15) can be represented with x_s and y_s:

$$P(X|Y) \propto \prod_{s,t:s>t,t\in N(s)} \psi_{st}(x_s, x_t) \prod_s \psi_s(x_s, y_s) \tag{16}$$

where

$$\psi_{st}(x_s, x_t) = \exp(-\rho_p(x_s, x_t))) \tag{17}$$

$$\psi_s(x_s, y_s) \propto \exp(-\rho_d(F(s, x_s, I))) \tag{18}$$

$\psi_{st}(x_s, x_t)$ is called compatibility matrix between node x_s and x_t, and $\psi_s(x_s, y_t)$ is called the local evidence for node x_s. If the disparity level is L, $\psi_{st}(x_s, x_t)$ is a $L \times L$ matrix and $\psi_s(x_s, y_s)$ is a L-length vector.

Belief Propagation is an iterative inference algorithm that propagates messages in the network. Let $m_{st}(x_s, x_t)$ be the message that node x_s sends to x_t, $m_s(x_s, y_s)$ be the message that observed node y_s sends to node x_s, $b_s(x_s)$ be the belief at node x_s. Note that $m_{st}(x_s, x_t)$, $m_s(x_s, y_s)$ and $b_s(x_s)$ are all 1D vectors. We simplify $m_{st}(x_s, x_t)$ as $m_{st}(x_t)$, and $m_s(x_s, y_s)$ as $m_s(x_s)$. There are

two kinds of BP algorithms with different message update rules: "max-product" and "sum-product", which maximize the joint posterior of the network, and the marginal posterior of each node, respectively. The standard "max-product" algorithm is shown below.

1. Initialize all messages as uniform distributions
2. Update messages iteratively for i=1:T

$$m_{st}^{i+1}(x_t) \leftarrow \kappa \max_{x_s} \psi_{st}(x_s, x_t) m_s^i(x_s) \prod_{x_k \in N(x_s) \setminus x_s} m_{ks}^i(x_s)$$

3. Compute beliefs

$$b_s(x_s) \leftarrow \kappa m_s(x_s) \prod_{x_k \in N(x_s)} m_{ks}(x_s)$$

$$x_s^{MAP} = \arg\max_{x_k} b_s(x_k)$$

For example, in Figure 3, the new message sent from node x_1 to x_2 is updated as: $m_{1,2}^{new} \leftarrow \kappa \max_{x_1} \psi_{12}(x_1, x_2) m_1 m_{3,1} m_{4,1} m_{5,1}$. The belief at node x_1 is computed as: $b_1 \leftarrow \kappa m_1 m_{2,1} m_{3,1} m_{4,1} m_{5,1}$ (the product of two messages is component-wise product). And κ is the normalization constant.

5 Integrating Multiple Cues

More constraints and priors (e.g., edges, corners, junctions, segmentation, visibility) can be incorporated to improve stereo matching. For instance, a segmentation-based stereo algorithm [22] has been recently proposed based on the assumption that the depth discontinuities occur on the boundary of the segmented regions. In [22], the segmentation results are used as hard constraints. In our work, we make use of image segmentation but incorporate segmentation results into our basic stereo model as soft constraints (prior) under a probabilistic framework.

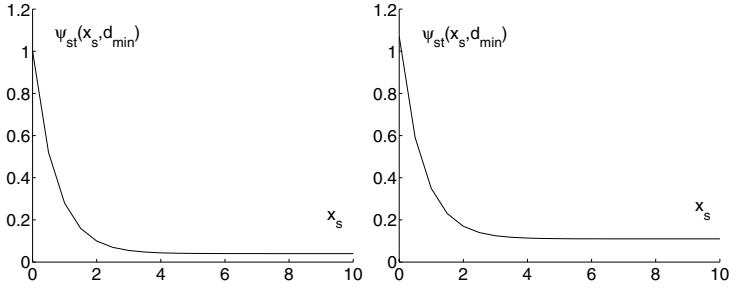
With additional cues, we extend the basic stereo model (15):

$$\begin{aligned} P(D, O, L|I) \propto & \prod_s \exp(-\rho_d(F(s, d_s, I)) \\ & \prod_s \prod_{t \in N(s)} \exp(-\varphi_c(d_s, d_t, l_{s,t})) \exp(-\rho_{pcue}(d_s, d_t)) \end{aligned} \tag{19}$$

where $\rho_{pcue}(d_s, d_t)$ encodes some constraints between sites. To integrate region similarities from image segmentation, we define $\rho_{pcue}(d_s, d_t)$ as:

$$\rho_{pcue}(d_s, d_t) = \rho_{seg}(d_s, d_t) = \begin{cases} 0 & seg(s) = seg(t) \\ \lambda_{seg} & seg(s) \neq seg(t) \end{cases} \tag{20}$$

where $seg(s)$ is the label of the segmentation result at site s. The larger the λ_{seg}, the more difficulty in passing the message between neighbor sites. In other



$e_p = 0.01, \sigma_p = 1.0, \lambda_{seg} = 0$ $seg(s) = seg(t)$ $\quad$ $e_p = 0.01, \sigma_p = 1.0, \lambda_{seg} = 0.05$ $seg(s) \neq seg(t)$

Fig. 4. Left image is the first row of $\psi_{st}(x_s, x_t)$ when node x_s and x_t are in same region. Right image is the first row of $\psi_{st}(x_s, x_t)$ when node x_s and x_t are in different regions

words, the influence from neighbors becomes smaller as λ_{seg} increases. In our experiments, the segmentation labels are produced by the Mean-Shift algorithm [7]. The execution time is usually just a few seconds in all images used in our experiments.

According to (15), the compatibility matrix $\psi_{st}(x_s, x_t)$ can be rewritten as:

$$\psi_{st}(x_s, x_t) = \exp(-\rho_p(x_s, x_t))) \exp(-\rho_{pcue}(x_s, x_t))) \tag{21}$$

Figure 4 shows the first rows of $\psi_{st}(x_s, x_t)$ when x_s and x_t are in same region and in different regions.

6 Experimental Results

In this paper, we evaluate the performance of our stereo algorithm using the quality measures proposed in [21] with those measures based on known ground truth data listed in Table 1. In particular, $B_{\bar{O}}$ represents the overall performance of a stereo algorithm.

Table 1. Quality measures based on known ground truth data

Percentage of bad matching pixels in non-occlusion regions $\bar{\mathtt{O}}$	$B_{\bar{\mathtt{O}}} = \frac{1}{N} \sum_{s \in \bar{\mathtt{O}}} (\lvert d(s) - d_T(s) \rvert > \delta_d)$
Percentage of bad matching pixels in textureless regions $\bar{\mathtt{T}}$	$B_{\bar{\mathtt{T}}} = \frac{1}{N} \sum_{s \in \bar{\mathtt{T}}} (\lvert d(s) - d_T(s) \rvert > \delta_d)$
Percentage of bad matching pixels in depth discontinuity regions $\mathtt{D}$	$B_{\mathtt{D}} = \frac{1}{N} \sum_{s \notin \mathtt{D}} (\lvert d(s) - d_T(s) \rvert > \delta_d)$

The test data set consists of four pairs of images: "Map", "Tsukuba", "Sawtooth" and "Venus" [21]. "Tsukuba" is a complicated indoor environment with

slanted surfaces and contains a number of integer valued disparities. Other pairs consist of mainly slanted planes.

Table 2 shows the results of applying our BP algorithm to all four pairs of images. It also lists the results of other stereo algorithms. This table is courtesy of Scharstein and Szeliski (see http://www.middlebury.edu/stereo/results.html for details). Our results with and without image segmentation incorporated into stereo matching are shown in the first and the second row, respectively.

For a complicated environment like "Tsukuba", incorporating image segmentation improves stereo matching significantly, with 40% error reduction in $B_{\bar{O}}$. In fact, our algorithm ranks as the best for "Tsukuba" and outperforms Graph Cut (with occlusion) [17] which was widely considered the state-of-art stereo matching algorithm. Our algorithm competes well with other stereo algorithms for the three other data sets, "Sawtooth", "Venus" and "Map". It is interesting to note that for these three data sets with simple slanted surfaces, incorporating image segmentation does not necessarily improve stereo matching, as seen from the first and second rows.

Figures 5 and 6 show the results obtained by our algorithm. The segmentation map is obtained by the Mean-Shift algorithm with default parameters suggested by [7]. Note that a fixed set of parameters $\{e_d = 0.05, \sigma_d = 0.6, e_p = 0.01, \sigma_p = 8\}$ are used in our BP algorithm for all image pairs. Obviously, this set of parameters is not the optimal for "Map" data because the disparity range of this data is almost twice that of "Tsukuba" data's disparity range.

The complexity of our BP algorithm is $O(L^2NT)$ where N is the number of pixels, L is the number of disparities, and T is the number of iterations. For the "Tsukuba" data, it took 288 seconds on a Pentium III 500 MHz PC. It is comparable or slightly better than the graph cut algorithm reported in [21].

The local oscillation phenomena of the BP algorithm also occurred in our experiments. A time average operation is executed after a fixed number of iterations: $m^t_{st}(x_t) = m^{t-1}_{st}(x_t) + m^t_{st}(x_t)$. This heuristic worked well in our experiments.

7 Discussion

Why BP works? The magic of the BP algorithm lies in its powerful message passing. A message presents the probability that the receiver should be at a disparity according to all information from the sender up to the current iteration. Message passing has two important properties. First, it is asymmetric. The entropy of the messages from high-confidence nodes to low-confidence nodes is smaller than the entropy of the messages from low-confidence nodes to high-confidence nodes. Second, it is adaptive. The influence of a message between a pair of nodes with larger divergence would be weakened more.

Therefore, BP's message passing provides a time-varying adaptive smoothing mechanism for stereo matching to deal with textureless regions and depth discontinuities naturally. In textureless regions, for example, the influence of a

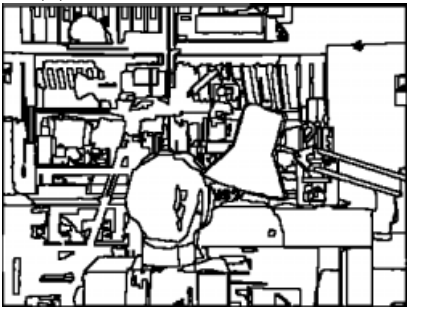
(a) Left (reference) Image

(b) Ground Truth

(c) Textureless regions

(d) Depth discontinuity regions

(e) Occlusion regions

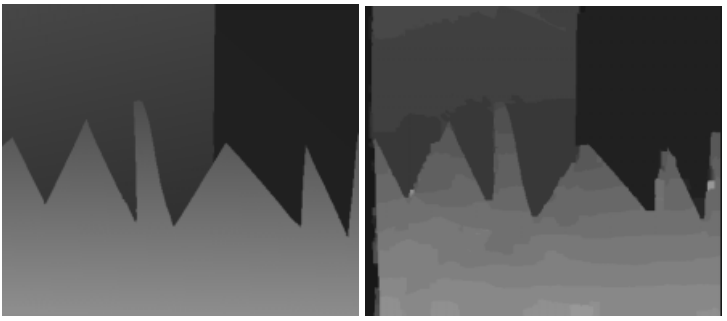


(f) Image segmentation

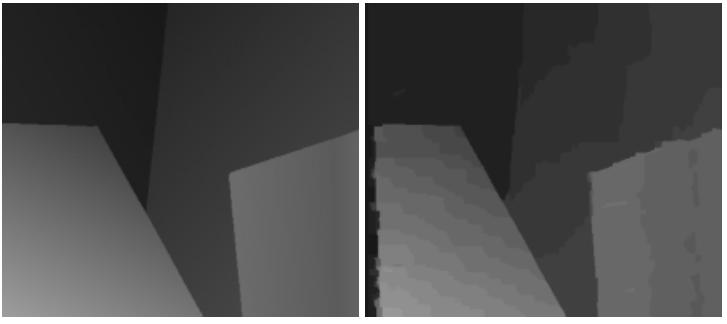
(g) "max-product" result

(h) "max-product" result with segmentation

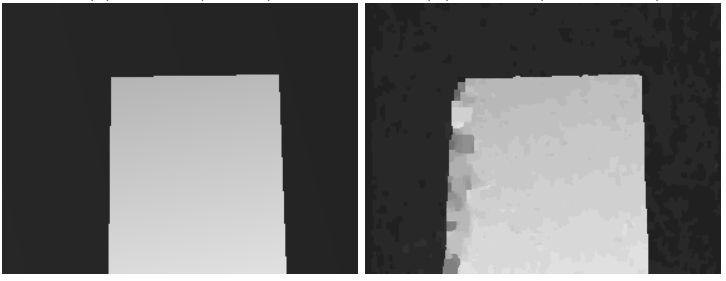
Fig. 5. The results on Tsukuba dataset. (a)-(e) are given.



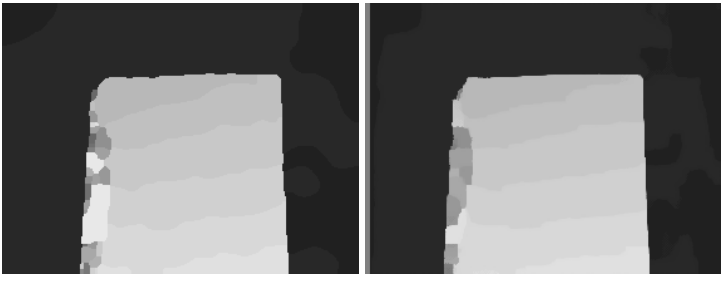
(a) Sawtooth (Truth) (b) Sawtooth (Our result)



(c) Venus (Truth) (d) Venus (Our result)



(e) Map (Truth) (f) Map (Max-product)



(g) Map (Bayes Diffusion) (h) Map (Sum-product)

Fig. 6. The results of Sawtooth and Venus based on the "max-product" algorithm are shown in (b) and (d). For the Map data, the "max-product" result is shown in (f). Bayesian diffusion results with $B_{\bar{O}} = 0.20$, $B_{DD} = 2.49$ are shown in (g), while "sum-product" results with $B_{\bar{O}} = 0.16$, $B_{DD} = 2.11$ are shown in (h).

Table 2. The performance of different stereo algorithms with fixed parameters on four test image pairs. An underlined number is the best in its category.

Algorithms	Tsukuba			Sawtooth			Venus			Map	
	$B_{\bar{0}}$	$B_{\bar{1}}$	$B_{\bar{D}}$	$B_{\bar{0}}$	$B_{\bar{1}}$	$B_{\bar{D}}$	$B_{\bar{0}}$	$B_{\bar{1}}$	$B_{\bar{D}}$	$B_{\bar{0}}$	$B_{\bar{D}}$
Belief prop. (seg)	1.15	0.42	6.31	0.98	0.30	4.83	1.00	0.76	9.13	0.84	5.27
Belief prop.	1.61	0.66	9.17	0.85	0.37	7.92	1.17	1.00	12.87	0.67	3.42
Graph cuts [21]	1.96	1.06	9.41	1.36	0.23	6.57	1.36	1.75	6.63	0.33	4.40
GC+occl. [17]	1.27	0.43	6.90	0.36	0.00	3.65	2.79	5.39	2.54	1.79	10.08
Graph cuts [6]	1.86	1.00	9.35	0.42	0.14	3.76	1.69	2.30	5.40	2.39	9.35
Realtime SAD [13]	4.25	4.47	15.05	1.32	0.35	9.21	1.53	1.80	12.33	0.81	11.35
Bay. diff. [21]	6.49	11.62	12.29	1.43	0.69	9.29	3.89	7.15	18.17	0.20	2.49
SSD+MF [21]	5.26	3.86	24.65	2.14	0.72	13.08	3.81	6.93	12.94	0.66	9.35
Dyn. prog. [21]	3.43	3.22	12.34	4.54	3.59	13.11	8.47	12.76	17.61	3.77	13.93
Scanl. opt. [21]	4.94	6.50	11.94	4.19	2.95	12.14	9.71	14.98	18.20	4.61	10.22

message can be passed far away. On the other hand, the influence in discontinuous regions will fall off quickly. Figure 7 shows this adaptive smoothing procedure in an example. In Figure 7, the image pair is modified from that used in [16] and [20]. A linear ramp in the direction of the baseline is used as the underlying intensity pattern. The disparity of background and foreground is 2 and 5, respectively. Unlike [16] or [20], a smaller pure textureless square is overlapped in the center of the foreground in the ramp1 pair.

We use entropy $H(b) = -\sum_i b_i log b_i$ to measure the confidence of the belief, and the symmetric version of the Kullback-Leiber(KL) divergence $KL_s(b^1 \| b^2) = \sum_i (b_i^1) - b_i^2) log(\frac{b_i^1}{b_i^2})$ to measure the difference between belief b^1 and b^2. Smaller entropy represents higher confidence of a belief. Larger divergence represents larger dissimilarity between beliefs. As shown in the figure, the entropy map of a belief represents the confidence of disparity estimation for each node. Clearly, the confidence of each node increases with each iteration. Note that the confidence in occlusion regions and corners is lower than that in other regions. This shows that the probabilistic method outputs not only a solution, but also its certainty. The divergence map of a belief shows where message-passing is stopped. The divergence map after convergence illustrates the ideal support regions.

Assumptions and future work. The Bayesian approaches have the advantage over energy minimization techniques that all assumptions need to be made explicitly. In fact, three important assumptions (2,3,10) are made in our model in order to apply BP. Although good experimental results are obtained with our model, it is worth investigating when these assumptions break. Many other future directions can also be pursued. Naturally, we plan to extend our work to multi-baseline stereo. We are also investigating how to improve stereo matching with other Bayesian inference techniques based on Markov networks such as generalized BP.

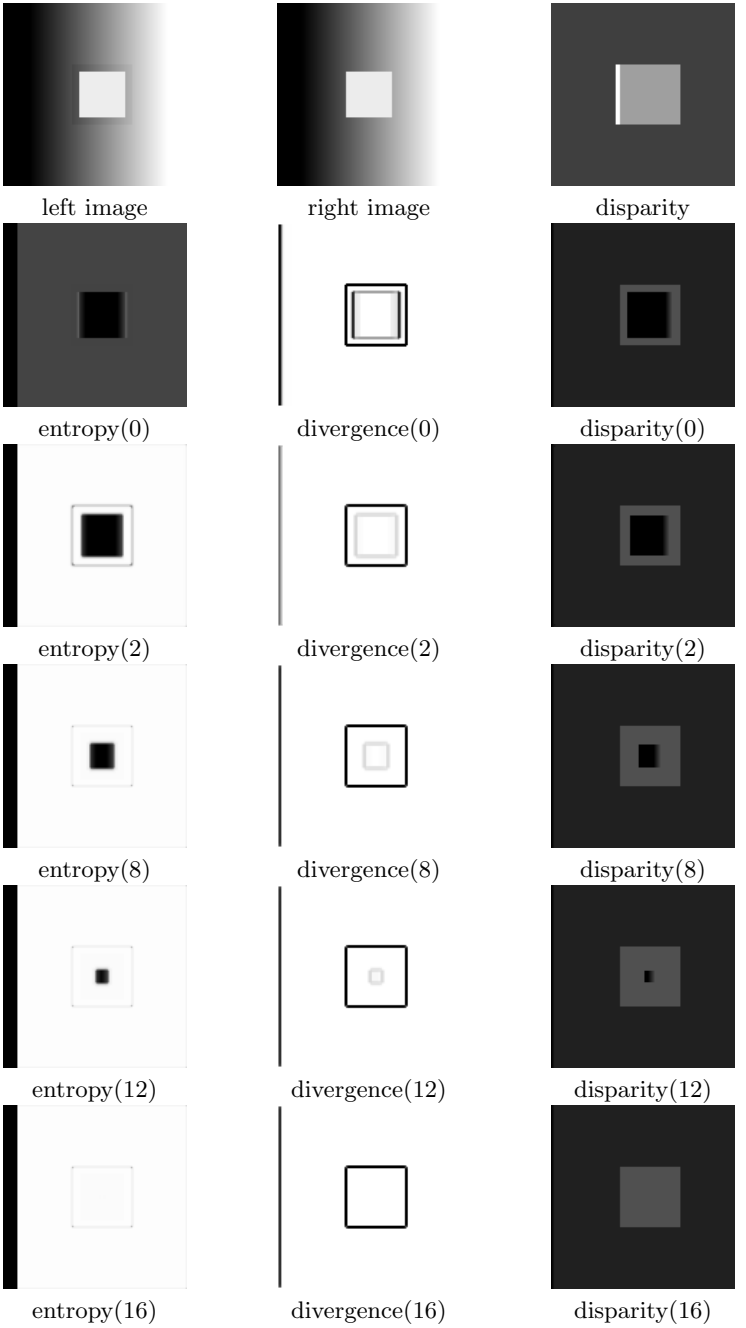


Fig. 7. Time-varying adaptive smoothing mechanism of the BP algorithm in stereo matching is illustrated from row 2 to row 6. The input image pair and the ground truth are shown in the first row. The number in the braces shows the iteration step.

References

1. P.N. Belhumeur. A bayesian-approach to binocular stereopsis. *IJCV*, 19(3):237–260, 1996.
2. S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *PAMI*, 20(4):401–406, 1998.
3. M.J. Black and A. Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *IJCV*, 19(1):57–91, 1996.
4. A. Blake and A Zisserman. *Visual reconstruction*. MIT Press, 1987.
5. A.F. Bobick and S.S. Intille. Large occlusion stereo. *IJCV*, 33(3):1–20, 1999.
6. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *ICCV*, 1999.
7. D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. *CVPR*, 1997.
8. I.J. Cox, S.L. Hingorani, S.B. Rao, and B.M. Maggs. A maximum-likelihood stereo algorithm. *CVIU*, 63(3):542–567, 1996.
9. W.T. Freeman, E.C. Pasztor, and O.T. Carmichael. Learning low-level vision. *IJCV*, 40(1):25–47, 2000.
10. D. Geiger and F. Girosi. Parallel and deterministic algorithms from mrfs: Surface reconstruction. *PAMI*, 13(5):401–412, 1991.
11. D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. *IJCV*, 14(3):211–226, 1995.
12. S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *PAMI*, 6(6):721–741, 1984.
13. H. Hirschmueller. Improvements in real-time correlation-based stereo vision. *IEEE Workshop on Stereo and Multi-Baseline Vision*, 2001.
14. B.K.P. Horn and M.J. Brooks. The variational approach to shape from shading. *CVGIP*, 33(2):174–208, 1986.
15. H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. *ECCV*, 1998.
16. T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *PAMI*, 16(9):920–932, 1994.
17. V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. *ICCV*, 2001.
18. S. Osher L.I. Rudin and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 27(60):259–268, 1992.
19. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, California, 1988.
20. D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. *IJCV*, 28(2):155–174, 1998.
21. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.
22. H. Tao, H.S. Sawhney, and R. Kumar. A global matching framework for stereo computation. *ICCV*, 2001.
23. O. Veksler. Stereo matching by compact windows via minimum ratio cycle. *ICCV*, 2001.
24. W. T. Yedidia, J. S. Freeman and Weiss Y. Bethe free energy, kikuchi approximations, and belief propagation algorithms. Technical Report TR-2001-16, Mitsubishi Electric Reseach, 2001.