

Stereo Matching Using Belief Propagation

Jian Sun, Nan-Ning Zheng, *Senior Member, IEEE*, and
Heung-Yeung Shum, *Senior Member, IEEE*

Abstract—In this paper, we formulate the stereo matching problem as a Markov network and solve it using Bayesian belief propagation. The stereo Markov network consists of three coupled Markov random fields that model the following: a smooth field for depth/disparity, a line process for depth discontinuity, and a binary process for occlusion. After eliminating the line process and the binary process by introducing two robust functions, we apply the belief propagation algorithm to obtain the maximum a posteriori (MAP) estimation in the Markov network. Other low-level visual cues (e.g., image segmentation) can also be easily incorporated in our stereo model to obtain better stereo results. Experiments demonstrate that our methods are comparable to the state-of-the-art stereo algorithms for many test cases.

Index Terms—Stereoscopic vision, belief propagation, Markov network, Bayesian inference.

1 INTRODUCTION

STEREO vision infers 3D scene geometry from two images with different viewpoints. This fundamental problem has been investigated for many years not only in computer vision but also in cognitive science and psychophysiology. Recent applications such as view synthesis and image-based rendering make stereo vision again an active research topic in computer vision.

Classical dense two-frame stereo matching computes a dense disparity or depth map from a pair of images under known camera configuration. In general, the scene is assumed Lambertian or intensity-consistent from different viewpoints, without specularities, reflective surfaces, or transparency. The known camera configuration can provide a powerful epipolar geometry constraint for matching. Stereo matching remains a difficult vision problem for the following reasons.

- **Noise.** There are always unavoidable light variations, image blurring, and sensor noise in image formation. A practical stereo algorithm must be robust.
- **Textureless regions.** This is also called the *aperture problem*. The intensity-consistency constraint is useless in textureless regions. Thus, information from highly textured regions needs to be propagated into textureless regions for stereo matching, e.g., by using spatial smoothness constraint.
- **Depth discontinuities.** The spatial smoothness constraint should be broken at object (depth) boundaries. In other words, information propagation should stop at depth discontinuities.
- **Occlusions.** Occluded pixels in one view should not be matched with pixels in the other view.

Clearly, stereo matching is an ill-posed problem with inherent ambiguities. The Bayesian approach provides a promising way for such ill-posed problems because it treats a task as an inference problem or finding the “best guess” solution. For stereo matching, we want to infer scene structure S given images I . The output from the Bayesian approach is not only a single solution but also a posterior probability distribution $P(S|I)$. By Bayes law, $P(S|I) \propto P(I|S)P(S)$, where $P(I|S)$ is the likelihood that encodes the process of forward image formation and $P(S)$ is the prior that encodes our assumptions on scene structure.

The Bayesian approach has many advantages when applied to stereo vision. It can encode various prior constraints, e.g., spatial smoothness, uniqueness, and the ordering constraint. It can also deal with uncertainties in stereo matching. Because the Bayesian approach states explicitly what assumptions are made, the strengths and the weaknesses of the proposed algorithm can be clearly examined. In addition to stereoscopic vision, people also use other cues to infer scene structure, e.g., shape from shading, shape from shadows, shape from focus, shape from silhouette, and shape from texture. The Bayesian approach provides a natural way to integrate the information from multiple sensors.

There are two contributions in this paper. First, we formulate stereo matching using three MRF's and subsequently estimate the optimal solution by a Bayesian Belief Propagation algorithm. Second, we propose a probabilistic framework to integrate additional information (e.g., segmentation) into the stereo algorithm.

The rest of paper is organized as follows: After reviewing related work in Section 2, we propose in Section 3 a novel stereo matching approach to explicitly model discontinuities, occlusions, and the disparity field in the Bayesian framework. In Section 4, Bayesian Belief Propagation is applied to infer the stereo matching. The basic stereo model is then extended in Section 5 to integrate other cues such as region similarity. The experimental results shown in Section 6 demonstrate that our model is effective and efficient. In Section 7, we adapt the stereo model for multiview stereo. Finally, we discuss in Section 8 why our stereo matching with belief propagation can produce results that are comparable to the state-of-the-art stereo algorithms.

• J. Sun and N.-N. Zheng are with Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China.
E-mail: sjf@aiar.xjtu.edu.cn, nnzheng@xjtu.edu.cn.

• H.-Y. Shum is with Microsoft Research Asia, Beijing 100080, China.
E-mail: hshum@microsoft.com.

Manuscript received 24 July 2002; revised 25 Feb. 2003; accepted 3 Mar. 2003.

Recommended for acceptance by J.M. Rehg.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 118202.

2 RELATED WORK

In this section, we review related stereo algorithms, especially those using the Bayesian approach. We refer the reader to a more detailed and updated taxonomy of dense, two-frame stereo correspondence algorithms by Scharstein and Szeliski [30]. A testbed for quantitative evaluation of stereo algorithms is also given in [30].

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 size, shape, and weight distribution of aggregation support for each pixel. An ideal support region should be bigger in textureless regions and should be suspended at depth discontinuities. The central problem of global algorithms is not only to define a good objective function but also to provide an effective computing method to find local or global minimum. In the taxonomy of Scharstein and Szeliski [30], a local method consists of matching cost computation, aggregation of cost, and disparity computation; a global method consists of matching cost computation and disparity optimization. From the Bayesian point of view, matching cost computation is a measurement or observation. The most common matching costs, e.g., squared intensity difference (SD), absolute intensity difference [20], normalized-cross correlation [28], [7], binary matching cost [25], rank transform [35], shifted absolute difference [3], are ways of computing the likelihood function. Different aggregation methods reflect different priors assumed on scene structure. For example, a fixed-window method implies a frontal-plane scene, and a 3D window method limits the disparity gradient. Obviously, the fixed window is invalid at depth discontinuities. Some improved window-based methods, such as adaptive windows [20] and shiftable windows [6], [33], [21] try to avoid windows that span depth discontinuities.

Bayesian methods (e.g., [13], [18], [2], [10], [6]) are global methods that model discontinuities and occlusion. Bayesian methods can be classified into two categories: dynamic programming-based or MRFs-based, depending on the computation model. Geiger et al. [13] and Ishikawa and Geiger [18] 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 where the global optimum is obtained by dynamic programming. Belhumeur [2] defined a set of priors from a simple scene to a complex scene. A simplified relationship between disparity and occlusion is used to solve scanline matching by dynamic programming. Unlike Geiger and Belhumeur who enforced a piecewise-smooth constraint, Cox et al. [10] and Bobick and Intille [6] did not require the smoothing prior. Assuming corresponding features are normally distributed and a fixed cost for occlusion, Cox proposed a dynamic programming solution using only the occlusion constraint and ordering constraints. Bobick and Intille incorporated the Ground Control Points constraint to reduce the sensitivity to occlusion cost and the computation complexity of Cox's method. These dynamic programming methods assume that the occlusion cost is the same in each scanline.

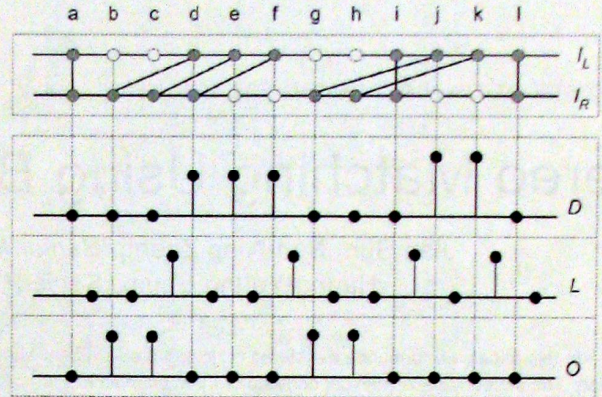


Fig. 1. A scene illustrates the geometric relationship among depth process D , discontinuity process L , and occlusion process O . Matched points between I_L (the reference view) and the right image I_R are connected by thick lines.

Ignoring the dependence between scanlines results in the characteristic "streaking" in the disparity maps.

Markov Random Fields (MRF) is a powerful tool to model spatial interaction. 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 [14], Mean-Field annealing [12], the Graduated Non-Convexity algorithm (GNC) [5], and Variational approximation [17]. Finding a solution by simulated annealing can often take an unacceptably long time although global optimization is achievable 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. Recently, the Graph Cut (GC) method [8] has been proposed based on the max flow algorithm in graph theory. This method is a fast efficient algorithm to find a local minimum for a MAP-MRF whose energy function is Potts or Generalized Potts.

The absence of an efficient stochastic computing method has made probabilistic models less attractive. In this paper, we formulate a probabilistic stereo model that can be efficiently solved by a Bayesian Belief Propagation algorithm.

3 BASIC STEREO MODEL

We model stereo matching by three coupled MRF's: D is the smooth disparity field defined on the image lattice of the reference view, L is a spatial line process located on the dual of the image lattice and represents 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. Fig. 1 illustrates these processes in the 1D case.

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, I_R is the left (reference) and right images, respectively, is:

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

Without occlusion, $\{D, L\}$ are coupled MRF's proposed by [14] to model a piecewise-smooth surface with two random fields: one representing the variable required to estimate, the other representing its discontinuities. Similar models such as the "weak membrane" model [5] in surface reconstruction and the "Mumford-Shah" model in image segmentation [26] have also been studied in computer vision. However, in image formation of stereo pairs, the piecewise-smooth scene is projected on a pair of stereo images. Some regions are only visible in one image. Each pixel in the occlusion region has no matching pixel in the other view. For example, in Fig. 1, points b, c, g, h from I_L cannot be matched in I_R . Adding occlusion process O into the piecewise-smooth model $\{D, L\}$ is therefore necessary.

3.1 Likelihood

We assume that the likelihood $P(I|D, O, L)$ is independent of L ,

$$P(I|D, O, L) = P(I|D, O) \quad (2)$$

because the observation (I) is pixel-based. Assuming that the observation noise follows an independent identical distribution (i.i.d.), we can define the likelihood $P(I|D, O)$ as:

$$P(I|D, O) \propto \prod_{s \notin O} \exp(-F(s, d_s, I)), \quad (3)$$

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 nonoccluded areas $\{s \notin O\}$ because likelihood in occluded areas cannot be well defined.

For the matching cost, we use Birchfield and Tomasi's pixel dissimilarity, which is provably insensitive to image sampling [3]:

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

where

$$\bar{d}(s, s', I) = \min\{|I_L(s) - I_R^-(s')|, |I_L(s) - I_R(s')|, |I_L(s) - I_R^+(s')|\},$$

s' is the matching pixel of s in the right view 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 the linearly interpolated intensity halfway between s' and its neighboring pixel to the right, $\bar{d}(s', s, I)$ is the symmetric version of $\bar{d}(s, s', I)$, and σ_f is the image noise variance to be estimated.

3.2 Prior

There is no simple statistical relationship between coupled fields $\{D, L\}$ and field O . The ordering constraint [1] assumes that the order of neighboring correspondences is always preserved. This ordering allows the construction of a dynamic programming scheme. However, this constraint may not always be true. For instance, this constraint is violated when a thin object is close to the viewer. As shown in Fig. 1, a thin object $\{j, k\}$ causes the order of points i and j in I_L to be different from that of their matched points in I_R .

In this paper, we ignore the statistical dependence between O and $\{D, L\}$ and assume that:

$$P(D, O, L) = P(D, L)P(O). \quad (4)$$

The Markov property asserts that the conditional probability of a site in the field depends only on its neighboring sites. Assuming D, L , and O follow the Markov property, by specifying the first order neighborhood system $G(s)$ and $N(s) = \{t | t > s, t \in G(s)\}$ of site s , the prior (4) can be expanded as:

$$P(D, L, O) \propto \prod_s \prod_{t \in N(s)} \exp(-\varphi_c(d_s, d_t, l_{s,t})) \prod_s \exp(-\eta_c(o_s)), \quad (5)$$

where $\varphi_c(d_s, d_t, l_{s,t})$ is the joint clique potential function of sites d_s, d_t (neighbor of d_s) and $l_{s,t}$. $l_{s,t}$ is the line variable between d_s and d_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 enforce the contextual constraints for stereo matching. To enforce spatial interactions between d_s and $l_{s,t}$, we define $\varphi_c(d_s, d_t, l_{s,t})$ as follows:

$$\varphi_c(d_s, d_t, l_{s,t}) = \varphi(d_s, d_t)(1 - l_{s,t}) + \gamma(l_{s,t}), \quad (6)$$

where $\varphi(d_s, d_t)$ penalizes the different assignments of neighboring sites when no discontinuity exists between them and $\gamma(l_{s,t})$ penalizes the occurrence of a discontinuity between sites s and t . Typically, $\gamma(0) = 0$.

By combining (3), (5), and (6), our basic stereo model (1) becomes:

$$\begin{aligned} \max_{\text{posterior}} P(D, O, L|I) &\propto \prod_{s \notin O} \exp(-F(s, d_s, I)) \prod_s \exp(-\eta_c(o_s)) \\ &\prod_s \prod_{t \in N(s)} \exp(-(\varphi(d_s, d_t)(1 - l_{s,t}) + \gamma(l_{s,t}))). \end{aligned} \quad (7)$$

4 APPROXIMATE INFERENCE BY BELIEF PROPAGATION

To find the MAP solution of (7), we need to:

- determine the forms and parameters of $\varphi(d_s, d_t)$, $\gamma(l_{s,t})$, and $\eta_c(o_s)$ and
- provide a tractable inference algorithm.

It is, however, nontrivial to specify or to learn appropriate forms and parameters of $\varphi(d_s, d_t)$, $\gamma(l_{s,t})$, and, especially, $\eta_c(o_s)$. Even if the forms and parameters are given, it is still difficult to find the MAP of a composition of a continuous MRFs D and two binary MRFs L and O . Although the Markov Chain Monte Carlo (MCMC) [14], [15] approach provides an effective way to explore a posterior distribution, the computational requirement makes MCMC impractical for stereo matching. The solution space of our model is $\Omega = \Omega_d \times \Omega_l \times \Omega_o$, where Ω_d , Ω_l , and Ω_o are the solution spaces of depth, discontinuity, and occlusion, respectively.

This is why we need to make some approximations on both the model and algorithm. In Section 4.1, the unification of line process and robust statistics [4] provides us a way to eliminate the binary random variable from our MAP problem. In Section 4.2, after converting MRFs to the corresponding Markov network, the approximate inference algorithm, a loopy belief propagation algorithm can be used to approximate the posterior probability for stereo matching.

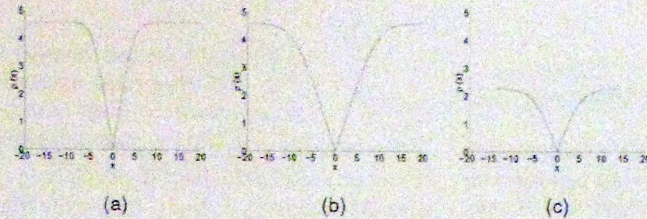


Fig. 2. (a) $\epsilon = 0.01, \sigma = 1.0$. (b) $\epsilon = 0.01, \sigma = 2.0$. (c) $\epsilon = 0.1, \sigma = 2.0$. Robust function $\rho(x) = -\ln((1 - \epsilon) \exp(-\frac{|x|}{\sigma}) + \epsilon)$ derived from TV model. Parameters σ and ϵ , respectively, control the sharpness and the upper-bound of the function.

4.1 Model Approximation: From Line Process to Outlier Process

Maximization of the posterior (7) can be rewritten as

$$\max_{D,L,O} P(D, L, O|I) = \max_D \left\{ \max_O \prod_s \exp(- (F(s, d_s, I)(1 - o_s) + \eta_c(o_s)o_s)) \right. \quad (8)$$

$$\left. \max_L \prod_s \prod_{t \in N(s)} \exp(- (\varphi(d_s, d_t)(1 - l_{s,t}) + \gamma(l_{s,t}))) \right\}$$

because the first two factors on the r.h.s of (7) are independent of L and the last factor on the r.h.s of (7) is independent of O .

Now, we relax the binary processes $l_{s,t}$ and o_s to analog processes $l_{s,t}^a$ and o_s^a ("outlier process" [4]) by allowing $0 \leq l_{s,t}^a \leq 1$ and $0 \leq o_s^a \leq 1$. For the first term in (8),

$$\max_O \prod_s \exp(- (F(s, d_s, I)(1 - o_s^a) + \eta_c(o_s^a)o_s^a)) \quad (9)$$

$$= \exp(- \min_o \sum_s (F(s, d_s, I)(1 - o_s^a) + \eta_c(o_s^a)o_s^a)),$$

where $\min_o \sum_s (F(s, d_s, I)(1 - o_s^a) + \eta_c(o_s^a)o_s^a)$ is the objective function of a robust estimator. The robust function of this robust estimator [4] is

$$\rho_d(d_s) = \min_{o_s^a} (F(s, d_s, I)(1 - o_s^a) + \eta_c(o_s^a)o_s^a). \quad (10)$$

For the second term in (8), we also have a robust function $\rho_p(d_s, d_t)$:

$$\rho_p(d_s, d_t) = \min_{l_{s,t}^a} (\varphi(d_s, d_t)(1 - l_{s,t}^a) + \gamma(l_{s,t}^a)). \quad (11)$$

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

$$P(D|I) \propto \prod_s \exp(-\rho_d(d_s)) \prod_s \prod_{t \in N(s)} \exp(-\rho_p(d_s, d_t)). \quad (12)$$

Thus, we not only eliminate two analog line processes via the outlier process but also model outliers in measurements. We convert the task of modeling the prior terms $\{\eta_c(o_s), \varphi(d_s, d_t), \gamma(l_{s,t})\}$ explicitly into defining two robust functions $\rho_d(d_s)$ and $\rho_p(d_s, d_t)$ that model occlusion and discontinuity implicitly.

In this paper, our robust functions are derived from the Total Variance (TV) model [23] with the potential function $\rho(x) = |x|$ because of its discontinuity preserving property. We truncate this potential function as our robust function.

$$\rho_d(d_s) = -\ln \left((1 - \epsilon_d) \exp \left(-\frac{|F(s, d_s, I)|}{\sigma_d} \right) + \epsilon_d \right). \quad (13)$$

$$\rho_p(d_s, d_t) = -\ln \left((1 - \epsilon_p) \exp \left(-\frac{|d_s - d_t|}{\sigma_p} \right) + \epsilon_p \right). \quad (14)$$

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

After approximating the model, the next task is to provide an effective and efficient inference algorithm. We describe below how the belief propagation algorithm is used to compute the MAP of the posterior distribution (12).

4.2 Algorithm Approximation: Loopy Belief Propagation

In the literature of probabilistic graph models [19], a Markov network is an undirected graph as shown in Fig. 3. Nodes $\{x_s\}$ are hidden variables and nodes $\{y_s\}$ are observed variables. By denoting $X = \{x_s\}$ and $Y = \{y_s\}$, the posterior $P(X|Y)$ can be factorized as:

$$P(X|Y) \propto \prod_s \psi_s(x_s, y_s) \prod_s \prod_{t \in N(s)} \psi_{st}(x_s, x_t), \quad (15)$$

where $\psi_{st}(x_s, x_t)$ is called the compatibility matrix between nodes x_s and x_t , and $\psi_s(x_s, y_s)$ is called the local evidence for node x_s . In fact, $\psi_s(x_s, y_s)$ is the observation probability $p(y_s|x_s)$. If the number of discrete states of x_s is L , $\psi_{st}(x_s, x_t)$ is an $L \times L$ matrix and $\psi_s(x_s, y_s)$ is a vector with L elements.

It can be observed that the form of our posterior (12) is same as the form of (15). If we define

$$\psi_{st}(x_s, x_t) = \exp(-\rho_p(x_s, x_t)). \quad (16)$$

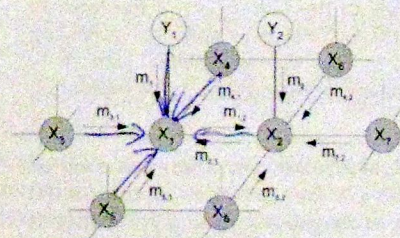
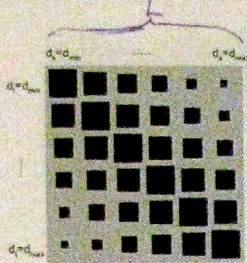


Fig. 3. Local message passing in a Markov Network. Gray nodes are hidden variables. White nodes are observable variables. In the "max-product" algorithm, the new message sent from node x_1 to x_2 is: $m_{1,2}^{new} \leftarrow \kappa \max_{x_1} \psi_{12}(x_1, x_2) m_{1,3} m_{1,4} m_{1,5}$. The belief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

$$\leftarrow m_{1,2}, m_{2,1}, m_{3,1}, m_{4,1}, m_{5,1}$$

belief ~ msg?



?
State: stay (less large)

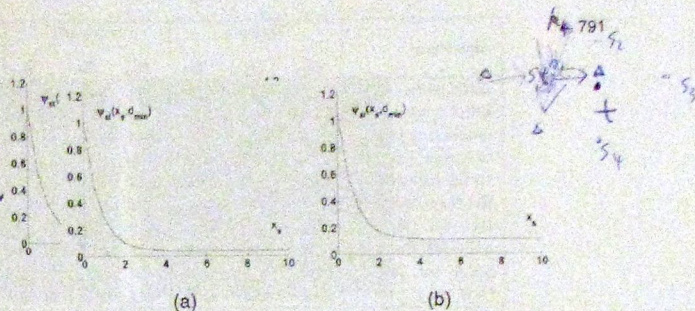


Fig. 4. Compatibility matrix $\psi_{st}(x_s, x_t)$. The range of disparity is $[d_{min}, d_{max}]$. A larger box represents a bigger value.

$$\psi_s(x_s, y_s) \propto \exp(-\rho_d(x_s)), \quad (17)$$

Fig. 5. (a) $seg(s) = seg(t)$. The left shows the first row of $\psi_{st}(x_s, x_t)$ when node x_s and x_t are in the same region. (b) $seg(s) \neq seg(t)$. The right shows the first row of $\psi_{st}(x_s, x_t)$ when node x_s and x_t are in different regions. ($\epsilon_p = 0.01, \sigma_p = 1.0, \lambda_{seg} = 0.05$).

our posterior (12) is exactly the posterior of a Markov network. Fig. 4 gives an illustration of $\psi_{st}(x_s, x_t)$ for our stereo model. Thus, finding the MAP of (12) is equal to finding the MAP of a Markov network.

For this Markov network, exact inference such as variable elimination is obviously intractable due to the large state space of D . Approximation methods include variational methods, sampling methods, bounded cutset conditioning, and parametric approximation methods [19]. In particular, loopy belief propagation is a linear time algorithm proportional to the number of hidden nodes. Loopy belief propagation applies Pearl's algorithm [27] to the graph that has loops. For graphs without loops, Pearl's algorithm is an exact inference method. For graph with loops, such as our Markov network for stereo matching, the belief propagation algorithm cannot guarantee the global optimal solution. Despite loops in the network, however, belief propagation has been applied successfully to some vision [11] and communication [34] problems recently.

Belief propagation (BP) 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 (in fact, $m_s(x_s, y_s) = \psi_s(x_s, y_s)$), and $b_s(x_s)$ 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 vectors with L elements. 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 $P(X|Y)$ of the network and the marginal posterior of each node $P(x_s|Y)$, respectively. The standard "max-product" algorithm is shown below:

1. Initialize all messages $m_{st}(x_t)$ as uniform distributions and messages $m_s(x_s) = \psi_s(x_s, y_s)$.

2. Update messages $m_{st}(x_t)$ iteratively for $i = 1:T$

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

Handwritten notes: x_s has L states, $x_s \in N(x_t)$, $x_k \in N(x_s)$, κ vary, expectation, move next.
3. Compute beliefs

$$b_s(x_s) \leftarrow \kappa m_s(x_s) \prod_{x_t \in N(x_s)} m_{ts}(x_s)$$

Handwritten notes: $x_s^{MAP} = \arg \max_{x_s} b_s(x_s)$, fixed for a spec, $x_k \in N(x_s)$, x_s choose a state, $x_k \in N(x_s)$, $b_s(x_s)$.

For example, in Fig. 3, the new message sent from node x_1 to x_2 is updated as: $m_{12}^{new} \leftarrow \kappa \max_{x_1} \psi_{12}(x_1, x_2) m_1 m_{3,1} m_{4,1} m_{5,1}$. 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 the component-wise product); κ is the normalization constant.

The computational complexity of a standard "max-product" BP algorithm is $O(TNL^2)$, where N is the number of pixels and T is the number of iterations. Most of the computation focuses on the multiplication of matrix $\psi_{st}(x_s, x_t)$ and vector $m_s(x_s) \prod_{x_k \in N(x_s) \setminus x_t} m_{ks}(x_s)$. However, in our experiments, some statistical properties of messages can be used to speed up belief propagation.

Propagation Speedup. It can be observed that each row of $\psi_{st}(x_s, x_t)$ is a unique peak distribution in our stereo model. In our experiments, most messages have unique peaks. We can exploit this property to identify unnecessary computation during iterations. We simplify matrix $\psi_{st}(x_s, x_t)$ as $[a_1^T, \dots, a_t^T]^T$, $m_s(x_s) \prod_{x_k \in N(x_s) \setminus x_t} m_{ks}(x_s)$ as b and $m_{st}^{i+1}(x_t)$ as c . The message update at one iteration is:

$$c(i) = \arg \max_j a_i(j) \cdot b(j). \quad (18)$$

$$m_s(x_s, y_s) = \psi_s(x_s, y_s) = m_s \bullet (x_s)$$

TABLE 1
Quantitative Statistics Based on Known Ground Truth Data

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

$\delta_d = 1$ in all our experiments.

Fig. 4