Available online at www.sciencedirect.com

ScienceDirect

Computer Vision and Image Understanding 106 (2007) 85–96

Computer Vision and Image Understanding

www.elsevier.com/locate/cviu

Bayesian stereo matching

Li Cheng [a,b,*], Terry Caelli [b]

[a] Department of Computing Science, University of Alberta, Edmonton, Alta., Canada T6G 2E8
[b] National ICT Australia, Research School of Information Science and Engineering, Australian National University, Canberra ACT 2601, Australia

Received 17 November 2004; accepted 8 September 2005
Available online 12 February 2007
Communicated by Arthur E.C. Pece

Abstract

A Bayesian framework is proposed for stereo vision where solutions to both the model parameters and the disparity map are posed in terms of predictions of latent variables, given the observed stereo images. A mixed sampling and deterministic strategy is adopted to balance between effectiveness and efficiency: the parameters are estimated via Markov Chain Monte Carlo sampling techniques and the Maximum A Posteriori (MAP) disparity map is inferred by a deterministic approximation algorithm. Additionally, a new method is provided to evaluate the partition function of the associated Markov random field model. Encouraging results are obtained on a standard set of stereo images as well as on synthetic forest images.

Keywords: Generative model; Stereo vision; Monte Carlo sampling; Bayesian analysis; Markov random field

1. Introduction

The goal of stereo matching is to infer the optimal disparity map for a given pair of images. Unfortunately, hand-crafting of model parameters is often necessary to ensure satisfactory results for specific image pairs [1]. A remedy is to adopt the Bayesian paradigm which naturally solves this problem of automatic parameter tuning, by treating both the unknown disparity map and the related parameters as random variables. The problem is then to infer the optimal distributions of the random variables. The merit of this scheme has been demonstrated in the related area of medical image processing by Higdon et al. [2]. However, the Bayesian approach is typically computationally demanding due to the use of sampling algorithms to explore the space of plausible distributions.

We propose the use of a generative Bayesian framework for stereo matching, which addresses the inference of disparity map and the estimation of parameters under a unified scheme. Further, efficient Markov Chain Monte Carlo (MCMC) methods [3] are proposed for parameter estimation, and a deterministic approximation algorithm, loopy belief propagation (LBP)[1] [6], is adopted to infer the disparity map.

Recently, a number of optimization methods have been used to solve the stereo problem. These include using simulated annealing [7], dynamic programming [8] and LBP [9] to infer the optimal disparity map. However, unlike the proposed method, existing models are not fully Bayesian, and their solution techniques are substantially different. The novel contributions of this work are threefold. First, stereo matching is explicitly addressed as a generative process, as illustrated in Fig. 1. Second, a Bayesian framework that naturally unifies the tasks of inferring the disparity

* Corresponding author. Present address: National ICT Australia, Research School of Information Science and Engineering, Australian National University, Canberra ACT 2601, Australia. Fax: +61 2 6230 7499.

E-mail addresses: Li.Cheng@anu.edu.au (L. Cheng), Terry.Caelli@anu.edu.au (T. Caelli).

[1] Two other papers of this issue [5] also use the LBP algorithm.

1077-3142/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2005.09.009

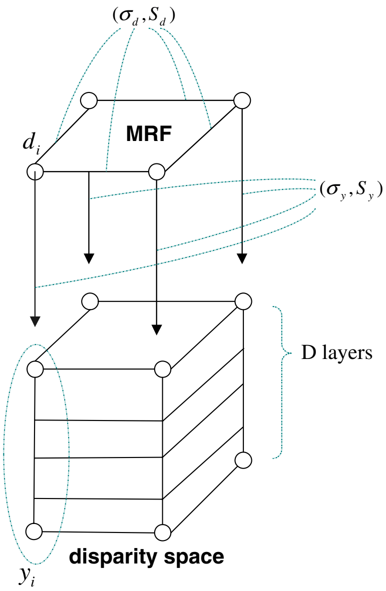

Fig. 1. A 2×2 2D lattice example that illustrates the proposed generative model for stereo matching. On the bottom, the 3D disparity space y is compiled by measuring the pixelwise dissimilarities of the left and right images with respect to shifts along the epipolar line. On the top, the disparity map d is modelled as a Markov random field. For a node i, given the latent disparity d_i, the pre-compiled observation y_i is independent of the rest of the disparity space y.

map and estimating the model parameters, is proposed. Third, a new method, based on the path sampling approach [10], is derived to evaluate the partition function of the underlying Markov random field (MRF). In particular, the proposed evaluation method is shown to bear theoretical advantages over both the coding and the pseudo-likelihood method [11]. Moreover, it greatly reduces the computational load when integrated into the MCMC samplers, and empirical experiments demonstrate the convergence behaviors of the proposed mixing strategy.

The Bayesian model is presented in Section 2, followed by a mixed updating strategy in Section 3. Details regarding the coding and the pseudo-likelihood methods are shown in Appendix C.1, and details of the proposed partition function evaluation method are presented in Appendix C.2. Finally, experiments are conducted in Section 4, with an empirical analysis of convergence behavior of the proposed approach addressed in Section 5.

2. The Generative Model

We assume a dense binocular stereo setting (e.g. [1]), where two views (left and right images, rectified to satisfy the epipolar constraint) of the same scene are presented. With the left image being the reference view, the task is to infer the disparity of each pixel, and to automatically estimate the model parameters for the image pair. This model, however, could be easily extended to more general scenarios.

Let $i = 1, \ldots, n$ index a 2D lattice of image pixels. Let $y = \{y_i\}$ denote a 3D disparity space with each y_i a vector of length D, where D is the range of possible disparity values. Essentially, y stores sufficient statistics about the input images, with each layer (see Fig. 1) storing the pixelwise dissimilarities of the two images, after shifting the left image horizontally a certain number of pixels. Therefore, y is referred to as the "*observed*" disparity space. The disparity map $d = \{d_i \in \{1, \ldots, D\}\}$ is modelled as a Markov random field (MRF) [12]. The proposed model consists of two components: the sensor model and the prior model. For the sensor model, $p(y|\, d, \sigma_y, s_y)$ captures the statistical dependencies of the observation y on the latent disparity MRF d, while the prior model $p(d|\sigma_d, s_d)$ addresses the neighboring dependencies within the disparity map. For convenience, denote the model parameters as $\theta = \{\sigma_y, s_y, \sigma_d, s_d\}$, with (σ_d, s_d) parameters of the prior model and (σ_y, s_y) parameters of the sensor model.

Because of the uncertainty of θ for different image pairs (see Fig. 1), Bayesian theory [13] treats θ as unknown and assigns a prior distribution for θ. By establishing the likelihood $p(y|\, d, \theta)$, the priors $p(d \mid \theta)$ and $p(\theta)$, the joint posterior is defined as

$$p(d, \theta|y) \propto p(y|d, \theta)p(d|\theta)p(\theta). \tag{1}$$

Our task is then twofold. First, we want to infer the MAP disparity map d^*:

$$d^* = \arg\max_d p(d|\theta^*, y), \tag{2}$$

where θ^* denotes the optimal parameter estimate. Second, we have to estimate the model parameters θ by its expectation

$$\theta^* = \int_\theta \theta p(\theta|d, y) \mathrm{d}\theta. \tag{3}$$

2.1. The Sensor and the Prior Models

Given the random variable $x \in \mathbb{R}^n$ and parameters (σ, s), we consider a class of density functions [2]

$$p(x|\sigma, s) = \frac{1}{\sigma^n z(s)} \exp\left\{-\frac{1}{s} u(x|\sigma, s)\right\}. \tag{4}$$

Here $u(x|\sigma, s) = \sum_i \rho(x_i/\sigma, s)$ is the energy function, and $z(s)$ is the normalization constant to ensure $p(x|\sigma,s)$ a valid density distribution. $\rho(\cdot, \cdot)$ is the potential function, with scale parameter $\sigma \in (0, \infty)$ and shape parameter $s \in (0, 2]$. We further decompose $x = \{x_i\}_{i=1}^n$ to represent a random field that could be either the MRF d or the disparity space y.

One reason for choosing this type of function is that the potential function, $\rho(\cdot, \cdot)$, unifies many existing function forms, both convex and non-convex, into one general representation [2]. In particular, it includes the generalized Gaussian distribution, when the potential function admits the following form,

$$\rho\left(\frac{x}{\sigma}, s\right) = |\frac{x}{\sigma}|^s, \tag{5}$$

when $s = 2$ we have the Gaussian distribution.

In Fig. 2, the two panels in each row show the effect of varying the shape parameter s, and the two panels in each

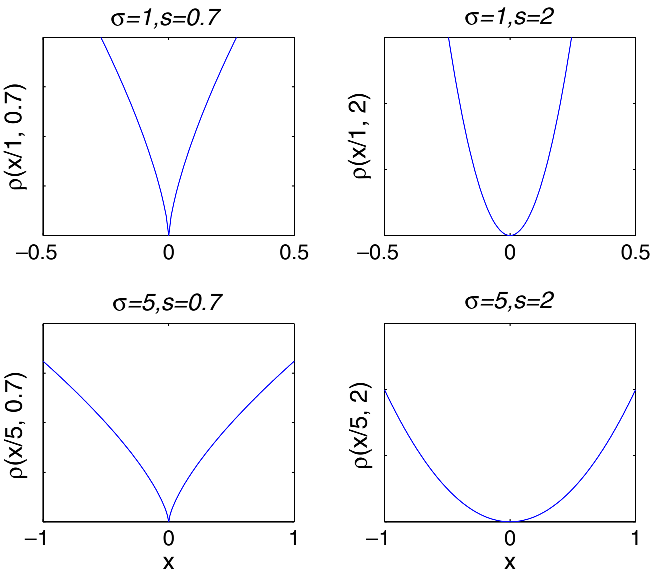

Fig. 2. Four examples of the potential function $\rho(x/\sigma, s)$ of the generalized Gaussian density function. Top-left panel shows a non-convex function when taking the shape parameter $s = 0.7$ and the scale parameter $\sigma = 1$. Top-right is a convex Gaussian potential function with $s = 2$ and $\sigma = 1$. The bottom two panels have similar interpretations, except that the scale parameters are changed ($\sigma = 5$).

column show the effect of adjusting the scale parameter σ, when the rest remains intact. More precisely, the space of generalized Gaussian functions spans a spectrum of feasible density functions, varying from sharper non-convex functions with $s < 1$ (called Laplace or double exponential function when $s = 1$), to smoother convex functions when $s \to 2$ (Gaussian if $s = 2$). Note that the shape parameter s plays the role of controlling the (non-) convexity of the function $p(x|\sigma, s)$. At the same time, the scale parameter σ captures the variance of data to which the function $p(x|\sigma, s)$ is fitted.

The sensor model is defined by the conditional likelihood of the disparity space given the disparity map

$$p(y|d, \sigma_y, s_y) \propto \frac{1}{\sigma_y^n} \exp\left\{-\frac{u(y|\sigma_y, s_y)}{s_y}\right\}, \tag{6}$$

where the energy function is $u(y|\sigma_y, s_y) = \sum_i \rho(y_i/\sigma_y, s_y)$. The conditional independent property of y, as illustrated in Fig. 1, makes it trivial to compute the partition function of the sensor model.

The prior model $p(d|\theta)$ of the MRF is

$$p(d|\sigma_d, s_d) = \frac{1}{z(s_d)\sigma_d^n} \exp\left\{-\frac{u(d|\sigma_d, s_d)}{s_d}\right\}, \tag{7}$$

where the energy is

$$u(d|\sigma_d, s_d) = \sum_{j\sim k} \rho\left(\frac{d_j - d_k}{\sigma_d}, s_d\right). \tag{8}$$

Here $\{j \sim k\}$ indexes the set of Markovian interaction of neighboring pixels over the MRF d. Notice that $z(s_d)$ in Eq. (7) is the partition function, which is known to be computationally intractable for MRFs [14].

3. The mixed strategy for inference and parameter estimation

Given the proposed generative model (Fig. 1) and the generalized Gaussian function, we form the parameter space ($\theta = \{\sigma_y, \sigma_d, s_y, s_d\}$). For a stereo pair we seek efficient procedures to infer the optimal disparity map d as well as the parameters θ. This poses a rather computationally challenging problem which is tackled by adopting a mixed strategy containing both deterministic approximation and stochastic sampling components.

3.1. LBP for approximate inference of d

As a deterministic approximation algorithm, the LBP algorithm is closely related to the fixed points of Bethe free energy which is well-studied in statistical physics [6], yet it is comparably easy to implement. In this section, we describe how to apply LBP approximation in our setting.

Define $\mathcal{N}_i$ as the set of neighboring nodes of i and $\mathcal{N}_j \setminus i$ as the set of neighboring nodes of j excluding node i. First, we derive the posterior over the MRF d, by assuming conditional independence of the nodes y_i given d_i, as

$$\begin{aligned} p(d|y,\theta) &\propto p(y|d,\theta)p(d|\theta) \\ &\propto \prod_i p(y_i|d_i,\theta) \prod_{j\in\mathcal{N}_i} \psi_e(d_i,d_j), \end{aligned} \tag{9}$$

where $\psi_e(d_i, d_j) \triangleq \exp\{-1/s_d \times \rho((d_j - d_i)/\sigma_d, s_d)\}$ models the interaction between d_i and d_j, which are the neighboring nodes in MRF d. The belief associated with node d_i is:

$$b(d_i) \propto p(y_i|d_i) \prod_{j\in\mathcal{N}_i} m_i^j(d_i), \tag{10}$$

where the message update rules are

$$m_i^j(d_i) \propto \sum_{d_j} \psi_e(d_i, d_j) p(y_j|d_j) \prod_{k\in\mathcal{N}_j\setminus i} m_j^k(d_j). \tag{11}$$

After computing the belief at node i, we have

$$\begin{aligned} b(d_i) &\propto \sum_{1,\ldots,i-1,i+1,\ldots,n} \prod_j p(y_j|d_j) \prod_{k\in\mathcal{N}_j} \psi_e(d_j, d_k) \\ &\propto p(d_i|y_i, \theta). \end{aligned} \tag{12}$$

This implies that the single belief $b(d_i)$ approximates the marginal probability $p(d_i|y_i, \theta)$ in MRF d. In networks without loops, beliefs are exactly inferred [6].

Second, it is easy to derive that the joint probabilities over the latent MRF d can be factorized as the product of beliefs over d

$$\begin{aligned} p(d|y,\theta) &\propto \prod_i p(y_i|d_i,\theta) \prod_{j\in\mathcal{N}_i} m_i^j(d_i) \\ &\propto \prod_i b(d_i). \end{aligned} \tag{13}$$

The LBP algorithm is thus implemented simply by (1) iteratively updating the messages by employing Eq. (11) over all node variables $\{d_i\}$ of the MRF, and (2) computing the beliefs from Eq. (10).

3.2. Updating σ_y and σ_d

Since there is no prior knowledge of σ_y, we assume a non-informative prior [13], which is defined in Appendix B and amounts to allocating equal weights to all possible hypotheses in the parameter space. From the proposed generative model (Fig. 1), we are set to derive the full conditional probability of σ_y as

$$\begin{aligned} p(\sigma_y^{s_y}|\cdot) &\propto p(\sigma_y^{s_y})p(y|d,\sigma_y,s_y) \\ &\propto \frac{1}{\sigma_y^{n+2}}\exp\left\{-\frac{1}{s_y}\sum_i \rho\left(\frac{y_i}{\sigma_y},s_y\right)\right\}, \end{aligned} \tag{14}$$

where the full conditional probability is defined in Appendix A. The scale parameter σ_y is gamma distributed as in [13]

$$(\sigma_y^{-s_y}|\cdot) \sim \gamma\left(\frac{n_y}{2},\frac{n_y v_y}{2}\right), \tag{15}$$

where n_y is the number of pixels and $v_y = \sum_i |y_i|^{s_y}/n_y$.$\sigma_d$ is updated similarly with

$$(\sigma_d^{-s_d}|\cdot) \sim \gamma\left(\frac{n_d}{2},\frac{n_d v_d}{2}\right), \tag{16}$$

where n_d is the number of edges and $v_d = \sum_{j\sim k}|d_j - d_k|^{s_d}/n_d$. The detailed derivations that lead to the aforementioned update rules are described in Appendix B.

3.3. Updating s_y

With no prior knowledge of s_y, we assume a uniform prior distribution over s_y. This allows us to derive the full conditional probability of s_y as

$$\begin{aligned} p(s_y|\cdot) &\propto p(s_y)p(y|d,\sigma_y,s_y) \\ &\propto \frac{1}{\sigma_y^{n}}\exp\left\{-\frac{1}{s_y}\sum_i \rho\left(\frac{y_i}{\sigma_y},s_y\right)\right\}. \end{aligned} \tag{17}$$

At step t, to ensure that s_y is drawn from the $(0,2]$ interval, we use the Metropolis sampling algorithm [3], with the proposal distribution

$$s_y^* \sim U[s_y^{(t-1)} - c, s_y^{(t-1)} + c]. \tag{18}$$

Here $U[\cdot,\cdot]$ denotes the bounded uniform distribution, $t-1$ refers to the previous step, and c is chosen such that s_y^* is accepted roughly half of the time. Hence, the candidate s_y^* is accepted with probability:

$$r = \min\{1, \alpha(s_y^{(t-1)}, s_y^*)\}, \tag{19}$$

where

$$\begin{aligned} \alpha(s_y^{(t-1)},s_y^*) &= \frac{p(s_y^*|\cdot)}{p(s_y^{(t-1)}|\cdot)} \\ &= \exp\left\{-\left(\frac{\sum\limits_i \rho\left(\frac{y_i}{\sigma_y},s_y^*\right)}{s_y^*} - \frac{\sum\limits_i \rho\left(\frac{y_i}{\sigma_y},s_y^{(t-1)}\right)}{s_y^{(t-1)}}\right)\right\}. \end{aligned} \tag{20}$$

In other words, $s_y^{(t)} = s_y^*$ with probability r, and $s_y^{(t)} = s_y^{(t-1)}$ with probability $1-r$.

3.4. Updating s_d

Due to the existence of the partition function $z(s_d)$, for the MRF d, updating the model parameter s_d is more involved. Algebraically, the full conditional probability of s_d is

$$\begin{aligned} p(s_d|\cdot) &\propto p(s_d)p(d|\sigma_d,s_d) \\ &\propto \frac{\exp\left\{-\frac{1}{s_d}\sum_{j\sim k}\rho\left(\frac{d_j-d_k}{\sigma_d},s_d\right)\right\}}{z(s_d)\sigma_d^n}. \end{aligned} \tag{21}$$

To compute the full conditional probability of s_d we have to evaluate the partition function $z(s_d)$—at least up to some constant of proportionality. Based on the path sampling paradigm, a new method is proposed to evaluate the log-ratio of partition functions (see Appendix C.2 for details). To update s_d, at step t, the candidate s_d^* is randomly accepted with probability

$$r = \min\{1, \alpha(s_d^{(t-1)}, s_d^*)\}, \tag{22}$$

where

$$\begin{aligned} \alpha(s_d^{(t-1)},s_d^*) =& \frac{z(s_d^{(t-1)})}{z(s_d^*)} \\ &\times \exp\left\{-\left(\frac{\sum_{j\sim k}\rho\left(\frac{d_j-d_k}{\sigma_d},s_d^*\right)}{s_d^*} - \frac{\sum_{j\sim k}\rho\left(\frac{d_j-d_k}{\sigma_d},s_d^{(t-1)}\right)}{s_d^{(t-1)}}\right)\right\}. \end{aligned} \tag{23}$$

Here $z(s_d^{(t-1)})$ and $z(s_d^*)$ correspond to the partition functions for the previous and proposed parameters of s_d, respectively. As described in Appendix C.2, we denote $\lambda(s_d^{(t-1)},s_d^*) \triangleq \log z(s_d^{(t-1)}) - \log z(s_d^*)$ and rewrite

$$\begin{aligned} \alpha(s_d^{(t-1)},s_d^*) = \exp\Bigg\{&-\lambda(s_d^{(t-1)},s_d^*) \\ &-\left(\frac{\sum_{j\sim k}\rho\left(\frac{d_j-d_k}{\sigma_d},s_d^*\right)}{s_d^*} - \frac{\sum_{j\sim k}\rho(\frac{d_j-d_k}{\sigma_d},s_d^{(t-1)})}{s_d^{(t-1)}}\right)\Bigg\}. \end{aligned} \tag{24}$$

3.5. The mixed updating strategy

An iterative procedure is employed to unify the deterministic approximate inference and the MCMC parameter estimation using a fixed sampling scheme:

1. Initialize $(d^{(0)}, \theta^{(0)})$, set $t = 0$.
2. At iteration t:
   - – Inference step: approximately infer the disparity map $d^{(t)}$ by LBP via Eqs. (10, 11).
   - – Estimation step: explore the $\theta^{(t)}$ distribution by drawing N_t cycles of MCMC kernel samples. The Gibbs sampling method is used to obtain *one cycle* of the kernels, as:
     - Sample σ_y according to Eq. (15),
     - Sample σ_d according to Eq. (16),
     - Sample s_y according to Eqs. (19) and (20),
     - Sample s_d according to Eqs. (22) and (23).
3. $t \leftarrow t + 1$, goto 2.

In practice $N_t = 4000$, and the algorithm terminates after 2–3 iterations, which are enough to ensure convergence.

Table 1
Table summarizes the rankings compared to existing methods on the four test pairs, where each row presents the same scenario as the corresponding row in Fig. 5

Test pairs	Tsukuba	Sawtooth	Venus	Map
Rankings on non-occluded regions	12	7	15	1
Rankings on textureless regions	14	9	16	
Rankings on discontinuous regions	17	10	14	1

The comparisons are conducted in March 2004.

4. Experimental results

Experiments are conducted on the well-known Middlebury testbed [1] with four stereo pairs: the “Tsukuba”, “Sawtooth”, “Venus” and the “Map” pairs (along with the evaluation methodology). In this evaluation methodology, the “bad-pixels” errors are defined as the “percentage of bad-pixels”, where each “bad-pixel” refers to a pixel where the absolute disparity error is greater than 1. In all, three types of ‘bad-pixels” errors are recorded: (1) the error accumulated over all pixels; (2) the error accumulated for the pixels in non-textured areas; and (3) the error accumulated for the pixels near depth discontinuities. In all three cases, only non-occluded pixels are considered (see Table 1).

According to the proposed update strategy described in Section 3.5, we first apply the Gaussian models to the testbed pairs with fixed model parameters. Two of them are shown (left view image only) in the top-left corners of Figs. 3, and 4, respectively. Figs. 3(b), 4(b), 3(c) and 4(c) show the obtained MAP disparity maps with different sets of Gaussian parameters. With fixed Gaussians, we obtain the inferred disparity maps in Fig. 3(b), (c), 4(b). By estimating the scale parameters $\{\sigma_d, \sigma_y\}$, we obtain improved disparity map estimates as shown in Fig. 4(c). As expected, we observe that the Gaussian models tended to oversmooth edges. By estimating θ from the data, we obtain the results in Figs. 3(d) and 4(d), where the inferred disparity maps preserve sharp disparities along

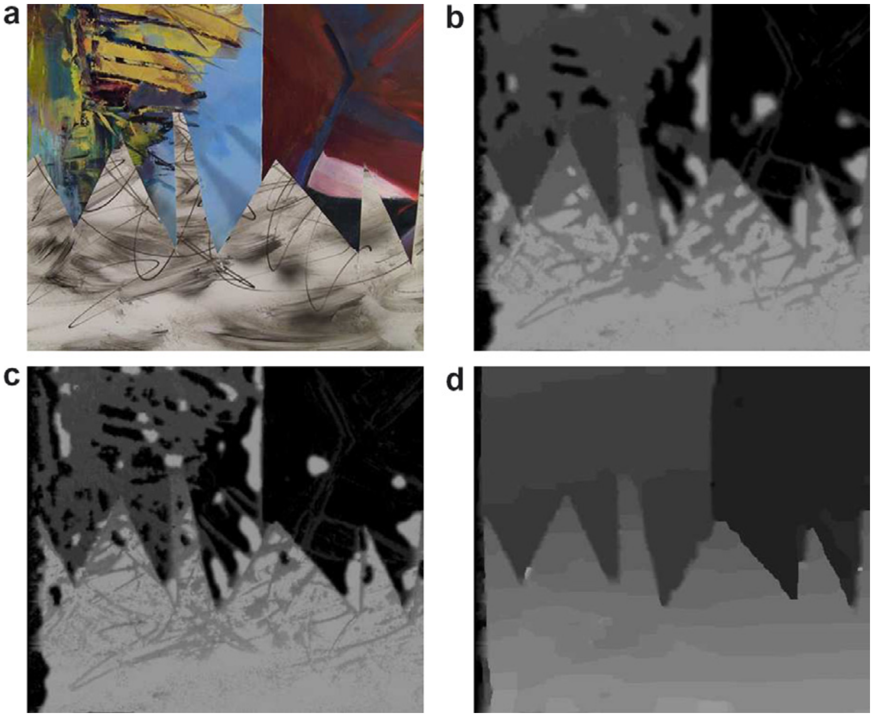

Fig. 3. Experimental results for the “Sawtooth” stereo pair. (a) Presents the left view of the stereo pair. The resultant disparity map is shown in (b) for fixed parameters $(s_d, \sigma_d, s_y, \sigma_y) = (2, 2, 2, 6)$. (c) is similarly obtained by fixing $(s_d, \sigma_d, s_y, \sigma_y) = (2, 5, 2, 10)$. Using the Monte Carlo samplers for parameter estimation, the model parameters converge to $(1.98, 0.26, 1.11, 3.07)$ after a few iterations, with the fixed-point inferred disparity map shown in (d). Notice that both the estimated parameters and the inferred disparity map converge regardless of different starting values of (d).

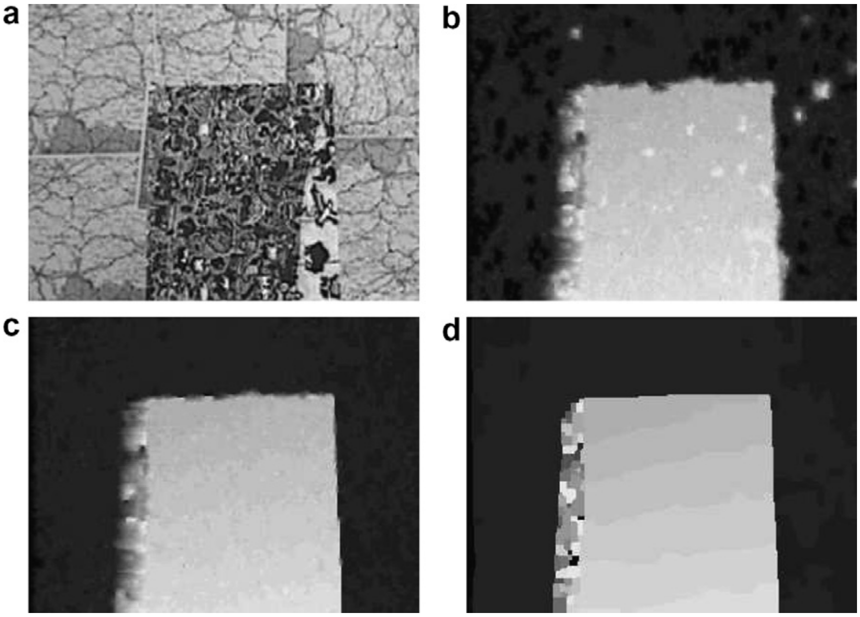

Fig. 4. Experiments on the "Map" stereo pair. (a) Presents the left view of the stereo pair. (b) Displays the resultant disparity map by fixing $(s_d, \sigma_d, s_y, \sigma_y) = (2, 2, 2, 6)$. (c) Is the inferred disparity map, by fixing (s_d, s_y) as Gaussians and estimating the scale parameters (σ_d, σ_y). By estimating all parameters $(s_d, \sigma_d, s_y, \sigma_y)$, the inferred disparity map is shown in (d).

the edges and corners. Notice the mean model parameters always converge, despite different initial values. Fig. 5 compares the evaluation results of four testbed pairs, where the first row is for the non-occluded regions, the second row for the textureless regions, and the third row corresponds to the discontinuity regions. The disparity and error maps of the testbed pairs are presented in Fig. 6. To summarize, the corresponding estimated $(s_d, \sigma_d, s_y, \sigma_y)$ mean values are (1.98, 0.48, 0.97, 1.45) for "Tsukuba", (1.98, 0.26, 1.11, 3.07) for "Sawtooth", (1.98, 0.59, 1.00, 2.76) for "Venus" and (0.93, 0.14, 0.95, 2.88) for "Map".

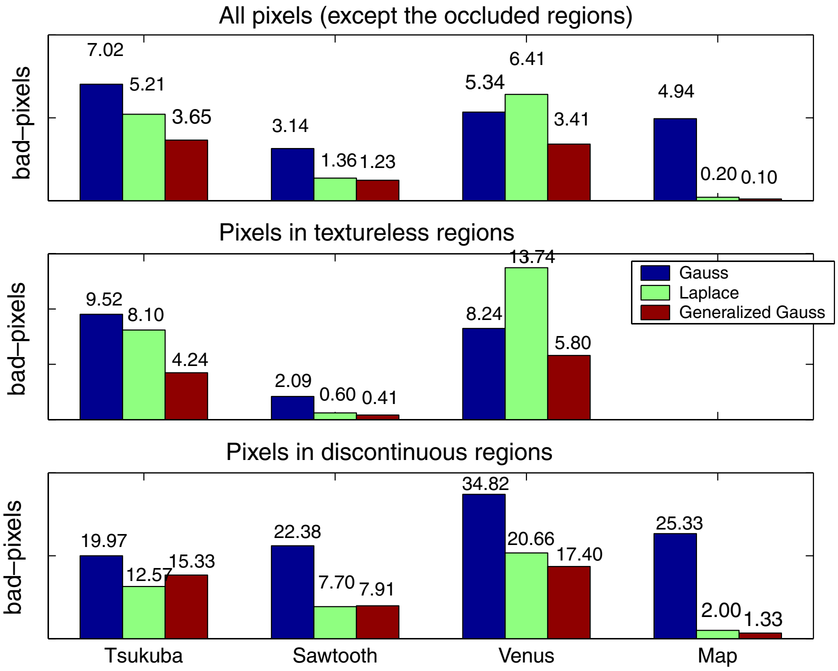

Fig. 5. This figure can be read as a three by four matrix where each element contains three bars comparing the Gaussian models (left in blue), the Laplacian models (middle in green) and the generalized Gaussian models (right in red) under the same scenario. Along the horizontal axes the four test pairs ("Tsukuba", "Sawtooth", "Venus" and "Map") are listed. The performance measure (vertical axes) corresponds to the "bad-pixels" errors for three types of regions: not occluded, textureless and discontinuous. See text for details.

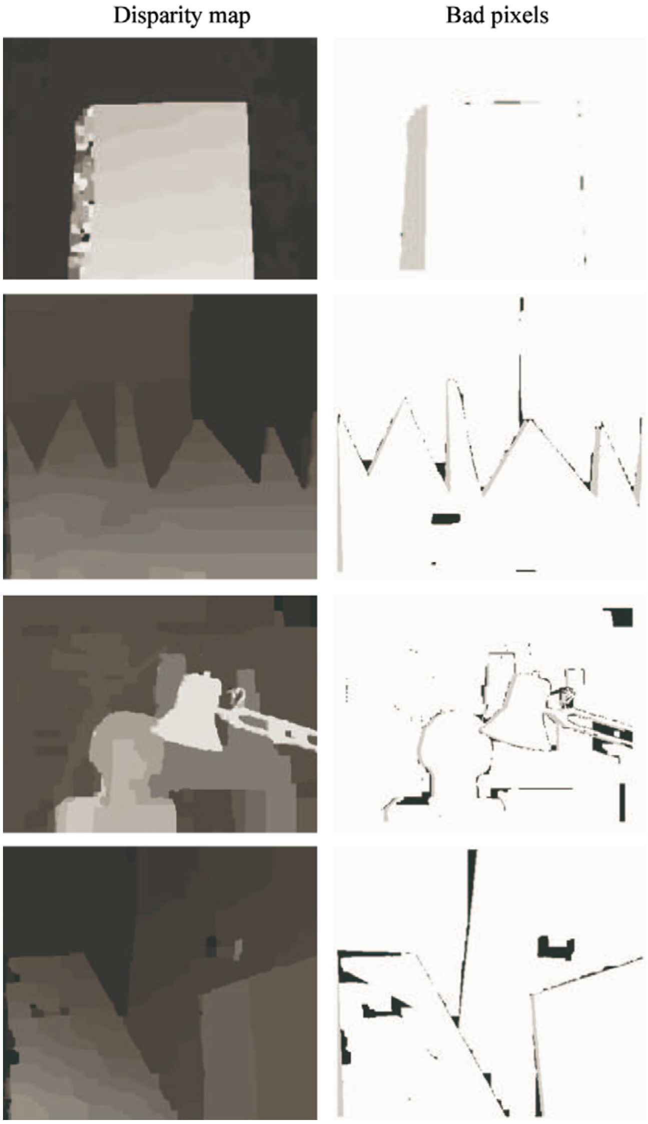

Fig. 6. In the first column, the resultant disparity maps are shown for the four image pairs by estimating θ, and the predicted disparity errors are shown in the second column. Here, black pixels are counted as errors, while gray pixels are those located in occluded patches and are not counted.

The performance of the Gaussian ($s_d = s_y = 2$), the Laplacians ($s_d = s_y = 1$) and the proposed generalized Gaussian models where θ is estimated from specific inputs, are compared in Fig. 5. In general, the Laplacian models perform better than the Gaussian models, largely due to their noise-resistance property. As expected, the generalized Gaussians perform best, since all model parameters θ are allowed to adapt to specific image pairs. The "Map" benefits most from the proposed approach, with a drastic reduction of errors and the best ranking. Improvements to the other three image pairs are relatively modest, probably due to the following reasons. First, to keep the model simple, one set of estimated model parameters was shared over all image pixels. Second, the proposed generative model is in a sense a simplification of the "true" model, where some important properties of the image formation process are ignored. Third, the LBP algorithm can fail in cases where the loops in the MRF are strongly correlated [6].

We have compared the proposed method with the closely related method reported in the literature, namely, that of Sun et al. [9], which also employs the LBP algorithm, but with a different statistical model and hand-tuned parameters. Empirical results show that the proposed method performs better on the "Map" pair but worse on the other three. Fig. 7 presents comparisons of the testbed pairs. The adaptability of the model parameters θ leads to superior performance observed in the "Map" pair, which is obviously different from the other three image pairs in term of the shape parameter s_d. The difficulty of obtaining a set of "good" hand-tuned parameters over generic image pairs was also observed by Sun et al. [9] (p. 9):

> 'Obviously, this set of parameters (that are good for the other three) is not the optimal for "Map" because the disparity range of this data is almost twice that of "Tsukuba".'

The inferior performance on the three remaining pairs in comparison to Sun et al. [9] are a result of the following. First, in Sun et al. [9] the image pairs are segmented and then integrated with the stereo matching results to boost performance. Second, Sun et al. consider a more complicated model that takes into account additional information such as the occlusion factor, while our proposed model is much simpler.

Due to our interest in applying computer vision to forestry inventory, we have conducted some experiments on a set of synthetic tree stand image pairs with varying degrees of overlapping canopies. Fig. 8 illustrates one representative example. We choose the disparity range to be

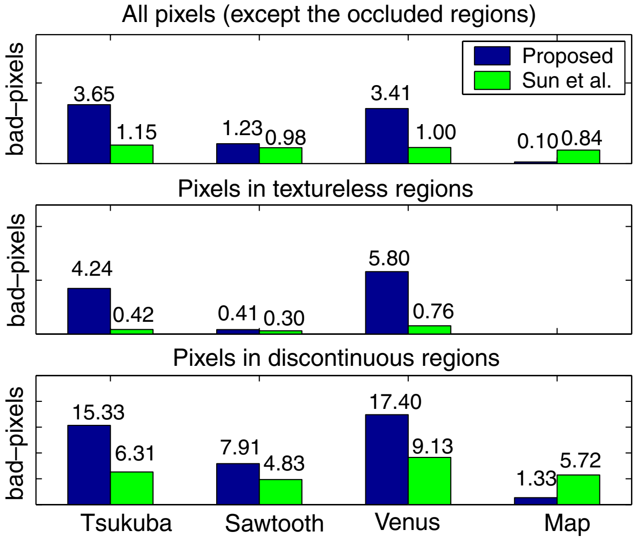

Fig. 7. Comparisons of our algorithm to Sun et al. [9] on the testbed pairs, using the "bad pixels" error score.

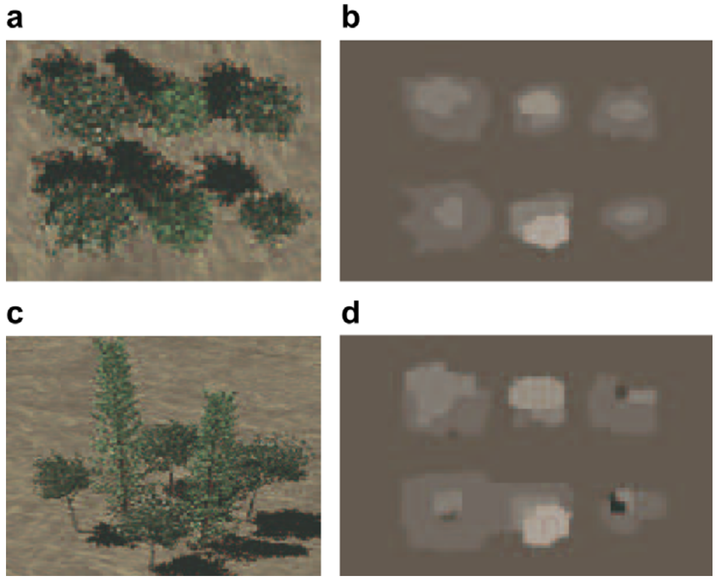

Fig. 8. Experimental result on one of the synthetic tree stand image pairs applying the proposed approach. The left view of the forest is presented in (a), which contains four aspens and two spruces. Its side view is also shown in (c). The true disparity map is shown in (b), and in comparison, (d) is the inferred disparity map. See text for details.

[0, . . ., 12], then scale it by 16 to form the grey-scale disparity map. The initial and the estimated $(s_d, \sigma_d, s_y, \sigma_y)$ hyperparameter values are $(2, 2, 2, 6)$ and $(1.26, 0.02, 0.80, 1.46)$, respectively. We adopt the root mean square (RMS) error as $(\frac{1}{n}\sum_{x,y}\text{diff}(x,y)^2)^{\frac{1}{2}}$, where diff(·) is the disparity differences and n is the number of pixels involved. The RMS error

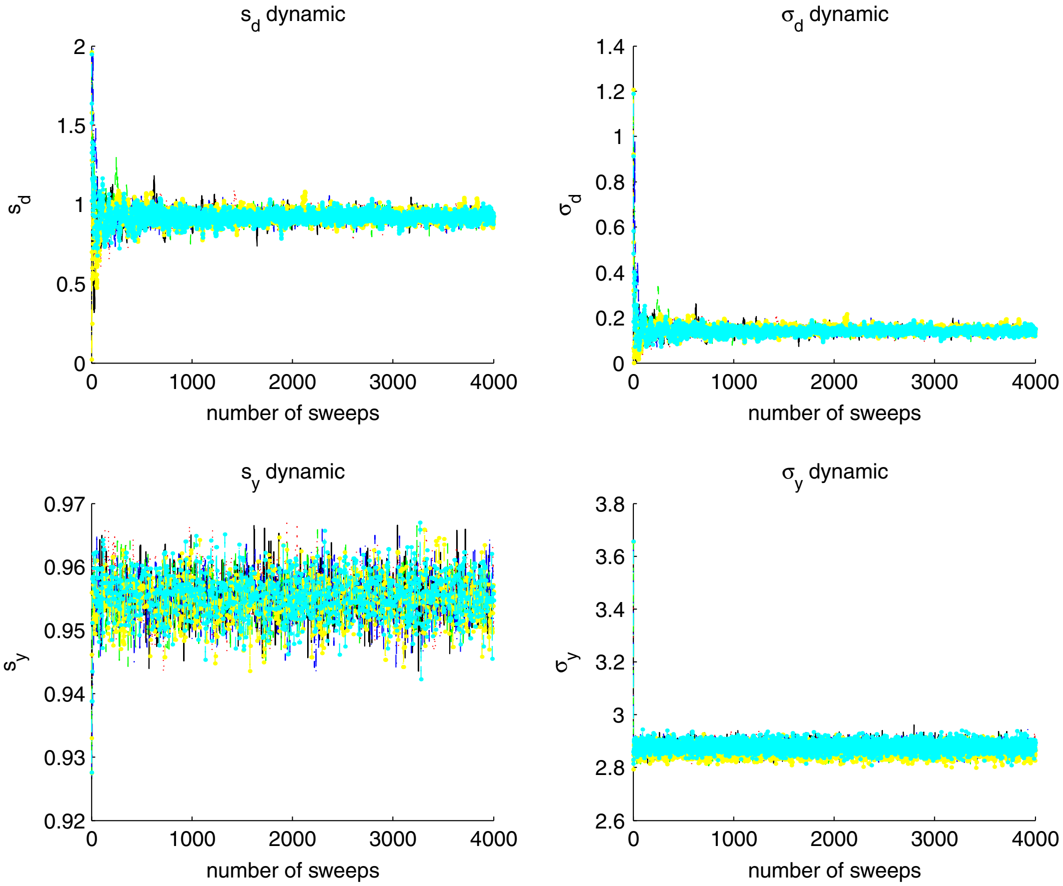

Fig. 9. The $\theta = \{s_d, \sigma_d, s_y, \sigma_y\}$ dynamics of six parallel sampled chains for the "Map" pair. See text for details.

(calculated only on the tree-related regions) of the inferred disparity map is 0.78.

In practice, we observe that 80 iterations of the LBP algorithm are enough to ensure convergence, which take from 1 to 5 min on an Intel Pentium 4 PC depending on the size and the maximum depth of the input images. Further, it takes several minutes for the sampling algorithms to generate the θ values. Typically we run this twice, which is observed to be enough to guarantee the convergence of both θ and d. As a result, the average running time is approximately 10 min.

5. Convergence monitoring

Ideally, the sampled chains asymptotically converge to the invariant distribution after a sufficient number of iterations, according to the ergodic property of MCMC samplers [3]. In practice, however, we have to monitor the convergence behavior of the sampled chains, and empirically detect the number of "burn-in" sweeps to be discarded in order to ensure that the rest of the chains are sampled from the invariant distribution. Our approach is to run parallel chains with different starting positions. Their converging statistics are measured by the "Gelman statistic" [3] which describes, at any time, the convergence behavior in terms of a scalar function by measuring how the ratio of maximum and minimum sample variances $\Sigma^2_{\max}/\Sigma^2_{\min}$ differs. This ratio is further computed from the between-chains variance and the within-chain variance of the multiple chains (Refer to Chapter 8, pp. 131–144 of [3] for the algorithmic details and related analysis.). Obviously, this "Gelman statistic" equals 1 at convergence.

In the experiments of the Middlebury testbed, for each image pair, multiple chains are run to ensure the convergence property. Fig. 9 presents the chain dynamics for the "Map" pair, where, from top-left to bottom-right, there are four panels showing the sampled chains dynamics for s_d, σ_d, s_y and σ_y, respectively. Each panel presents six parallel chains starting from the following positions: $\{0.2, 9, 0.2, 27\}$, $\{0.5, 8, 0.5, 26\}$, $\{0.8, 7, 0.8, 23\}$, $\{1, 5, 1, 20\}$, $\{1.5, 4, 1.5, 15\}$, and $\{2, 2, 2, 10\}$, respectively. After a short period of burn-in, these chains start to mix. The mixing or convergence behavior is then monitored by the "Gelman statistic" and is shown in Fig. 10. Empirically, these chains appear to mix well after 2000 iterations. We also observe similar convergence results for the rest of the image pairs. Based on these results, we, in practice, run one chain of 4000 sweeps (iterations), and discard the first half as the burn-in period, as suggested in [3].

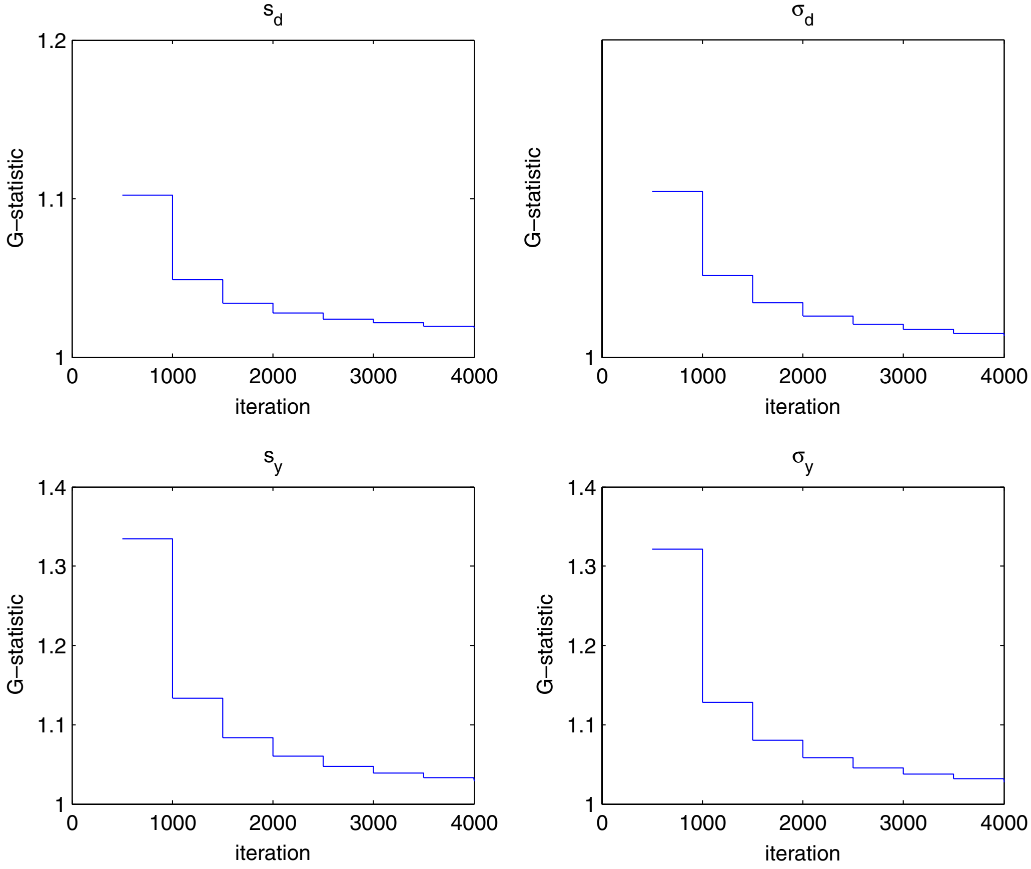

Fig. 10. The "Gelman statistic" of $\theta = \{s_d, \sigma_d, s_y, \sigma_y\}$ for six parallel sampled chains, respectively, on the "Map" pair. See text for details.

6. Discussion and future work

A Bayesian perspective to stereo computation enables the estimation of model parameters together with the inference of the MAP disparity map within a unified framework. A rather simplified model is proposed for the stereo matching task, which does not explicitly address the occlusion factor. Nevertheless, experimental results show convincing evidence that the estimated model parameters do consistently produce more accurate disparity maps.

The proposed generative model is rather flexible and can be extended in many ways to boost its performance. One natural direction is to go beyond the homogeneous and isotropic assumption of the MRF which applies one set of model parameters θ over the entire image. For example, pixels with similar (inferred) depth values can be spatially clustered together, so that θ could be tied only within clusters. This strategy is in line with the Swendsen–Wang type block-sampling techniques [15]. Furthermore, we can explicitly address the discontinuity and occlusion factors within such a framework, as did in Sun et al. [9].

Acknowledgments

We thank Arthur E.C. Pece for his constructive comments and his efforts to improve the format of this paper. This research was supported by grants from the Natural Science and Engineering Research Council of Canada and the Alberta Research Council of Canada.

Appendix A. Full conditional probability

Given a graphical model with node (variable) set $V = \{v\}$, let $pa[v]$ be the parent nodes of v, $ch[v]$ be the children nodes of v, and $V\backslash v$ be all nodes except v. The joint density can be factorized as

$$p(V) = \prod_{v \in V} p(v|pa[v]). \tag{A.1}$$

The full conditional probability for each node is conditioned only on its Markov Blanket [14]. This amounts to

$$\begin{aligned} p(v|V \setminus v) &\propto p(v, V \setminus v) \\ &= p(v|pa[v]) \prod_{w \in ch[v]} p(w|pa[w]). \end{aligned} \tag{A.2}$$

Appendix B. Updating σ_y

First of all, consider the specific case where $s_y = 2$. Define the prior $p(\sigma_y^2)$ as the inverse gamma density function [13]

$$p(\sigma_y^2) = \left(\frac{\sigma_{y0}^2}{\sigma_y^2}\right)^{v_0/2+1} \exp\left\{-\frac{v_0\sigma_{y0}^2}{2\sigma_y^2}\right\}. \tag{B.1}$$

where v_0 is confined to a small positive value, to ensure $p(\sigma_y^2)$ a non-informative prior. We then rearrange the likelihood as

$$p(y|d, \sigma_y, s_y) \propto (\sigma_y^2)^{-\frac{n}{2}} \exp\left\{-\frac{nv}{2\sigma_y^2}\right\}, \tag{B.2}$$

where v is the sufficient statistic:[2]

$$v = \frac{\sum_i y_i^2}{n}. \tag{B.3}$$

From the prior, the likelihood, and Eq. (A.2), the full conditional probability of σ_y^2 follows

$$\begin{aligned} p(\sigma_y^2|\cdot) &\propto p(\sigma_y^2) p(y|d, \sigma_y, s_y) \\ &\propto (\sigma_y^{-2})^{\frac{v_0+n}{2}+1} \exp\left\{-\frac{v_0\sigma_{y0}^2 + nv}{2\sigma_y^2}\right\}. \end{aligned} \tag{B.4}$$

However, the degrees of freedom v_0 in the prior $p(\sigma_y^2)$ are far smaller than the degrees of freedom n in the likelihood $p(y|d, \sigma_y, s_y)$. For the sake of simplicity, the posterior is then approximated by setting $v_0 = 0$, resulting in the following posterior

$$(\sigma_y^{-2}|\cdot) \sim \gamma\left(\frac{n}{2}, \frac{nv}{2}\right). \tag{B.5}$$

Based on similar derivations, for the generalized Gaussian functions, we have

$$(\sigma_y^{-s_y}|\cdot) \sim \gamma\left(\frac{n_y}{2}, \frac{n_y v_y}{2}\right) \tag{B.6}$$

where n_y is the number of pixels, and $v_y = (1/n_y) \times \sum_i |y_i|^{s_y}$.

Appendix C. Evaluating the partition functions

C.1. Connection to the Coding [12,11] and the Pseudo-likelihood Methods [11]

We provide an example to illustrate why the partition function is difficult to evaluate. We then proceed to introduce the coding method and the pseudo-likelihood method, respectively. Finally, we make some comments which lead to our proposed method to evaluate the partition functions.

Example (the partition function). Let a discrete MRF $X = \{x_i \in \{1, \ldots, D\}, \forall i\}$ be defined over a 4×1 lattice, which consists of four nodes indexed by $i \in \{1, \ldots, n = 4\}$, and three edges $(i \sim j) \in \{(1,2), (2,3), (3,4)\}$. Given the parameter $\xi \in \mathbb{R}^K$, we define the conditional probability of X admitting a configuration x (for example, $(x_1, x_2, x_3, x_4) = (1, 1, D, D)$) as

[2] Define a real value function $T = f(X)$ where X is the observations that are generated by the distribution $p(X|\xi)$, where ξ denotes the associated parameter. We says T is the sufficient statistic if the conditional distribution $p(\xi|T, X) = p(\xi|T)$.

$$p(x|\xi) = \frac{\exp\{\phi(\xi,x_1,x_2) + \phi(\xi,x_2,x_3) + \phi(\xi,x_3,x_4)\}}{z(\xi)}, \tag{C.1}$$

where $z(\xi) = \sum_{x_1,x_2,x_3,x_4} \exp\{\phi(\xi,x_1,x_2) + \phi(\xi,x_2,x_3) + \phi(\xi,x_3,x_4)\}$ is the partition function for this MRF, and $\phi(\xi,x_i,\ x_j) : \mathbb{R}^K \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the sufficient statistic defined over the edge $(i \sim j)$.

As shown in this example, the partition function $z(\xi)$ is evaluated by marginalization over all possible configurations of X. As a consequence, the computational complexity grows exponentially (on the order of $O(D^n)$) with respect to n, the size of the MRF. Due to this inherit difficulty of evaluating $z(\xi)$, two types of approaches have been developed.

The coding/pseudo-likelihood methods of Besag [12,11] are representatives of the first type of approaches that avoid evaluating $z(\xi)$ at all. Continuing with Example 1, the coding method partitions the nodes into two sets such that no two nodes from the same set are connected. Consequently, all nodes in the same set are conditionally independent providing that the remaining nodes are known [16]. Within one set, the conditional probability can thus be factorized as products of independent local probabilities $p(x_1,x_3 \mid \xi) = p(x_1 \mid \xi)p(x_3 \mid \xi)$, where $p(x_i|\xi) = \exp\{\overline{\phi}(\xi,x_i)\}/z_i(\xi)$, $z_i(\xi) = \sum_{x_i} \exp\{\overline{\phi}(\xi,x_i)\}$, and $\overline{\phi}(\xi,x_i)$ is the local sufficient statistic defined on node i, such that

$$\overline{\phi}(\xi,x_i) \approx \sum_{\forall j: j\sim i} \phi(\xi,x_i,x_j)/2. \tag{C.2}$$

The conditional probability of the other set is computed similarly. Then we iterate between the two sets until we have convergence. Instead of the conditional update scheme proposed in the coding method, the pseudo-likelihood method approximates the true conditional probability (Eq. (C.1)) over all node variables $x = \{x_i\}$ directly as a product of independent local probabilities

$$p(x|\xi) \approx \prod_{i=1}^{4} p(x_i|\xi), \tag{C.3}$$

which essentially decomposes the joint features on edges to local ones on single nodes. Obviously, this can lead to a valid approximation of the original joint density probability only if the node variables $\{x_i\}$ are weakly correlated (such that Eq. (C.3) holds).

However, the reduction of computation demands relies on the independence assumption which essentially discards the statistical dependencies among neighboring nodes of the image. Further, the partition function $z(\xi)$ cannot even be evaluated, since the objective function $p(x \mid \xi)$ is essentially changed by the approximation in Eq. (C.3). On the contrary, the second type of approaches attempts to evaluate the partition function, when dealing with parameter estimation in the MRF. One such method is proposed by Higdon et al. [2]. Here, we propose an alternative and efficient method, based on the idea of path sampling [10].

C.2. Evaluating the ratios of partition functions $z(s_d)$

We choose to estimate the partition function ($z(s_d)$ in Eq. (7), $s_d \in (0..2]$), as similarly done in Higdon et al. [2]. One possible method is to derive variational algorithms to bound the Bethe free energy [6]. This turns out to have many difficulties, since the Bethe free energy is (a) not a convex function, and (b) only an approximation of the original objective function $\log z(s_d)$. Instead, we adopt the path sampling scheme [10]. To simplify the notation, we denote $s \equiv s_d$ in this section.

We are specifically interested in estimating the log-ratio of the partition function. Gelman et al. [10] provides a framework which unifies various approaches toward this goal. Without loss of generality, we follow the notation of Gelman et al. [10]. The Gibbs prior can be written as $p(d \mid s) = q(d|s)/z(s)$, where d denotes the collection of node variables in the MRF, $z(s)$ is the partition function with s taking the range of (0..2]. In details, we have

$$q(d|s) = \frac{1}{\sigma_d^n} \exp\left\{ -\frac{1}{s} \sum_{j\sim k} \left| \frac{d_j - d_k}{\sigma_d} \right|^s \right\}, \tag{C.4}$$

the un-normalized density function with respect to certain fixed s, while $p(d|s)$ is normalized. Similar to Eq. (7), the partition function could be expressed as a summation of $q(d|s)$ over all possible configurations, $z(s) = \sum_d q(d|s)$. Then by taking the logarithm of $z(s)$ and differentiating with respect to s, we can easily obtain the identity (Eq. (6) in [10]):

$$\frac{\partial}{\partial s} \log z(s) = E_{d|s}\left[\frac{\partial}{\partial s} \log q(d|s) \right] \tag{C.5}$$

where $E_{d|s}$ is the expectation with respect to $q(d|s)$. To simplify the equations, we can further derive

$$\begin{aligned} U(d,s) &= \frac{\partial}{\partial s} \log q(d|s) \\ &= \frac{\partial}{\partial s} \left\{ -n \log \sigma_d - \frac{1}{s} \sum_{j\sim k} \rho\left(\frac{d_j - d_k}{\sigma_d}, s \right) \right\} \\ &= \sum_{j\sim k, d_j \neq d_k} \left(\frac{1}{s} \left| \frac{d_j - d_k}{\sigma_d} \right|^s \left(\frac{1}{s} - \log \left| \frac{d_j - d_k}{\sigma_d} \right| \right) \right). \end{aligned} \tag{C.6}$$

Note that here we only compute those j,k pixels such that $d_j \neq d_k$, because the others do not contribute to this equation.

Define the log-ratio and, by applying Eqs. (C.5 and C.6), transform it into an integral form

$$\begin{aligned} \lambda(a,b) &= \log\left[\frac{z(b)}{z(a)} \right] \\ &= \int_a^b E_{d|s}[U(d,\xi)] \mathrm{d}s \end{aligned} \tag{C.7}$$

where $0 < a < b \leqslant 2$.

As pointed out in Gelman et al. [10], Eq. (C.7) can be approximated by numerical integration. Define j_a and j_b as the indexes such that $s_{j_a} \leqslant a < s_{j_a+1} < \cdots < s_{j_b-1} <$

$b \leqslant s_{j_b}$. Using trapezoid rule, we have the numerical approximation (Eq. (15) of [10]):

$$\begin{aligned}\hat{\lambda}(a,b) = &\frac{1}{2}(s_{j_a+1} - a)(\hat{U}_{j_a+1} + \hat{U}_a)\\ &+ \frac{1}{2}\sum_{j=j_a+1}^{j_b-2}(s_{j+1} - s_j)(\hat{U}_{j+1} + \hat{U}_j)\\ &+ \frac{1}{2}(b - s_{j_b-1})(\hat{U}_b + \hat{U}_{j_b-1})\end{aligned}$$

where $\hat{U}_a$ and $\hat{U}_b$ could be obtained by interpolation, and $\hat{U}_i$ is the average of the values of $U(d_i, s_i)$ for all simulations drawn from s_i.

Based on the above derivations, we propose the following steps to estimate the log-ratio of $z(s_d)$, $\forall\ s_d \in (0..2]$:

1. For specific s_d^l, s_d^r values, interpolate existing $\log\{z_l/z_r\}$ values to obtain an approximate guess of $\log\{z_l/z_r\}$,
2. Divide the parameter space $z(s_d)$ into evenly spaced intervals $l = 1, \ldots, L$ and define $z_1 = z(s_d = l)$, $z_L = z(s_d = r)$, with fixed σ_d value,
3. Compute the log-ratio $\lambda(l, r)$ via trapezoid rule.

References

[1] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision 47 (2002) 7–42.

[2] D. Higdon, J. Bowsher, V. Johnson, T. Turkington, D. Gilland, R. Jaszczak, Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data, IEEE Transactions on Medical Imaging 16 (1997) 516–526.

[3] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman and Hall, London, 1996.

[5] J. Coughlan, H. Shen, Dynamic quantization for belief propagation in sparse spaces, Computer Vision and Image Understanding (2006) Special issue, doi:10.1016/j.cviu.2005.09.008.

[6] J. Yedidia, W.T. Freeman, Y. Weiss, Understanding belief propagation and its generalizations, in: Gerhard Lakemeyer (ed.), Exploring Artificial Intelligence in the New Millennium, 2003, pp. 239–236.

[7] S.T. Barnard, Stochastic stereo matching over scale, International Journal of Computer Vision 3 (1989) 17–32.

[8] Peter N. Belhumeur, A Bayesian approach to binocular stereopsis, International Journal of Computer Vision 19 (3) (1996) 237–262.

[9] J. Sun, N. Zheng, H. Shum, Stereo matching using belief propagation, IEEE Transactions on Pattern Recognition and Machine Intelligence 25 (7) (2003) 787–800.

[10] A. Gelman, X. Meng, Simulating normalizing constants: From importance sampling to bridge sampling to path sampling, Statistical Science 13 (1998) 163–185.

[11] J. Besag, Statistical analysis of non-lattice data, Statistician 24 (1975) 179–195.

[12] J. Besag, Spatial interaction and the statistical analysis of lattice systems, Journal of the Royal Statistical Society B 36 (2) (1974) 192–225.

[13] A. Gelman, J. Carlin, H. Stern, D. Rubin, Bayesian Data Analysis, Chapman and Hall, London, 1997.

[14] Michael Jordan, Chris Bishop, An Introduction to Graphical Models, in preparation.

[15] R.H. Swendsen, J.S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, Physical Review Letters 58 (1987) 86–88.

[16] S.L. Lauritzen, Graphical Models, Clarendon Press, Oxford, 1996.