

Estimating Optimal Parameters for MRF Stereo from a Single Image Pair

Li Zhang, *Member, IEEE Computer Society*, and Steven M. Seitz, *Member, IEEE*

Abstract—This paper presents a novel approach for estimating the parameters for MRF-based stereo algorithms. This approach is based on a new formulation of stereo as a maximum a posterior (MAP) problem in which both a disparity map and MRF parameters are estimated from the stereo pair itself. We present an iterative algorithm for the MAP estimation that alternates between estimating the parameters while fixing the disparity map and estimating the disparity map while fixing the parameters. The estimated parameters include robust truncation thresholds for both data and neighborhood terms, as well as a regularization weight. The regularization weight can be either a constant for the whole image or spatially-varying, depending on local intensity gradients. In the latter case, the weights for intensity gradients are also estimated. Our approach works as a wrapper for existing stereo algorithms based on graph cuts or belief propagation, automatically tuning their parameters to improve performance without requiring the stereo code to be modified. Experiments demonstrate that our approach moves a baseline belief propagation stereo algorithm up six slots in the Middlebury rankings.

Index Terms—Stereo matching, parameter setting, Markov Random Fields.



1 INTRODUCTION

STEREO matching has been one of the core challenges in computer vision for decades. See [1], [2] for an excellent survey and evaluation of the current state of the art. Many of the current best-performing techniques are based on Markov Random Field (MRF) formulations [3] that balance a data matching term with a regularization term and are solved using graph cuts [4], [5] or belief propagation [6], [7]. (For a comparative study of these MRF methods and source code, see [8].) Virtually all of these techniques require users to set hand-tuned parameters, e.g., regularization weight, by trial and error on a set of images. In this paper, we argue that different stereo pairs require different parameter settings for optimal performance and we seek an automated method to estimate those parameters for each pair of images.

To see the effect of parameter setting on stereo matching, we estimated disparity maps, $D = \{d_i\}$, for Tsukuba and Map image pairs [2] by minimizing the following energy

$$\sum_{i \in \mathcal{I}} U(d_i) + \lambda \sum_{(i,j) \in \mathcal{G}} V(d_i, d_j), \quad (1)$$

where \mathcal{I} is the set of pixels, \mathcal{G} is the set of graph edges connecting adjacent pixels, U measures similarity between matching pixels, and V is a regularization term that encourages neighboring pixels to have similar disparities. We minimize (1) using an existing MRF solver [7] and plot the error rate of the disparity estimation versus ground truth as a function of λ , as shown in Fig. 1. The figure shows that, for the

same algorithm, the optimal regularization weight λ varies across different stereo pairs. As shown later in the paper, λ and other MRF parameters, e.g., robust truncation thresholds, are related to the statistics of image noise and variation of scene structures and can all be estimated from a single stereo pair. Furthermore, we also show that neighboring pixel intensity difference [4] can be conveniently incorporated into our formulation to encourage the disparity discontinuities to be aligned with intensity edges and the relevant parameters can be estimated automatically.

To estimate the MRF parameters, we interpret them using a probabilistic model that reformulates stereo matching as a maximum a posterior (MAP) problem for both the disparity map and the MRF parameters. Under this formulation, we develop an alternating optimization algorithm that computes both the disparity map and the parameters. Although the derivation of the algorithm has numerous equations and is somewhat involved, the resulting technique itself is quite simple (summarized as a recipe in Fig. 5). Specifically, our technique serves as a *wrapper* for existing MRF stereo matching algorithms that solves for the optimal parameters for each image pair. Our routine uses the output of the stereo matcher to update the parameter values, which are in turn fed back into the stereo matching procedure—it can interface with many stereo implementations without modification. Therefore, we emphasize that the goal of this paper is *not* a specific stereo algorithm that performs better than existing algorithms. Rather, we introduce a methodology that boosts the performance of MRF-based stereo algorithms. In particular, we demonstrate our method on both graph cuts with the Potts model and belief propagation with truncated absolute difference.

The rest of the paper is organized as follows: After reviewing related work in Section 2, we first give the intuition for our parameter estimation technique in Section 3. We then formulate the idea as an MAP problem in Section 4, propose an optimization algorithm in Section 5, and extend it to estimate the weights that depend on intensity gradients in

• L. Zhang is with the Computer Science Department, Columbia University, 450 Mudd Hall, 500 West 120 Street, New York, NY 10027.
E-mail: lizhang@cs.columbia.edu.

• S.M. Seitz is with the Department of Computer Science and Engineering, University of Washington, Box 352350, Seattle, WA 98195-2350.
E-mail: seitz@cs.washington.edu.

Manuscript received 27 Feb. 2006; revised 12 June 2006; accepted 27 June 2006; published online 13 Dec. 2006.

Recommended for acceptance by R. Zabih.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0191-0206.

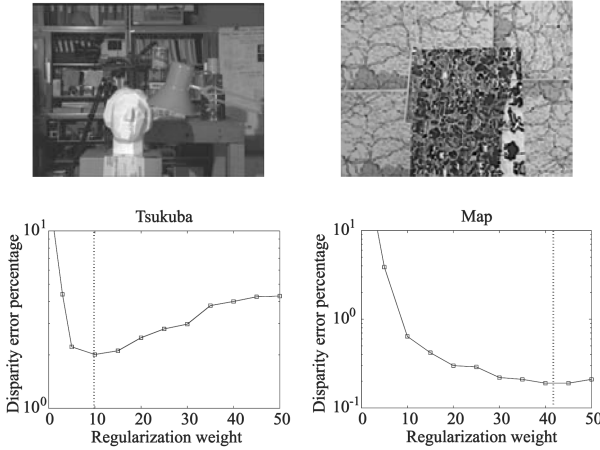


Fig. 1. Some stereo pairs require more regularization than others, as shown in the above graphs that plot error as a function of regularization weight λ . The parameters shown above (dotted vertical lines) were computed automatically using our algorithm.

Section 7. Finally, we show experimental results in Section 8 and discuss future research directions in Section 9.

2 PREVIOUS WORK

An early stereo method that requires no parameter setting is the adaptive window method of Kanade and Okutomi [9], which depends on proper initialization for good performance. The only prior work that addressed the problem of computing MRF parameters (aka *hyperparameters*) for stereo matching is by Cheng and Caelli [10]. While their approach is an important first step, they relied on a restricted MRF model from the image restoration literature [11] and did not support key features of the leading stereo algorithms, e.g., occlusion modeling and gradient-dependent regularization. (Other MRF models in the image restoration literature, e.g., [12], [13], [14], also have this limitation when applied to stereo matching.) In contrast, we designed our approach to support these features in order to interface with many of the leading stereo algorithms. Toward this end, we show that the Potts model and the truncated absolute distance commonly used in leading stereo algorithms [4], [7] correspond to a binary distribution and a mixture of an exponential distribution and an outlier process, respectively. These distributions can be extended to incorporate static cues, e.g., intensity edges, to further improve stereo matching performance. We use hidden variables to model occlusions and other outliers and apply expectation maximization (EM) to infer the hidden variables and estimate the mixture models. Because we use EM instead of MCMC, our approach is also simpler and more efficient compared to [10]. Finally, we benchmark our approach on the Middlebury database [2] and show that it dramatically improves the performance of a leading algorithm with the *recommended* hand-tuned parameters (as opposed to showing improvement over randomly chosen parameters [10]).

In this work, we use insights from statistical learning to improve vision algorithms. Our work is therefore related to Freeman et al. [15] who formulate super-resolution as MRF inference based on training images and apply belief propagation to obtain good results. Similarly, with training images, Freeman and Torralba [16] infer 3D scene structure from a single image. Unlike Freeman et al.'s approach, our

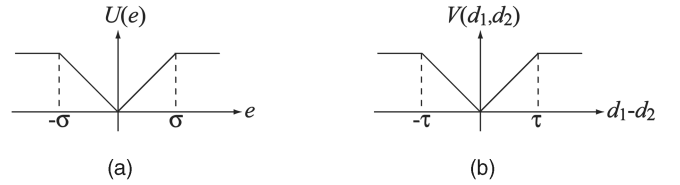


Fig. 2. Illustration of the truncated absolute difference metric for the (a) pixel similarity measure and (b) regularization term in (2).

method doesn't require training images—MRF parameters are estimated from the stereo pair itself.

3 INTUITION

In this section, we describe our basic idea for parameter estimation in MRF-based stereo. In the energy function in (1), U measures similarity between matching pixels and V encourages neighboring pixels to have similar disparities. Many functional forms have been proposed for U and V , including squared differences, absolute differences, and other robust metrics [1]. In this paper, we first focus on setting parameters for truncated absolute difference (TAD) because it is a popular choice of top performing stereo algorithms [4], [7] and it has several good properties. It is derived from total variation [17], thus preserving discontinuities. It can be efficiently computed via distance transform [7] in belief propagation. Also, it satisfies the *metric* property required by the traditional α -expansion algorithm in graph cut [4]. More recent graph cut algorithms relax this assumption, though.¹ Specifically,

$$\begin{aligned} U(d_i) &= \min(|e(d_i)|, \sigma), \\ V(d_i, d_j) &= \min(|d_i - d_j|, \tau), \end{aligned} \quad (2)$$

where $e(d_i) = I(x_i, y_i) - J(x_i - d_i, y_i)$ is the intensity difference between matching pixels in the image pair I and J and σ and τ are truncation thresholds, as illustrated in Fig. 2. Our method can be used to set parameters for other metrics as well and we give a derivation for the Potts model later in this paper.

To best set the parameters σ , τ , and λ for a stereo pair, we need to know how well the corresponding pixels in two images can be matched and how similar the neighboring disparities are in a statistical sense. However, without knowing the disparity map, those two questions cannot be answered. This dilemma explains why existing MRF-based stereo algorithms require users to set parameters manually.

To resolve this dilemma, let's first consider the case in which we know the disparity maps. In Figs. 3a and 3c, using the ground truth disparities from the Middlebury Web site [2], we plot the histograms of pixel matching errors and neighboring disparity differences for the Tsukuba stereo pair. In Figs. 3b and 3d, we show the same histograms in log-scale. Since the log-scale histograms are not straight lines or quadratic curves, it means that the probability of pixel matching errors and that of neighboring disparity

1. Bokyov et al. introduced the α -expansion algorithm in [4], which requires that V is a metric. Kolmogorov and Zabih [18] generalized V to a slightly larger class. More recently, Rother et al. [19] generalized it still further to allow some hard constraints. In addition, they provided a way to use the expansion algorithm for an arbitrary choice of V at the price of no longer guaranteeing that the expansion move is optimal (they show, however, that it will not increase the energy).

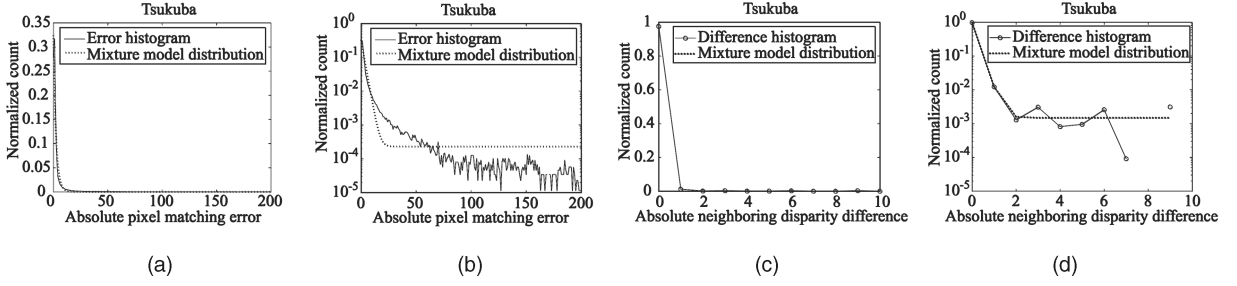


Fig. 3. Histograms of errors between corresponding pixels in two images in (a) linear and (b) log scale. Superimposed on the plots is the fitted mixture model. (c) and (d) show histograms and models for neighboring disparity difference.

differences are not simple exponential or Gaussian distributions. The heavy tail in the matching error histogram is due to occlusion and violation of brightness constancy; the heavy tail in the neighboring disparity difference histogram is due to disparity discontinuities. Those histograms can be approximated by two segments, a mixture of an exponential distribution and a uniform outlier process. Figs. 3a and 3c show the probability distribution of fitted mixture models overlaid on the histograms. The fit is quite accurate: The errors are around 10^{-3} , only noticeable in the log-scale graphs in Figs. 3b and 3d. From the shapes of the fitted distributions, we can recover the optimal set of MRF parameters, as we describe later in the paper.

In practice, however, ground truth disparities are unknown and we propose an iterative algorithm that alternates between estimating MRF parameters from the current histograms and estimating disparities using the current MRF parameters. The algorithm iterates until the estimated disparity map yields histograms that agree with the MRF parameters or a fixed number of iterations is reached. In the next section, we present the details of this method by casting the problem in a probabilistic framework.

4 A PROBABILISTIC MIXTURE MODEL FOR STEREO

In this section, we present the mixture models for the histograms of pixel matching errors and neighboring disparity differences and formulate stereo matching probabilistically, based on those mixture models.

4.1 Matching Likelihood

Given an image pair I and J and the disparity map D , we define the mixture model for pixel matching error as follows. We assign each pixel i in I a hidden binary random variable γ_i , indicating whether the corresponding scene point is visible in J .² Let $e(d_i) = I(x_i, y_i) - J(x_i - d_i, y_i)$. We define the mixture model for $e(d_i)$ as

$$P(e(d_i)|d_i, \gamma_i) = \begin{cases} \zeta e^{-\mu|e(d_i)|}, & \gamma_i = 1 \\ \frac{1}{N}, & \gamma_i = 0, \end{cases} \quad (3)$$

where μ is the decay rate for the exponential distribution, $|e(d_i)|$ takes discrete values, $\{0, 1, \dots, N-1\}$, and $\zeta = \frac{1 - \exp(-\mu)}{1 - \exp(-\mu N)}$ is a normalization factor. We define the mixture probability

2. For brevity, we refer to γ_i as a *visibility* variable, but it can also account for differences in brightness, e.g., due to specularity.

$$P(\gamma_i = 1) = \alpha, \quad (4)$$

where α is the fraction of pixels in I that are also visible in J . Summing over γ_i gives the marginal matching likelihood

$$P(e(d_i)|d_i) = \alpha \zeta e^{-\mu|e(d_i)|} + (1 - \alpha) \frac{1}{N}. \quad (5)$$

4.2 Disparity Prior

Define $\Delta d_g = d_i - d_j$ to be the disparity difference on the graph edge g connecting adjacent pixels i and j . Similarly, as for pixel matching probability, we assign each edge g a binary random variable θ_g , indicating whether the edge is continuous. We define the mixture model for Δd_g as

$$P(\Delta d_g|\theta_g) = \begin{cases} \eta e^{-\nu|\Delta d_g|}, & \theta_g = 1 \\ \frac{1}{L}, & \theta_g = 0, \end{cases} \quad (6)$$

where ν is the decay rate, $|\Delta d_g| \in \{0, 1, \dots, L-1\}$, and $\eta = \frac{1 - \exp(-\nu)}{1 - \exp(-\nu L)}$. We define the mixture probability

$$P(\theta_g = 1) = \beta, \quad (7)$$

where β is the fraction of continuous edges in I . The marginal distribution is

$$P(\Delta d_g) = \beta \eta e^{-\nu|\Delta d_g|} + (1 - \beta) \frac{1}{L}. \quad (8)$$

4.3 Stereo as a MAP Problem

Now, we formulate stereo matching as a MAP problem based on the two defined mixture distributions. Given an image pair, I and J , our probabilistic model consists of a disparity function $D = \{d_i\}$ over I and two sets of random variables $\Gamma = \{\gamma_i\}$ and $\Theta = \{\theta_g\}$ for pixel visibility and edge connectivity, respectively.³

We seek to estimate D , α , μ , β , and ν , given I and J , by maximizing

$$P(D, \alpha, \mu, \beta, \nu | I, J) = \frac{P(I, J, D | \alpha, \mu, \beta, \nu) P(\alpha, \mu, \beta, \nu)}{P(I, J)} \propto P(I, J, D | \alpha, \mu, \beta, \nu), \quad (9)$$

where the prior on $(\alpha, \mu, \beta, \nu)$ is assumed to be uniform. To compute $P(I, J, D | \alpha, \mu, \beta, \nu)$, we first factor it as

$$P(I, J, D | \alpha, \mu, \beta, \nu) = P(I, J | D, \alpha, \mu) P(D | \beta, \nu) \quad (10)$$

by assuming that I and J are independent of the disparity prior parameters β and ν and that D is independent of

3. This model is called “three coupled MRF’s” in [6].

matching likelihood parameters α and μ . Then, we compute $P(I, J|D, \alpha, \mu)$ and $P(D|\beta, \nu)$ by marginalizing over visibility variables Γ and continuity variables Θ , respectively, as follows:

$$P(I, J|D, \alpha, \mu) = \prod_i P(e(d_i)|d_i, \alpha, \mu), \quad (11)$$

which assumes $P(\Gamma|D, \alpha) = P(\Gamma|\alpha)$, ignoring the dependence of visibility on geometry for computational convenience. This assumption allows us to evaluate the matching likelihood by comparing pixel intensities without checking the global visibility, a simplification that is shared by many stereo algorithms. Similarly,

$$P(D|\beta, \nu) = \prod_g P(\Delta d_g|\beta, \nu), \quad (12)$$

which assumes independence between Δd_g , also for computational convenience. The independence assumptions in (11) and (12) allow us to efficiently estimate parameters α , μ , β , and ν , using the Expectation Maximization algorithm, as shown in Section 5.

Putting (11) and (12) together, we obtain

$$\begin{aligned} & P(D, \alpha, \mu, \beta, \nu|I, J) \\ & \propto \underbrace{\prod_i P(e(d_i)|d_i, \alpha, \mu)}_{P(I, J|D, \alpha, \mu)} \underbrace{\prod_g P(\Delta d_g|\beta, \nu)}_{P(D|\beta, \nu)}. \end{aligned} \quad (13)$$

Given disparity map D , we can estimate α and μ by maximizing the marginal data likelihood $P(I, J|D, \alpha, \mu)$ and we can estimate β and ν by maximizing the marginal prior distribution $P(D|\beta, \nu)$. Also, we can estimate D by maximizing the likelihood and the prior *jointly*. Next, we propose an alternating optimization algorithm for this maximization and relate the probabilistic model parameters $(\alpha, \mu, \beta, \nu)$ to (σ, τ, λ) in (1) and (2).

5 AN ALTERNATING OPTIMIZATION

In this section, we present an alternating algorithm to maximize (13). In (13), given D , maximizing $P(I, J|D, \alpha, \mu)$ leads to an estimate of (α, μ) . This maximization is equivalent to fitting the mixture model of (3) to the histogram of pixel matching errors in Fig. 3a, and the EM algorithm is well-suited for this type of model fitting. Similarly, we can also use the EM algorithm to estimate (β, ν) by maximizing $P(D|\beta, \nu)$. From $(\alpha, \mu, \beta, \nu)$, we then compute the optimal MRF parameters (σ, τ, λ) , which in turn are used to update D .

5.1 Estimating β and ν Given D

Given D , to estimate the parameters β and ν using the EM algorithm, we first compute

$$L = \max_g \{|\Delta d_g|\} + 1, \quad (14)$$

where L is the number of possible neighboring disparity differences from 0 to $\max_g \{|\Delta d_g|\}$. Then, we compute the conditional probability of θ_g as

$$\omega_g \stackrel{\text{def}}{=} P(\theta_g = 1|\Delta d_g, \beta, \nu) = \frac{\beta \eta e^{-\nu|\Delta d_g|}}{\beta \eta e^{-\nu|\Delta d_g|} + \frac{1-\beta}{L}}. \quad (15)$$

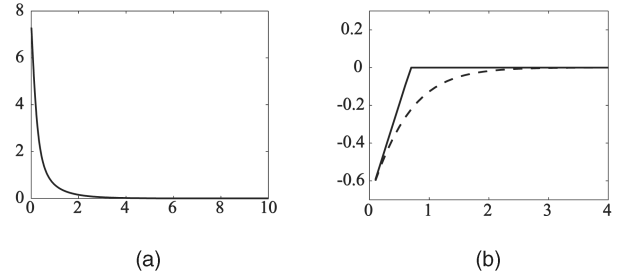


Fig. 4. (a) A graph for (18) for $L = 16$. (b) An illustration of (24) (solid curve) as an upper bound for (23) (dotted curve) for $(a, b, c) = (1, 2, 1)$.

Finally, we estimate β and ν by maximizing the expected log-probability $E_{\theta_g}[\log P(\Delta d_g, \theta_g|\beta, \nu)]$, computed as

$$\begin{aligned} & E_{\theta_g}[\log P(\Delta d_g, \theta_g|\beta, \nu)] \\ & = \sum_g \omega_g \log P(\Delta d_g, \theta_g = 1|\beta, \nu) \\ & \quad + (1 - \omega_g) \log P(\Delta d_g, \theta_g = 0|\beta, \nu) \\ & = \sum_g \omega_g (\log(\beta \eta) - \nu|\Delta d_g|) + (1 - \omega_g) \log \frac{1-\beta}{L}. \end{aligned} \quad (16)$$

By setting the partial derivatives of (16) with respect to β and ν to be zero, we obtain the following estimation of β and ν .

$$\beta = \frac{1}{|\mathcal{G}|} \sum_g \omega_g, \quad (17)$$

where $|\mathcal{G}|$ is the number of edges in \mathcal{G} and ν is the solution of the equation

$$\frac{1}{e^\nu - 1} - \frac{L}{e^{L\nu} - 1} = \frac{\sum_g \omega_g |\Delta d_g|}{\sum_g \omega_g}. \quad (18)$$

Let $f(\nu; L) = \frac{1}{e^\nu - 1} - \frac{L}{e^{L\nu} - 1}$ be the left-hand side of (18). f monotonically decreases from $\frac{L-1}{2}$ to 0 over $[0, \infty)$, as shown in Fig. 4a. When L is large, the second term in $f(\nu; L)$, $\frac{L}{e^{L\nu} - 1}$, is negligible and the equation has a closed form solution $\nu_0 = \log(\frac{1}{y} + 1)$, where y is the right-hand side of (18). When L is small, we start from $\nu = \nu_0$ and refine ν using the Newton-Raphson method.

5.2 Estimating α and μ Given D

The EM algorithm can also be used to estimate α and μ . Given D , we first compute

$$N = \max_i \{e(d_i)\} + 1, \quad (19)$$

where N is the number of possible pixel matching errors from 0 to $\max_i \{e(d_i)\}$. Then, we can estimate α and μ in the same way as we estimate β and ν using (15), (17), and (18) with the following variable replacement:

$$(\beta, \nu, \eta, L, g, \mathcal{G}, \Delta d_g) \rightarrow (\alpha, \mu, \zeta, N, i, \mathcal{I}, e(d_i)). \quad (20)$$

5.3 Estimating D Given α, μ, β , and ν

In this section, we describe how the estimated values of α, μ, β , and ν are used for stereo matching. Given α, μ, β , and ν , we wish to maximize (13) by minimizing

$$\begin{aligned}\Psi &\stackrel{\text{def}}{=} -\log P(D, \alpha, \mu, \beta, \nu | I, J) \\ &= \sum_i \rho_d(d_i; \alpha, \mu) + \sum_g \rho_p(\Delta d_g; \beta, \nu),\end{aligned}\quad (21)$$

where

$$\begin{aligned}\rho_d(d_i; \alpha, \mu) &= -\log\left(\alpha \zeta e^{-\mu|e(d_i)|} + \frac{1-\alpha}{N}\right), \\ \rho_p(\Delta d_g; \beta, \nu) &= -\log\left(\beta \eta e^{-\nu|\Delta d_g|} + \frac{1-\beta}{L}\right).\end{aligned}\quad (22)$$

Equation (21) can be minimized directly using existing techniques. For example, Sun et al. [6] use belief propagation to minimize a form of (21).

5.3.1 From Mixture Model to Regularized Energy

Although (21) can be optimized directly, it is not in a form that existing efficient MRF solvers [4], [7] require. Recall that our objective is to interface with and boost the performance of existing stereo algorithms and we therefore want to convert (21) to the form of (1) and (2). We notice that a function of the form

$$h(x; a, b, c) = -\log(a \exp(-b|x|) + c) \quad (23)$$

is tightly upper bounded by

$$\bar{h}(x; r, s, t) = \min(s|x|, t) + r, \quad (24)$$

where $s = \frac{ab}{a+c}$, $t = \log(\frac{a+c}{c})$, and $r = -\log(a+c)$, as shown in Fig. 4b. Therefore, minimizing (21) can also be approximately achieved by minimizing

$$\bar{\Psi} = \sum_i \min(s_d |e(d_i)|, t_d) + \sum_g \min(s_p |\Delta d_g|, t_p) - C, \quad (25)$$

where

$$\begin{aligned}s_d &= \frac{\alpha \zeta \mu}{\alpha \zeta + (1-\alpha)\frac{1}{N}} & t_d &= \log\left(1 + \frac{\alpha \zeta N}{1-\alpha}\right) \\ s_p &= \frac{\beta \eta \nu}{\beta \eta + (1-\beta)\frac{1}{L}} & t_p &= \log\left(1 + \frac{\beta \eta L}{1-\beta}\right) \\ C &= |\mathcal{I}| \log\left(\alpha \zeta + \frac{1-\alpha}{N}\right) + |\mathcal{G}| \log\left(\beta \eta + \frac{1-\beta}{L}\right).\end{aligned}\quad (26)$$

To further simplify the problem, let $\sigma = \frac{t_d}{s_d}$, $\tau = \frac{t_p}{s_p}$, and $\lambda = \frac{s_p}{s_d}$, and define

$$\bar{\Psi}' = \sum_i \min(|e(d_i)|, \sigma) + \lambda \sum_g \min(|\Delta d_g|, \tau). \quad (27)$$

$\bar{\Psi}'$ differs from $\bar{\Psi}$ by an affine transform, which does not affect the estimation of D . Equation (27) is the objective function used in [4], [7] for stereo matching.

Our iterative algorithm is summarized in Fig. 5. We typically start with $\alpha = \beta = 0.5$, $\mu = \nu = 1.0$, $N = 255$, and set L be the maximum disparity plus 1, although robust convergence is observed with various initial values, as shown in Section 8. STEREO-MATCHING could be any stereo algorithm that works with (27).

Initialize $(\alpha, \mu, N, \beta, \nu, L)$, and iterate

- Compute s_d, t_d, s_p , and t_p using Eq. (26)
- Set $\sigma = \frac{t_d}{s_d}$, $\tau = \frac{t_p}{s_p}$, and $\lambda = \frac{s_p}{s_d}$
- Compute D by STEREO-MATCHING with Eq. (27)
- Update L, β , and ν by iterating EM Eqs. (14,15,17,18)
- Update N, α , and μ by iterating EM Eqs. (19,15,17,18)

with the variable replacement defined in Eq. (20)

Until convergence or a fixed number of iterations.

Fig. 5. Algorithm for estimating stereo matching parameters.

Although we develop our method based on mixture distributions that correspond to a truncated absolute difference metric, we emphasize that our method is general and can be applied to other metrics as well by modeling the matching likelihood and/or disparity prior with other types of distributions. For example, notice that, if $|x|$ is replaced by x^2 in (23) and (24), (24) is still a good upper bound for (23). This suggests that, if we use a mixture of Gaussian (instead of exponential) distribution and uniform outlier to model pixel matching errors, we can have a truncated quadratic metric for the data term (with certain technical modifications in the EM fitting steps). In Section 6, we replace the mixture distribution for neighboring disparity differences with a binary distribution and transform our algorithm in Fig. 5 to handle the Potts model as a disparity prior.

6 PARAMETER SETTING FOR THE POTTS MODEL

To set parameters for the Potts model, we assume the neighboring disparity differences have a binary distribution, defined as:

$$P(\Delta d_g) = \begin{cases} \beta, & \Delta d_g = 0 \\ 1 - \beta, & \Delta d_g \neq 0. \end{cases} \quad (28)$$

This binary distribution is simpler than the mixture distribution in (8), assumed for the truncated absolute difference measure. Given the disparity field $D = \{d_g\}$, the maximum likelihood estimation for parameter β is simply

$$\beta = \frac{|\{\Delta d_g = 0\}|}{|\mathcal{G}|}, \quad (29)$$

which is the percentage of neighboring pixels that have the same disparity values. This constitutes one step of our alternating optimization method.

For the other step, given the parameter β , the disparity field D is estimated by minimizing (21) with ρ_p computed as follows: $\rho_p(\Delta d_g; \beta) = -\log(P(\Delta d_g|\beta))$ with $P(\Delta d_g|\beta)$ defined as in (28). To connect ρ_p to the Potts model, we rewrite $P(\Delta d_g|\beta)$ using the delta notation, $\delta(\cdot)$, as

$$P(\Delta d_g) = \beta \delta(\Delta d_g) + (1 - \beta)(1 - \delta(\Delta d_g)). \quad (30)$$

Using (30), we have

$$\begin{aligned}\rho_p(\Delta d_g; \beta) &= -\delta(\Delta d_g) \log \beta - (1 - \delta(\Delta d_g)) \log(1 - \beta) \\ &= s_p(1 - \delta(\Delta d)) + \log \frac{1}{\beta},\end{aligned}\quad (31)$$

where

$$s_p = \log \frac{\beta}{1 - \beta}. \quad (32)$$

Notice that $\rho_p(\Delta d_g; \beta)$ in this form differs from the Potts model only by a constant $\log \frac{1}{\beta}$, which does not affect the optimization. Therefore, our parameter estimation method in Fig. 5, developed for the truncated absolute difference metric, also applies to the Potts model, with the following modifications: Use (29) to estimate β instead of the EM equations (14), (15), (17), and (18); use (32) to estimate s_p instead of (26); ignore t_p and τ .

7 INTENSITY GRADIENT CUES

Recent stereo algorithms use static cues, such as color segments [6], [20] and intensity edges [21], [22], to improve performance. Here, we show that neighboring pixel intensity difference [4] can be conveniently incorporated into our formulation to encourage the disparity discontinuities to be aligned with intensity edges and the relevant weighting parameters can be estimated automatically.

Define ΔI_g to be the intensity difference between the two pixels connected by a graph edge g . To relate ΔI_g to the continuity of the disparity map, we treat ΔI_g as a random variable and define a corresponding mixture distribution. We require the mixture distribution of ΔI_g to share the same hidden variable θ_g of Δd_g . Specifically,

$$P(\Delta I_g | \theta_g) = \begin{cases} \xi e^{-\kappa |\Delta I_g|}, & \theta_g = 1, \\ \frac{1}{K}, & \theta_g = 0, \end{cases} \quad (33)$$

where κ is the decay rate,⁴ $|\Delta I_g| \in \{0, 1, \dots, K-1\}$, and $\xi = \frac{1 - \exp(-\kappa)}{1 - \exp(-\kappa K)}$. This model has the following property: If a graph edge is continuous, both the color and the disparity differences are encouraged to be small; if a graph edge is discontinuous, the color and disparity differences are unconstrained. The corresponding marginal distribution is

$$P(\Delta I_g, \Delta d_g) = \beta \xi \eta e^{-(\kappa |\Delta I_g| + \nu |\Delta d_g|)} + (1 - \beta) \frac{1}{KL}. \quad (34)$$

Given I , ΔI , and J , our goal is to recover D , α , μ , β , ν , and κ , by maximizing

$$\begin{aligned} & P(D, \alpha, \mu, \beta, \kappa, \nu | I, \Delta I, J) \\ & \propto \underbrace{\prod_i P(e(d_i) | d_i, \alpha, \mu)}_{P(I, J | D, \alpha, \mu)} \underbrace{\prod_g P(\Delta I_g, \Delta d_g | \beta, \kappa, \nu)}_{P(\Delta I, D | \beta, \kappa, \nu)}. \end{aligned} \quad (35)$$

The alternating algorithm in Section 5 can still be applied with a minor change. The estimation of α and μ is the same as before. The estimation of β , κ , and ν can be done as follows. Initially, we set $K = \max_g \{\Delta I_g\} + 1$. For each

iteration, we first update L as in (14). Then, we compute the condition probability of θ_g as

$$\omega_g \stackrel{\text{def}}{=} \frac{\beta \xi \eta e^{-(\kappa |\Delta I_g| + \nu |\Delta d_g|)}}{\beta \xi \eta e^{-(\kappa |\Delta I_g| + \nu |\Delta d_g|)} + \frac{1 - \beta}{KL}}. \quad (36)$$

Finally, we update β and ν using (17) and (18), respectively, and update κ also using (18) with the following variable replacement: $(\nu, L, \Delta d_g) \rightarrow (\kappa, K, \Delta I_g)$.

After estimating $(\alpha, \mu, \beta, \nu, \kappa)$, we estimate D by minimizing (21) with ρ_p depending on ΔI_g . Specifically,

$$\rho_p(d_i; \Delta I_g, \beta, \kappa, \nu) = -\log \left(\beta \xi \eta e^{-\kappa |\Delta I_g|} e^{-\nu |\Delta d_g|} + \frac{1 - \beta}{KL} \right). \quad (37)$$

Accordingly, s_p and t_p also depend on ΔI_g .

$$s_p = \frac{\beta \xi \eta \nu e^{-\kappa |\Delta I_g|}}{\beta \xi \eta e^{-\kappa |\Delta I_g|} + \frac{1 - \beta}{KL}} \quad t_p = \log \left(1 + \frac{\beta \xi \eta K L e^{-\kappa |\Delta I_g|}}{1 - \beta} \right). \quad (38)$$

In (38), s_p approaches 0 in proportion to $e^{-\kappa |\Delta I_g|}$ when ΔI_g is large; s_p approaches $\frac{\nu}{1 + \frac{1 - \beta}{KL \beta \xi \eta}}$ when ΔI_g is near 0. Therefore, the regularization weight $\lambda = \frac{s_p}{s_d}$ varies over the image: large in uniform areas and small across color edges.

8 EXPERIMENTAL RESULTS

We implemented the EM algorithm to estimate the parameters for two different MRF stereo algorithms, in particular, graph cuts (GC) with the Potts model [4] and belief propagation (BP) with truncated absolute difference (TAD) [7]. We first describe our test images and then report our results using GC, followed by the results using BP.

8.1 Image Selection

The Middlebury Web site [2] is a well-known resource for benchmarking stereo algorithms. There are currently two collections of benchmark data sets—an older collection and a newer one. In Fig. 6, we show the performance evaluation of the leading stereo algorithms reported on the Web site using the old benchmark collection.⁵ From left to right, the 3D structures of the underlying scenes in these benchmarks become simpler: Tsukuba (multiple depth layers with irregular boundaries), Sawtooth (three planes with irregular boundaries), Venus (four planes with regular boundaries), and Map (two layers with regular boundaries). Notice that most algorithms perform well on either complex scenes or simpler ones, but not both. We believe that this inconsistency is partly due to the fact that each algorithm uses the same set of parameters for all the scenes of different complexity. The number 1 algorithm [24] in Fig. 6 performs very well for all the cases, partly because one of its parameters is automatically adapted for different cases. Therefore, we decided to mainly use these four image pairs to test our parameter setting method. The Middlebury Web site also evaluates algorithms using the new benchmark collection, which mostly contain complex scene geometry. There, different algorithms perform much more consistently: The top algorithms tend to give top results across different data sets. We show that our method can converge to reasonable parameters for these data sets as well and

4. The exponential part of this distribution is inspired by [23], in which a Laplacian distribution is used to model derivative filter responses of natural images.

5. The overall rank is obtained by sorting the sum of all the individual ranks, as per personal communication with Daniel Scharstein.

Algorithm name	Overall ranking	Tsukuba all untex. disc.			Sawtooth all untex. disc.			Venus all untex. disc.			Map all disc.
Sym. BP + occl.	1	2	3	2	1	1	1	4	3	6	1 1
Patch-based	2	1	1	1	5	1	5	2	3	2	8 10
Segm.-based GC	3	6	5	7	6	1	6	1	1	1	28 33
Graph+segm.	4	10	3	9	4	1	3	3	2	3	33 37
GC + mean shift	5	3	9	4	17	11	7	7	9	7	23 29
Segm.+glob.vis.	6	8	10	12	2	1	2	8	11	14	30 35
Belief prop.	7	4	7	3	14	22	13	11	10	20	22 14
Layered	8	14	17	15	7	1	8	19	29	5	13 13
2-pass DP	9	13	13	14	10	10	14	9	12	13	20 21
Region-Progress.	10	11	11	13	3	1	4	10	16	15	29 36
GC+occl.	11	5	2	5	13	14	17	22	27	11	18 16
MultiCam GC	12	18	23	8	12	1	19	16	19	12	10 12
GC+occl.	13	7	8	6	8	1	9	31	32	4	31 24
Improved Coop.	14	15	14	19	20	17	21	12	13	26	6 7
Adapt. weights	15	12	12	10	18	20	15	13	8	8	27 31
Symbiotic	16	23	22	21	15	15	23	5	5	17	16 18
Disc. pres.	17	17	19	20	19	13	16	21	22	19	11 6
Var. win.	18	21	21	23	21	19	22	17	14	24	4 4
Graph cuts	19	20	18	18	22	11	18	25	26	16	9 8
Reliability-DP	20	9	15	11	16	24	11	27	24	25	17 17
Multiw. cut	21	37	33	38	11	26	12	6	6	18	5 5
Graph cuts	22	19	16	16	9	16	10	24	23	10	35 22
Tree DP	23	16	6	17	25	29	20	15	17	9	26 30
4-State DP	24	31	28	34	24	17	31	14	7	29	7 11
Comp. win.	25	26	26	26	28	25	24	23	20	23	12 9
Realtime	26	30	31	30	23	23	25	20	18	21	21 27
Cooperative	27	27	27	28	29	33	30	30	30	37	3 2
Relax+occl.	28	35	34	35	27	21	36	18	15	32	14 15
Bay. diff.	29	36	39	24	26	27	26	33	35	31	2 3
Stoch. diff.	30	28	30	32	33	30	27	28	25	34	25 20

Fig. 6. The performance evaluation of the top 30 stereo algorithms on the Middlebury Web site as of 26 February 2006. The complexity of 3D structures of the underlying scenes in these benchmarks varies significantly, starting from Tsukuba (multiple depth layers with irregular boundaries) on the left to Map (two layers with regular boundaries) on the right. Most algorithms perform well on either complex scenes or simpler ones, but not both. This inconsistency is partly due to the fact that each algorithm uses the same set of parameters for all the scenes of different complexity.

discuss how our method may be extended to further improve algorithm performance on these data sets in Section 9.

8.2 Results on Graph Cuts with the Potts Model

For the experiments in this section, we used the α -expansion graph cut stereo matching code [4] with the Potts model as the neighborhood terms. In all experiments, we alternated between EM and GC six times. In each alternation, alpha expansion was applied to all depth layers for three passes in total. The cost for EM is negligible and the total runtime varies from 20 to 40 seconds for different stereo pairs.

8.2.1 Convergence

In our first experiment, we tested our algorithm on the four old Middlebury benchmarks. In Figs. 7 and 8, we show the disparity maps and corresponding (σ, λ) ⁶ at iterations 1, 2, 4, 6 for the four cases, starting from noisy and over-smoothed initializations, respectively. As the algorithm proceeds, the parameters are tuned and the disparity maps improve. Note that this experiment shows that the solution is relatively invariant to the initial conditions.

8.2.2 Optimality

To quantify the optimality of the tuned parameters, in our second experiment, for each case in Experiment 1, we fix σ but vary λ from 1 to 80 and estimate disparity using the same GC algorithm. We plot the error as a function of λ in

6. Recall that the Potts model has only parameters σ and λ but no τ .

Fig. 9. The bars indicate our estimated values whose corresponding error rates are quite close to the minimum error rates of the graphs in all four benchmarks.

8.3 Results on Belief Propagation with Truncated Distance

For the experiments in this section, we implemented belief propagation using distance transforms [7]⁷ as our baseline stereo matcher. In all experiments, we alternated between EM and BP six times. In each alternation, BP was executed for 60 iterations and each iteration takes about 1 second. The cost for EM is negligible and the total runtime is about 6 minutes.

8.3.1 Convergence

In our third experiment, we tested our algorithm on the four old and two new Middlebury benchmarks. In Fig. 10, we show the disparity maps and corresponding (σ, τ, λ) at iterations 1, 2, 4, and 6, obtained by using our method without the intensity gradient cue.⁸ In Fig. 11, we show the disparity maps and corresponding $(\alpha, \mu, \beta, \nu, \kappa)$ at the same iterations, using our method with the intensity gradient cue. In both figures, the initial regularization is weak and the disparity

7. Among the three acceleration techniques proposed in [7], we have only used the distance transform technique.

8. The initial values in Fig. 10 are not round numbers as in Figs. 7 and 8 because they are converted from the initial values $\alpha = \beta = 0.5$, $\mu = \nu = 1.0$, $N = 255$, and $L = \text{MAX_DISPARITY} + 1$. They are slightly different because MAX_DISPARITY is different for different scenes.

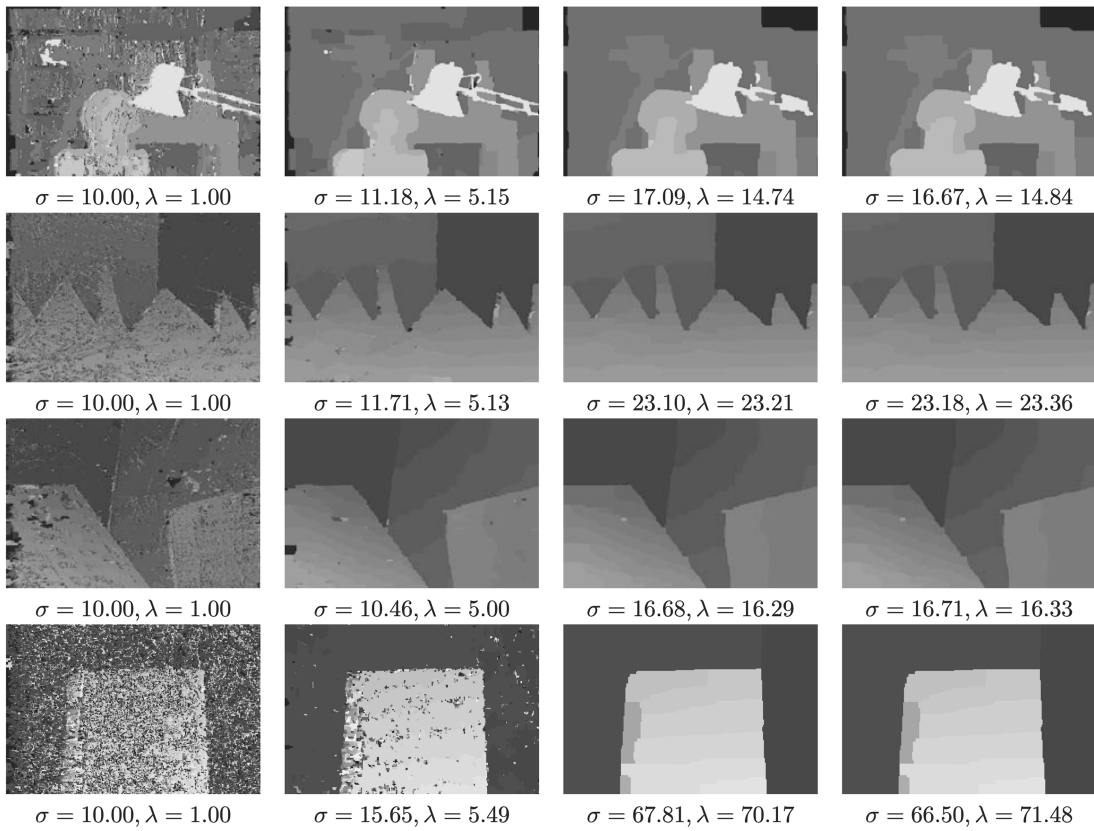


Fig. 7. Convergence of our parameter estimation for GC stereo with the Potts model on the four Middlebury benchmarks starting from noisy disparity maps. The four columns correspond to iterations 1, 2, 4, and 6.

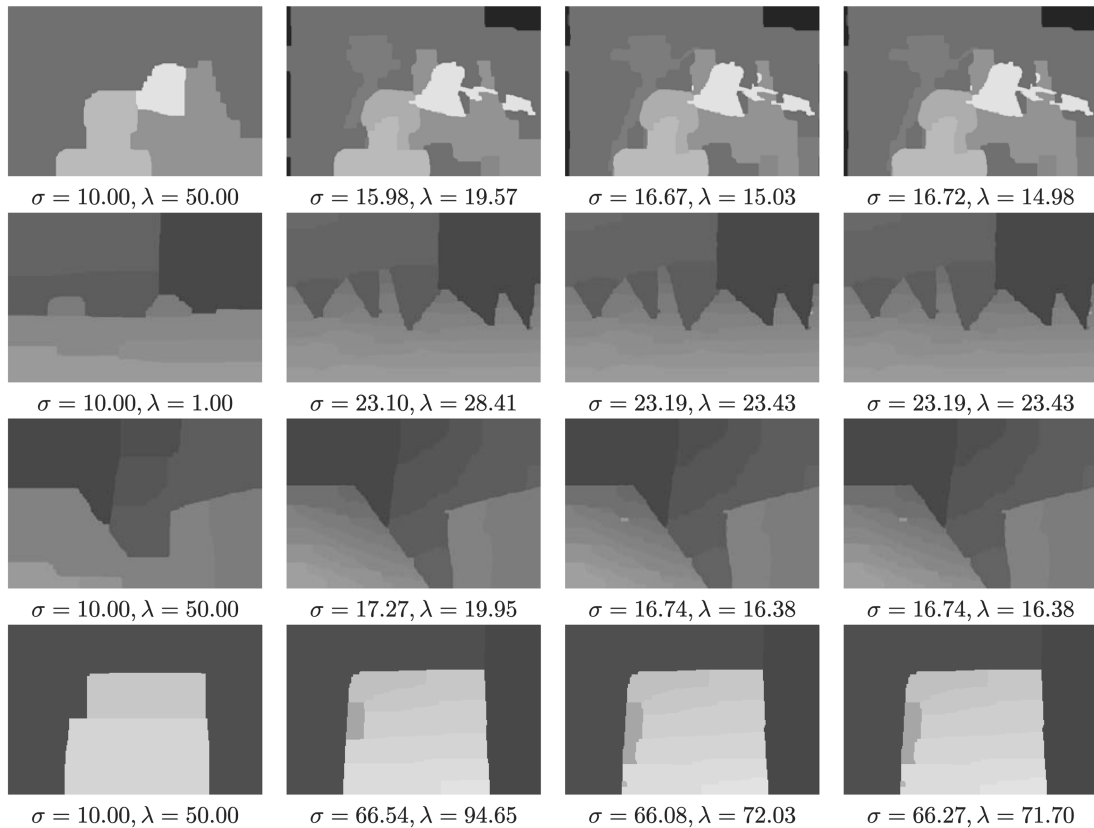


Fig. 8. Convergence of our parameter estimation for GC stereo with the Potts model on the four Middlebury benchmarks starting from oversmoothed disparity maps. The four columns correspond to iterations 1, 2, 4, and 6.

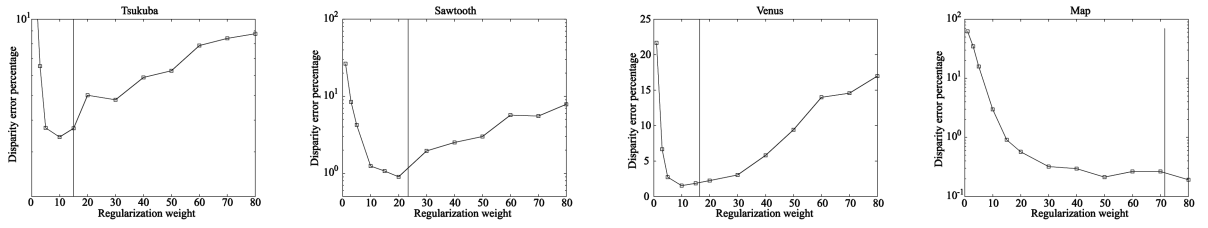


Fig. 9. Optimality of our parameter estimation for GC stereo with the Potts model. The four graphs show error rate with respect to ground truth as a function of regularization weight λ while fixing σ . The vertical solid lines are our estimation for λ .

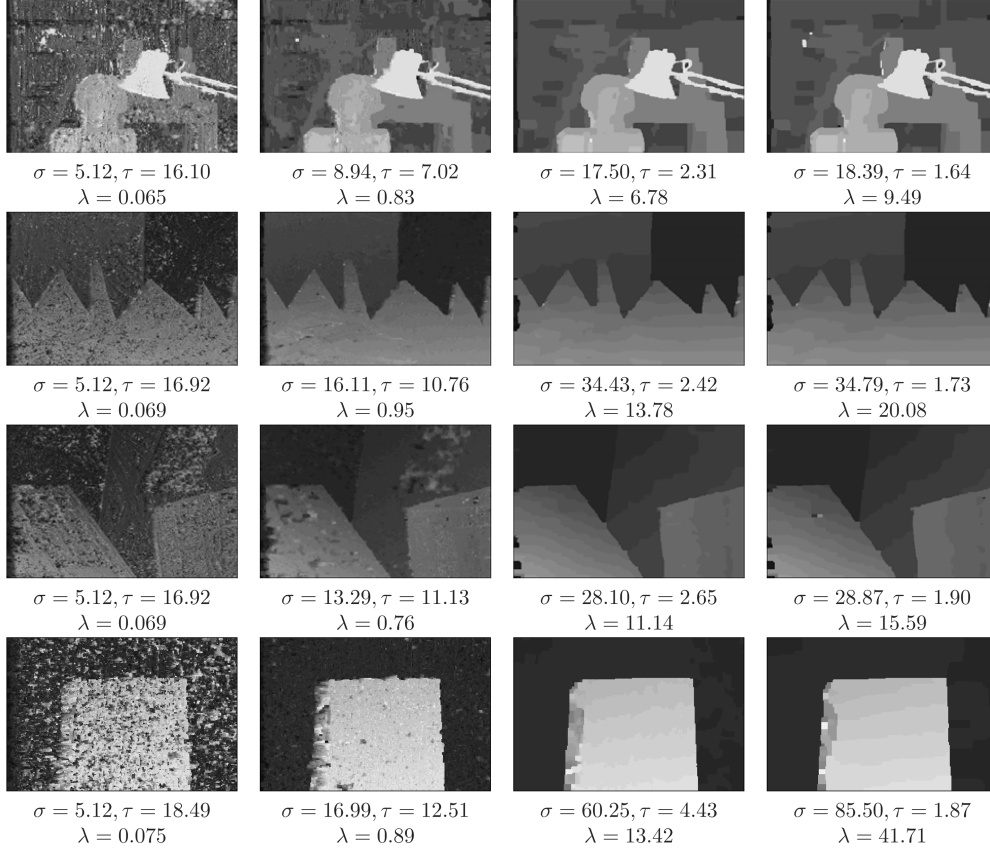


Fig. 10. Convergence of our parameter estimation for BP stereo with the TAD metric on the four Middlebury benchmarks. The four columns correspond to iterations 1, 2, 4, and 6.

map is noisy. As the algorithm proceeds, the regularization increases and the disparity map becomes cleaner.

In our fourth experiment, we repeated Experiment 3, but with different initial values. We show the initial and final (σ, τ, λ) , including those of Experiment 3, in the top five rows in the tables in Fig. 12. Despite the variation of scales in initial parameters, the final parameters are consistent, showing robust convergence of our algorithm. We also compared the final disparity maps using the ground truth and showed the error rate in the last column. The error rates are also consistent. In addition to trying different initial values, we also tried starting with the ground truth disparity maps and estimated the parameters. Then, we estimated the disparity map while keeping those parameters fixed. The error rates are shown in the bottom right corner. Both the error rates and the parameters are close to the results obtained without knowing ground truth. However, the parameters computed from ground truth disparities don't result in disparity maps

with lower error rates. This unintuitive fact is because real scenes are not perfectly described by our MRF model, as discussed in Section 9.

8.3.2 Optimality

In our fifth experiment, for each case in Experiment 4, we fix σ and τ but vary λ from 1 to 50 and estimate disparity using BP. We plot the error as a function of λ in Fig. 13. The vertical dotted lines indicate our estimated values, whose corresponding error rates are quite close to the minimum error rates of the graphs in all four benchmarks.

8.3.3 Improvement

In our sixth experiment, we show how our automatic parameter setting method can improve over choosing fixed parameters manually. We first run BP with the fixed parameters suggested in [7] and the result is shown in the fourth row ("Fixed") in Fig. 14. We then compare this result

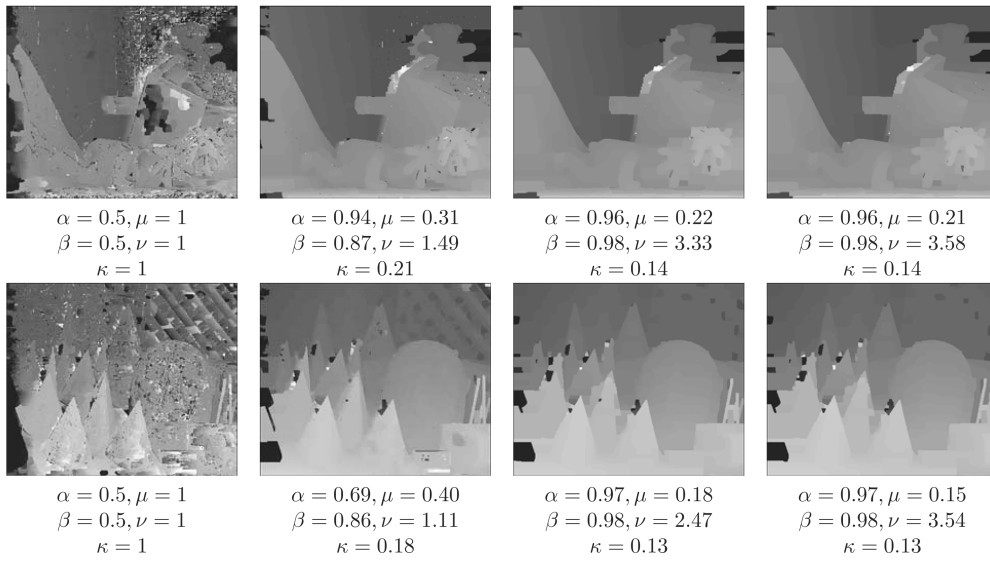


Fig. 11. Convergence of our parameter estimation for BP stereo with the TAD metric on the two new Middlebury benchmarks. The four columns correspond to iterations 1, 2, 4 and 6. The intensity gradient cue is used in this experiment.

Tsukuba							Sawtooth						
Initial parameters			Final parameters			Disparity error(%)	Initial parameters			Final parameters			Disparity error(%)
σ	τ	λ	σ	τ	λ		σ	τ	λ	σ	τ	λ	
5.12	2.60	0.91	18.49	1.61	9.75	2.23	5.12	2.82	0.93	34.79	1.72	20.12	0.97
33.66	2.60	9.42	18.53	1.60	9.84	2.26	33.66	2.82	9.65	34.78	1.72	20.10	0.97
1.11	2.60	0.18	18.47	1.62	9.67	2.15	1.11	2.82	0.19	34.44	1.72	20.07	0.99
5.12	16.10	0.065	18.39	1.64	9.49	2.12	5.12	16.92	0.069	34.79	1.73	20.08	0.97
5.12	0.59	4.71	18.52	1.60	9.83	2.24	5.12	0.64	4.79	34.78	1.72	20.08	0.98
Ground truth disparities			17.44	1.44	10.83	2.29	Ground truth disparities			31.72	1.59	21.62	0.99
Venus							Map						
Initial parameters			Final parameters			Disparity error(%)	Initial parameters			Final parameters			Disparity error(%)
σ	τ	λ	σ	τ	λ		σ	τ	λ	σ	τ	λ	
5.12	2.82	0.93	28.88	1.85	15.82	1.34	5.12	3.15	0.96	86.39	1.81	43.08	0.20
33.66	2.82	9.65	28.88	1.84	15.90	1.33	33.66	3.15	9.89	87.01	1.81	42.98	0.19
1.11	2.82	0.19	28.89	1.84	15.88	1.34	1.11	3.15	0.19	85.71	1.81	43.05	0.20
5.12	16.92	0.069	28.87	1.90	15.59	1.33	5.12	18.50	0.075	85.50	1.87	41.71	0.18
5.12	0.64	4.79	28.88	1.84	15.84	1.33	5.12	0.71	4.87	86.55	1.81	43.01	0.20
Ground truth disparities			26.54	1.75	15.38	1.42	Ground truth disparities			60.58	1.81	46.80	0.19

Fig. 12. Convergence of our parameter estimation for BP stereo with the TAD metric from five different initializations. The last rows use ground truth disparity maps to estimate parameters and then compute new disparity maps with these estimated parameters.

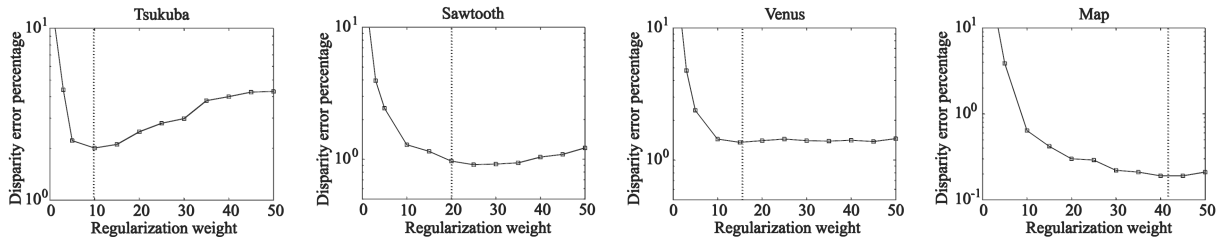


Fig. 13. Optimality of our parameter estimation for BP stereo with TAD metric. The four graphs show error rate with respect to ground truth as a function of regularization weight λ while fixing (σ, τ) . The vertical dotted lines are our estimation for λ .

with the result in the third experiment where we solve for (σ, τ, λ) , shown in the third row (“Adaptive”) in Fig. 14. Our adaptive method shows similar results to the fixed parameters for Sawtooth and Venus, but dramatic improvement on Map. As reported in [6], the Map pair requires different parameter settings than the other three data sets. Our algorithm automatically finds appropriate parameters without any user intervention. For the Tsukuba data set, the result is slightly worse than the results with the fixed manually

chosen parameters. The reason is that a user exploits the ground truth or his perception as a reference when setting the parameters, while our algorithm estimates the parameters based on model fitting. When the ground truth is not the optimal solution of our model, the estimated disparity map can deviate from the ground truth. Overall, our automatic parameter setting technique improves the ranking of the baseline algorithm by six slots on the Middlebury benchmarks.

Parameter setting	Overall ranking	Tsukuba						Sawtooth						Venus						Map			
		all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all
Adaptive+Grad	9	1.87	20	0.67	14	7.13	9	0.83	14	0.32	23	3.48	9	1.53	20	0.92	12	10.37	21	0.20	2	2.20	1
Fixed+Grad 2	14	1.84	18	1.05	17	9.87	21	0.87	14	0.28	21	5.78	18	1.22	17	1.05	13	13.39	25	0.20	2	2.49	3
Adaptive	17	2.12	21	1.36	21	10.76	21	0.97	14	0.31	23	6.79	19	1.33	18	1.13	14	14.65	28	0.18	2	2.20	1
Fixed	23	1.84	18	1.33	20	10.02	21	1.24	21	0.32	23	7.18	23	1.34	18	1.18	15	15.17	28	0.38	14	3.47	7
Fixed+Grad 1	36	7.68	37	5.76	33	11.79	21	5.92	40	0.30	22	13.12	29	6.90	39	3.20	30	14.11	27	20.19	41	35.41	41

Fig. 14. Performance comparison of fixed and adaptive BP stereo with TAD metric. Fixed: $(\sigma, \tau, \lambda) = (10, 2, 10)$. Adaptive: estimated (σ, τ, λ) . Fixed + Grad 1: estimated $(\alpha, \mu, \beta, \nu)$, fixed $\kappa = 1$. Fixed + Grad 2: estimated $(\alpha, \mu, \beta, \nu)$, fixed $\kappa = 0.01$. Adaptive + Grad: estimated $(\alpha, \mu, \beta, \nu, \kappa)$.

The improvements so far do not include the intensity gradient cue proposed in Section 7. Now, we consider this cue. First, we use the estimated values for $(\alpha, \mu, \beta, \nu)$, but set $\kappa = 1$ and 0.01, respectively. The results for the two κ values are shown in the fifth row (“Fixed + grad 1”) and the second row (“Fixed + grad 2”), respectively. $\kappa = 1$ is apparently too large and $\kappa = 0.01$ is better. If we estimate $(\alpha, \mu, \beta, \nu, \kappa)$ together, as described in Section 7, we get the first row (“Adaptive + grad”) of Fig. 14. As expected, error rates in discontinuity regions, shown in the columns under “disc,” are consistently reduced. Overall, our parameter estimation technique raises the rank of the baseline algorithm (with intensity gradient cue) by five slots and the resulting adaptive algorithm is ranked ninth among all stereo algorithms in the Middlebury rankings.

9 DISCUSSION

In this paper, we presented a parameter estimation method for MRF stereo. Our method converges consistently and significantly improves a baseline stereo algorithm. Our method works as a wrapper that interfaces with many stereo algorithms without requiring any changes to those algorithms. Here, we discuss some ideas for future work.

First, our model gives higher energy to the ground truth than to the estimated disparity maps [25]. One of the reasons is that we model visibility *photometrically*, but not *geometrically*. In other words, we assume visibility variables Γ are independent of D . Although this independence assumption simplifies our computation, it also causes errors in the results. The same problem is also observed in [26] for computing optical flow. One topic of future work is to model the occlusion process more precisely, such as in [24], and estimate its parameters accordingly.

Second, we use histograms of pixel matching errors and neighboring disparity differences for a whole image, assuming the mixture models don’t vary across the image. This assumption may be valid for the matching errors, which are largely due to sensor noise, but it may not be accurate for the disparity maps, which may have spatially varying smoothness. For example, it is difficult to evaluate whether the Teddy benchmark is smoother than the Cones or vice versa, based on the whole image. However, if we consider image patches of 100×100 pixels at different locations, the answer is much clearer. It is therefore an interesting problem to devise an MRF stereo model with spatially varying regularization. As this model will potentially involve many more parameters, automatic parameter tuning is even more important in this case, compared to the spatially constant regularization case.

Third, graph cuts and belief propagation have also shown great promise in other low-level vision problems, e.g., segmentation [27], [28], matting [29], and optical flow [26]. Estimating parameters for MRF models in those problems will help to create more automated vision systems in a large variety of applications.

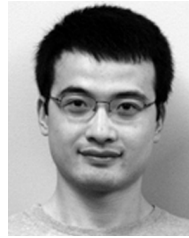
ACKNOWLEDGMENTS

The authors would like to thank Aaron Hertzmann and Richard Szeliski for helpful discussions. This work was supported in part by US National Science Foundation grant IIS-0049095, a US Office of Naval Research YIP award, and Microsoft Corporation. The work on this paper was done when L. Zhang was a PhD student at the University of Washington.

REFERENCES

- [1] D. Scharstein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms,” *Int’l J. Computer Vision*, vol. 47, no. 1, pp. 7-42, 2002.
- [2] “Middlebury Stereo Vision Page,” <http://cat.middlebury.edu/stereo/>, 2002.
- [3] S. Geman and G. Geman, “Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.
- [4] Y. Boykov, O. Veksler, and R. Zabih, “Fast Approximate Energy Minimization via Graph Cuts,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222-1239, Nov. 2001.
- [5] V. Kolmogorov and R. Zabih, “Computing Visual Correspondence with Occlusions Using Graph Cuts,” *Proc. Int’l Conf. Computer Vision*, pp. 508-515, 2001.
- [6] J. Sun, H.-Y. Shum, and N.N. Zheng, “Stereo Matching Using Belief Propagation,” *Proc. European Conf. Computer Vision*, pp. 510-524, 2002.
- [7] P.F. Felzenszwalb and D.P. Huttenlocher, “Efficient Belief Propagation for Early Vision,” *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 261-268, 2004.
- [8] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, “A Comparative Study of Energy Minimization Methods for Markov Random Fields,” *Proc. European Conf. Computer Vision*, pp. 19-26, 2006.
- [9] T. Kanade and M. Okutomi, “A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920-932, Sept. 1994.
- [10] L. Cheng and T. Caelli, “Bayesian Stereo Matching,” *Proc. Conf. Computer Vision and Pattern Recognition Workshop*, pp. 192-192, 2004.
- [11] D.M. Higdon, J.E. Bowsher, V.E. Johnson, T.G. Turkington, D.R. Gilland, and R.J. Jaszcak, “Fully Bayesian Estimation of Gibbs Hyperparameters for Emission Computed Tomography Data,” *IEEE Trans. Medical Imaging*, vol. 16, no. 5, pp. 516-26, 1997.
- [12] A. Mohammad-Djafari, “Joint Estimation of Parameters and Hyperparameters in a Bayesian Approach of Solving Inverse Problems,” *Proc. Int’l Conf. Image Processing*, pp. 473-477, 1996.
- [13] Z. Zhou, R.M. Leahy, and J. Qi, “Approximate Maximum Likelihood Hyperparameter Estimation for Gibbs Priors,” *IEEE Trans. Image Processing*, vol. 6, no. 6, pp. 844-861, 1997.

- [14] S.S. Saquib, C.A. Bouman, and K. Sauer, "ML Parameter Estimation for Markov Random Fields with Applications to Bayesian Tomography," *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1029-1044, 1998.
- [15] W.T. Freeman, E.C. Pasztor, and O.T. Carmichael, "Learning Low-Level Vision," *Int'l J. Computer Vision*, vol. 40, no. 1, pp. 25-47, 2000.
- [16] W. Freeman and A. Torralba, "Shape Recipes: Scene Representations that Refer to the Image," *Proc. Conf. Neural Information Processing Systems*, pp. 1335-1342, 2003.
- [17] L.I. Rudin and S. Osher, "Total Variation Based Image Restoration with Free Local Constraints," *Proc. Int'l Conf. Image Processing*, pp. 31-35, 1994.
- [18] V. Kolmogorov and R. Zabih, "What Energy Functions Can Be Minimized via Graph Cuts?," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147-159, Feb. 2004.
- [19] C. Rother, S. Kumar, V. Kolmogorov, and A. Blake, "Digital Tapestry," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 589-596, 2005.
- [20] H. Tao, H.S. Sawhney, and R. Kumar, "Dynamic Depth Recovery from Multiple Synchronized Video Streams," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 118-124, 2001.
- [21] P. Fua, "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features," *Machine Vision and Applications*, vol. 6, pp. 35-49, 1993.
- [22] A.F. Bobick and S.S. Intille, "Large Occlusion Stereo," *Int'l J. Computer Vision*, vol. 33, no. 3, pp. 181-200, 1999.
- [23] Y. Weiss, "Deriving Intrinsic Images from Image Sequences," *Proc. Int'l Conf. Computer Vision*, pp. 68-75, 2001.
- [24] J. Sun, Y. Li, S. Kang, and H.-Y. Shum, "Symmetric Stereo Matching for Occlusion Handling," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 399-406, 2005.
- [25] M.F. Tappen and W.T. Freeman, "Comparison of Graph Cuts with Belief Propagation for Stereo, Using Identical MRF Parameters," *Proc. Int'l Conf. Computer Vision*, pp. 900-906, 2003.
- [26] C.L. Zitnick, N. Jojic, and S.B. Kang, "Consistent Segmentation for Optical Flow Estimation," *Proc. Int'l Conf. Computer Vision*, pp. 1308-1315, 2005.
- [27] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy Snapping," *Proc. ACM SIGGRAPH*, pp. 303-308, 2004.
- [28] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts," *Proc. ACM SIGGRAPH*, pp. 309-314, 2004.
- [29] J. Wang and M.F. Cohen, "An Iterative Optimization Approach for Unified Image Segmentation and Matting," *Proc. Int'l Conf. Computer Vision*, pp. 936-943, 2005.



Li Zhang obtained the BE degree in automation at Tsinghua University in Beijing, People's Republic China, in 1998. He then became a PhD student at the Robotics Institute at Carnegie Mellon University. After two years, he moved to the Department of Computer Science and Engineering at the University of Washington, where he received the PhD degree in 2005. He is now a postdoctoral research scientist at Columbia University. His research interests

include computational imaging and vision, computer graphics and animation, and their various applications. He is a member of the IEEE Computer Society.



Steven M. Seitz received the BA degree in computer science and mathematics from the University of California, Berkeley, in 1991 and the PhD degree in computer sciences from the University of Wisconsin, Madison, in 1997. He holds the Short-Dooley Associate Professorship in the Department of Computer Science and Engineering at the University of Washington. Following his doctoral work, he spent one year at Microsoft Research and, subsequently, two

years as an assistant professor in the Robotics Institute at Carnegie Mellon University. He joined the faculty at the University of Washington in July 2000. He was twice awarded the David Marr Prize for the best paper at the International Conference of Computer Vision, and has received an US National Science Foundation CAREER Award, a US Office National Research Young Investigator Award, and an Alfred P. Sloan Fellowship. He is interested in problems in computer vision and computer graphics. His current research focuses on capturing the structure, appearance, and behavior of the real world from digital imagery. He is a member of the IEEE.

► **For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.**