

Parameter Estimation for MRF Stereo

Li Zhang Steven M. Seitz
University of Washington

Abstract

This paper presents a novel approach for estimating parameters for MRF-based stereo algorithms. This approach is based on a new formulation of stereo as a maximum a posteriori (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. Experiments indicate that our approach, as a wrapper for existing stereo algorithms, moves a baseline belief propagation stereo algorithm up six slots in the Middlebury rankings.

1. Introduction

Stereo matching has been one of the core challenges in computer vision for decades. Many of the current best-performing techniques are based on Markov Random Field (MRF) formulations [8] that balance a data matching term with a regularization term and are solved using Graph Cuts [2, 11] or Belief Propagation [16, 4]. Virtually all of these techniques require users to properly set hard-coded 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 [15] 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 Eq. (1) using an existing MRF solver [4] and plot the

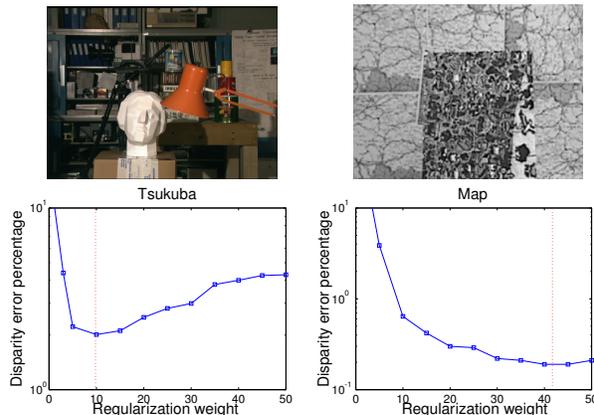


Figure 1. Some stereo pairs require more regularization than others, as shown in the above graphs that plot error as a function of regularization weight λ . The parameters shown above (dotted vertical lines) were computed automatically using our algorithm.

error rate of the disparity estimation versus ground truth as a function of λ , as shown in Figure 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 [2] 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 posteriori (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. Our approach 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. Although we use a baseline stereo algorithm based on belief propagation, any MRF-based stereo algorithm could be used instead (e.g., graph cuts, dynamic programming).

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 Section 6. Finally, we show experimental results in Section 7, and discuss future research directions in Section 8.

2. Previous work

An early stereo method that requires no parameter setting is the adaptive window method of Kanade and Okutomi [10], which requires proper initialization for good performance. The only prior work that addressed the problem of computing MRF parameters (aka *hyper-parameters*) for stereo matching is by Cheng and Caelli [3]. While their approach is an important first step, they relied on a restricted MRF model from the image restoration literature [9], 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, 19, 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—our approach operates as an auxiliary routine that does not require modifying existing stereo code. Towards this end, we show that the truncated absolute distance commonly used in leading stereo algorithms [2, 4] corresponds to a mixture of an exponential distribution and an outlier process. 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 [3]. Finally, we benchmark our approach on the Middlebury database, 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 [3]).

In this work, we use insights from statistical learning to improve vision algorithms. Our work is therefore related to Freeman et al. [6] 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 [5] infer 3D scene structure from a single image. Unlike Freeman et al.’s approach, our

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 for MRF-based stereo. In the energy function in Eq. (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 many other robust metrics [15]. In this paper, we focus on truncated absolute difference because it is a popular choice of top performing stereo algorithms [2, 4] and it has several good properties. First, it is derived from total variation [13], thus preserving discontinuities. Second, it does not have frontal parallel bias and satisfies the *metric* property required by the α -expansion algorithm in graph cut [2]. Third, it can be efficiently computed via distance transform [4] in belief propagation. Specifically,

$$\begin{aligned} U(d_i) &= \min(|I(x_i, y_i) - J(x_i - d_i, y_i)|, \sigma) \\ V(d_i, d_j) &= \min(|d_i - d_j|, \tau) \end{aligned} \tag{2}$$

where I and J are the image pairs, and σ and τ are truncation thresholds.

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 can not 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 Figure 2(a,c), using the ground truth disparities from the Middlebury website [15], we plot the histograms of pixel matching errors and neighboring disparity differences for the Tsukuba stereo pair. In Figure 2(b,d), 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 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. Figure 2(a,c) shows 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 Figure 2(b,d). From the shapes of the fitted distributions, we can recover the optimal set of MRF parameters, as we describe later in the paper.

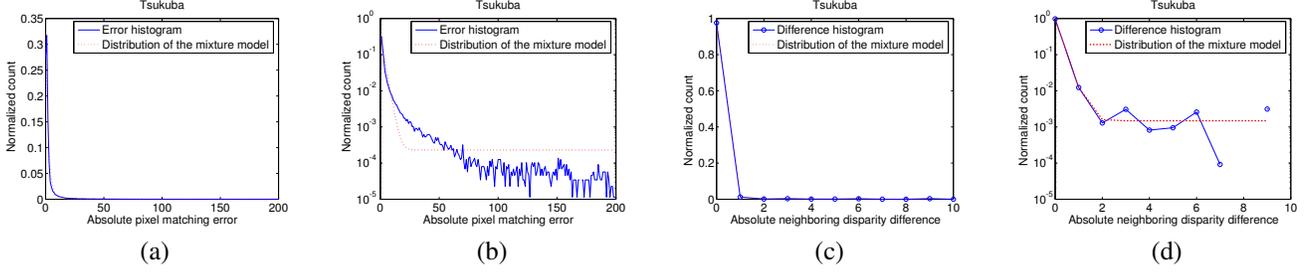


Figure 2. Histograms of errors between corresponding pixels in two images in linear (a) and log (b) scale. Superimposed on the plots is the fitted mixture model. (c) and (d) show histograms and models for neighboring disparity difference.

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 .¹ 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

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

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

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

¹For brevity, we refer to γ_i as a *visibility* variable, but it can also account for differences in brightness, e.g., due to specularity.

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 percent 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 field $D = \{d_i\}$ over I and two sets of random variables $\Gamma = \{\gamma_i\}$ and $\Theta = \{\theta_g\}$ for pixel visibility and edge connectivity, respectively.²

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

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

where the prior on $(\alpha, \mu, \beta, \nu)$ is assumed to be uniform. We can factor $P(I, J, D | \alpha, \mu, \beta, \nu)$ as

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

and 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.

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

²This model is called ‘‘three coupled MRF’’s in [16].

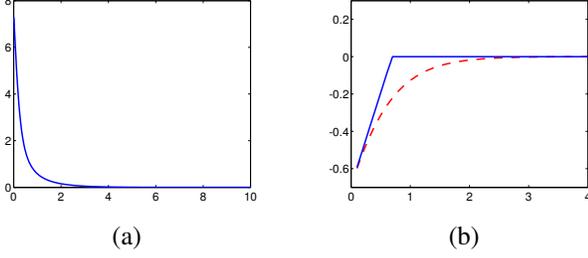


Figure 3. (a) A graph for Eq. (18) for $L = 16$. (b) An illustration of Eq. (24) (solid curve) as an upper bound for Eq. (23) (dotted curve) for $(a, b, c) = (1, 2, 1)$.

which assumes $P(\Gamma|D, \alpha) = P(\Gamma|\alpha)$, ignoring the dependence of visibility on geometry for computational convenience. Similarly,

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

which assumes the independence between Δd_g also for computational convenience.

Putting Eqs. (11,12) together, we obtain

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

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 Eqs. (1,2).

5 An alternating optimization

In this section, we present an alternating algorithm to maximize Eq. (13). Given D , we apply the EM algorithm to estimate (α, μ) by maximizing $P(I, J|D, \alpha, \mu)$ and estimate (β, ν) by maximizing $P(D|\beta, \nu)$. From $(\alpha, \mu, \beta, \nu)$, we then compute the optimal MRF parameters (σ, τ, λ) .

5.1 Estimating β and ν given D

The EM algorithm is well suited for estimating parameters for mixture models. Given D , we first compute

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

Then we compute the conditional probability of θ_g as

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

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

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

By maximizing Eq. (16), we obtain the following relation

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

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 Eq. (18). f monotonically decreases from $\frac{L-1}{2}$ to 0 over $[0, \infty)$, as shown in Figure 3(a). When L is large, the second term in $f(\nu; L)$, $\frac{L}{e^{L\nu} - 1}$, is negligible, and the equation has a close-form solution $\nu_0 = \log(\frac{1}{y} + 1)$, where y is the right hand side of Eq. (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

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

Then we can estimate α and μ in the same way as we estimate β and ν using Eqs.(15,17,18) 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 Eq. (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

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

Eq. (21) can be minimized directly using existing techniques. For example, Sun et al. [16] use Belief Propagation to minimize a form of Eq. (21).

5.3.1 From mixture model to regularized energy

Although Eq. (21) can be optimized directly, it is not in a form that efficient MRF solvers [2, 4] assume. Recall that our objective is to interface with and boost the performance of existing stereo algorithms, and we therefore want to convert Eq. (21) to the form of Eqs. (1,2). 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 Figure 3(b). Therefore, minimizing Eq. (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

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

To further simplify the problem, we 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 . Eq. (27) is the objective function used in [2, 4] for stereo matching.

Now, we summarize our algorithm as follows.

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 $L, \beta,$ and ν by iterating EM Eqs. (14,15,17,18)
- Update $N, \alpha,$ 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.

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 7. STEREO-MATCHING could be any stereo algorithm that works with Eq. (27).

6 Intensity gradient cues

Recent stereo algorithms use static cues, such as color segments [17, 16] and color edges [7, 1], to improve performance. Here, we show that neighboring pixel intensity difference [2] 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,

$$P(\Delta I_g | \theta_g) = \begin{cases} \xi e^{-\kappa |\Delta I_g|}, & \theta_g = 1. \\ \frac{1}{K}, & \theta_g = 0. \end{cases} \quad (28)$$

where κ is the decay rate, $|\Delta I_g| \in \{0, 1, \dots, K-1\}$, 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

$$P(\Delta I_g, \Delta d_g) = \beta \xi \eta e^{-(\kappa |\Delta I_g| + \nu |\Delta d_g|)} + (1-\beta) \frac{1}{KL} \quad (29)$$

Given $I, \Delta I,$ and J , our goal is to recover $D, \alpha, \mu, \beta, \nu,$ and κ , by maximizing

$$\begin{aligned} & P(D, \alpha, \mu, \beta, \kappa, \nu | I, \Delta I, J) \\ & \propto \underbrace{\prod_i P(e(d_i) | d_i, \alpha, \mu)}_{P(I, J | D, \alpha, \mu)} \underbrace{\prod_g P(\Delta I_g, \Delta d_g | \beta, \kappa, \nu)}_{P(\Delta I, D | \beta, \kappa, \nu)} \end{aligned} \quad (30)$$

The alternating algorithm in Section 5 can still be applied with minor a change. The estimation of α and μ is the same as before. The estimation of $\beta, \kappa,$ and ν can be done as follows. Initially, we set $K = \max\{\Delta I_g\} + 1$. For each iteration, we first update L as in Eq. (14). Then we compute the condition probability of θ_g as

$$\omega_g \stackrel{\text{def}}{=} \frac{\beta \xi \eta e^{-(\kappa |\Delta I_g| + \nu |\Delta d_g|)}}{\beta \xi \eta e^{-(\kappa |\Delta I_g| + \nu |\Delta d_g|)} + \frac{1-\beta}{KL}} \quad (31)$$

Finally, we update β and ν using Eqs. (17) and (18), respectively, and update κ also using Eq. (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 Eq. (21) with ρ_p depending on ΔI_g . Specifically,

$$\rho_p(d_i; \Delta I_g, \beta, \kappa, \nu) = -\log(\beta \xi \eta e^{-\kappa |\Delta I_g|} e^{-\nu |\Delta d_g|} + \frac{1-\beta}{KL}) \quad (32)$$

Accordingly, s_p and t_p also depend on ΔI_g .

$$s_p = \frac{\beta \xi \eta \nu e^{-\kappa |\Delta I_g|}}{\beta \xi \eta e^{-\kappa |\Delta I_g|} + \frac{1-\beta}{KL}} \quad t_p = \log\left(1 + \frac{\beta \xi \eta K L e^{-\kappa |\Delta I_g|}}{1-\beta}\right) \quad (33)$$

In Eq. (33), 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.

7. Results

We implemented the EM algorithm to estimate the MRF parameters, and implemented belief propagation (BP) using distance transform [4] 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 run-time is about 6 minutes.

7.1 Convergence

In our first experiment, we tested our algorithm on the four Middlebury benchmarks. In Figure 4, we show the disparity maps and corresponding (σ, τ, λ) at iteration 1, 2, 4, 6 for the four cases. Initially, the regularization is weak and the disparity map is noisy. As the algorithm proceeds, the regularization increases and the disparity map becomes cleaner.

In our second experiment, we repeated experiment 1, but with different initial values. We show the initial and final (σ, τ, λ) , including those of experiment 1, in the top five rows in the tables in Table 1. 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 8.

7.2 Optimality

In our third experiment, for each case in experiment 1, 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 Figure 5. The red bar indicate our estimated values, whose corresponding error rates are quite close to the minimum error rates of the graphs in all the four benchmarks.

7.3 Improvement

In our fourth 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 [4] and the result is shown in the fourth row ("Fixed") in Table 2. Then we compare this result with the result in the first experiment where we solve for (σ, τ, λ) , shown in the third row ("Adaptive") in Table 2. Our adaptive method shows similar results to the fixed parameters for Sawtooth and Venus, but dramatic improve-

ment on Map. As reported in [16], the Map pair requires different parameter settings than the other three datasets. 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 could 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.

The experiments so far do not include the intensity gradient cue proposed in Section 6. 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 6, we get the first row ("Adaptive+grad") of Table 2. 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 six slots, and the resulting adaptive algorithm is ranked the *fifth* among all stereo algorithms in the Middlebury rankings, as of April 24, 2005.

8. 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 ground truth than to the estimated disparity maps [18]. 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 . One piece of future work would be to model the occlusion process more precisely. 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.

Acknowledgments

We would like to thank Aaron Hertzmann and Richard Szeliski for helpful discussions. This work was supported in part by National Science Foundation grant IIS-0049095,

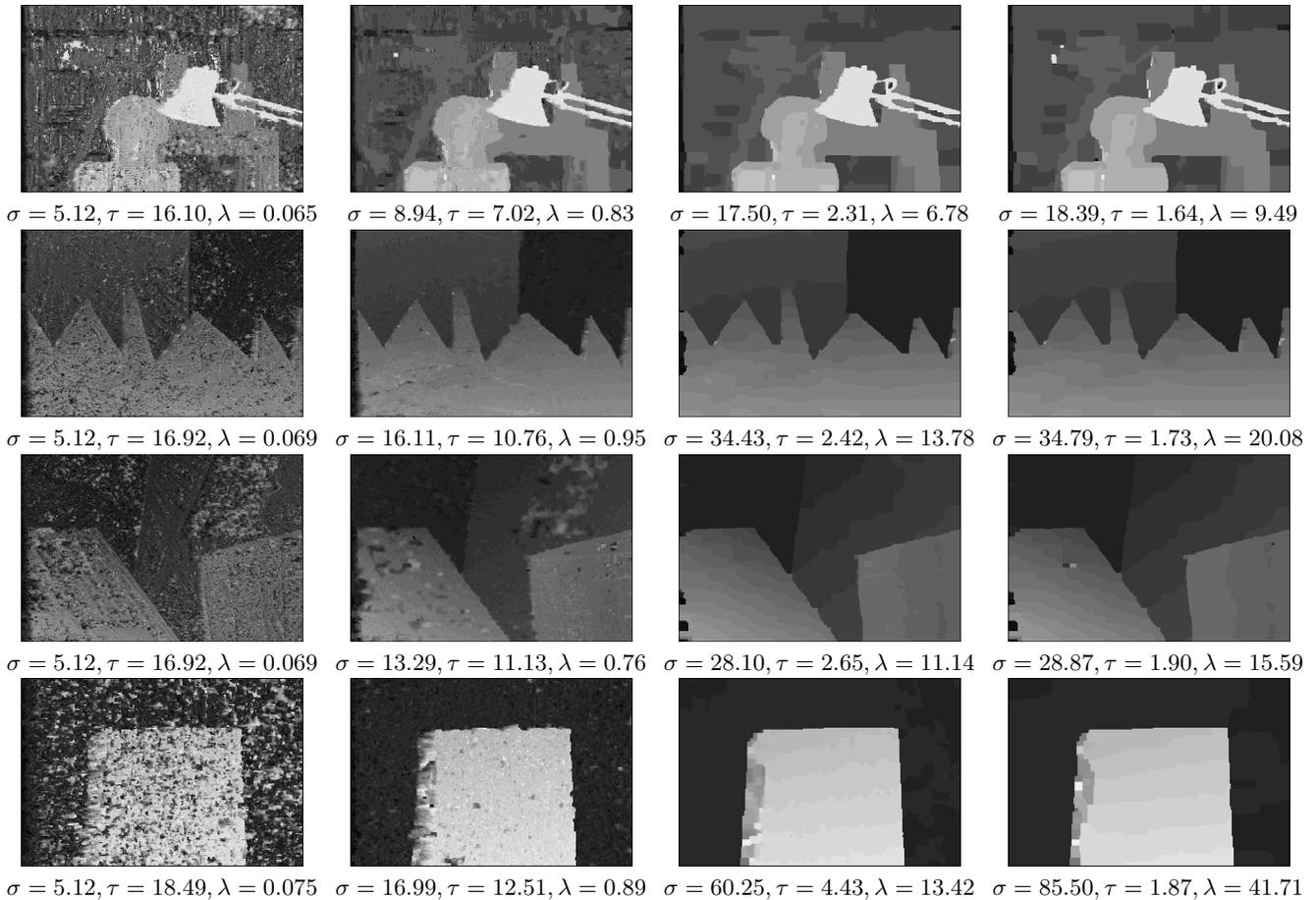


Figure 4. Convergence on the four Middlebury benchmarks. The four columns correspond to iteration 1, 2, 4, and 6.

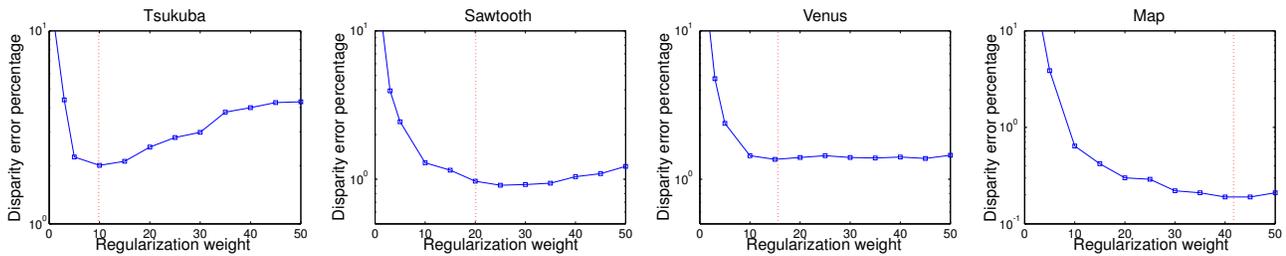


Figure 5. Graphs of error rate with respect to ground truth as a function of regularization weight λ while fixing (σ, τ) . The vertical dotted lines are our estimation for λ .

an Office of Naval Research YIP award, and Microsoft Corporation.

References

- [1] A. F. Bobick and S. S. Intille. Large occlusion stereo. *IJCV*, 33(3):181–200, 1999.
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on PAMI*, 23(11):1222–1239, 2001.
- [3] L. Cheng and T. Caelli. Bayesian stereo matching. In *Proc. CVPR Workshop*, pages 192–192, 2004.
- [4] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. In *Proc. CVPR*, pages 261–268, 2004.
- [5] W. Freeman and A. Torralba. Shape recipes: Scene representations that refer to the image. In *Proc. NIPS.*, pages 1335–1342, 2003.
- [6] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *IJCV*, 40(1):25–47, 2000.
- [7] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6:35–49, 1993.
- [8] S. Geman and G. Geman. Stochastic relaxation, gibbs dis-

Tsukuba						
Initial para			Final para			disparity error(%)
σ	τ	λ	σ	τ	λ	
5.12	2.60	0.91	18.49	1.61	9.75	2.23
33.66	2.60	9.42	18.53	1.60	9.84	2.26
1.11	2.60	0.18	18.47	1.62	9.67	2.15
5.12	16.10	0.065	18.39	1.64	9.49	2.12
5.12	0.59	4.71	18.52	1.60	9.83	2.24
Ground truth disparities			17.44	1.44	10.83	2.29

Sawtooth						
Initial para			Final para			disparity error(%)
σ	τ	λ	σ	τ	λ	
5.12	2.82	0.93	34.79	1.72	20.12	0.97
33.66	2.82	9.65	34.78	1.72	20.10	0.97
1.11	2.82	0.19	34.44	1.72	20.07	0.99
5.12	16.92	0.069	34.79	1.73	20.08	0.97
5.12	0.64	4.79	34.78	1.72	20.08	0.98
Ground truth disparities			31.72	1.59	21.62	0.99

Venus						
Initial para			Final para			disparity error(%)
σ	τ	λ	σ	τ	λ	
5.12	2.82	0.93	28.88	1.85	15.82	1.34
33.66	2.82	9.65	28.88	1.84	15.90	1.33
1.11	2.82	0.19	28.89	1.84	15.88	1.34
5.12	16.92	0.069	28.87	1.90	15.59	1.33
5.12	0.64	4.79	28.88	1.84	15.84	1.33
Ground truth disparities			26.54	1.75	15.38	1.42

Map						
Initial para			Final para			disparity error(%)
σ	τ	λ	σ	τ	λ	
5.12	3.15	0.96	86.39	1.81	43.08	0.20
33.66	3.15	9.89	87.01	1.81	42.98	0.19
1.11	3.15	0.19	85.71	1.81	43.05	0.20
5.12	18.50	0.075	85.50	1.87	41.71	0.18
5.12	0.71	4.87	86.55	1.81	43.01	0.20
Ground truth disparities			60.58	1.81	46.80	0.19

Table 1. Convergence of our algorithm under five different initializations. The last rows use ground truth disparity maps to estimate parameters and then compute new disparity maps with these estimated parameters.

Parameter setting	Overall ranking	Tsukuba			Sawtooth			Venus			Map	
		all	untex.	disc.	all	untex.	disc.	all	untex.	disc.	all	disc.
Adaptive+Grad	5	1.87 ¹⁶	0.67 ¹⁰	7.13 ⁷	0.83 ¹³	0.32 ¹⁹	3.48 ⁷	1.53 ¹⁵	0.92 ⁸	10.37 ¹⁷	0.20 ³	2.20 ¹
Fixed+Grad 2	11	1.84 ¹⁴	1.05 ¹³	9.87 ¹⁷	0.87 ¹³	0.28 ¹⁷	5.78 ¹⁵	1.22 ¹²	1.05 ⁹	13.39 ²¹	0.20 ³	2.49 ³
Adaptive	13	2.12 ¹⁷	1.36 ¹⁷	10.76 ¹⁷	0.97 ¹³	0.31 ¹⁹	6.79 ¹⁶	1.33 ¹³	1.13 ¹⁰	14.65 ²⁴	0.18 ²	2.20 ¹
Fixed	19	1.84 ¹⁴	1.33 ¹⁶	10.02 ¹⁷	1.24 ¹⁸	0.32 ¹⁹	7.18 ¹⁹	1.34 ¹³	1.18 ¹¹	15.17 ²⁴	0.38 ¹²	3.47 ⁷
Fixed+Grad 1	31	7.68 ³²	5.76 ²⁸	11.79 ¹⁷	5.92 ³⁵	0.30 ¹⁸	13.12 ²⁵	6.90 ³⁴	3.20 ²⁵	14.11 ²³	20.19 ³⁶	35.41 ³⁶

Table 2. Performance comparison of fixed and adaptive stereo solvers. 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 rankings are as of 4/24/2005.

tribution and the bayesian restoration of images. *IEEE Trans. on PAMI*, 6:721–741, 1984.

[9] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak. Fully bayesian estimation of gibbs hyperparameters for emission computed tomography data. *IEEE Trans. on Medical Imaging*, 16(5):516–26, 1997.

[10] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Trans. on PAMI*, 16(9):920–932, 1994.

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

[12] A. Mohammad-Djafari. Joint estimation of parameters and hyperparameters in a bayesian approach of solving inverse problems. In *Proc. ICIP*, pages 473–477, 1996.

[13] L. I. Rudin and S. Osher. Total variation based image restoration with free local constraints. In *Proc. ICIP*, pages 31–35, 1994.

[14] S. S. Saquib, C. A. Bouman, and K. Sauer. ML paramter estimation for markov random fields with applications to bayesian tomography. *IEEE Trans. on Image Processing*, 7(7):1029–1044, 1998.

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

[16] J. Sun, H.-Y. Shum, and N. N. Zheng. Stereo matching using belief progation. In *Proc. ECCV*, pages 510–524, 2002.

[17] H. Tao, H. S. Sawhney, and R. Kumar. Dynamic depth recovery from multiple synchronized video streams. In *Proc. CVPR*, pages 118–124, 2001.

[18] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *Proc. ICCV*, pages 900–906, 2003.

[19] Z. Zhou, R. M. Leahy, and J. Qi. Approximate maximum likelihood hyperparameter estimation for gibbs priors. *IEEE Trans. on Image Processing*, 6(6):844–861, 1997.