Learning Conditional Random Fields for Stereo

Daniel Scharstein
Middlebury College
Middlebury, VT, USA
schar@middlebury.edu

Chris Pal
University of Massachusetts Amherst
Amherst, MA, USA
pal@cs.umass.edu

Abstract

State-of-the-art stereo vision algorithms utilize color changes as important cues for object boundaries. Most methods impose heuristic restrictions or priors on disparities, for example by modulating local smoothness costs with intensity gradients. In this paper we seek to replace such heuristics with explicit probabilistic models of disparities and intensities learned from real images. We have constructed a large number of stereo datasets with ground-truth disparities, and we use a subset of these datasets to learn the parameters of Conditional Random Fields (CRFs). We present experimental results illustrating the potential of our approach for automatically learning the parameters of models with richer structure than standard hand-tuned MRF models.

1. Introduction and related work

In recent years, machine learning methods have been successfully applied to a large number of computer vision problems, including recognition, super-resolution, inpainting, texture segmentation, denoising, and context labeling. Stereo has remained an exception because of the lack of sufficient training data with ground-truth disparities. While a few datasets with known disparities are available, they have mainly been used for benchmarking of stereo methods [1].

The goal of this paper is to replace the heuristic cues used in previous approaches with probabilistic models derived from real images. To obtain a sufficient amount of training data, we used the structured-lighting approach of [2] to construct a database of 30 stereo pairs with ground-truth disparities, which we provide for use by other researchers at http://vision.middlebury.edu/stereo/data/.

In this paper we present a Conditional Random Field (CRF) model for stereo vision and derive a gradient-based learning approach that leverages efficient graph-cut minimization methods and our ground-truth database. We then explore the characteristics and properties of a number of different models when learning model parameters.

Among the few existing learning approaches for stereo, the most prominent is the work by Zhang and Seitz [3], who iteratively estimate the global parameters of a Markov Random Field (MRF) stereo method from the previous disparity estimates, and thus do not rely on ground-truth data. Kong and Tao [4] learn to categorize matching errors of local methods using the Middlebury images. Kolmogorov et al. [5] construct MRF models for binary segmentation using locally learned Gaussian Mixture Models (GMMs) for foreground and background colors.

While learning approaches have been sparse, there has nevertheless been much recent progress in stereo vision. Breakthroughs have been achieved along two avenues.

First, global optimization methods have become practical with the emergence of powerful optimization techniques. Considered too slow when first proposed, global methods currently dominate the top of the Middlebury stereo rankings. In particular, MRF models for stereo have become popular since high-quality approximate solutions can be obtained efficiently using graph cuts (GC) [6–8] and belief propagation (BP) [9–11]. Tappen and Freeman [12] compare GC and BP for stereo; Szeliski et al. [13] compare a larger set of MRF minimization techniques and provide software that we use in our implementation.

The second breakthrough has been the realization of the importance of intensity changes as a cue for object boundaries, i.e., disparity discontinuities. Taken to an extreme, this translates into the assumption that disparity jumps always coincide with color edges, which is the basis of a large number of recent segmentation-based stereo methods [10, 14–19]. Such methods start with a color segmentation and then assume that within each segment disparities are constant, planar, or vary smoothly. This assumption works surprisingly well if the segments are small enough.

Segmentation is not the only way to utilize this monocular cue; many pixel-based global methods also change the smoothness cost (i.e., penalty for a disparity change) if the local intensity gradient is high [1, 6, 9, 20]. This is the approach taken here, where we learn the relationship between intensity gradient and smoothness cost from real images.

The probabilistic models we develop in Section 2 below are Conditional Random Fields (CRFs). CRFs are obtained from the conditional distribution defined for a subset of random variables in a Markov Random Field. The parameters of a CRF can then be optimized for a given dataset based on the corresponding conditional likelihood.

The CRF approach was first articulated for sequence processing problems [21]. In the linear models commonly used in language processing, the feature expectations required for gradient-based CRF optimization can be computed efficiently [22]. For many graphical models with more complex structure, however, approximate inference methods must be used. Dynamic conditional random fields [23] use a factorized set of variables at each segment of a linear-chain CRF, yielding a shallow lattice-structured model. Approximate inference and learning methods include loopy belief propagation and tree-based reparameterization [24]. Kumar and Hebert [25] optimize the parameters of lattice-structured binary CRFs using a pseudolikelihood approach. Other work [26] has investigated the discriminative optimization of a lattice-structured joint random field models using autoregression over the pseudolikelihood. These learning methods use spatially localized approximations to the true global distributions needed for learning. In contrast, our approach described in Section 3 uses fast graph-cuts based methods to compute good approximate global minimizations that correspond to *most-probable-explanation* (MPE) [27] estimates. These estimates are then used to create approximate model expectations required for gradient based learning of model parameters.

Finally, related work on analyzing motion parallax has used priors on the probability that an object at a given depth is visible [28]. As we use a conditional model with unnormalized factors, there is no explicit prior on the distributions for disparities in our framework. It is possible within a discriminative framework to introduce local potential functions that depend only upon local disparity values and play a similar role to a prior. The influence of such knowledge, however, can often also be achieved through the parameters of the local cost and pairwise potentials.

2. CRFs for Stereo

We define the disparity of pixel $p \in \mathcal{P}$, the set of all pixels in the reference (left) image, as a random variable d_p with N discrete but ordered states. Assuming rectified images, d_p represents the horizontal shift in pixels with respect to the other image. We define $\mathbf{c}_p$ as a vector of N continuous random variables representing the matching cost for each discrete disparity level. In this paper we use the sampling-insensitive cost of [29], which is the minimum distance between the linearly interpolated left and right scanlines over $x \pm 1/2$ at each pixel. For color images we use the sum of this measure over all color bands. We de-
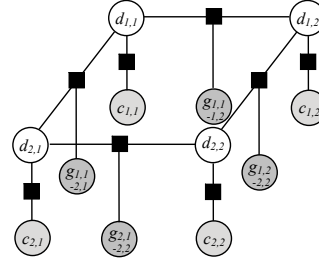

Figure 1. The repeating unit used in our graphical model.

fine g_{pq} as an M-state random variable for discretized color gradients between neighboring pixels $(p, q) \in \mathcal{N}$, where $\mathcal{N}$ is the standard 4-neighborhood. We compute the gradients as the RMS difference between color bands.

We then construct conditional random fields for disparities $\mathcal{D} = \{d\}$, matching costs $\mathcal{C} = \{\mathbf{c}\}$, and gradients $\mathcal{G} = \{g\}$ with the following form

$$P(\mathcal{D}|\mathcal{C}, \mathcal{G}) = \frac{1}{Z(\mathcal{C}, \mathcal{G})} \prod_{p \in \mathcal{P}} \Phi(d_p, \mathbf{c}_p) \prod_{(p,q) \in \mathcal{N}} \Psi(d_p, d_q, g_{pq}), \tag{1}$$

where the (conditional) partition function $Z(\mathcal{C}, \mathcal{G})$ is obtained by summing over all possible disparity values

$$Z(\mathcal{C}, \mathcal{G}) = \sum_{d \in \mathcal{D}} \prod_{p \in \mathcal{P}} \Phi(d_p, \mathbf{c}_p) \prod_{(p,q) \in \mathcal{N}} \Psi(d_p, d_q, g_{pq}). \tag{2}$$

The potential function Φ models the agreement of disparites and intensities, while Ψ jointly models the smoothness of neighboring disparities and the color gradient between them. Figure 1 illustrates the factorization we use for our models using a factor graph [30].

The CRF described by (1) represents a general formulation. We now present the specific model used in our exploration here in more detail. First, we take the negative log probability of our model and obtain U and V terms analogous to the data and smoothness terms commonly used in other energy-based stereo techniques,

$$U(d_p, \mathbf{c}_p) = -\log \Phi(d_p, \mathbf{c}_p), \tag{3}$$
$$V(d_p, d_q, g_{pq}) = -\log \Psi(d_p, d_q, g_{pq}). \tag{4}$$

Our goal is to minimize the sum of the negative log conditional probabilities

$$-\log P(\mathcal{D}|\mathcal{C}, \mathcal{G}) = \log Z(\mathcal{C}, \mathcal{G}) + \sum_{p \in \mathcal{P}} U(d_p, \mathbf{c}_p)$$
$$+ \sum_{(p,q) \in \mathcal{N}} V(d_p, d_q, g_{pq}). \tag{5}$$

Note that our formulation, unlike other energy-based stereo approaches, explicitly accounts for the partition function.

We express cost terms U and pairwise smoothness terms V using a linear combination of feature functions f_u, f_v,

$$U(d_p, \mathbf{c}_p) = \sum_u \theta_u f_u(d_p, \mathbf{c}_p), \qquad (6)$$

$$V(d_p, d_q, g_{pq}) = \sum_v \theta_v f_v(d_p, d_q, g_{pq}), \qquad (7)$$

where θ_u, θ_v are the parameters of our model. The notation follows the usual format for specifying the potential functions of CRFs [21, 22], and the linear form allows us to derive an intuitive gradient-based minimization procedure for parameter estimation. Here, we define $f_u(d_p, \mathbf{c}_p)$ to return $c_p[d_p]$ (i.e., the matching cost at pixel p given disparity d_p) if $u = d_p$ and 0 otherwise. For our experiments in Section 5 we fix all θ_u to one, yielding $U(d_p, \mathbf{c}_p) = c_p[d_p]$, since the focus of this paper is on exploring the impact of learning parameters that modulate disparity smoothness by image intensity gradients. However, we shall derive learning equations in the next section using the more general formulation including the θ_u parameters.

For the V terms we use a gradient-modulated Potts model [1, 6] to express the relationship between color changes and depth changes. Unlike existing approaches that use a single gradient threshold, we use parameters $\theta_{v=1}, \ldots, \theta_{v=k}$ as modulation costs associated with each of K different gradient bins. We define our binary feature functions such that

$$V(d_p, d_q, g_{pq}) = \begin{cases} 0 & \text{if } d_p = d_q \\ \theta_{v=1} & \text{if } d_p \neq d_q \text{ and } g_{pq} = 1 \\ \ldots & \ldots \\ \theta_{v=k} & \text{if } d_p \neq d_q \text{ and } g_{pq} = k, \end{cases} \qquad (8)$$

where the discrete gradient variable g_{pq} represents the interval containing the real-valued gradient. In Section 5 below we explore a spectrum of gradient discretization schemes ranging from one to six intervals, with interval breakpoints from the set $\{0, 2, 4, 8, 12, 16, \infty\}$.

A gradient-modulated Potts model is one of the simplest ways of relating smoothness and color changes. In this paper we learn CRFs using models with simple structure. While we expect and indeed find that the benefits of models in this class are limited, our approach is easily generalized to more sophisticated models. In future work we plan to learn more general smoothness terms, e.g., dependent upon disparity changes $|d_p - d_q|$.

3. Learning

To simplify our exposition, we define feature vectors for each each pixel location p and edge pq as $F_p(d_p, \mathbf{c}_p) = \{f_u(d_p, \mathbf{c}_p)\}$ and $F_{pq}(d_p, d_q, g_{pq}) = \{f_v(d_p, d_q, g_{pq})\}$.

Similar to [31], we define *global* feature functions as

$$\mathbf{F}_u(\mathcal{D}, \mathcal{C}) = \sum_{p \in \mathcal{P}} F_q(d_p, \mathbf{c}_q),$$

$$\mathbf{F}_v(\mathcal{D}, \mathcal{G}) = \sum_{(p,q) \in \mathcal{N}} F_{pq}(d_p, d_q, g_{pq}). \qquad (9)$$

We wish to optimize the parameters $\Theta = [\Theta_v; \Theta_u]$ of our CRF for the log conditional probability of the data, which can now be expressed as

$$\mathcal{L}(\Theta) = \sum_i \log p(D_i | C_i, G_i; \Theta)$$

$$= \sum_i \left(\Theta^T \mathbf{F}(\mathcal{D}_i, \mathcal{G}_i, \mathcal{C}_i) - \log Z(\mathcal{C}_i, \mathcal{G}_i) \right) \qquad (10)$$

for $i = 1 \ldots N$ training images and with $\mathbf{F}(\mathcal{D}, \mathcal{G}, \mathcal{C}) = [\mathbf{F}_v(\mathcal{D}, \mathcal{G}); \mathbf{F}_u(\mathcal{D}, \mathcal{C})]$. Under this construction, the analytic gradient with respect to parameters Θ can be expressed as

$$\nabla \mathcal{L} \propto \left\langle \mathbf{F}(\mathcal{D}, \mathcal{G}, \mathcal{C}) \right\rangle_{\tilde{p}(\mathcal{D}, \mathcal{G}, \mathcal{C})} - \left\langle \left\langle \mathbf{F}(\mathcal{D}, \mathcal{G}, \mathcal{C}) \right\rangle_{p(\mathcal{D}|\mathcal{C}, \mathcal{G})} \right\rangle_{\tilde{p}(\mathcal{C}, \mathcal{G})}, \qquad (11)$$

where $\langle \cdot \rangle_p$ denotes the expectation under the probability distribution $p(\cdot)$, and $\tilde{p}(\cdot)$ denotes the empirical distribution of variables in the argument. The first term in (11) is computed by evaluating our feature functions across our ground truth disparities, image gradients and costs. The second term in (11) arises from the gradient of the partition function. Its outer expectation is computed by observing the costs $\mathcal{C}$ and gradients $\mathcal{G}$ for each image and computing the inner expectation from the corresponding $p(D|C, G; \Theta)$. In linear CRFs this can be done efficiently with a forward-backward pass analogous to the well-known algorithm used for HMMs. Sha and Pereira [31] provide a review of methods for optimizing CRFs when this expectation can be computed exactly. However, here the expectation involving $p(D|C, G)$ is intractable due to the lattice structure of our model, and therefore must be approximated.

To achieve this approximation we use the fact that for an observed $\mathcal{C}$ and $\mathcal{G}$ in (5) the log partition function is constant. We can thus use the fast alpha-expansion graph-cuts algorithm [6, 13] to minimize our function for the first two terms involving Vs and Us. This allows us to obtain a good approximation to the *most probable explanation* (MPE) under the conditional probability distribution defined by our model with the current settings of parameters.

Other work [23] has found that loopy belief propagation can be used effectively to compute approximate marginals and from them, approximate expectations. However, here we use graph cuts for approximate MPE inference since it is faster, and also because the results of [13] suggest that graph cuts finds lower energy solutions than BP. Once we

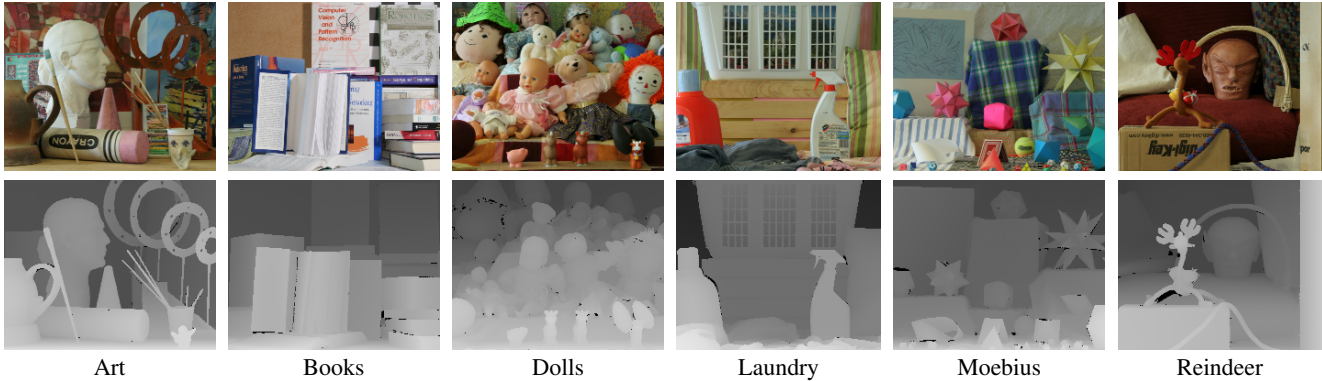| Art | Books | Dolls | Laundry | Moebius | Reindeer |

Figure 2. The six datasets used in this paper. Shown is the left image of each pair and the corresponding ground-truth disparities.

have an MPE estimate from running graph cuts we use it to compute our expectation in a manner similar to the empirical distribution. Training a lattice-structured model using the approach described here is thus a generalization of Viterbi path-based methods described in [32]. For our learning experiments we use straightforward gradient-based updates with a variable learning rate.

4. Datasets

In order to obtain a significant amount of training data for stereo learning approaches, we have created 30 new stereo datasets with ground-truth disparities using an automated version of the structured-lighting technique of [2]. Our datasets are available for use by other researchers at `http://vision.middlebury.edu/stereo/data/`. Each dataset consists of 7 rectified views taken from equidistant points along a line, as well as ground-truth disparity maps for viewpoints 2 and 6. The images are about 1300×1100 pixels (cropped to the overlapping field of view), with about 150 different integer disparities present. Each set of 7 views was taken with three different exposures and under three different lighting conditions, for a total of 9 different images from each viewpoint.

For the work reported in this paper we only use the six datasets shown in Figure 2: Art, Books, Dolls, Laundry, Moebius and Reindeer. As input images we use a single image pair (views 2 and 6) taken with the same exposure and lighting. In future work we plan to ultilize the other views and the additional datasets for learning from much larger training sets. To make the images tractable by the graph-cut stereo matcher, we downsample the original images to one third of their size, resulting in images of roughly 460×370 pixels with a disparity range of 80 pixels. The resulting images are still more challenging than standard stereo benchmarks such as the Middlebury Teddy and Cones images, due to their larger disparity range and higher percentage of untextured surfaces.
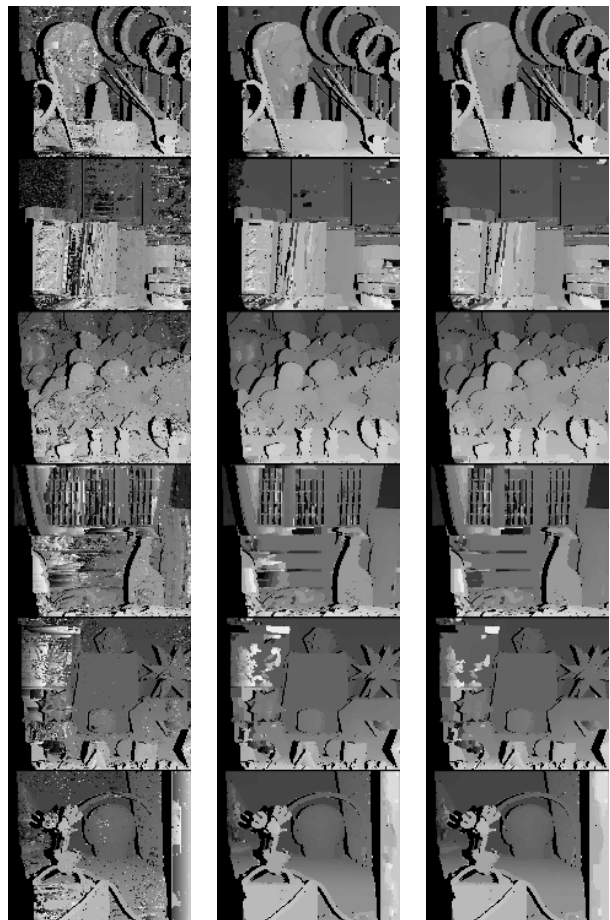

Figure 3. Disparity maps of the entire training set for $K = 3$ parameters after 0, 10, and 20 iterations. Occluded areas are masked.

5. Experiments

In this section we first examine the convergence of learning for models with different numbers of parameters θ_v, using all six datasets as training set. We then use a leave-one-out approach to evaluate the performance of the learned parameters on a new dataset. Finally, we examine how the
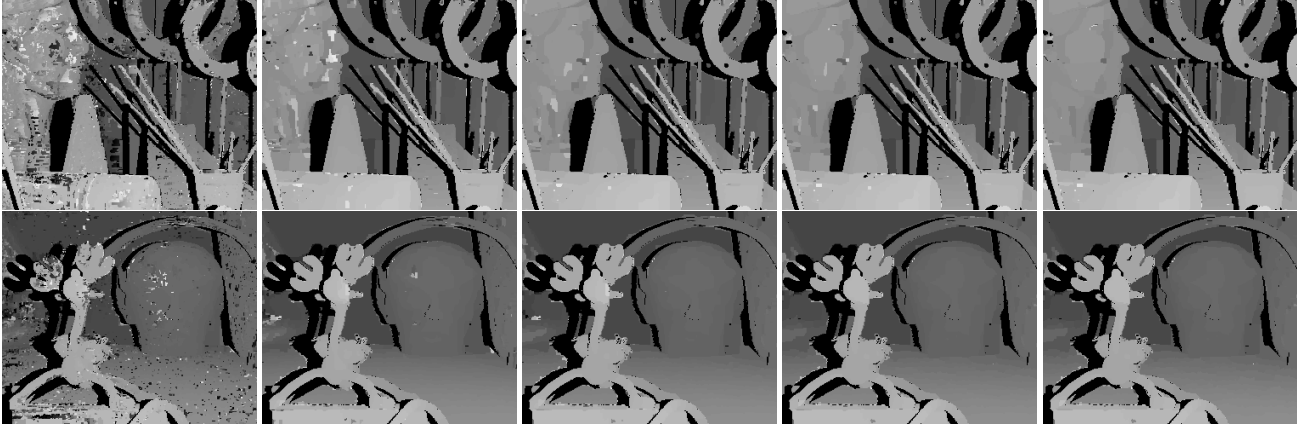
Figure 4. Two zoomed views of the disparity maps for $K=3$ parameters and learning on all six data sets after 0, 5, 10, 15, and 20 iterations. Occluded areas are masked.
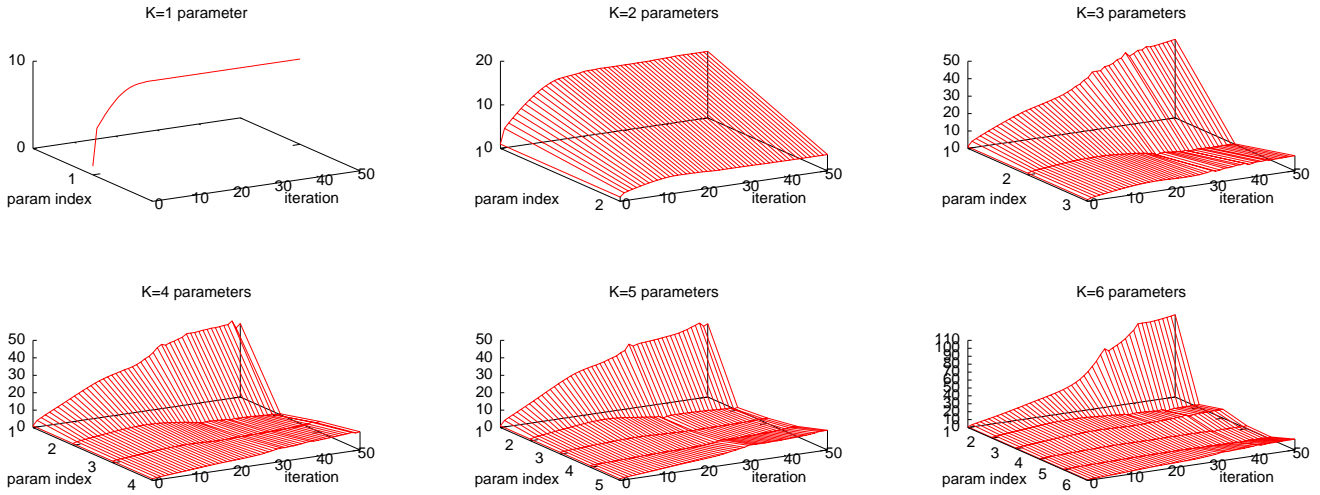

Figure 5. Learning on all six datasets using models with different numbers of parameters θ. See Table 1 for the learned parameter values.

learned parameters generalize to other datasets.

For our experiments we use a straightforward gradient-based optimization procedure: we start with a small learning rate (10^{-4}) and increase it by a small factor unless the norm of the gradient increases dramatically, in which case we backtrack and decrease the learning rate.

As mentioned in Section 2, in our experiments here we focus on learning the θ_v parameters of the pairwise V potentials, while holding the U potentials fixed with all $\theta_u = 1$. To simplify notation, we abbreviate $\theta_{v=1}, \theta_{v=2}, \ldots$ with $\theta_1, \theta_2, \ldots$ below.

It is important to account for the fact that we do not model occlusions in our CRF. It is well-known that spurious minimal-cost matches in occluded areas can cause artifacts in the inferred disparity maps. We therefore use our ground-truth data to mask out the contributions of variables in occluded regions to our gradient computation during training.

Intervals	0-2	2-4	4-8	8-12	12-16	16-∞
$\{\theta_k\}, K=1$	9.8					
$\{\theta_k\}, K=2$	15.3			3.7		
$\{\theta_k\}, K=3$	45.1		0.3	8.7		
$\{\theta_k\}, K=4$	42.2		0.5	5.6		10.4
$\{\theta_k\}, K=5$	42.0		1.6	3.1	5.9	11.3
$\{\theta_k\}, K=6$	104	3.9	11.2	3.8	3.0	13.7

Table 1. The gradient bins for $K = 1 \ldots 6$ parameters and the parameter values θ_k learned over all six datasets.

5.1. Convergence

We experiment with learning models possessing different numbers of parameters $\{\theta_k\}$, for $K = 1$ (i.e., a single global smoothness weight) to $K = 6$ (i.e., a parameter for each of 6 gradient bins). We first demonstrate the effective-
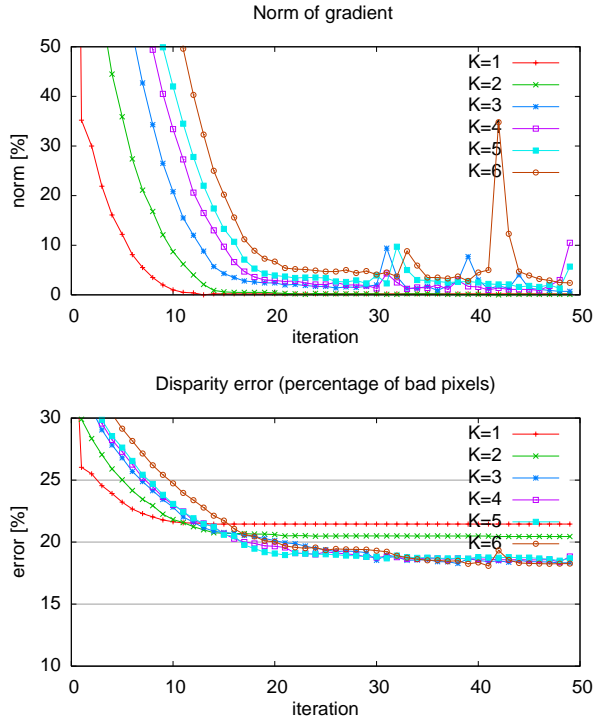
Figure 6. Gradient norm (top) and disparity errors (bottom) during learning on all 6 datasets.
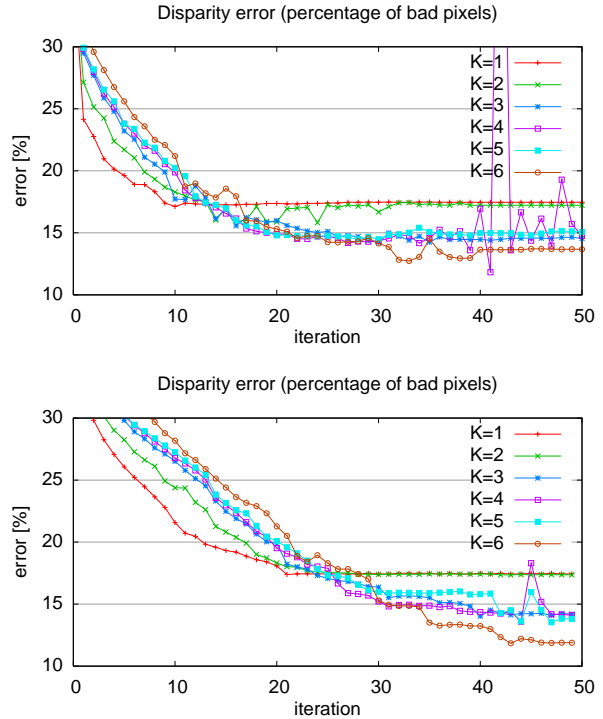

Figure 7. Results of leave-one-out learning on the Moebius dataset. Top: Moebius disparity errors using the parameters obtained during learning from the other 5 datasets. Bottom: Moebius disparity errors using the parameters learned from the dataset itself.

ness of the learning by training on all six datasets. It is useful to visualize the disparities predicted by the model over each iteration of learning. Figures 3 and 4 show how the disparity maps change during training. For clarity we have masked the occluded regions in black in these plots, since our model will assign arbitrary disparities in these areas. Figure 5 shows convergence of the individual parameters over 50 iterations. Table 1 shows the discretization strategy we use for image gradients as well as the final values of the learned parameters.

Figure 6 (top) shows how the gradient (i.e., difference between empirical and model expectation) converges to zero during learning, which indicates that the learning converges to a global minimum. Note that convergence is faster for fewer parameters. Figure 6 (bottom) shows the disparity errors during learning. Again, models with fewer parameters converge more quickly, thus yielding lower errors faster. However, the models with more parameters eventually outperform the simpler models. In Figure 6 (top) we observe that there appears to be an initial phase (e.g., during the first 25 iterations) where the norm of the approximate gradient monotonically decreases during the optimization. After this point, models with larger numbers of parameters appear to have less stability. This effect may be as a result of noisy gradient approximations due to our use of graph-cut-derived MPEs for the model expectation term of our gradient.

5.2. Performance of learned parameters

We now use 5 of the 6 datasets for training, and evaluate the disparity error of the remaining dataset using the parameters obtained during training. Figure 7 shows the results for the Moebius dataset. The top plot shows the errors during leave-one-out training. One can observe a similar trend as in figure Figure 6 (bottom), namely that the errors decrease during learning, and that the more complex models eventually outperform the simpler models. For comparison, the bottom plot in Figure 7 shows the errors when using the Moebius dataset itself for training. In this case finding a low-gradient solution means that we have effectively matched the distribution of disparity changes and associated intensity gradients of the ground-truth image. Not surprisingly, this results in lower errors, but not significantly lower than in the top plot — which indicates that the parameters learned from the other 5 images generalize reasonably well.

Figure 8 shows the equivalent plots for a different dataset, Reindeer. Again we show the errors during leave-one-out training at the top and those during training on the dataset itself on the bottom. Here we get slightly different results. First, the leave-one-out results no longer indicate
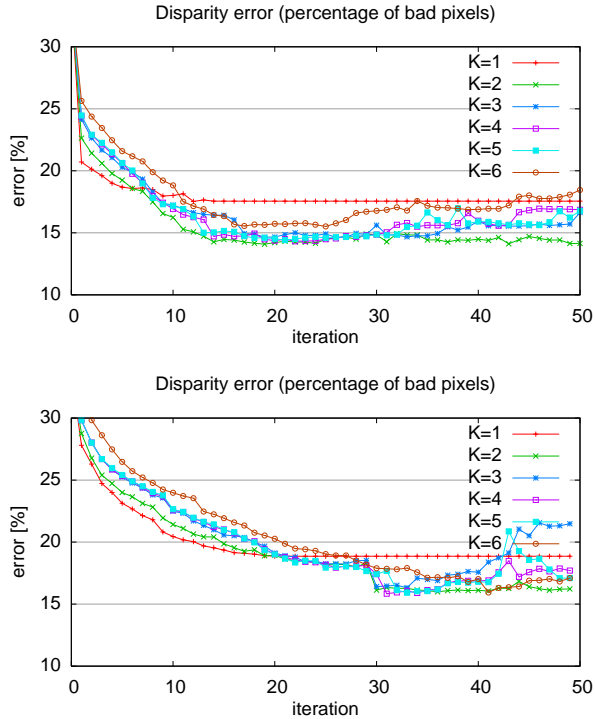
Figure 8. Results of leave-one-out learning on the Reindeer dataset. Top: Disparity errors using the parameters obtained during learning from the other 5 datasets. Bottom: Disparity errors using the parameters learned from the dataset itself.

	Tsukuba	Venus	Teddy	Cones	Average
$K=1$	3.0	1.3	11.1	10.8	6.6
$K=2$	2.2	1.6	11.3	10.7	6.5
$K=3$	3.1	2.6	16.4	19.6	10.4
$K=4$	3.0	2.5	17.3	21.5	11.1
$K=5$	2.8	2.1	16.4	21.2	10.6
$K=6$	3.1	2.7	14.5	16.8	9.3
GC	1.9	1.8	16.5	7.7	7.0
BP+occl	1.0	0.2	6.5	4.8	3.1

Table 2. A comparison of models with different numbers of parameters K trained on our ground-truth data but evaluated on the Middlebury data set. The last two rows are the performance of the graph cut implementation of [1] and the symmetric BP method with occlusion model by Sun *et al.* [10].

indicates that the learned parameters of those models are tuned more finely to the characteristics of the training data and generalize less well to datasets that are quite different. We include the BP method, which is currently ranked third, to show the potential of explicit occlusion models. We plan to adopt our learning approach to such models next.

6. Discussion and conclusion

Our work makes a number of contributions. We provide a large database of ground-truth stereo datasets that, for the first time, enables supervised learning methods in stereo. We also develop a novel conditional random field (CRF) model for stereo, and present an approximate but efficient gradient-based learning procedure. This procedure leverages the effectiveness of graph-cut-based energy minimization to solve a *most-probable-explanation* (MPE) problem. The specific model we experimentally investigate in this paper is a gradient-modulated Potts model with a varying number of gradient bins.

Our experiments show that models with more parameters can better capture the relationship between image gradients and disparity jumps, usually resulting in reduced disparity errors. On the other hand, our simple scheme using fixed gradient bins becomes more sensitive to brightness and contrast changes as the number of bins increases. This may be one of the reasons that the more complex learned models generalize less well to other datasets. Previous applications of CRFs in text processing have included a Gaussian prior on parameters to mitigate model overfitting [31]. Such techniques may be worth exploration for stereo. However, we believe more promising extensions to the approach here include learning across larger datasets and more robust gradient binning schemes.

Based on our results, we feel that our proposed framework has great potential. We believe the most promising avenues for future work include: (1) including an occlu-

that performance increases with the number of parameters. In fact the model with $K=2$ does best in the end. But the results in the bottom plot (where we train the parameters on the test data itself) show that this is not necessarily a problem of insufficient generalization, but rather that learning the best parameters (which amounts to matching the smoothness properties of the ground truth) might not always yield to lower matching errors. On the other hand, this could also be due to noisy gradient approximations as mentioned earlier.

5.3. Performance on standard benchmarks

Finally, we examine how well the parameters learned from our six datasets generalize to other stereo images. Table 2 shows the disparity errors on the Middlebury benchmark consisting of the Tsukuba, Venus, Teddy, and Cones images. We compare these errors with those of the graph cuts (GC) method [1], which uses a hand-tuned MRF model with two gradient bins, and the state-of-the-art method by Sun et al. [10], which uses BP and an explicit occlusion model. Our average results for $K=1$ and $K=2$ are slightly better than those of GC, and would result in a similar ranking as the GC method in the Middlebury evaluation. The fact that the errors for the more complex models are higher

sion model, (2) learning more general forms of the pairwise V potentials, and (3) improving the approximate gradient computations.

Acknowledgments

References

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

[2] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proc. CVPR*, volume I, pages 195–202, 2003.

[3] L. Zhang and S. Seitz. Parameter estimation for MRF stereo. In *Proc. CVPR*, volume II, pages 288–295, 2005.

[4] D. Kong and H. Tao. A method for learning matching errors in stereo computation. In *Proc. BMVC*, 2004.

[5] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother. Probabilistic fusion of stereo with color and contrast for bilayer segmentation. *IEEE TPAMI*, 28(9):1480–1492, 2006.

[6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE TPAMI*, 23(11):1222–1239, 2001.

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

[8] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? In *Proc. ECCV*, volume III, pages 65–81, 2002.

[9] J. Sun, N. Zheng, and H. Shum. Stereo matching using belief propagation. *IEEE TPAMI*, 25(7):787–800, 2003.

[10] J. Sun, Y. Li, S. Kang, and H.-Y. Shum. Symmetric stereo matching for occlusion handling. In *Proc. CVPR*, volume II, pages 399–406, 2005.

[11] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70(1):41–54, 2006.

[12] M. Tappen and W. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *Proc. ICCV*, pages 900–907, 2003.

[13] 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. In *Proc. ECCV*, volume 2, pages 19–26, 2006.

[14] H. Tao, H. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *Proc. ICCV*, volume I, pages 532–539, 2001.

[15] Y. Zhang and C. Kambhamettu. Stereo matching with segmentation-based cooperation. In *Proc. ECCV*, volume II, pages 556–571, 2002.

[16] M. Bleyer and M. Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. In *Proc. ICIP*, 2004.

[17] L. Hong and G. Chen. Segment-based stereo matching using graph cuts. In *Proc. CVPR*, volume I, pages 74–81, 2004.

[18] Y. Wei and L. Quan. Region-based progressive stereo matching. In *Proc. CVPR*, volume I, pages 106–113, 2004.

[19] L. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM ToG*, 23(3):600–608, 2004.

[20] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. ECCV*, volume III, pages 82–96, 2002.

[21] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. ICML*, pages 282–289, 2001.

[22] C. Sutton and A. McCallum. An introduction to conditional random fields for relational learning. In L. Getoor and B. Taskar, editors, *Introduction to Statistical Relational Learning*. MIT Press, 2006.

[23] C. Sutton, K. Rohanimanesh, and A. McCallum. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. In *Proc. ICML*, 2004.

[24] M. Wainwright, T. Jaakkola, and A. Willsky. Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Trans. Info. Theory*, 45(9):1120–1146, 2003.

[25] S. Kumar and M. Hebert. Man-made structure detection in natural images using a causal multiscale random field. In *Proc. CVPR*, volume I, pages 119–126, 2003.

[26] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In *Proc. ECCV*, pages 428–441, 2004.

[27] R. Cowell, A. Dawid, S. Lauritzen, and D. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Springer, 2003.

[28] D. Rivait and M. Langer. Spatiotemporal power spectra of motion parallax: the case of cluttered 3D scenes. In *IS&T/SPIE Symp. on El. Imaging*, 2007. To appear.

[29] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE TPAMI*, 20(4):401–406, 1998.

[30] F. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Info. Theory*, 47(2):498–519, 2001.

[31] F. Sha and F. Pereira. Shallow parsing with conditional random fields. In *Proc. NAACL*, pages 134–141, 2003.

[32] M. Collins. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proc. EMNLP*, 2002.