Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation

Abstract

A wide range of low level vision problems have been formulated in terms of finding the most probable assignment of a Markov Random Field (or equivalently the lowest energy configuration). Perhaps the most successful example is in the case of stereo vision. For the stereo problem, it has been shown that finding the global optimum is NP hard but good results have been obtained using a number of approximate optimization algorithms.

In this paper we show that for standard benchmark stereo pairs, the *global* optimum can be found in a few minutes using a variant of the belief propagation (BP) algorithm. We extend previous theoretical results on reweighted belief propagation to account for possible ties in the beliefs and using these results we obtain easily checkable conditions that guarantee that the BP disparities are the global optima. We verify experimentally that these conditions are met for the standard benchmark stereo pairs and discuss the implications of our results for further progress in stereo.

1 Introduction

Considerable progress in stereo vision has been achieved by formulating the problem in terms of energy minimization [3, 10, 9, 13, 6]. To illustrate the power of energy based methods, Figure 1(a) shows the "tsukuba" image, a benchmark image for stereo vision. Figure 1(b) shows the output of a standard SSD based algorithm followed by a "min" filter [11]. As can be seen, the use of a single window size is problematic — windows that are too small may not have enough structure in them to resolve the correspondence and windows that are too large cause noticeable artifacts at the occlusion boundaries and completely miss thin structures.

The energy based methods, in contrast, do not use a window of analysis. Rather, an energy function is defined which has a local term that meaures the goodness of a correspondence at a single pixel and a pairwise term that penalizes differences in disparity between neighboring pixels. The disparity is then found by running an algorithm that attempts to minimize this energy function. Perhaps the two most successful energy minimizers for this problem are Graph Cuts [3, 9] and belief propagation [6, 10, 13]. Figure 1(c) shows the output of the Graph Cuts algorithm as implemented in [10] (visually similar results are also obtained using belief propagation). Unlike the normalized correlation output, the energy minimization approach is able to provide sharp boundaries and preserves thin structures.

Despite the success of these methods, their output is still not perfect. For example, in Figure 1(c), the video camera is chopped in half. One can think of two different approaches for improving the output: (1) changing the energy function and (2) finding a different minimizer for the same energy function. Deciding between these two approaches is currently difficult because both belief propagation and Graph Cuts are only guaranteed to find *local* minima of the energy function. We do not know if a better optimzer would find a better solution.

Obviously, if we had a method that is capable of finding the *global* optimum of the energy function we would have a better idea of how to proceed. Unfortunately, it has been shown that for energy functions typically used in stereo, finding the global optimum is NP complete [3]. While this makes it extremely unlikely that we will be able to find the global optimum for *all* images in polynomial time, it leaves open the option for finding the global optimum for *some* images. In this paper, we show that a modification of belief propagation provides such an algorithm. In particular, we show that for the standard stereo benchmark images, the global minimum can be found in a few minutes per image.

1.1 Linear Programming Relaxation and Reweighted Belief Propagation

The algorithm we will use for finding the global optimum is called "Tree Reweighted Belief Propagation" (TRBP) [15, 17, 16, 7, 8]. The algorithm is closely related to *Linear Programming Relaxations*, a standard approach in computer science for approximating combinatorial problems [1]. In this section, we give a very brief introduction to this connection and refer the reader to [15, 7] for more details.

Denote by x the disparity image, $E(x_i, x_j)$, the pairwise compatability cost and $E(x_i)$ is local data cost. Then the global minimum of the energy, x^* is:

(a) **(b)** **(c)**

Figure 1: **(a)** A single frame from the "tsukuba" benchmark image. **(b)** The output of a standard SSD based algorithm, and **(c)** the output of an energy based method. The energy based method performs noticeably better but still misses some structure. Is the problem due to poor optimization or a bad energy function?

$$x^* = \arg\min_x \sum_{i,j} E_{ij}(x_i, x_j) + \sum_i E_i(x_i) \tag{1}$$

By introducing indicator variables, $q_i(x_i), q_{ij}(x_i, x_j)$ we can reformulate this as an integer program:

$$\text{minimize: } J(\{q\}) = \tag{2}$$

$$= \sum_{i,j} \sum_{x_i,x_j} q_{ij}(x_i, x_j) E_{ij}(x_i, x_j) + \sum_i \sum_{x_i} q_i(x_i) E_i(x_i)$$

subject to:

$$q_{ij}(x_i, x_j) \in \{0, 1\} \tag{3}$$

$$\sum_{x_i,x_j} q_{ij}(x_i, x_j) = 1 \tag{4}$$

$$\sum_{x_i} q_{ij}(x_i, x_j) = q_j(x_j) \tag{5}$$

The *Linear Programming (LP) Relaxation* involves replacing the hard constraint $q_{ij}(x_i, x_j) \in \{0, 1\}$ with the relaxed constraint $q_{ij}(x_i, x_j) \in [0, 1]$. This relaxed problem can now be solved in polynomial time using an LP solver, and if all the relaxed variables $q_{ij}(x_i, x_j)$ in the LP solution happen to be nonfractional (i.e. they satisfy the hard constraint $q_{ij}(x_i, x_j) \in \{0, 1\}$) then we have found the global optimum of the energy.

Note that there is usually no way to know in advance whether the LP solution will be fractional or not — we just have to run it on a particular problem and check. Unfortunately, as [4, 7] have observed, running standard LP solvers on the stereo problem is not practical with today's computing hardware. This is simply due to the large number of variables and constraints. In most stereo problems, the disparities are discretized to 30 possible states, so that the pairwise indicator $q_{ij}(x_i, x_j)$ is 30×30 matrix and for a

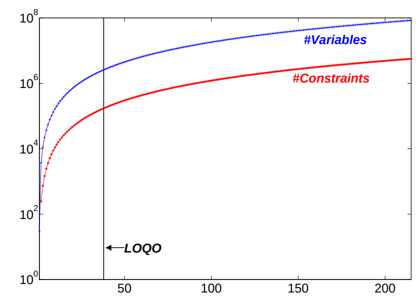

Figure 2: The number of variables and constraints in the LP relaxation as a function of the size of the image. With current hardware, the largest problem that can be solved with an interior-point method is about 40×40

.

200×200 image there are $2 \times 200 \times 200$ of these matrices. Figure 2 shows the number of variables and constraints in the LP problem as a function of the size of the image. We have found that the largest image we can solve with a standard, interior-point LP solver [14] is 39×39 and we estimated that we would need about 3000 Gigabytes of memory to run such a solver on a full size image.

Wainwright and colleagues suggested a different way of solving LP relaxations, which is related to the Lagrangian dual of the LP problem. Their algorithm, TRBP, is a variant of BP that differs slightly in the message update equations. Define pairwise potentials $\Psi_{ij}(x_i, x_j) = \exp(-E_{ij}(x_i, x_j))$ and singleton potentials $\Psi_i(x_i) = \exp(-E_i(x_i))$, the TRBP algorithm iterates the following equation:

$$m_{ij}(x_j) \leftarrow \alpha \max_{x_i} \Psi_{ij}^{1/\rho_{ij}}(x_i, x_j) \Psi_i(x_i) \frac{\prod_{k \in N_i \setminus j} m_{ki}^{\rho_{ki}}(x_i)}{m_{ji}^{1-\rho_{ji}}(x_i)} \tag{6}$$

where α is a normalization constant. After one has found a fixed point of these message update equations, the pairwise and singleton beliefs are defined as:

$$b_i(x_i) = \alpha \Psi_i(x_i) \prod_{j \in N_i} m_{ji}^{\rho_{ji}}(x_i) \quad (7)$$

$$b_{ij}(x_i, x_j) = \alpha \Psi_i(x_i) \Psi_j(x_j) \Psi_{ij}^{1/\rho_{ij}}(x_i, x_j) \cdot \quad (8)$$
$$\cdot \quad \frac{\prod_{k \in N_i \setminus j} m_{ki}^{\rho_{ki}}(x_i)}{m_{ji}^{1-\rho_{ji}}(x_i)} \frac{\prod_{k \in N_j \setminus i} m_{kj}^{\rho_{kj}}(x_j)}{m_{ij}^{1-\rho_{ij}}(x_j)}$$

The edge weights ρ_{ij} depend on the graph topology and for a grid graph they need to be strictly less than one. Specifically, we used $\rho_{ij} = 0.5$ for all edges. Note that for $\rho_{ij} = 1$, TRBP reduces to simple BP. Note also that the memory requirements of TRBP are similar to those of BP and are far less than second order LP solvers.

Wainwright et al. showed that under certain conditions it is possible to transform the pairwise beliefs and singleton beliefs $b_{ij}(x_i, x_j), b_i(x_i)$ into indicator variables $q_{ij}(x_i, x_j), q_i(x_i)$ so that these indicator variables will be a solution to the LP problem.

We restate here two results given by Wainwright et al. [15] that we will need for our results.

Max-Marginalization Lemma: A necessary and sufficient condition for messages to be fixed point of TRBP is that the beliefs defined from them using equation 7,8 satisfy max-marginalization:

$$\max_{x_j} b_{ij}(x_i, x_j) = b_i(x_i) \quad (9)$$

This lemma can be proven directly by showing that equations 7,8, and equation 9 are equivalent algebraically to equations 7,8 and equation 6 in fixed-point.

TRBP=MAP Theorem: Let $b_i(x_i)$ be beliefs calculated from fixed-point messages of TRBP. If there are no ties in these beliefs — for every i the maximum of $b_i(x_i)$ is attained at a unique value x_i^* — then x^* is the global minimum of the energy function.

The proof is given in [15, 16].

Kolmogorov [7] applied TRBP to the stereo problem and found that the conditions of the TRBP=MAP theorem *do not* hold for the standard benchmark images. That is, he found that many of the nodes had "ties" in them. He suggested a heuristic for breaking these ties and obtained lower energy than Graph Cuts using this heuristic, but emphasized that these disparities *are not* necessarily the global minimum of the energy function.

2 Theory

We now derive a novel result that enables us to obtain MAP solutions even when the TRBP beliefs have ties.

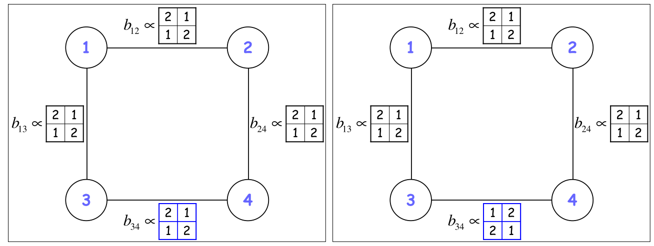

Figure 3: An illustration of our TRBP with ties theorem. For the four nodes on the left, it is possible to find an assignment x^* that maximizes the pairwise and singleton beliefs. Our theorem proves that this means that x^* is the global optimum. For the four nodes on the right it is impossible to find such an assignment.

TRBP with ties Theorem: Let $b_i(x_i)$ be beliefs calculated from fixed-point messages of TRBP (possibly with ties). If there exists an assignment x^* such that for every connected pair of pixels ij, $b_{ij}(x_i^*, x_j^*)$ maximizes the pairwise belief $b_{ij}(x_i^*, x_j^*) = \max b_{ij}(x_i, x_j)$ and for every pixel i, $b_i(x_i^*)$ maximizes the singleton belief, then x^* is the global minimum of the energy function.

Figure 3 illustrates the theorem. In figure 3a, there are four pixels for which ties exist, and it is easy to see that choosing $x^* = 0$ will maximize the local singleton and pairwise beliefs. On the other hand, in figure 3b it is easy to show that there exists no assignment x^* which will maximize all the pairwise beliefs. This is because the pairwise beliefs create a *frustrated cycle*. To satisfy the first three beliefs requires $x_1 = x_2 = x_3 = x_4$ while satisfying the fourth pairwise beliefs requires $x_4 \neq x_3$.

Proof: We define a new "regularized" energy function:

$$\tilde{E}(x; \lambda) = E(x) + \lambda \left[\sum_i (1 - \delta(x_i - x_i^*)) \right] \quad (10)$$

We will show that x^* minimizes $\tilde{E}(x; \lambda)$ for arbitrarily small λ and this implies (since $\sum_i [1 - \delta(x_i - x_i^*)]$ is bounded) that x^* is the global minimum of $E(x)$.

We will use the TRBP=MAP theorem to prove that x^* is the global minimum of $\tilde{E}$. Specifically, we will show that the same fixed-point mesages that were obtained by running TRBP to minimize E will also be fixed-point messages if we were to run TRBP to minimize the regularized energy $\tilde{E}$. To show this, consider the potentials $\tilde{\Psi}_{ij}(x_i, x_j), \tilde{\Psi}_i(x_i)$ that would be defined if we were to run TRBP on the regularized problem. The pairwise potentials will not change $\tilde{\Psi}_{ij}(x_i, x_j) = \Psi_{ij}(x_i, x_j)$ and the singleton potentials would be $\tilde{\Psi}_i(x_i) = \Psi_i(x_i) e^{-\lambda(1-\delta(x_i - x_i^*))}$

Now, we take the fixed point TRBP messages and define the pairwise and singleton beliefs $\tilde{b}_{ij}(x_i, x_j), \tilde{b}_i(x_i)$ using equation 7,8. This implies that:

$$\begin{aligned}\tilde{b}_{ij}(x_i, x_j) &= b_{ij}(x_i, x_j)e^{-\lambda(1-\delta(x_i-x_i^*))}e^{-\lambda(1-\delta(x_j-x_j^*))} \\ \tilde{b}_i(x_i) &= b_i(x_i)e^{-\lambda(1-\delta(x_i-x_i^*))} \end{aligned} \tag{11}$$

Note that *for any* λ, $\tilde{b}_{ij}(x_i, x_j)$ is *uniquely* maximized by x_i^*, x_j^*, and $\tilde{b}_i(x_i)$ is *uniquely* maximized by x_i^*. This is because any other assignment that maximized the non-regularized beliefs $b_{ij}(x_i, x_j)$ now has strictly smaller beliefs after being multiplied by $e^{-\lambda(1-\delta(x_i-x_i^*))}$,$e^{-\lambda(1-\delta(x_j-x_j^*))}$. Since $b_{ij}(x_i, x_j)$ satisfy max-marginalization, so does $\tilde{b}_{ij}(x_i, x_j)$. Hence by the max-marginal lemma the messages are also fixed-points of TRBP for the regularized problem. Furthermore, for any value of λ the TRBP beliefs for the regularized problem have a unique maximizing value $x*$ so by the TRBP=MAP theorem, x^* is the global minimum of the regularized problem for any value of λ. This completes the proof.

A naive way of searching for x^* would be to perform an exhaustive enumeration over all pixels in which there are ties in the beliefs. This is of course exponential in the number of pixels with ties. Fortunately, one can do this much more efficiently by taking advantage of the structure of the graph. In fact, as the following observation proves, when the set of pixels in which there are ties forms a singly connected set, there is no need to perform any search.

Observation (1): Let X_T denote the set of pixels for which there are ties in the TRBP beliefs and X_{NT} the set of pixels for which the TRBP beliefs are uniquely maximized by x_{NT}^*. If the graph of pixels X_T contains no cycles, then there exists an extension of x_{NT}^* to x^* that satisfies the conditions of the TRBP with ties theorem and hence x^* is the global minimum of the energy.

Proof: The proof is based on a single pass construction for the extension. For simplicity of exposition we describe the construction in the case where the set of pixels with ties X_T is a chain, and we label these nodes $x_1, x_2, x_3 \cdots x_n$. We choose an assignment for x_1^* as one of the maximizing values of $b_1(x_1)$. We then choose an assignment for x_2 as $x_2^* = \arg\max_{x_2} b_{12}(x_1^*, x_2)$ (when $\arg\max_{x_2} b_{12}(x_1^*, x_2)$ is not unique, arbitrarily set x_2^* to one of these maximizing states). The fact that the beliefs satisfy max-marginalization (eq. 9) guarantees that x_2^* also maximizes $b_2(x_2)$ and that $b_{12}(x_1^*, x_2^*)$ maximizes $b_{12}(x_1, x_2)$. We now continue and choose $x_3^* = \arg\max_{x_3} b_{23}(x_2^*, x_3)$ and continue in this fashion until we have extended x^*. By construction, for any pair of nodes ij in the chain x_i^*, x_j^* maximize the pairwise beliefs and the singleton beliefs. Now consider a pair of nodes ij for which i is in the chain and j is not. Due to the max-marginalization property, x_i^*, x_j^* must also maximize b_{ij}.

Observation (2): Let X_T denote the set of pixels for which there are ties in the TRBP beliefs and X_{NT} the set of pixels for which the TRBP beliefs are uniquely maximized by x_{NT}^*. Define a new cost function $C(x_T) = \sum_{i,j \in T} C_{ij}(x_i, x_j)$, with $C(i,j) = 0$ if x_i, x_j maximize b_{ij} and ϵ otherwise. If $\min_{x_T} C(x_T) = 0$, then there exists an extension of x_{NT}^* to x^* that satisfies the conditions of the TRBP with ties theorem and hence x^* is the global minimum of the energy.

Note that $C(x_T)$ is a cost involving only pairwise costs among the tied pixels and we can minimize it by finding the MAP assignment in an undirected graphical model. The complexity of minimizing $C(x_T)$ will typically be far less than a full exponential enumeration over all tied pixels and depends on the clique size in the junction tree [5].

3 Experiments

We ran TRBP on the images in the Middlebury stereo benchmark set [12]. We used the same energy function used by Tappen and Freeman [10]. The local cost is based on the Birchfield-Tomasi matching cost [2] and the pairwise energy penalizes for neighboring pixels having different disparities. The amount of penalty depends on the intensity difference between the two pixels — the smaller the intensity difference the larger the penalty. We used the BP code provided by Marshall Tappen and modified the code to perform TRBP rather than BP. We ran TRBP until the beliefs numerically converged, but no more than 5000 iterations.

Figure 4 shows the results. At each pixel we show the disparity that had the highest TRBP belief. Red pixels are those for which the TRBP belief had a tie. In all cases the ties occur at depth edges and the beliefs are undecided between the disparities at the two sides of the edge. As observed by [7], for each image there are some red pixels and hence the TRBP=MAP therorem does not hold. Note that for the “tsukuba” and “map” images, the set of red pixels does not contain any cycles so that the TRBP with ties theorem guarantees that one can easily find assignments for the red pixels and obtain the global minimum. For the “venus” synthetic image, we found that four of the red pixels actually form a “frustrated” cycle. We solved this by conditioning on one of the red pixels. We ran TRBP a few times, either with clamping the disparity at that pixel to a specific value or constraining the disparity at that pixel to be different from a specific value. In all constrained runs, the conditions of the TRBP with ties theorem were satisfied so that we obtained the global minimum of all constrained problems. Finally, the global minimum of the original problem was chosen to be the minimum of the constrained problems.

Figure 5 compares the output of Graph Cuts, BP and the global optimum for the three images. The outputs of Graph Cuts and BP are replotted from [10]. The global minimum produces smoother solutions but does not solve many of the problems present in the approximate solutions. For

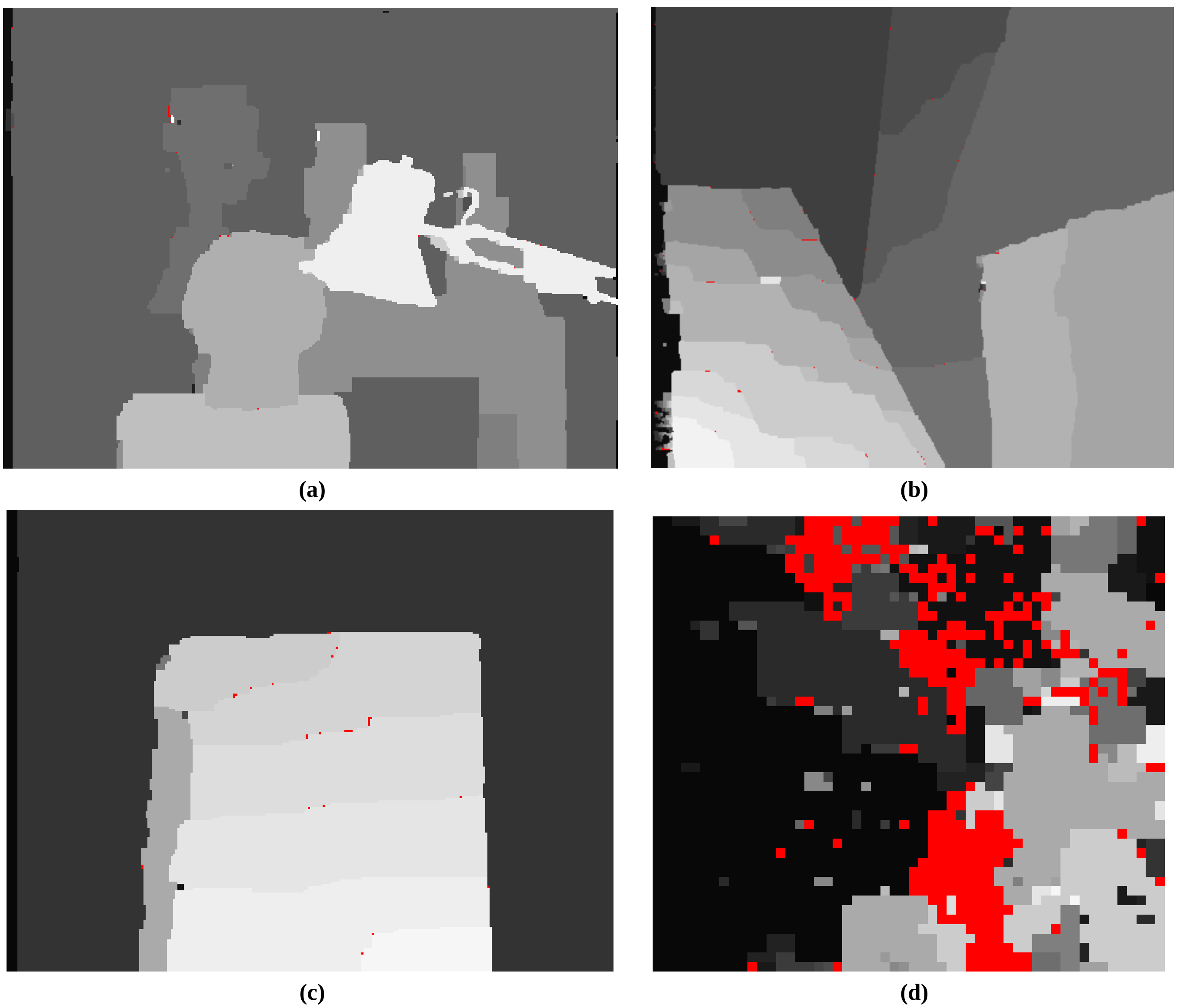

Figure 4: The global optima for three images from the Middlebury stereo benchmark set: **(a)** tsukuba, **(b)** venus, **(c)** map. Red pixels are those for which ties exist in the TRBP beliefs. Our theoretical result show that despite these ties, the assignment of the non red pixels is optimal. **(d)** TRBP applied to white noise images. Note that approximately 15% of the pixels in the image are tied.

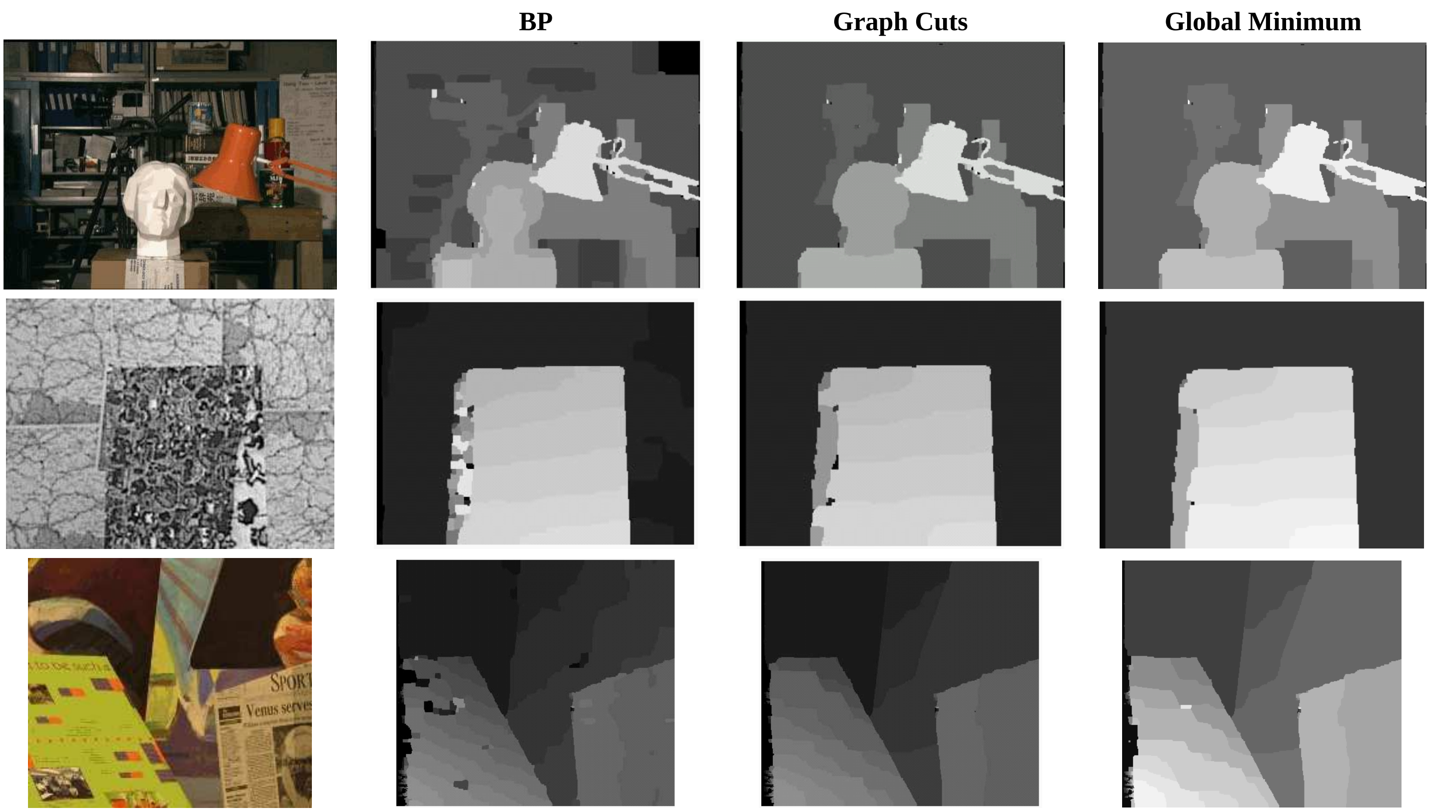

Figure 5: Comparison of the solutions obtained by BP, Graph Cuts and the global minimum using identical energy functions. The BP and Graph Cuts solutions are replotted from [10]. The global minimum produces smoother solutions but does not solve many of the problems present in the approximate solutions.

	BP	Graph Cuts	Global minimum
Tsukuba	775	663	644
Map	442	383	372
Venus	1501	1442	1399

Table 1: The energies for the three images.

MRF parameters			**Energy**		
T	s	P	*Graph Cuts*	*Belief Propagation*	*Global minimum*
0	50	1	366289	385652	365501
4	50	2	372693	405933	371993
4	50	4	375979	427580	375734
8	50	2	384342	442665	382160
8	50	4	399455	518615	399383

Table 2: The energies obtained by *Graph Cuts* and *BP* compared to the *global minimum energy* found using *TRBP* on the “map” sequence.

example, in the tsukuba image, the global minimum also “misses” half of the video camera and erroneously fills in the thin structure holding up the lamp.

Table 1 compares the global minimum of the energies for the three images to those found using Graph Cuts and belief propagation. It can be seen that the Graph Cuts solution is closer to the global minimum but always higher.

Table 2 compares the global minimum of the energies for the three images to those found using Graph Cuts and belief propagation. It can be seen that the Graph Cuts solution is closer to the global minimum but always higher.

4 Discussion

Finding the global minimum of the stereo energy function for any image is NP-complete. Nevertheless, in this paper we have shown that using tree reweighted belief propagation, it is possible to find the global minimum for several standard benchmark images. TRBP is similar to linear programming relaxations in that it is guaranteed to find the global optimum when the beliefs are non-fractional, but unlike general purpose LP solvers, it can be applied to these large problems in a matter of minutes. We have extended the theory of TRBP to allow for the case of ties in the beliefs and derived easily verified conditions for the TRBP beliefs with ties to give the global minimum. We have verified experimentally that these conditions indeed hold on the standard benchmark images.

MRF parameters			**Energy**		
T	s	P	*Graph Cuts*	*Belief Propagation*	*Global minimum*
0	50	1	602355	633092	600086
4	50	2	645865	758117	643946
4	50	4	696251	899215	
8	50	2	663845	775085	662543
8	50	4	739000	941129	737699

Table 3: The energies obtained by *Graph Cuts* and *BP* compared to the *global minimum energy* found using *TRBP* on the ”tsukuba” sequence.

As can be seen, the global minimum of the energy function does not solve many of the problems in the BP or Graph Cuts solutions. This suggests that the problem is not in the optimization algorithm but rather in the energy function. A promising problem for future research is to learn better energy functions from ground truth data.

References

[1] D. Bertismas and J. Ttsitskikilis. *Introduction to linear optimization*. Athena Scientific, 1997.

[2] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):401–406, 1998.

[3] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. In *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 1999.

[4] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.

[5] R. Cowell. Advanced inference in Bayesian networks. In M.I. Jordan, editor, *Learning in Graphical Models*. MIT Press, 1998.

[6] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. In *Proceedings of IEEE CVPR*, pages 261–268, 2004.

[7] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. 2005. submitted to PAMI, `research.microsoft.com/~vnk`.

[8] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. In *Proceedings AI Stats*, 2005.

[9] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.

[10] M.Tappen and W.T. Freeman. Graph cuts with belief propagation for stereo, using identical mrf parameters. In *ICCV*, 2003.

[11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Compter Vision*, 2002.

[12] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, volume 1, pages 195–202, 2003.

[13] J. Sun, H.-Y. Shum, and N.-N. Zheng. Stereo matching using belief propagation. In *ECCV*, volume 2, pages 510–524, 2002.

[14] R. Vanderbei. LOQO. `www.princeton.edu/~rvdb`.

[15] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. In *Allerton Conference on Communication, Control, and Computing*, 2002.

[16] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Exact MAP estimates by (hyper)tree agreement. In *Advances in Neural Processing Systems (NIPS)*. MIT Press, 2003.

[17] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Tree-reweighted belief propagation and approximate ml estimation by pseudo-moment matching. In *9th Workshop on Artificial Intelligence and Statistics*, 2003.