Markov Random Fields with Efficient Approximations

Yuri Boykov
yura@cs.cornell.edu

Olga Veksler
olga@cs.cornell.edu

Ramin Zabih
rdz@cs.cornell.edu

Computer Science Department
Cornell University
Ithaca, NY 14853

Abstract

Markov Random Fields (MRF's) can be used for a wide variety of vision problems. In this paper we address the estimation of first-order MRF's with a particular clique potential that resembles a well. We show that the maximum a posteriori *estimate of such an MRF can be obtained by solving a multiway cut problem on a graph. This allows the application of near linear-time algorithms for computing provably good approximations. We formulate the visual correspondence problem as an MRF in our framework, and show that this yields quite promising results on real data with ground truth.*

1 Introduction

Many early vision problems require estimating some spatially varying quantity (such as intensity, texture or disparity) from noisy measurements. These problems can be naturally formulated in a Bayesian framework using Markov Random Fields [4]. In this framework, the task is to find the maximum *a posteriori* (MAP) estimate of the underlying quantity. Bayes' rule states that the posterior probability $\Pr(f|O)$ of the hypothesis f given the observations O is proportional to the product of the likelihood $\Pr(O|f)$ and the prior probability $\Pr(f)$. The likelihood models the sensor noise, and the prior describes preferences among different hypotheses.

In this paper, we investigate MAP estimation of a special class of first-order Markov Random Fields. These MRF's, which we will call Well MRF's, have Gibbs clique potentials with a particular form that resembles a well. We begin by describing these Well MRF's, and giving an energy function that has a global minimum at the MAP estimate. In section 3 we show that the global minimum of this energy function can be obtained by finding a minimum multiway cut on a graph. Section 4 formulates the visual correspondence problem as a Well MRF and suggests a greedy method for finding a multiway cut.

We demonstrate the effectiveness of our approach for computing stereo depth in section 5. For example, we have bench-marked several algorithms using real images where the University of Tsukuba has produced dense ground truth. Our method produces an incorrect result at under 3% of the pixels, while correlation-based methods produce approximately 10% errors.

2 Markov Random Fields

Markov Random Fields were first introduced into vision by Geman and Geman [4], and have been widely used (see [6] for a particularly readable textbook). The MRF framework can express a wide variety of spatially varying priors, which accounts for much of its popularity. In early vision it is commonly assumed that the underlying quantity is smooth, either piecewise [4] or globally [5].

An MRF has several components: a set $\mathcal{P} = \{1, \ldots, m\}$ of sites p, which will be pixels; a neighborhood system $\mathcal{N} = \{ \mathcal{N}_p \mid p \in \mathcal{P} \}$ where each $\mathcal{N}_p$ is a subset of pixels in $\mathcal{P}$ describing the neighbors of p; and a field (or set) of random variables $F = \{ F_p \mid p \in \mathcal{P} \}$.

Each random variable F_p takes a value f_p in some set $\mathcal{L} = \{l_1, \ldots, l_k\}$ of the possible labels (for example, the possible intensities or disparities). Following [6] a joint event $\{F_1 = f_1, \ldots, F_m = f_m\}$ is abbreviated as $F = f$ where $f = \{ f_p \mid p \in \mathcal{P} \}$ is a *configuration* of F, corresponding to a realization of the field. For simplicity, we will write $\Pr(F = f)$ as $\Pr(f)$ and $\Pr(F_p = f_p)$ as $\Pr(f_p)$. In order to be an MRF, the random variables in the field F must satisfy

$$\Pr(f_p | f_{S-\{p\}}) = \Pr(f_p | f_{\mathcal{N}_p}), \quad \forall p \in \mathcal{P}.$$

This condition states that each random variable F_p depends on other random variables in F only through its neighbors in $F_{\mathcal{N}_p} = \{F_q | \, q \in \mathcal{N}_p\}$.

The key result concerning Markov Random Fields is the Hammersley-Clifford theorem. This states that the probability of a particular configuration $\Pr(f) \propto \exp(-\sum_C V_C(f))$, where the sum is over all cliques in the neighborhood system $\mathcal{N}$. Here, V_C is a *clique potential*, which describes the prior probability of a particular realization of the elements of the clique C.

We will restrict our attention to first-order MRF's where each $\mathcal{N}_p$ can contain only north, east, south and west neighbors of pixel p. The cliques of a first-order MRF are either single pixels or ordered pairs of neighboring pixels. Single pixel clique potentials provide a prior bias towards a particular label for a particular pixel. In the majority of vision applications it is reasonable to assume there are no single pixel clique potentials, leaving

$$\Pr(f) \propto \exp\left(- \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p} V_{(p,q)}(f_p, f_q) \right).$$

In general, the field F is not directly observable in the experiment. We have to estimate its realized configuration f based on the observation O, which is related to f by means of the likelihood function $\Pr(O|f)$. In the context of the image restoration problem the observation O is the joint event $\{I_p = i_p\}$ over all $p \in \mathcal{P}$ where I_p denotes the observable intensity at pixel p and i_p is its particular realization. If F_p denotes the true intensity at p then assuming i.i.d. noise

$$\Pr(O|f) = \prod_{p \in \mathcal{P}} g(i_p, f_p)$$

where $g(i_p, f_p) = \Pr(I_p = i_p | F_p = f_p)$ represents the sensor noise model.

We will make a slightly more general assumption. We will assume that the likelihood can be written as

$$\Pr(O|f) = \prod_{p \in \mathcal{P}} g(i, p, f_p), \tag{1}$$

where i is a configuration of some field I that can be directly observed and g is a sensor noise distribution $(0 \leq g \leq 1)$. An example of the likelihood function with this general structure can be found in section 4.

We wish to obtain the configuration $f \in \mathcal{L} \times \ldots \times \mathcal{L} = \mathcal{L}^m$ that maximizes the posterior probability $\Pr(f|O)$. Bayes' law tells us that $\Pr(f|O) \propto \Pr(O|f)\Pr(f)$. It follows that our MAP estimate f should minimize the posterior energy function

$$E(f) = \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{N}_p} V_{(p,q)}(f_p, f_q) - \sum_{p \in \mathcal{P}} \ln\left(g(i, p, f_p)\right).$$

2.1 Well MRF's

In this paper we consider first-order MRF's with a special form of clique potentials that resembles a well. If $\delta(\cdot)$ represents the unit impulse function, then $u\,(1 - \delta(\cdot))$ is a well with "depth" u. A first-order MRF is called a *Well MRF* if its single clique potentials are zeros and clique potential for any pair of neighboring pixels p and q is

$$V_{(p,q)}(f_p, f_q) = u_{\{p,q\}} \cdot (1 - \delta(f_p - f_q))$$

where the coefficient $u_{\{p,q\}} \geq 0$ specifies the depth of the well. Note that $\{p, q\}$ is a set, not a tuple, so $V_{(p,q)}(f_p, f_q) = V_{(q,p)}(f_q, f_p)$. Well MRF's are thus isotropic (i.e., independent of orientation).

The prior probability of a Well MRF is thus

$$\Pr(f) \propto \exp\left(-\sum_{\{p,q\}\in\mathcal{E}_{\mathcal{N}}} 2u_{\{p,q\}}(1-\delta(f_p - f_q))\right)$$

where $\mathcal{E}_{\mathcal{N}}$ is the set of distinct $\{p, q\}$ such that $q \in \mathcal{N}_p$. Each term in the summation above equals $2u_{\{p,q\}}$ if p and q have different labels ($f_p \neq f_q$) and zero otherwise. The coefficient $u_{\{p,q\}}$ can be interpreted as a cost of a "discontinuity" between p and q, that is, the penalty for assigning different labels to neighboring pixels p and q. The sum in the exponent above is proportional to the total cost of discontinuities in f. Thus, the prior probability $\Pr(f)$ is larger for configurations f with fewer discontinuities.

The posterior energy function of a Well MRF is

$$\begin{aligned} E(f) \;&=\; \sum_{\{p,q\}\in\mathcal{E}_{\mathcal{N}}} 2u_{\{p,q\}}(1-\delta(f_p - f_q)) \\ &\quad - \sum_{p\in\mathcal{P}} \ln\left(g(i, p, f_p)\right). \end{aligned} \tag{2}$$

The MAP estimate f minimizes $E(f)$. Thus, it should both agree with the observed data and have a small number of discontinuities.

Note that the clique potential of a Well MRF resembles a robust estimator, in that it has a fixed maximum value (in the language of robust statistics, it is re-descending). Most vision applications of MRF's follow [4] by introducing a line process that explicitly models discontinuities. [2] showed that if spatial restrictions on discontinuities are ignored, the line process can be eliminated by using a robust penalty function.[1] We take a somewhat similar approach, by using a re-descending clique potential instead of a line process.

3 Optimizing the energy function

In this section we show that minimizing the energy function $E(f)$ in (2) over $f \in \mathcal{L}^m$ is equivalent to solving a multiway cut problem on a certain graph. In section 3.1 we give another formulation of the posterior energy minimization problem that is equivalent to (2). This formulation, shown in equation (3), reduces the search space for f and simplifies our transition to the graph problem. Then in section 3.2 we construct a particular graph, and prove that solving the multiway cut problem on this graph is equivalent to minimizing the energy function of equation (3).

3.1 Reformulating the energy function

We want to find $f^* \in \mathcal{L}^m$ that minimizes $E(f)$ in (2). It is straightforward to reduce the search space for f^*. Assuming $E(f^*)$ is finite, we can always find some constant $K(p)$ for each pixel p satisfying

$$-\ln(g(i, p, f_p^*)) < K(p).$$

For example, if no better argument is available we can always take $K(p) = K = E(f)$ where f is any fixed configuration of F such that $E(f)$ is finite.

For a given collection of constants $K(p)$ we define

$$\mathcal{L}_p \;=\; \{l \in \mathcal{L} : \; -\ln(g(i, p, l)) < K(p)\}$$

for each pixel p in $\mathcal{P}$. Each $\mathcal{L}_p$ prunes out a set of labels which cannot be assigned to p in the optimal solution. For example, if we take $K(p) = E(f)$ as suggested above, then for $l \notin \mathcal{L}_p$ a single sensor noise term

[1] See [1] for further analysis of the relationship between MRF's and robust estimation.

$-\ln(g(i,p,l))$ in (2) will exceed the total value of the posterior energy function $E(f)$ at some configuration f. Since $f_p^* \in \mathcal{L}_p$ then each $\mathcal{L}_p$ is a nonempty set. Define also $\bar{\mathcal{L}} = \mathcal{L}_1 \times \ldots \times \mathcal{L}_m$. Since $f^* \in \bar{\mathcal{L}}$, our search can be restricted to the set $\bar{\mathcal{L}}$.

It is possible to rewrite $-\ln(g(i,p,f_p))$ as

$$\bar{K}(p) + \sum_{\substack{l \in \mathcal{L}_p \\ l \neq f_p}} (\ln(g(i,p,l)) + K(p))$$

where $\bar{K}(p)$ is some constant that does not depend on f_p. It follows that minimizing $E(f)$ in (2) is equivalent to minimizing

$$\begin{aligned} \bar{E}(f) \quad &= \sum_{\{p,q\} \in \mathcal{E}_{\mathcal{N}}} 2u_{\{p,q\}}(1 - \delta(f_p - f_q)) \\ &\quad + \sum_{p \in \mathcal{P}} \sum_{\substack{l \in \mathcal{L}_p \\ l \neq f_p}} h(i,p,l) \end{aligned} \tag{3}$$

where $h(i,p,l) = \ln(g(i,p,l)) + K(p)$ and the minimization takes place over $f \in \bar{\mathcal{L}}$. Note that $h(i,p,l) > 0$ for any $p \in \mathcal{P}$ and for any $l \in \mathcal{L}_p$.

3.2 Multiway cut formulation

Consider a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ with non-negative edge weights, along with a set of terminal vertices $\mathcal{L} \subset \mathcal{V}$. A subset of edges $\mathcal{C} \subset \mathcal{E}$ is called a *multiway cut* if the removal of $\mathcal{C}$ from $\mathcal{G}$ completely separates all terminal vertices from each other. The cost of the cut $\mathcal{C}$ is denoted by $|\mathcal{C}|$ and equals the sum of its edge weights. The *multiway cut problem* is to find $\mathcal{C}$ which has the minimum cost.

The multiway cut problem reduces to the standard max flow-min cut problem in the case of 2 terminals. With 3 or more terminals it is known to be NP-complete [3]. Fortunately, [3] also gives an almost linear time method for computing a provably good approximation. The approximation produced is optimal to within a factor of 2. It requires running a min cut algorithm $|\mathcal{L}|$ times. The worst-case complexity of computing a min cut is worse than linear, but in practice modern algorithms run in near-linear time.

We now show that the minimization problem in (3) is equivalent to a multiway cut problem. We begin by constructing $\mathcal{G}$. We take $\mathcal{V} = \mathcal{P} \cup \mathcal{L}$. This means that $\mathcal{G}$ contains two types of vertices: *p-vertices* (pixels) and *l-vertices* (labels). Note that l-vertices will serve as terminals for our multiway cut problem. Two p-vertices are connected by an edge if and only if the corresponding pixels are neighbors in $\mathcal{N}$. Therefore, the set $\mathcal{E}_{\mathcal{N}}$ corresponds to the set of edges between p-vertices. We will refer to elements of $\mathcal{E}_{\mathcal{N}}$ as *n-links*. Each n-link $\{p,q\} \in \mathcal{E}_{\mathcal{N}}$ is assigned a weight

$$w_{\{p,q\}} = 2u_{\{p,q\}}. \tag{4}$$

A p-vertex is connected by an edge to an l-vertex if and only if $l \in \mathcal{L}_p$. An edge $\{p,l\}$ that connects a p-vertex with a terminal (an l-vertex) will be called a *t-link* and the set of all such edges will be denoted by $\mathcal{E}_{\mathcal{T}}$. Each t-link $\{p,l\} \in \mathcal{E}_{\mathcal{T}}$ is assigned a weight

$$w_{\{p,l\}} = h(i,p,l) + \sum_{q \in \mathcal{N}_p} w_{\{p,q\}}. \tag{5}$$

Note that each p-vertex is connected to at least one terminal since $\mathcal{L}_p$ is non-empty. No edge connects terminals directly to each other. Therefore, $\mathcal{E} = \mathcal{E}_{\mathcal{N}} \cup \mathcal{E}_{\mathcal{T}}$. Figure 1 shows the general structure of the graph $\mathcal{G}$.

Since a multiway cut separates all terminals it can leave at most one t-link at each p-vertex. A multiway cut $\mathcal{C}$ is called *feasible* if each p-vertex is left with exactly one t-link. Each feasible multiway cut $\mathcal{C}$ corresponds to some configuration $f^{\mathcal{C}}$ in $\bar{\mathcal{L}}$ in an obvious manner: simply assign the label l to all pixels p which are t-linked to the l-vertex.

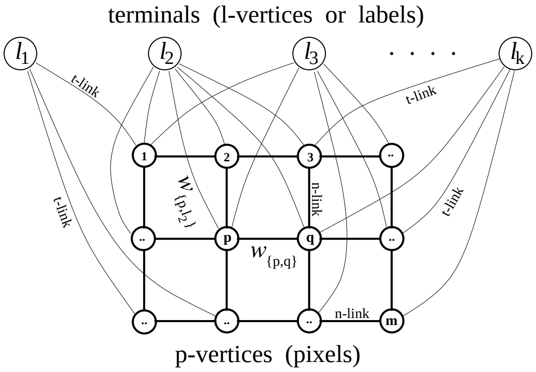


Figure 1: An example of the graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ where the terminals are $\mathcal{L} = \{l_1, \ldots, l_k\}$ and p-vertices are elements of $\mathcal{P} = \{1, \ldots, p, q, \ldots, m\}$. Each p-vertex is connected to at least one terminal.

Lemma 1 *A minimum cost multiway cut $\mathcal{C}$ on $\mathcal{G}$ for terminals $\mathcal{L}$ must be feasible.*

PROOF: Due to equation (5), each t-link $\{p, l\}$ has a weight larger then the sum of weights of all n-links adjacent to the p-vertex. If a multiway cut of minimum cost is not feasible then there exists some p-vertex with no t-link left. In such a case we will obtain a smaller cut by returning to the graph one t-link $\{p, l\}$ for an arbitrary $l \in \mathcal{L}_p$ and cutting all n-links adjacent to this p-vertex. ∎

Theorem 1 *If $\mathcal{C}$ is a minimum cost multiway cut on $\mathcal{G}$, then $f^{\mathcal{C}}$ minimizes $\bar{E}$.*

PROOF: Lemma 1 allows to concentrate on feasible multiway cuts only. Note that distinct feasible multiway cuts $\mathcal{C}1$ and $\mathcal{C}2$ can induce the same configuration $f^{\mathcal{C}1} = f^{\mathcal{C}2}$. However, among all distinct feasible cuts corresponding to the same configuration $f \in \bar{\mathcal{L}}$ we can always find an *irreducible* one that does not sever n-links between two p-vertices connected to the same terminal. It follows that there is a one to one correspondence between configurations f in $\bar{\mathcal{L}}$ and irreducible feasible multiway cuts on the $\mathcal{G}$.

Obviously, the minimum multiway cut should be both feasible and irreducible. To conclude the theorem it suffices to show that the cost of any irreducible feasible multiway cut $\mathcal{C}$ satisfies $|\mathcal{C}| = A + \bar{E}(f^{\mathcal{C}})$, where A is the same constant for all irreducible feasible multiway cuts. Since $\mathcal{C}$ is feasible, the sum of the weights for t-links in $\mathcal{C}$ is equal to

$$\sum_{p \in \mathcal{P}} \sum_{\substack{l \in \mathcal{L}_p \\ l \neq f_p^{\mathcal{C}}}} w_{\{p,l\}}.$$

Since $\mathcal{C}$ is irreducible, the sum of weights for the n-links in the cut is equal to

$$\sum_{\{p,q\} \in \mathcal{E}_{\mathcal{N}}} w_{\{p,q\}} (1 - \delta(f_p^{\mathcal{C}} - f_q^{\mathcal{C}})).$$

The theorem now follows from (4) and (5). ∎

4 Computing Visual Correspondence

We now describe how our framework can be applied to the visual correspondence problem, which is the basis of stereo and motion. Given two images of the same scene, a pixel in one image corresponds to a pixel in the other if both pixels are projections along lines of sight of the same physical scene element. The problem is to determine this correspondence between pixels of two images.

We begin by showing how to formulate the correspondence problem as a Well MRF, and thus as a multiway cut problem.[2] We arbitrarily select one of the images to be the primary image. Let $\mathcal{P}$ denote the set of pixels in the primary image and $\mathcal{S}$ denote a set of pixels of the second image. The quantity to be estimated is the *disparity* configuration $d = \{ d_p \mid p \in \mathcal{P} \}$ on the primary image where each d_p establishes the correspondence between the pixel p in the primary image and the pixel $s = p \oplus d_p$ in the second image.[3]

We assume that each d_p has a value in $\mathcal{L}$, which is a finite set of possible disparities. For simplicity, we consider configurations $d \in \mathcal{L}^m$. (This allows double-assignments, since distinct pixels p and q in $\mathcal{P}$ can correspond to the same pixel $p \oplus d_p = q \oplus d_q$.) The information available consists of the observed intensities of pixels in both images. Let $I_{\mathcal{P}} = \{ I_p \mid p \in \mathcal{P} \}$ and $I_{\mathcal{S}} = \{ I_s \mid s \in \mathcal{S} \}$ be the random fields of intensities in the primary and in the second images. Assume also that i_p denotes the observed value of intensity I_p.

4.1 Incorporating context

Note that the intensities of pixels in $\mathcal{P}$ contain information that can significantly bias our assessment of disparities without even considering the second image. For example, two neighboring pixels p and q in $\mathcal{P}$ are much more likely to have the same disparity if we know that $i_p \approx i_q$. Most methods for computing correspondence do not make use of this kind of contextual information. An exception is [7], which describes a method also based on MRF's. In their approach, intensity edges were used to bias the line process. They allow discontinuities to form without penalty on intensity edges. While our MRF's do not use a line process, we can easily incorporate contextual information into our framework.

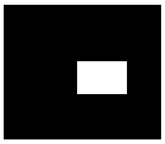
Formally, we assume that the conditional distribution $\Pr'(d) = \Pr(d \mid I_{\mathcal{P}})$ is a distribution of a Well MRF on $\mathcal{P}$ with neighborhood system $\mathcal{N}$. $\Pr'(d)$ can be viewed as a "prior" distribution of d before the information in the second image is disclosed. Conditioning on $I_{\mathcal{P}}$ allows to choose well clique potential "depths" $u_{\{p,q\}}$ according to

$$u_{\{p,q\}} = U(|i_p - i_q|), \qquad \forall \{p,q\} \in \mathcal{E}_{\mathcal{N}}. \tag{6}$$

Each $u_{\{p,q\}}$ represents a penalty for assigning different disparities to neighboring pixels p and q in $\mathcal{P}$. The value of the penalty $u_{\{p,q\}}$ should be smaller for pairs $\{p,q\}$ with larger intensity differences $|i_p - i_q|$. In practice we use an empirically selected decreasing function $U(\cdot)$. Note that instead of (6) we can set the coefficients $u_{\{p,q\}}$ according to an output of an edge detector on the primary image. For example, $u_{\{p,q\}}$ can be made small for pairs $\{p,q\}$ where an intensity edge was detected and large otherwise. Segmentation of the primary image can also be used.

The following example shows the importance of contextual information. Consider the pair of synthetic images below, with a uniformly white rectangle in front of a uniformly black background.

Primary image ($I_{\mathcal{P}}$)



Second Image ($I_{\mathcal{S}}$)

There is a one pixel horizontal shift in the location of the rectangle, and there is no noise. Without noise, the problem of estimating $d = \{ d_p \mid p \in \mathcal{P} \}$ is reduced to maximizing the prior $\Pr'(d)$ under the constraint that pixel p in $\mathcal{P}$ can be assigned a pixel $p \oplus d_p$ in $\mathcal{S}$ only if they have the same intensity.

If $u_{\{p,q\}}$ is the same for all pairs of neighbors $\{p,q\}$ in $\mathcal{P}$ then $\Pr'(d)$ is maximized at the disparity configuration shown either in the left or in the middle pictures below depending on the exact height of the rectangle.

[2] [8] recently gave a very different formulation of the multi-camera stereo problem as a maximum flow problem.

[3] To be precise, $p \oplus d_p$ stands for the pixel in $\mathcal{S}$ whose 2D coordinates are obtained by adding the disparity d_p to the 2D coordinates of p.

$u_{\{p,q\}} = const$ $\quad u_{\{p,q\}} = const$ $\quad u_{\{p,q\}} \neq const$

Suppose now that the penalty $u_{\{p,q\}}$ is much smaller if $i_p \neq i_q$ than it is if $i_p = i_q$. In this case the maximum of $\Pr'(d)$ is achieved at the disparity configuration shown in the right picture. This result is much closer to human perception.

4.2 Sensor noise

The sensor noise is the difference in intensities between corresponding pixels. We assume that the likelihood function is

$$\Pr'(I_{\mathcal{S}} \,|\, d) = \Pr(I_{\mathcal{S}} \,|\, d, I_{\mathcal{P}}) \propto \prod_{p \in \mathcal{P}} g(i_{p \oplus d_p} | i_p) \tag{7}$$

where d is the true disparity correspondence. Here, $g(i_s \,|\, i_p)$ is the conditional distribution of intensity at pixel s in the second image given the intensity at pixel p in the primary image if the two pixels are known to correspond. The function g is determined by the sensor noise model, and typically $g(i_s \,|\, i_p)$ is a symmetric distribution centered at i_p.

Obviously, $g(i_{p \oplus d_p} | i_p)$ can be rewritten as $g(i, p, d_p)$ and therefore the noise model in (7) is consistent with equation (1). Note that the main idea behind assumption (7) is that sensor noise is independent.

4.3 Implementation

Equations (6) and (7) describe how to use Well MRF's for visual correspondence. The prior distribution $\Pr'(d)$ of the disparity configuration d is determined by the clique potentials given in (6), and the likelihood function $\Pr'(O|\,d)$ consistent with (1) is determined by equation (7). Now the multiway cut approach explained in section 3 can be used to find the MAP estimate of d for any pair of stereo images.

Assuming that the reduction explained in section 3.1 has been made, we have a complete description of the graph $\mathcal{G}$ whose terminals $\mathcal{L}$ we wish to separate by a minimum multiway cut. While the multiway minimum cut problem is NP-complete, there exist provably good approximations with near linear running time [3], and this is an area of active research. We have developed a simple greedy method with almost linear running time.

Each multiway cut $\mathcal{C}$ can be uniquely represented by a collection of completely disjoint subgraphs $\mathcal{G}_{\mathcal{C}} = \{\, \mathcal{G}_l = \langle \mathcal{V}_l, \mathcal{E}_l \rangle \mid l \in \mathcal{L} \,\}$ such that $l \in \mathcal{V}_l$, $p \in \mathcal{V}_l$ implies $l \in \mathcal{L}_p$, and $\mathcal{E}_l$ consists of all edges in $\mathcal{G}$ that connect vertices in $\mathcal{V}_l$. As an initial solution we take a trivial collection $\mathcal{G}_{\mathcal{C}}$ where $\mathcal{G}_l = \langle \{l\}, \emptyset \rangle$. At each iteration we would like to obtain a new collection $\mathcal{G}_{\mathcal{C}}$ that corresponds to a cut with lower cost.

There are two steps at each iteration. At the first step we select some $l \in \mathcal{L}$ in a certain order and *expand* $\mathcal{G}_l$ by adding in $\mathcal{V}_l$ all vertices p in $\mathcal{G}$ such that $\mathcal{L}_p$ contains l and which are not contained in $\mathcal{G}_\lambda$ for $\lambda \neq l$.

At the second step of each iteration we run a standard min cut algorithm for terminal l against other terminals in $\mathcal{L}$. In some arbitrary order we select one $\lambda \neq l$. Then we reallocate pixels in $\mathcal{V}_l \cup \mathcal{V}_\lambda$ between the terminals l and λ trying to obtain a smaller cut. More specifically, we solve a standard two terminals min cut problem on a graph $\mathcal{G}_{\{l,\lambda\}}$ with vertices $\mathcal{V}_{\{l,\lambda\}} = \mathcal{V}_l \cup \mathcal{V}_\lambda$. The set of edges $\mathcal{E}_{\{l,\lambda\}}$ includes $\mathcal{E}_l \cup \mathcal{E}_\lambda$ and all other edges in $\mathcal{E}$ that connect vertices in $\mathcal{V}_{\{l,\lambda\}}$. The output is a new pair of subgraphs $\mathcal{G}_l$ and $\mathcal{G}_\lambda$ with a smaller cut.

At each iteration we obtain a smaller multiway cut. In the current implementation we iterate through all labels $l \in \mathcal{L}$ only once. It can be easily checked that the algorithm is quadratic in the number of labels and has the same almost linear time complexity in the number of nodes as a standard min cut algorithm.

5 Experimental results

In this section we give some experimental results on stereo data that use our greedy multiway cut algorithm. For simplicity, we have used a uniform noise model for g. We also used a two-valued function $U(|i_p - i_q|)$, which has a large value if i_p is close to i_q, and a small value otherwise. The parameter values used for the algorithms in the experiments in this section were determined by hand. We used the parameters that gave the results with the best overall appearance. Empirically, our method's performance does not appear to depend strongly upon the precise choices of parameters.

We have bench-marked several methods using a real image pair with dense ground truth. We obtained an image pair from the University of Tsukuba Multiview Image Database for which the ground truth disparity is known at every pixel. The image and the ground truth are shown in figure 2, along with the results from our method and an image showing the pixels where our answers are incorrect.

Having ground truth allows a statistical analysis of algorithm performance. The table below shows the number of correct answers that are obtained by various methods. There appear to be some discretization errors in the ground truth, so it is worth concentrating on errors larger than ± 1 disparity.

Method	Total errors	Errors $> \pm 1$
Well MRF	8.6	2.8
LOG-filtered L_1	19.9	9.0
Normalized correlation	24.7	10.0
MLMHV	24.5	11.0

We have also run our method on a number of standard benchmark images. The results are shown in figure 3. Various details in the images (such as the front parking meter in the meter image and the sign in the shrub image) are sharp and accurately localized.

6 Conclusions

We have described a class of MRF's whose MAP estimate can be efficiently approximated. These Well MRF's can be applied to a variety of problems in computer vision. We have demonstrated that a Well MRF formulation of the correspondence problem yields very promising experimental results.

Acknowledgments

We thank E. Tardos and D. Shmoys for providing important references and for insightful remarks on the content of the paper. We are also grateful to Dr. Y. Ohta and Dr. Y. Nakamura for supplying the ground truth imagery from the University of Tsukuba. This research has been supported by DARPA under contract DAAL01-97-K-0104, monitored by ARL.

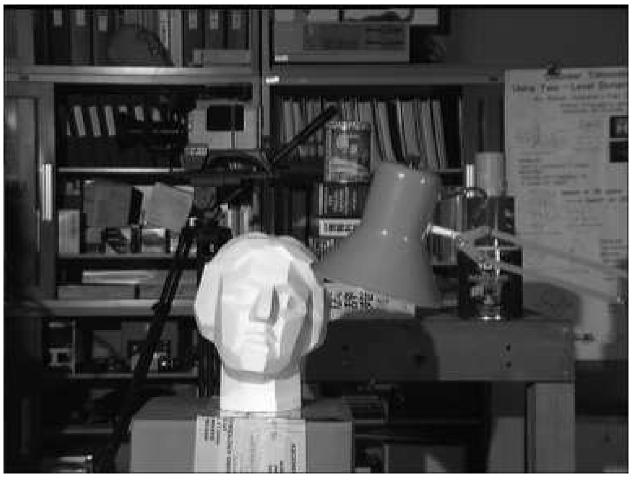
References

[1] Michael Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–92, July 1996.

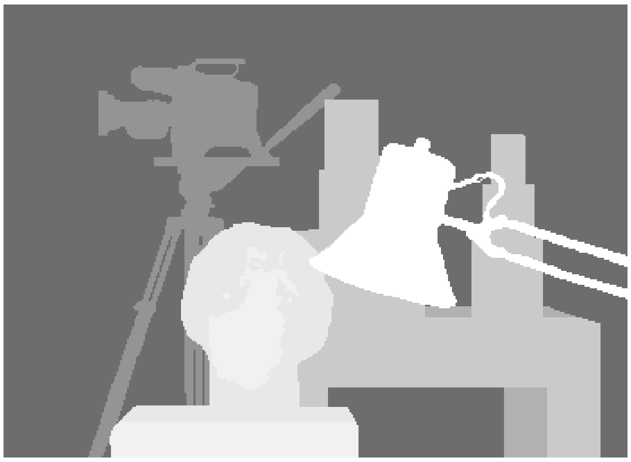
[2] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, 1987.

[3] E. Dahlhaus, D. S. Johnson, C.H. Papadimitriou, P. D. Seymour, and M. Yannakakis. The complexity of multiway cuts. In *24th Annual ACM Symposium on Theory of Computing*, pages 241–251, 1992.

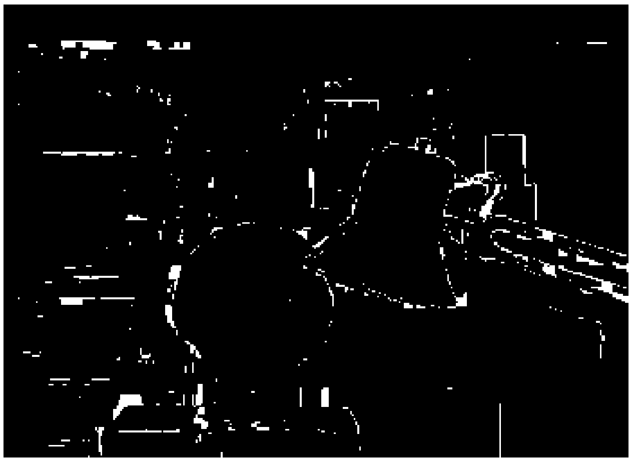
[4] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.



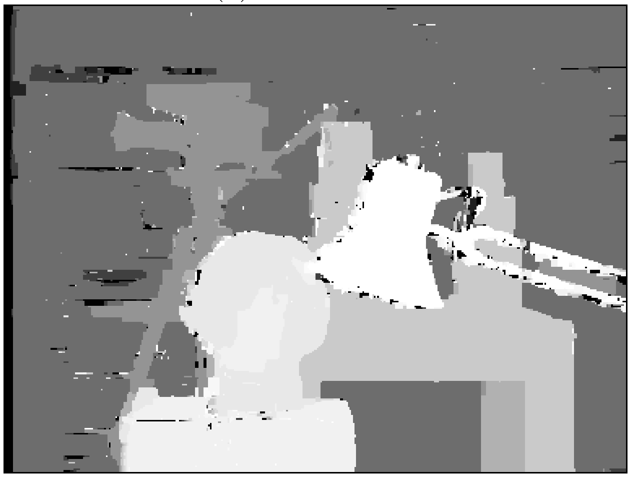
(a) Scene



(b) Ground truth



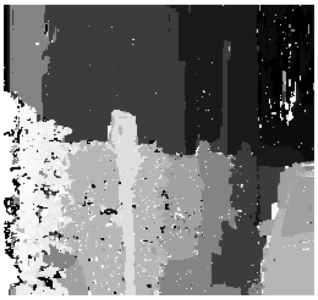
(c) Errors $> \pm 1$



(d) Well MRF results

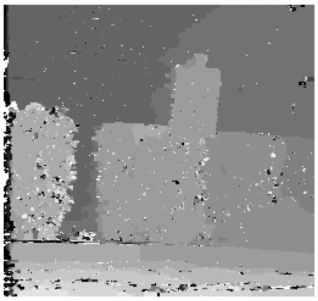
Figure 2: Ground truth results

Meter image



Well MRF results

Shrub image



Well MRF results

Figure 3: Results on other benchmark images

[5] Berthold Horn and Brian Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[6] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.

[7] T. Poggio, E. Gamble, and J. Little. Parallel integration of vision modules. *Science*, 242:436–440, October 1988. See also E. Gamble and T. Poggio, MIT AI Memo 970.

[8] Sébastien Roy and Ingemar Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *6th International Conference on Computer Vision*, 1998.