Markov Random Fields with Efficient Approximations

Yuri Boykov Olga Veksler Ramin Zabih
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 focus on MRF's with two-valued clique potentials, which form a generalized Potts model. We show that the maximum a posteriori *estimate of such an MRF can be obtained by solving a multiway minimum cut problem on a graph. We develop efficient algorithms for computing good approximations to the minimum multiway cut. The visual correspondence problem can be formulated as an MRF in our framework; this yields quite promising results on real data with ground truth. We also apply our techniques to MRF's with linear clique potentials.*

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 [6]. 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 focus on MAP estimation of a class of Markov Random Fields which generalizes the Potts model [11]. These MRF's have Gibbs clique potentials with a particular form that resembles a well. We begin by describing the generalized Potts model, 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, and give a greedy method for computing a multiway cut. Section 4 formulates the visual correspondence problem as a generalized Potts model.

We demonstrate the effectiveness of our approach for computing stereo depth in section 5. For example, we have benchmarked several algorithms using real images with dense ground truth. Our method produces an incorrect result at under 3% of the pixels, while correlation-based methods produce approximately 10% errors. In section 6 we describe some related work where graph cuts are applied to vision problems, and we show that our techniques can be used to efficiently compute the MAP estimate of an MRF with linear clique potentials.

2 Markov Random Fields

Markov Random Fields were first introduced into vision by Geman and Geman [6]. The MRF framework can express a wide variety of spatially varying priors. 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 [9] 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. 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, 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 MRF's whose clique potentials involve pairs of neighboring pixels, so

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

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 image restoration, 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 ($0 \le g \le 1$). An example of a 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 The generalized Potts model

We now consider MRF's with 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 *Generalized Potts Model MRF* (GPM-MRF) has a clique potential for any pair of neighboring pixels p and q given by

$$V_{(p,q)}(f_p, f_q) = u_{\{p,q\}} \cdot (1 - \delta(f_p - f_q)), \tag{2}$$

where the coefficient $u_{\{p,q\}} \ge 0$ specifies the depth of the well. If $u_{\{p,q\}}$ is a constant, then this yields the Potts model [11]. Note that $\{p, q\}$ is a set, not a tuple, so $V_{(p,q)}(f_p, f_q) = V_{(q,p)}(f_q, f_p)$. These MRF's are thus isotropic (i.e., independent of orientation).

The prior probability of a GPM-MRF is therefore

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

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 ($f_p \ne f_q$) 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. The prior probability $\Pr(f)$ is therefore larger for configurations f with fewer discontinuities.

The posterior energy function of a GPM-MRF is

$$\begin{aligned} E(f) \quad = \quad & \sum_{\{p,q\}\in\mathcal{E}_\mathcal{N}} 2u_{\{p,q\}}(1 - \delta(f_p - f_q)) \\ & - \sum_{p\in\mathcal{P}} \ln\left(g(i, p, f_p)\right). \end{aligned} \tag{3}$$

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 such an MRF resembles a robust estimator, in that it has a fixed maximum value (in the language of robust statistics, it is redescending). Most vision applications of MRF's follow [6] by introducing a line process that explicitly models discontinuities. [1] showed that if spatial restrictions on discontinuities are ignored, the line process can be eliminated by using a robust penalty function. We take a related 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 (3) 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 (3). This formulation, shown in equation (4), 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 (4). In section 3.3 we describe an algorithm for approximating the minimum multiway cut.

3.1 Reformulating the energy function

We want to find $f^* \in \mathcal{L}^m$ that minimizes $E(f)$ in (3). 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 $-\ln(g(i,p,l))$ in (3) will exceed the total value of the posterior energy function $E(f)$ at some configuration f. Each $\mathcal{L}_p$ is a nonempty set, since it contains f_p^*. Our search can be restricted to the set $\bar{\mathcal{L}} = \mathcal{L}_1 \times \ldots \times \mathcal{L}_m$, since f^* is in $\bar{\mathcal{L}}$.

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 (3) is equivalent to minimizing

$$\begin{aligned} \bar{E}(f) \ = \ & \sum_{\{p,q\} \in \mathcal{E}_\mathcal{N}} 2u_{\{p,q\}}(1 - \delta(f_p - f_q)) \\ & + \sum_{p \in \mathcal{P}} \sum_{\substack{l \in \mathcal{L}_p \\ l \neq f_p}} h(i,p,l) \end{aligned} \qquad (4)$$

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 terminals are completely separated in the *induced* graph $\mathcal{G}(\mathcal{C}) = \langle \mathcal{V}, \mathcal{E} - \mathcal{C} \rangle$. 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 the minimum cost multiway cut.

We now show that the minimization problem in (4) 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

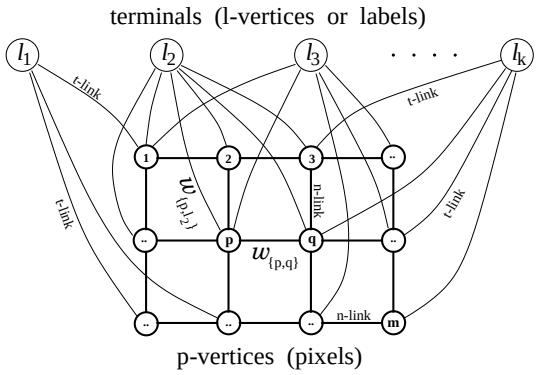


Figure 1: An example of the graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ where the terminals are $\mathcal{L} = \{l_1, \ldots, l_k\}$ and p-vertices are elements of $\mathcal{P} = \{1, \ldots, p, q, \ldots, m\}$. Each p-vertex is connected to at least one terminal.

$$w_{\{p,q\}} = 2u_{\{p,q\}}. \qquad (5)$$

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

$$w_{\{p,l\}} = h(i,p,l) + \sum_{q \in \mathcal{N}_p} w_{\{p,q\}}. \qquad (6)$$

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 in $\mathcal{G}(\mathcal{C})$.

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

PROOF: Due to equation (6), 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 $E(f)$ in (3).*

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 feasible cuts corresponding to the same $f \in \bar{\mathcal{L}}$ there is a unique *irreducible* cut $\mathcal{C}$, where *irreducible* means that it does not sever n-links between two p-vertices connected to the same terminal in $\mathcal{G}(\mathcal{C})$. 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

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

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

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

The theorem now follows from (5) and (6). ■

3.3 Multiway cut minimization

While the general multiway minimum cut problem is NP-complete, there are provably good approximations with near linear running time [3], and this is an area of active research. Approximating cuts, however, should be used carefully. If $\mathcal{C}$ approximates the minimum multiway cut on $\mathcal{G}$ within some known bounds, the value of $E(f^{\mathcal{C}})$ might not be within the same bounds with respect to the exact minimum of the posterior energy in (3). For example, cuts produced by the algorithm in [3] are not guaranteed to be even feasible. Here we describe a method that greedily reduces the cost of multiway cuts on $\mathcal{G}$. Our algorithm generates a cut $\mathcal{C}$ such that $f^{\mathcal{C}}$ is a local minimum of the posterior energy function (3) in a certain strong sense.

Our algorithm considers only irreducible feasible cuts on $\mathcal{G}$. Any such cut can be uniquely represented by a feasible partition $\mathcal{P}_{\mathcal{V}} = \{ \mathcal{V}_l \mid l \in \mathcal{L} \}$ of the set $\mathcal{V}$ where $l \in \mathcal{V}_l$ and $p \in \mathcal{V}_l$ implies $l \in \mathcal{L}_p$. An irreducible feasible cut $\mathcal{C}$ corresponds to $\mathcal{P}_{\mathcal{V}}$ where each $\mathcal{V}_l$ contains l plus all pixels $p \in \mathcal{V}$ connected to l in $\mathcal{G}(\mathcal{C})$. As an initial solution we can take any irreducible feasible cut. For example, consider a cut where $\mathcal{V}_l = \{l\} \cup \{ p \in \mathcal{V} \mid l = \min \mathcal{L}_p \}$.

At each iteration we consider a fixed pair of distinct labels $\{l, \lambda\} \subset \mathcal{L}$. The basic operation is to improve the current cut $\mathcal{C}$, that is the current feasible partition $\mathcal{P}_{\mathcal{V}}$, by reallocating pixels in $\mathcal{V}_l \cup \mathcal{V}_\lambda$ between the terminals l and λ More specifically, we solve a standard two terminal min cut problem on a graph $\mathcal{G}_{\{l,\lambda\}} = \langle \mathcal{V}_{\{l,\lambda\}}, \mathcal{E}_{\{l,\lambda\}} \rangle$, where $\mathcal{V}_{\{l,\lambda\}} = \mathcal{V}_l \cup \mathcal{V}_\lambda$ and $\mathcal{E}_{\{l,\lambda\}}$ includes all edges in $\mathcal{E}$ that connect vertices in $\mathcal{V}_{\{l,\lambda\}}$. The optimal cut on $\mathcal{G}_{\{l,\lambda\}}$ divides the pixels in $\mathcal{V}_{\{l,\lambda\}}$ between the terminals l and λ and, thus, generates the new sets $\mathcal{V}'_l$ and $\mathcal{V}'_\lambda$. This yields a new feasible partition $\mathcal{P}'_{\mathcal{V}}$ corresponding to an irreducible feasible cut $\mathcal{C}'$ such that $|\mathcal{C}'| \leq |\mathcal{C}|$. If the inequality is strict, we call the iteration successful and accept the new cut $\mathcal{C}'$. If not, we stick to the old cut $\mathcal{C}$.

At each iteration we take a new pair of terminals until all distinct pairs were considered. Then, we start a new cycle of iterations and consider the pairs of terminals all over again. The algorithm stops when no successful iterations were made in a cycle. The obtained multiway cut $\mathcal{C}$ yields $f^{\mathcal{C}}$ with the following property: the value of the energy function $E(f^{\mathcal{C}})$ cannot be decreased by switching *any* subset of pixels with one common label l to any other common label λ. This means that $f^{\mathcal{C}}$ achieves a local minimum of E in a richer "move space" than the obvious one where a move changes the label of a single pixel. We are currently developing a more sophisticated algorithm which achieves an even stronger local minimum, where the energy function cannot be decreased by switching any set of pixels to a common label λ.

Each cycle of the algorithm is quadratic in the number of labels and has the same effectively linear time complexity in the number of nodes as a standard min cut algorithm. We do not have any bounds on the number of cycles it takes to complete the algorithm. Nevertheless, in the visual correspondence applications we considered the algorithm produced promising results even after the first cycle. In section 5 we show these one cycle results.

4 Computing Visual Correspondence

We now describe how these techniques 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 GPM-MRF, and thus as a multiway cut problem. 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.[1]

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 [10], 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 GPM-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 $I_{\mathcal{P}}$ permits clique potential "depths"

$$u_{\{p,q\}} = U(|i_p - i_q|), \qquad \forall \{p,q\} \in \mathcal{E}_{\mathcal{N}}. \tag{7}$$

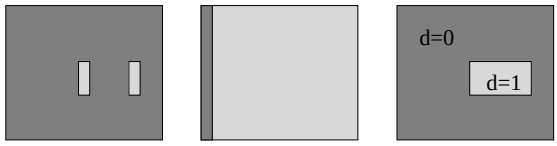
Each $u_{\{p,q\}}$ represents a penalty for assigning different disparities to neighboring pixels p and q in $\mathcal{P}$. The value of the penalty $u_{\{p,q\}}$ should be smaller for pairs $\{p,q\}$ with larger intensity differences $|i_p - i_q|$. In practice we use an empirically selected decreasing function $U(\cdot)$. Note that instead of (7) 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.



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

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 like-

[1] 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.

lihood function is

$$\Pr'(I_{\mathcal{S}} \mid d) = \Pr(I_{\mathcal{S}} \mid d, I_{\mathcal{P}}) \propto \prod_{p \in \mathcal{P}} g(i_{p \oplus d_p} | i_p) \qquad (8)$$

where d is the true disparity correspondence. Here, $g(i_s \mid i_p)$ 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 $g(i_s \mid i_p)$ 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 (8) is consistent with equation (1). Note that the main idea behind assumption (8) is that sensor noise is independent.

Equations (7) and (8) complete the description of our GPM-MRF model for visual correspondence. Now the multiway cut approach of section 3 can be used to estimate a disparity configuration d.

5 Experimental results

In this section we give some experimental results on stereo data that use our greedy multiway cut algorithm of section 3.3. 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 benchmarked 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 $> \pm 1$
GPM-MRF	8.6	2.8
LOG-filtered L_1	19.9	9.0
Normalized correlation	24.7	10.0
MLMHV [2]	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 Related work and extensions

There has been a significant amount of recent work on computer vision applications of max flow-min cut. If there are only two possible labels, the multiway cut problem simplifies to the traditional max flow-min cut. This allows the MAP estimate to be computed very efficiently, as was shown by Greig, Porteus and Seheult in [7]. Our solution can be viewed as a generalization of their result beyond binary images. Other generalizations with quite different properties have been recently proposed by Ferrari *et al.* [4, 5].

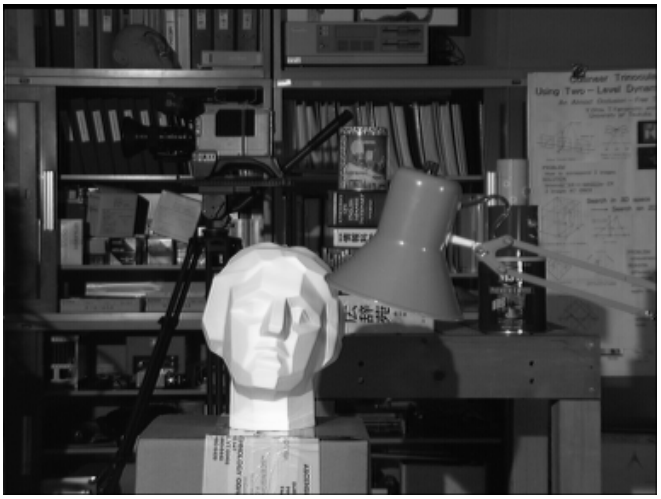
Recently, Roy and Cox [12] gave a formulation of the multi-camera stereo problem as a standard two terminal min cut problem. The approach in [12] is quite different from our work. Their problem formulation does not use energy minimization; they describe their method as a generalization of dynamic programming, while we use the MAP-MRF framework. In fact, a graph with a structure similar to that of [12] can be used to obtain the exact MAP estimate of the following MRF.

Suppose that the clique potentials V of equation (2) are replaced by $\tilde{V}_{(p,q)}(f_p, f_q) = u_{\{p,q\}} \cdot |f_p - f_q|$. We require that the label set $\mathcal{L} = \{l_1, \ldots, l_k\}$ consists of consecutive integers. Such MRF's appear suitable for image restoration and stereo matching (assuming known epipolar geometry), but not for motion. The MAP estimate of such an MRF can be obtained by minimizing over $f \in \mathcal{L}^m$ the posterior energy function

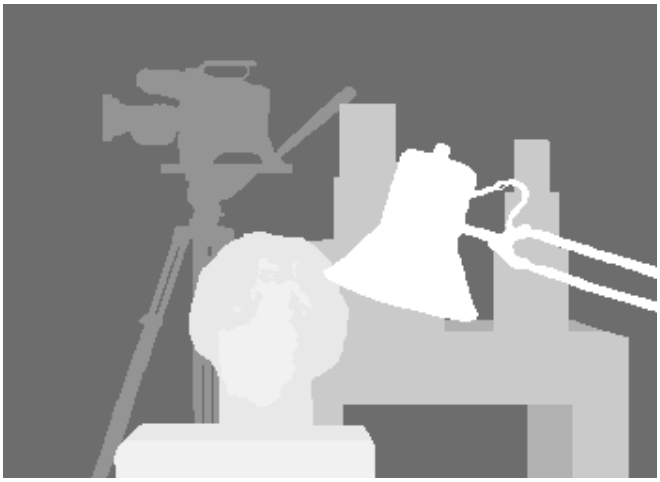
$$\tilde{E}(f) = \sum_{\{p,q\} \in \mathcal{E}_{\mathcal{N}}} 2u_{\{p,q\}} |f_p - f_q| \; - \sum_{p \in \mathcal{P}} \ln \left(g(i, p, f_p) \right).$$

To minimize $\tilde{E}(f)$ we apply our graph techniques of section 3.2. Consider a graph $\tilde{\mathcal{G}}$ defined as follows. There are two terminals: the source R and the sink S. For each pixel p we create a set of vertices $p_1, \ldots, p_{k-1}$. We connect them by t-links $\{t_1^p, \ldots, t_k^p\}$ where $t_1^p = \{R, p_1\}$, $t_j^p = \{p_{j-1}, p_j\}$, and $t_k^p = \{p_{k-1}, S\}$. For each pair of neighboring pixels p, q and for each $j \in \{1, \ldots, k-1\}$ we create an n-link $\{p_j, q_j\}$ with weight $w_{\{p,q\}} = 2u_{\{p,q\}}$. Each t-link t_j^p is assigned a weight $K_p - \ln(g(i, p, l_j))$ where K_p is any constant such that $K_p > (k-1) \sum_{q \in \mathcal{N}_p} w_{\{p,q\}}$.

A cut on the graph $\tilde{\mathcal{G}}$ will break at least one t-link for each pixel; we call a cut *feasible* if it breaks exactly one t-link for each pixel. Each feasible cut $\mathcal{C}$

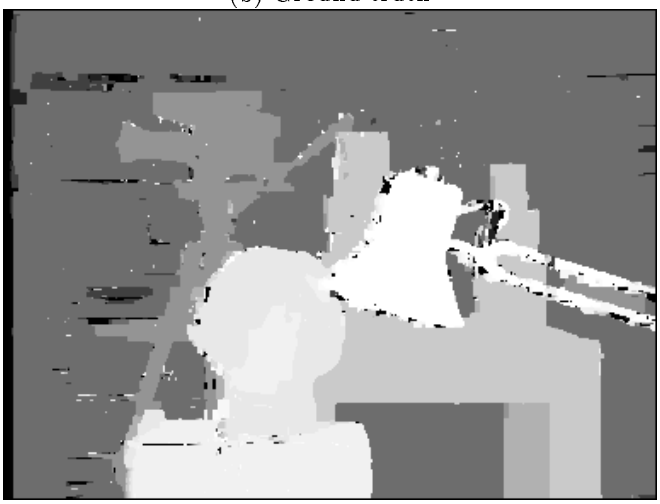


(a) Scene



(b) Ground truth

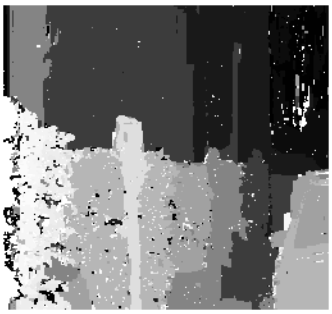
(c) Errors $> \pm 1$



(d) GPM-MRF results

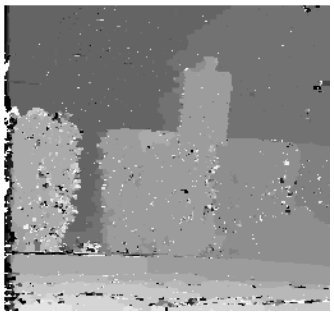
Figure 2: Ground truth results

Meter image



GPM-MRF results

Shrub image



GPM-MRF results

Figure 3: Benchmark results. For more images see `http://www.cs.cornell.edu/home/yura`.

corresponds to some configuration $f^{\mathcal{C}}$ where for each pixel p we take $f_p^{\mathcal{C}} = l_j$ if the t-link t_j^p is cut by $\mathcal{C}$.

Lemma 2 *A minimum cut $\mathcal{C}$ on $\tilde{\mathcal{G}}$ must be feasible.*

PROOF: Suppose that t_a^p and t_b^p are cut. Then we can find a smaller cut by restoring t_b^p and breaking n-links $\{p_j, q_j\}$ for all $q \in \mathcal{N}_p$ and $1 \leq j \leq k-1$. The cost of the cut will decrease at least by $K_p - \ln(g(i,p,l_b)) - (k-1)\sum_{q\in\mathcal{N}_p} w_{\{p,q\}}$ which is strictly positive due to our choice of K_p. ■

Theorem 2 *If $\mathcal{C}$ is a minimum cut on $\tilde{\mathcal{G}}$, then $f^{\mathcal{C}}$ minimizes the posterior energy function $\tilde{E}(f)$.*

PROOF: We follow the same path as in the proof of Theorem 1. A cut $\mathcal{C}$ is called *irreducible* if it does not sever n-links between vertices connected to the same terminal in $\tilde{\mathcal{G}}(\mathcal{C})$. It is easy to show that there is a one to one correspondence between the set of all irreducible feasible cuts and the set of all configurations $f \in \mathcal{L}^m$. Since the minimum cut has to be both feasible and irreducible it remains to show that the cost of any irreducible feasible cut $\mathcal{C}$ satisfies $|\mathcal{C}| = A + \tilde{E}(f^{\mathcal{C}})$. If $\mathcal{C}$ is feasible, the cost of cutting t-links is $\sum_{p\in\mathcal{P}} \left(K_p - \ln(g(i,p,f_p^{\mathcal{C}}))\right)$. If $\mathcal{C}$ is also irreducibile the cost of cutting n-links is $\sum_{\{p,q\}\in\mathcal{E}_{\mathcal{N}}} w_{\{p,q\}} |f_p^{\mathcal{C}} - f_q^{\mathcal{C}}|$. ■

The difference between $\tilde{\mathcal{G}}$ and the graph in [12] lies in the link weights. Our choice of edge weights guarantees the optimality property of Theorem 2. In contrast, the weights in [12] lack theoretical justification. As a result, their algorithm does not appear to have any optimality properties.

Note that Ishikawa and Geiger [8] describe an image segmentation technique that finds the global minimum of an energy function closely related to $\tilde{E}(f)$. Their solution, developed independently before ours, finds a minimum cut on a graph similar to $\tilde{\mathcal{G}}$ except for some details. For example, their graph is directed and has some infinite capacity links, while we employ an undirected graph. We also emphasize the use of contextual information for selecting penalties $u_{\{p,q\}}$ as described in section 4.1. Our experiments suggest that this may significantly improve the results.

Acknowledgments

We thank J. Kleinberg, D. Shmoys and E. Tardos 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 ONR.

References

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

[2] I. Cox, S. Hingorani, S. Rao, and B. Maggs. A maximum likelihood stereo algorithm. *Computer Vision, Graphics and Image Processing*, 63(3):542–567, 1996.

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

[4] P. Ferrari, A. Frigessi, and P. de Sá. Fast approximate maximum a posteriori restoration of multicolour images. *Journal of the Royal Statistical Society, Series B*, 57(3):485–500, 1995.

[5] P. Ferrari, M. Gubitoso, and E. Neves. Restoration of multicolor images. Available from http://www.ime.usp.br/ pablo, December 1997.

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

[7] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279, 1989.

[8] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.

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

[10] 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.

[11] R. Potts. Some generalized order-disorder transformation. *Proceedings of the Cambridge Philosophical Society*, 48(106), 1952.

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