An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision

Yuri Boykov, *Member*, *IEEE*, and Vladimir Kolmogorov, *Member*, *IEEE*

Abstract—After [15], [31], [19], [8], [25], [5], minimum cut/maximum flow algorithms on graphs emerged as an increasingly useful tool for exact or approximate energy minimization in low-level vision. The combinatorial optimization literature provides many min-cut/max-flow algorithms with different polynomial time complexity. Their practical efficiency, however, has to date been studied mainly outside the scope of computer vision. The goal of this paper is to provide an experimental comparison of the efficiency of min-cut/max flow algorithms for applications in vision. We compare the running times of several standard algorithms, as well as a new algorithm that we have recently developed. The algorithms we study include both Goldberg-Tarjan style "push-relabel" methods and algorithms based on Ford-Fulkerson style "augmenting paths." We benchmark these algorithms on a number of typical graphs in the contexts of image restoration, stereo, and segmentation. In many cases, our new algorithm works several times faster than any of the other methods, making near real-time performance possible. An implementation of our max-flow/min-cut algorithm is available upon request for research purposes.

Index Terms—Energy minimization, graph algorithms, minimum cut, maximum flow, image restoration, segmentation, stereo, multicamera scene reconstruction.

1 Introduction

Greig et al. [15] were the first to discover that powerful min-cut/max-flow algorithms from combinatorial optimization can be used to minimize certain important energy functions in vision. The energies addressed by Greig et al. and by most later graph-based methods (e.g., [32], [18], [4], [17], [8], [2], [30], [39], [21], [36], [38], [6], [23], [24], [9], [26]) can be represented as[1]

$$E(L) = \sum_{p \in \mathcal{P}} D_p(L_p) + \sum_{(p,q) \in \mathcal{N}} V_{p,q}(L_p, L_q), \qquad (1)$$

where $L = \{L_p \,|p \in \mathcal{P}\}$ is a labeling of image $\mathcal{P}$, $D_p(\cdot)$ is a data penalty function, $V_{p,q}$ is an interaction potential, and $\mathcal{N}$ is a set of all pairs of neighboring pixels. An example of image labeling is shown in Fig. 1. Typically, data penalties $D_p(\cdot)$ indicate individual label-preferences of pixels based on observed intensities and prespecified likelihood function. Interaction potentials $V_{p,q}$ encourage spatial coherence by penalizing discontinuities between neighboring pixels. The papers above show that, to date, graph-based energy minimization methods arguably provide some of the most accurate solutions for the specified applications. For example, consider two recent evaluations of stereo algorithms using real imagery with dense ground truth [34], [37].

1. Greig et al. [15] consider energy (1) in the context of maximum a posteriori estimation of Markov Random Fields (MAP-MRF).

- *Y. Boykov is with the Computer Science Department, the University of Western Ontario, London, Ontario N6A 5B7, Canada. E-mail: yuri@csd.uwo.ca.*
- *V. Kolmogorov is with Microsoft Research, 7 J.J. Thomson Ave., Cambridge CB3 0FB, UK. E-mail: vnk@microsoft.com.*

Manuscript received 4 June 2003; revised 16 Feb. 2004; accepted 25 Feb. 2004. Recommended for acceptance by A. Rangarjan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0120-0603.

Greig et al. constructed a two terminal graph such that the minimum cost cut of the graph gives a globally optimal binary labeling L in case of the Potts model of interaction in (1). Previously, exact minimization of energies like (1) was not possible and such energies were approached mainly with iterative algorithms like simulated annealing. In fact, Greig et al. used their result to show that, in practice, simulated annealing reaches solutions very far from the global minimum even in a very simple example of binary image restoration.

Unfortunately, the graph cut technique in Greig et al. remained unnoticed for almost 10 years mainly because binary image restoration looked very limited as an application. Early attempts to use combinatorial graph cut algorithms in vision were restricted to image clustering [40]. In the late 1990s, a large number of new computer vision techniques appeared that figured how to use min-cut/max-flow algorithms on graphs for solving more interesting nonbinary problems. Roy and Cox [32] were the first to use these algorithms to compute multicamera stereo. Later, [18], [4] showed that, with the right edge weights on a graph similar to that used in [32], one can minimize a fairly general energy function (1) in a multilabel case with linear interaction penalties. This graph construction was further generalized to handle arbitrary convex cliques in [19]. Another general case of multilabel energies where interaction penalty is a *metric* (on the space of labels) was studied in [4], [8]. Their α-expansion algorithm finds provably good approximate solutions by iteratively running min-cut/max-flow algorithms on appropriate graphs. The case of *metric* interactions includes many kinds of "robust" cliques that are frequently preferred in practice.

Several recent papers studied theoretical properties of graph constructions used in vision. The question of what energy functions can be minimized via graph cuts was addressed in [25]. This work provided a simple, necessary,

0162-8828/04/$20.00 © 2004 IEEE Published by the IEEE Computer Society

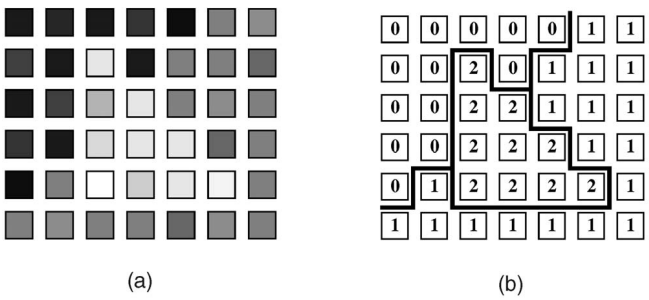


Fig. 1. An example of image labeling. An image in (a) is a set of pixels $\mathcal{P}$ with observed intensities I_p for each $p \in \mathcal{P}$. A labeling L shown in (b) assigns some label $L_p \in \{0, 1, 2\}$ to each pixel $p \in \mathcal{P}$. Such labels can represent depth (in stereo), object index (in segmentation), original intensity (in image restoration), or other pixel properties. Normally, graph-based methods assume that a set of feasible labels at each pixel is finite. Thick lines in (b) show labeling discontinuities between neighboring pixels.

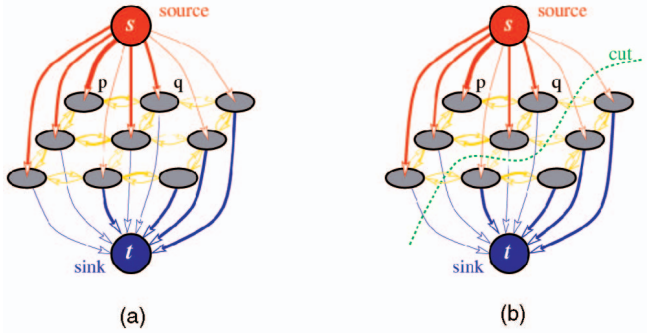


Fig. 2. Example of a directed capacitated graph. Edge costs are reflected by their thickness. A similar graph-cut construction was first used in vision by Greig et al. [15] for binary image restoration. (a) A graph $\mathcal{G}$. (b) A cut on $\mathcal{G}$.

and sufficient condition on such functions. However, the results in [25] apply only to energy functions of binary variables with double and triple cliques. In fact, the full potential of graph-cut techniques in multilabel cases is still not entirely understood.

Geometric properties of segments produced by graph-cut methods were investigated in [3]. This work studied *cut metric* on regular grid-graphs and showed that discrete topology of graph-cuts can approximate any continuous Riemannian metric space. The results in [3] established a link between two standard energy minimization approaches frequently used in vision: combinatorial graph-cut methods and geometric methods based on level-sets (e.g., [35], [29], [33], [28]).

A growing number of publications in vision use graph-based energy minimization techniques for applications like image segmentation [18], [39], [21], [5], restoration [15], stereo [32], [4], [17], [23], [24], [9], shape reconstruction [36], object recognition [2], augmented reality [38], texture synthesis [26], and others. The graphs corresponding to these applications are usually huge 2D or 3D grids and min-cut/max-flow algorithm efficiency is an issue that cannot be ignored.

The main goal of this paper is to experimentally compare the running time of several min-cut/max-flow algorithms on graphs typical for applications in vision. In Section 2, we provide basic facts about graphs, min-cut and max-flow problems, and some standard combinatorial optimization algorithms for them. We consider both Goldberg-Tarjan style *push-relabel* algorithms [14] as well as methods based on *augmenting paths* a la Ford-Fulkerson [13]. Note that, in the course of our experiments with standard augmenting path techniques, we developed some new algorithmic ideas that significantly boosted empirical performance on grid-graphs in vision. Section 3 describes our new min-cut/max-flow algorithm. In Section 4, we tested this new augmenting-path style algorithm as well as three standard algorithms: the H_PRF and Q_PRF versions of the "push-relabel" method [14], [10] and the Dinic algorithm [12] that also uses augmenting paths. We selected several examples in image restoration, stereo, and segmentation where different forms of energy (1) are minimized via graph structures originally described in [15], [18], [4], [8], [23], [24], [6]. Such (or very similar) graphs are used in all computer vision papers known to us that use graph cut algorithms. In many interesting cases, our new algorithm was significantly faster than the standard min-cut/max-flow techniques from combinatorial optimization. More detailed conclusions are presented in Section 5.

2 Background on Graphs

In this section, we review some basic facts about graphs in the context of energy minimization methods in vision. A directed weighted (capacitated) graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ consists of a set of nodes $\mathcal{V}$ and a set of directed edges $\mathcal{E}$ that connect them. Usually, the nodes correspond to pixels, voxels, or other features. A graph normally contains some additional special nodes that are called terminals. In the context of vision, terminals correspond to the set of labels that can be assigned to pixels. We will concentrate on the case of graphs with two terminals. Then, the terminals are usually called the *source*, s, and the *sink*, t. In Fig. 2a, we show a simple example of a two terminal graph (due to Greig et al. [15]) that can be used to minimize the Potts case of energy (1) on a 3×3 image with two labels. There is some variation in the structure of graphs used in other energy minimization methods in vision. However, most of them are based on regular 2D or 3D grid graphs such as the one in Fig. 2a. This is a simple consequence of the fact that, normally, graph nodes represent regular image pixels or voxels.

All edges in the graph are assigned some weight or cost. A cost of a directed edge (p, q) may differ from the cost of the reverse edge (q, p). In fact, the ability to assign different edge weights for (p, q) and (q, p) is important for many graph-based applications in vision. Normally, there are two types of edges in the graph: n-links and t-links. N-links connect pairs of neighboring pixels or voxels. Thus, they represent a neighborhood system in the image. The cost of n-links corresponds to a penalty for discontinuity between the pixels. These costs are usually derived from the pixel interaction term $V_{p,q}$ in energy (1). T-links connect pixels with terminals (labels). The cost of a t-link connecting a pixel and a terminal corresponds to a penalty for assigning the corresponding label to the pixel. This cost is normally derived from the data term D_p in the energy (1).

2.1 Min-Cut and Max-Flow Problems

An s/t cut C on a graph with two terminals is a partitioning of the nodes in the graph into two disjoint subsets $\mathcal{S}$ and $\mathcal{T}$ such that the source s is in $\mathcal{S}$ and the sink t is in $\mathcal{T}$. For simplicity, throughout this paper, we refer to s/t cuts as just *cuts*. Fig. 2b shows one example of a cut. In combinatorial optimization, the cost of a cut $C = \{\mathcal{S}, \mathcal{T}\}$ is defined as the

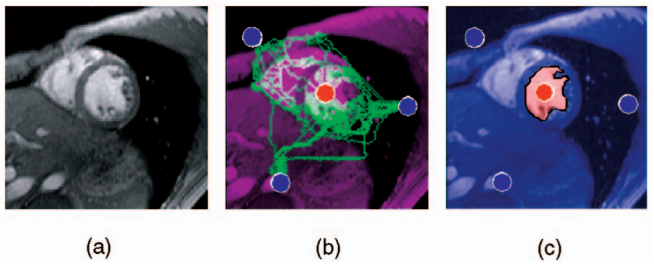


Fig. 3. Graph cut/flow example in the context of image segmentation in Section 4.4. Red and blue seeds are "hard-wired" to the source s and the sink t, correspondingly. As usual, the cost of edges between the pixels (graph nodes) is set to low values in places with high intensity contrast. Thus, cuts along object boundaries in the image should be cheaper. Weak edges also work as "bottlenecks" for a flow. In (b), we show a maximum flow from s to t. In fact, it saturates graph edges corresponding to a minimum cut boundary in (c). (a) Original image. (b) A maximum flow. (c) A minimum cut.

sum of the costs of "boundary" edges (p, q) where $p \in \mathcal{S}$ and $q \in \mathcal{T}$. Note that cut cost is "directed" as it sums up weights of directed edges specifically from $\mathcal{S}$ to $\mathcal{T}$. The *minimum cut* problem on a graph is to find a cut that has the minimum cost among all cuts.

One of the fundamental results in combinatorial optimization is that the minimum s/t cut problem can be solved by finding a *maximum flow* from the source s to the sink t. Loosely speaking, maximum flow is the maximum "amount of water" that can be sent from the source to the sink by interpreting graph edges as directed "pipes" with capacities equal to edge weights. The theorem of Ford and Fulkerson [13] states that a maximum flow from s to t saturates a set of edges in the graph dividing the nodes into two disjoint parts $\{\mathcal{S}, \mathcal{T}\}$ corresponding to a minimum cut. Thus, min-cut and max-flow problems are equivalent. In fact, the maximum flow value is equal to the cost of the minimum cut. The "duality" relationship between maximum flow and minimum cut problems is illustrated in Fig. 3 in the context of image segmentation. Max-flow displayed in Fig. 3a saturates the edges in the min-cut boundary in Fig. 3b.

We can intuitively show how min-cut (or max-flow) on a graph may help with energy minimization over image labelings. Consider an example in Fig. 2. The graph corresponds to a 3×3 image. Any s/t cut partitions the nodes into disjoint groups each containing exactly one terminal. Therefore, any cut corresponds to some assignment of pixels (nodes) to labels (terminals). If edge weights are appropriately set based on parameters of an energy, a minimum cost cut will correspond to a labeling with the minimum value of this energy.[2]

2.2 Standard Algorithms in Combinatorial Optimization

An important fact in combinatorial optimization is that there are polynomial algorithms for min-cut/max-flow problems on directed weighted graphs with two terminals. Most of the algorithms belong to one of the following two groups: Goldberg-Tarjan style "push-relabel" methods [14] and algorithms based on Ford-Fulkerson style "augmenting paths" [13].

2. Different graph-based energy minimization methods may use different graph constructions, as well as different rules for converting graph cuts into image labelings. Details for each method are described in the original publications.

Standard augmenting paths-based algorithms, such as the Dinic algorithm [12], work by pushing flow along non-saturated paths from the source to the sink until the maximum flow in the graph $\mathcal{G}$ is reached. A typical augmenting path algorithm stores information about the distribution of the current $s \to t$ flow f among the edges of $\mathcal{G}$ using a *residual graph* $\mathcal{G}_f$. The topology of $\mathcal{G}_f$ is identical to $\mathcal{G}$, but the capacity of an edge in $\mathcal{G}_f$ reflects the residual capacity of the same edge in $\mathcal{G}$ given the amount of flow already in the edge. At the initialization, there is no flow from the source to the sink (f = 0) and edge capacities in the residual graph $\mathcal{G}_0$ are equal to the original capacities in $\mathcal{G}$. At each new iteration, the algorithm finds the shortest $s \to t$ path along nonsaturated edges of the residual graph. If a path is found, then the algorithm *augments* it by pushing the maximum possible flow df that saturates at least one of the edges in the path. The residual capacities of edges in the path are reduced by df while the residual capacities of the reverse edges are increased by df. Each augmentation increases the total flow from the source to the sink $f = f + df$. The maximum flow is reached when any $s \to t$ path crosses at least one saturated edge in the residual graph $\mathcal{G}_f$.

The Dinic algorithm uses breadth-first search to find the shortest paths from s to t on the residual graph $\mathcal{G}_f$. After all shortest paths of a fixed length k are saturated, the algorithm starts the breadth-first search for $s \to t$ paths of length $k + 1$ from scratch. Note that the use of shortest paths is an important factor that improves theoretical running time complexities for algorithms based on augmenting paths. The worst-case running time complexity for the Dinic algorithm is $O(mn^2)$, where n is the number of nodes and m is the number of edges in the graph.

Push-relabel algorithms [14] use quite a different approach. They do not maintain a valid flow during the operation; there are "active" nodes that have a positive "flow excess." Instead, the algorithms maintain a labeling of nodes giving a low bound estimate on the distance to the sink along nonsaturated edges. The algorithms attempt to "push" excess flows toward nodes with smaller estimated distance to the sink. Typically, the "push" operation is applied to active nodes with the largest distance (label) or based on FIFO selection strategy. The distances (labels) progressively increase as edges are saturated by push operations. Undeliverable flows are eventually drained back to the source. We recommend our favorite textbook on basic graph theory and algorithms [11] for more details on push-relabel and augmenting path methods.

Note that the most interesting applications of graph cuts to vision use directed N-D grids with locally connected nodes. It is also typical that a large portion of the nodes is connected to the terminals. Unfortunately, these conditions rule out many specialized min-cut/max-flow algorithms that are designed for some restricted classes of graphs. Examples of interesting but inapplicable methods include randomized techniques for dense undirected graphs [20], methods for planar graphs assuming small number of terminal connections [27], [16], and others.

3 New Min-Cut/Max-Flow Algorithm

In this section, we present a new algorithm developed during our attempts to improve empirical performance of standard augmenting path techniques on graphs in vision.

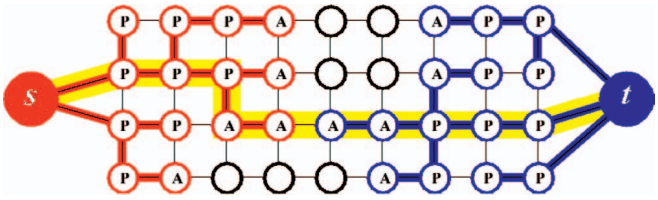


Fig. 4. Example of the search trees S (red nodes) and T (blue nodes) at the end of the growth stage when a path (yellow line) from the source s to the sink t is found. Active and passive nodes are labeled by letters A and P, correspondingly. Free nodes appear in black.

Normally, see Section 2.2, augmenting path-based methods start a new breadth-first search for $s \rightarrow t$ paths as soon as all paths of a given length are exhausted. In the context of graphs in computer vision, building a breadth-first search tree typically involves scanning the majority of image pixels. Practically speaking, it could be a very expensive operation if it has to be performed too often. Indeed, our real-data experiments in vision confirmed that rebuilding a search tree on graphs makes standard augmenting path techniques perform poorly in practice. We developed several ideas that improved empirical performance of augmenting path techniques on graphs in computer vision.

The new min-cut/max-flow algorithm presented here belongs to the group of algorithms based on augmenting paths. Similar to Dinic [12], it builds search trees for detecting augmenting paths. In fact, we build two search trees, one from the source and the other from the sink.[3] The other difference is that we reuse these trees and never start building them from scratch. The drawback of our approach is that the augmenting paths found are not necessarily shortest augmenting path; thus, the time complexity of the shortest augmenting path is no longer valid. The trivial upper bound on the number of augmentations for our algorithm is the cost of the minimum cut $|C|$, which results in the worst-case complexity $O(mn^2|C|)$. Theoretically speaking, this is worse than the complexities of the standard algorithms discussed in Section 2.2. However, experimental comparison in Section 4 shows that, on typical problem instances in vision, our algorithm significantly outperforms standard algorithms.

3.1 Algorithm's Overview

Fig. 4 illustrates our basic terminology. We maintain two nonoverlapping search trees S and T with roots at the source s and the sink t, correspondingly. In tree S, all edges from each parent node to its children are nonsaturated, while, in tree T, edges from children to their parents are nonsaturated. The nodes that are not in S or T are called "free." We have

$$S \subset \mathcal{V}, \quad s \in S, \quad T \subset \mathcal{V}, \quad t \in T, \quad S \cap T = \emptyset.$$

The nodes in the search trees S and T can be either "active" or "passive." The active nodes represent the outer border in each tree, while the passive nodes are internal. The point is that active nodes allow trees to "grow" by acquiring new children (along nonsaturated edges) from a set of free nodes. The passive nodes cannot grow as they are completely blocked by other nodes from the same tree. It is also important that active nodes may come in contact with the nodes from the other tree. An augmenting path is found as soon as an active node in one of the trees detects a neighboring node that belongs to the other tree.

The algorithm iteratively repeats the following three stages:

- "growth" stage: search trees S and T grow until they touch giving an $s \rightarrow t$ path,
- "augmentation" stage: the found path is augmented, search tree(s) break into forest(s), and
- "adoption" stage: trees S and T are restored.

At the growth stage, the search trees expand. The active nodes explore adjacent nonsaturated edges and acquire new children from a set of free nodes. The newly acquired nodes become active members of the corresponding search trees. As soon as all neighbors of a given active node are explored, the active node becomes passive. The growth stage terminates if an active node encounters a neighboring node that belongs to the opposite tree. In this case, we detect a path from the source to the sink, as shown in Fig. 4.

The augmentation stage augments the path found at the growth stage. Since we push through the largest flow possible, some edge(s) in the path become saturated. Thus, some of the nodes in the trees S and T may become "orphans," that is, the edges linking them to their parents are no longer valid (they are saturated). In fact, the augmentation phase may split the search trees S and T into forests. The source s and the sink t are still roots of two of the trees, while orphans form roots of all other trees.

The goal of the adoption stage is to restore the single-tree structure of sets S and T with roots in the source and the sink. At this stage, we try to find a new valid parent for each orphan. A new parent should belong to the same set, S or T, as the orphan. A parent should also be connected through a nonsaturated edge. If there is no qualifying parent, we remove the orphan from S or T and make it a free node. We also declare all its former children orphans. The stage terminates when no orphans are left and, thus, the search tree structures of S and T are restored. Since some orphan nodes in S and T may become free, the adoption stage results in contraction of these sets.

After the adoption stage is completed, the algorithm returns to the growth stage. The algorithm terminates when the search trees S and T cannot grow (no active nodes) and the trees are separated by saturated edges. This implies that a maximum flow is achieved. The corresponding minimum cut can be determined by $\mathcal{S} = S$ and $\mathcal{T} = T$.[4]

3.2 Details of Implementation

Assume that we have a directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$. As with any augmenting path algorithm, we will maintain a flow f and the residual graph G_f (see Section 2.2). We will keep the lists of all active nodes, A, and all orphans, O. The general structure of the algorithm is:

3. Note that, in the earlier publication [7], we used a single tree rooted at the source that searched for the sink. The two-trees version presented here treats the terminals symmetrically. Experimentally, the new algorithm consistently outperforms the one in [7].

4. Strictly speaking, this is is true only if there are no free nodes upon termination, i.e., $S \cup T = \mathcal{V}$. If there are isolated free nodes in the end, then minimum cut can be determined as $\{S, \mathcal{V} - S\}$ or $\{\mathcal{V} - T, T\}$. Both solutions have the same cost.

```
initialize:   S = {s}, T = {t}, A = {s,t}, O = ∅
while true
        grow S or T to find an augmenting path
             P from s to t
        if P = ∅ terminate
        augment on P
        adopt orphans
end while
```

The details of the *growth*, *augmentation*, and *adoption* stages are described below. It is convenient to store content of search trees S and T via flags $TREE(p)$ indicating the affiliation of each node p so that

$$TREE(p) = \begin{cases} S & \text{if } p \in S \\ T & \text{if } p \in T \\ \emptyset & \text{if } p \text{ is free.} \end{cases}$$

If node p belongs to one of the search trees, then the information about its parent will be stored as $PARENT(p)$. Roots of the search trees (the source and the sink), orphans, and all free nodes have no parents, i.e., $PARENT(p) = \emptyset$. We will also use notation $tree\_cap(p \to q)$ to describe the residual capacity of either edge (p,q) if $TREE(p) = S$ or edge (q,p) if $TREE(p) = T$. These edges should be nonsaturated in order for node p to be a valid parent of its child q depending on the search tree.

3.2.1 Growth Stage

At this stage, active nodes acquire new children from a set of free nodes.

```
while A ≠ ∅
    pick an active node p ∈ A
    for every neighbor q such that
          tree_cap(p → q) > 0
    if TREE(q) = ∅ then add q to search tree as an
          active node:
        TREE(q) := TREE(p), PARENT(q) := p,
        A := A ∪ {q}
    if TREE(q) ≠ ∅ and TREE(q) ≠ TREE(p)
        return P = PATH_{s→t}
    end for
    remove p from A
end while
return P = ∅
```

3.2.2 Augmentation Stage

The input for this stage is a path P from s to t. Note that the orphan set is empty in the beginning of the stage, but there might be some orphans in the end since at least one edge in P becomes saturated.

```
find the bottleneck capacity Δ on P
update the residual graph by pushing flow Δ
      through P
for each edge (p,q) in P that becomes saturated
    if TREE(p) = TREE(q) = S then set
       PARENT(q) := ∅ and O := O ∪ {q}
    if TREE(p) = TREE(q) = T then set
       PARENT(p) := ∅ and O := O ∪ {p}
end for
```

3.2.3 Adoption Stage

During this stage, all orphan nodes in O are processed until O becomes empty. Each node p being processed tries to find a new valid parent within the same search tree; in case of success, p remains in the tree but with a new parent; otherwise, it becomes a free node and all its children are added to O.

```
while O ≠ ∅
    pick an orphan node p ∈ O and remove it
         from O
    process p
end while
```

The operation "process p" consists of the following steps: First, we are trying to find a new valid parent for p among its neighbors. A valid parent q should satisfy: $TREE(q) = TREE(p)$, $tree\_cap(q \to p) > 0$, and the "origin" of q should be either source or sink. Note that the last condition is necessary because, during the adoption stage, some of the nodes in the search trees S or T may originate from orphans.

If node p finds a new valid parent q, then we set $PARENT(p) = q$. In this case, p remains in its search tree and the active (or passive) status of p remains unchanged. If p does not find a valid parent, then p becomes a free node and the following operations are performed:

- Scan all neighbors q of p such that $TREE(q) = TREE(p)$:
  - If $tree\_cap(q \to p) > 0$, add q to the active set A
  - If $PARENT(q) = p$, add q to the set of orphans O and set $PARENT(q) := \emptyset$
- $TREE(p) := \emptyset$, $A := A - \{p\}$.

Note that, as p becomes free, all its neighbors connected through nonsaturated edges should become active. It may happen that some neighbor q did not qualify as a valid parent during the adoption stage because it did not originate from the source or the sink. However, this node could be a valid parent after the adoption stage is finished. At this point, q must have active status as it is located next to a free node p.

3.3 Algorithm Tuning

The proof of correctness of the algorithm presented above is straightforward (see [22]). At the same time, our description leaves many free choices in implementing certain details. For example, we found that the order of processing active nodes and orphans may have a significant effect on the algorithm's running time. Our preferred processing method is a minor variation of "First-In-First-Out." In this case, the growth stage can be described as a breadth-first search. This guarantees that at least the first path from the source to the sink is the shortest. Note that the search tree may change unpredictably during the adoption stage. Thus, we cannot guarantee anything about paths found after the first one.

There are several additional free choices in implementing the adoption stage. For example, as an orphan looks for a new parent, it has to make sure that a given candidate is connected to the source or to the sink. We found that "marking" nodes confirmed to be connected to the source at a given adoption stage helps to speed up the algorithm. In this case, other orphans do not have to trace the roots of their potential parents all the way to the terminals. We also found that keeping distance-to-source information in addition to these

"marks" allows orphans to select new parents that are closer to the source. This further helps with the algorithm's speed because we get shorter paths.

We used a fixed tuning of our algorithm in all experiments of Section 4. Complete details of this tuning can be found in [22]. A library with our implementation is available upon request for research purposes. The general goal of tuning was to make augmenting paths as short as possible. Note that augmenting paths on graphs in vision can be easily visualized. In the majority of cases, such graphs are regular grids of nodes that correspond to image pixels. Then, augmenting paths and the whole graph flow can be meaningfully displayed (e.g., Fig. 3b). We can also display the search trees at different stages. This allows a very intuitive way of tuning max-flow methods in vision.

4 Experimental Tests on Applications in Vision

In this section, we experimentally test min-cut/max-flow algorithms for three different applications in computer vision: image restoration (Section 4.2), stereo (Section 4.3), and object segmentation (Section 4.4). We chose formulations where certain appropriate versions of energy (1) can be minimized via graph cuts. The corresponding graph structures were previously described by [15], [18], [4], [8], [23], [24], [5] in detail. These (or very similar) structures are used in all computer vision applications with graph cuts (that we are aware of) to date.

4.1 Experimental Setup

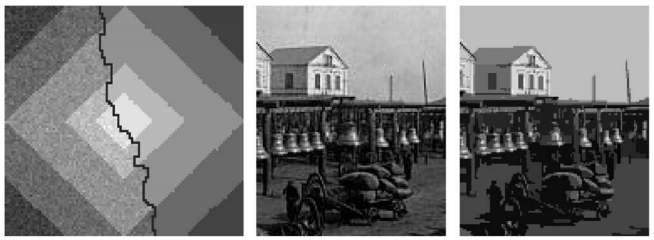
Note that we could not test all known min-cut/max-flow algorithms. In our experimental tests on graph-based energy minimization methods in vision, we compared the new algorithm in Section 3 and the following standard min-cut/max-flow algorithms outlined in Section 2.2:

- **DINIC**: Algorithm of Dinic [12].
- **H_PRF**: Push-Relabel algorithm [14] with the highest level selection rule.
- **Q_PRF**: Push-Relabel algorithm [14] with the queue-based selection rule.

Many previous experimental tests, including the results in [10], show that the last two algorithms work consistently better than a large number of other min-cut/max-flow algorithms of combinatorial optimization. The theoretical worst-case complexities for these "push-relabel" algorithms are $O(n^3)$ for Q_PRF and $O(n^2\sqrt{m})$ for H_PRF.

For DINIC, H_PRF, and Q_PRF we used the implementations written by Cherkassky and Goldberg [10], except that we converted them from C to C++ style and modified the interface (i.e., functions for creating a graph). Both H_PRF and Q_PRF use global and gap relabeling heuristics. Our algorithm was implemented in C++. We selected a tuning described in Section 3.3 with more details available in [22]. We did not make any machine specific optimization (such as pipeline-friendly instruction scheduling or cache-friendly memory usage).

Experiments in Sections 4.2 and 4.4 were performed on a 1.4GHz Pentium IV PC (2GB RAM, 8KB L1 cache, 256KB L2 cache) and experiments in Section 4.3 were performed on an UltraSPARC II workstation with four 450 MHz processors and 4GB RAM. In the former case, we used Microsoft Visual C++ 6.0 compiler, Windows NT platform, and, in the latter case, GNU C++ compiler, version 3.2.2 with the flag "-O5," SunOS 5.8 platform. To get system time, we used the ftime() function in Unix and the _ftime() function in Windows. Although these functions do not measure process computation time, we felt that they were appropriate since we got very consistent results (within 1 percent) when running tests multiple times.



(a) (b) (c)

Fig. 5. Image restoration examples. (a) *Diamond* restoration. (b) Original *Bell Quad*. (c) "Restored" *Bell Quad*.

4.2 Image Restoration

Image restoration is a representative early vision problem. The goal is to restore original pixel intensities from the observed noisy data. Some examples of image restoration are shown in Fig. 5. The problem can be very easily formulated in terms of energy (1) minimization. In fact, many other low-level vision problems can be represented by the same energies. We chose the context of image restoration mainly for its simplicity.

In this section, we consider two examples of energy (1) based on the Potts and linear models of interaction, correspondingly. Besides image restoration [15], graph methods for minimizing Potts energy were used in segmentation [21], stereo [4], [8], object recognition [2], shape reconstruction [36], and augmented reality [38]. Linear interaction energies were used in stereo [32] and segmentation [18]. Minimization of the linear interaction energy is based on graphs that are quite different from what is used for the Potts model. At the same time, there is very little variation between the graphs in different applications when the same type of energy is used. They mainly differ in their specific edge cost settings while the topological properties of graphs are almost identical once the energy model is fixed.

4.2.1 Potts Model

The Potts energy that we use for image restoration is

$$E(I) = \sum_{p \in \mathcal{P}} ||I_p - I_p^o|| + \sum_{(p,q) \in \mathcal{N}} K_{(p,q)} \cdot T(I_p \neq I_q), \qquad (2)$$

where $I = \{I_p \,|p \in \mathcal{P}\}$ is a vector of unknown "true" intensities of pixels in image $\mathcal{P}$ and $I^o = \{I_p^o \,|p \in \mathcal{P}\}$ are observed intensities corrupted by noise. The Potts interactions are specified by penalties $K_{(p,q)}$ for intensity discontinuities between neighboring pixels. Function $T(\cdot)$ is 1 if the condition inside the parentheses is true and 0 otherwise. In the case of two labels, the Potts energy can be minimized exactly using the graph cut method of Greig et al. [15].

We consider image restoration with multiple labels where the problem becomes NP hard. We use the iterative α-$expansion$ method in [8] which is guaranteed to find a

TABLE 1

method	input: Diamond, 210 labels							input: Bell Quad, 244 labels					
	35x35	50x50	70x70	100x100	141x141	200x200	282x282	44x44	62x62	87x87	125x125	176x176	250x250
DINIC	0.39	0.77	3.42	4.19	13.85	43.00	136.76	1.32	4.97	13.49	37.81	101.39	259.19
H_PRF	0.17	0.34	1.16	1.68	4.69	12.97	32.74	0.31	0.72	1.72	3.85	8.24	18.69
Q_PRF	0.16	0.35	1.24	1.70	5.14	14.09	40.83	0.20	1.00	1.70	4.31	10.65	25.04
Our	0.16	0.20	0.71	0.74	2.21	4.49	12.14	0.19	0.48	0.98	2.11	4.84	10.47

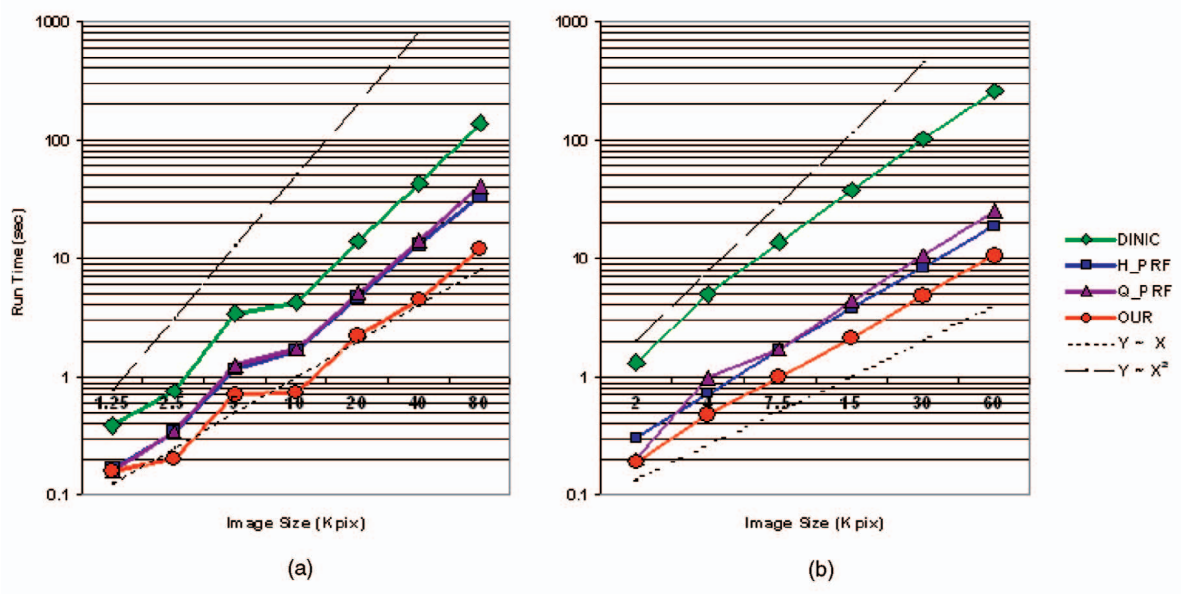


Fig. 6. Running times for the α-expansion algorithm [8]. The results are obtained in the context of image restoration with the Potts model (see Section 4.2.1). In two examples (a) and (b), we fixed the number of allowed labels but varied image size in order to estimate empirical complexities of tested min-cut/max-flow algorithms. Images of smaller size were obtained by subsampling. Our running time plots are presented in logarithmic scale. Note that empirical complexities of each algorithm can be estimated from slopes of each plot. Dashed lines provide reference slopes for linear and quadratic growth. All max-flow/min-cut algorithms gave near-linear (with respect to image size) performance in these experiments. (a) *Diamond*, 210 labels. (b) *Bell Quad*, 244 labels.

solution within a factor of two from the global minimum of the Potts energy. At a given iteration, [8] allows any subset of pixels to switch to a fixed label α. In fact, the algorithm finds an optimal subset of pixels that gives the largest decrease in the energy. The computation is done via graph cuts using some generalization of the basic graph structure in [15] (see Fig. 2). The algorithm repeatedly cycles through all possible labels α until no further improvement is possible.

In Table 1, the running times (in seconds, 1.4 GHz Pentium IV) when different max-flow/min-cut algorithms are employed in the basic step of each α-expansion. Each table corresponds to one of the original images shown in Fig. 5. The number of allowed labels is 210 (*Diamond*) and 244 (*Bell Quad*), correspondingly. We run the algorithms on images at different resolutions. At each column, we state the exact size (H x W) in pixels. Note that the total number of pixels increases by a factor of two from left to right. See Fig. 6 for logarithmic scale plots.

Note that the running times above correspond to the end of the first cycle of the α-expansion method in [8] when all labels were expanded once. The relative speeds of different max-flow/min-cut algorithms do not change much when the energy minimization is run to convergence. The number of cycles it takes to converge can vary from 1 to 3 for different resolutions/images. Thus, the running times to convergence are hard to compare between the columns and we do not present them. In fact, restoration results are quite good even after the first iteration. In most cases, additional iterations do not improve the actual output much. Fig. 5a shows the result of the Potts model restoration of the *Diamond* image (100×100) after the first cycle of iterations.

4.2.2 Linear Interaction Energy

Here, we consider image restoration with "linear" interaction energy. Fig. 5c shows one restoration result that we obtained in our experiments with this energy. The linear interaction energy can be written as

$$E(I) = \sum_{p \in \mathcal{P}} ||I_p - I_p^o|| + \sum_{(p,q) \in \mathcal{N}} A_{(p,q)} \cdot |I_p - I_q|, \qquad (3)$$

where constants $A_{(p,q)}$ describe the relative importance of interactions between neighboring pixels p and q. If the set of labels is finite and ordered, then this energy can be minimized exactly using either of the two almost identical graph-based methods developed in [18], [4]. In fact, these methods use graphs that are very similar to the one introduced by [32], [31] in the context of multicamera stereo. The graphs are constructed by consecutively connecting multiple layers of image-grids. Each layer corresponds to one label. The two terminals are connected only to the first and the last layers. Note that the topological structure of these graphs is noticeably different from the Potts model graphs, especially when the number of labels (layers) is large.

TABLE 2

method	**input:** Diamond, 54 labels						**input:** Bell Quad, 32 labels					
	35x35	50x50	70x70	100x100	141x141	200x200	44x44	62x62	87x87	125x125	176x176	250x250
DINIC	1.34	4.13	8.86	18.25	34.84	57.09	0.55	1.25	2.77	6.89	15.69	31.91
H_PRF	0.47	1.30	3.03	7.48	17.53	43.58	0.48	1.25	2.75	7.42	17.69	38.81
Q_PRF	0.55	1.16	3.05	6.50	12.77	22.48	0.27	0.56	1.55	2.39	6.78	10.36
Our	0.17	0.33	0.63	1.41	2.88	5.98	0.13	0.27	0.52	1.09	2.33	4.84

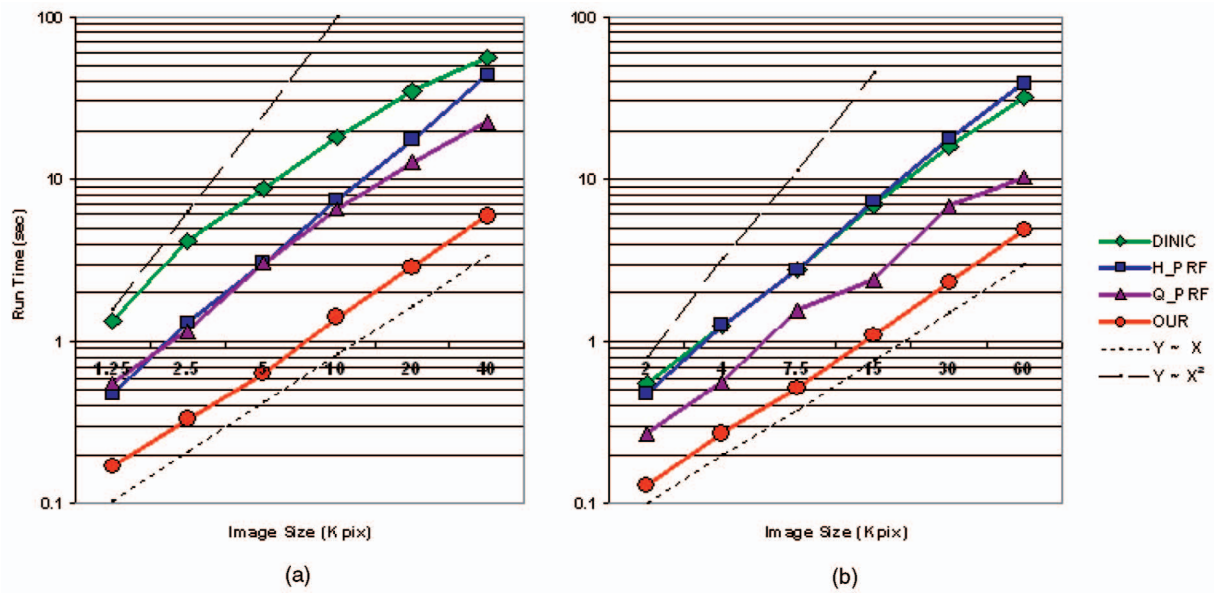


Fig. 7. Running times for "multilayered" graphs (e.g., [31], [19]). as functions of image size. The results results are obtained in the context of image restoration with linear interaction potentials (see Section 4.2.2). Here, we fixed the number of allowed labels (graph layers) and tested the empirical complexities of min-cut/max-flow algorithms with respect to image size. Images of smaller size were obtained by subsampling. The running time plots are presented in logarithmic scale where the empirical complexities of algorithms can be estimated from slopes of each plot. Dashed lines provide references for linear and quadratic growth slopes. All max-flow/min-cut algorithms gave near-linear (with respect to image size) performance in these experiments. (a) *Diamond*, 54 labels. (b) *Bell Quad*, 32 labels.

Table 2 shows the running times (in seconds on 1.4 GHz, Pentium IV) that different min-cut/max-flow algorithms took to compute the exact minimum of the linear interactions energy (3). We used the same *Diamond* and *Bell Quad* images as in the Potts energy tests. We run the algorithms on images at different resolution. At each column, we state the exact size (height and width) in pixels. Note that the total number of pixels increases by a factor of two from left to right. Also, see Figs. 7a and 7b for logarithmic scale plots.

The structure of linear interaction graph directly depends on the number of labels.[5] In fact, if there are only two labels then the graph is identical to the Potts model graph. However, both, size and topological properties of the linear interaction graphs change as the number of labels (layers) gets larger and larger. In Table 3, we compare the running times of the algorithms for various numbers of allowed labels (layers). We consider the same two images, *Diamond* and *Bell Quad*. In each case, the size of the corresponding image is fixed. At each column, we state the number of allowed labels $\mathcal{L}$. The number of labels increases by a factor of two from left to right. See Figs. 8a and 8b for logarithmic scale plots.

5. Note that, in Section 4.2.1, we tested the multilabel Potts energy minimization algorithm [8] where the number of labels affects the number of iterations but has no effect on the graph structures.

Our experiments with linear interaction graphs show that most of the tested max-flow/min-cut algorithms are close to linear both with respect to increase in image size and in the number of labels. At the same time, none of the algorithms behaved linearly with respect to the number of labels despite the fact that the size of graphs linearly depends on the number of labels. Our algorithm is a winner in absolute speed as, in most of the tests, it is 2-4 times faster than the second best method. However, our algorithm's dynamics with respect to increase in the number of labels is not favorable. For example, Q_PRF gets very close to the speed of our method in case of $\mathcal{L} = 250$ (*Bell Quad*) even though our algorithm was two times faster than Q_PRF when the number of labels was $\mathcal{L} = 32$.

4.3 Stereo

Stereo is another classical vision problem where graph-based energy minimization methods have been successfully applied. The goal of stereo is to compute the correspondence between pixels of two or more images of the same scene obtained by cameras with slightly different view points. We consider three graph-based methods for solving this problem: pixel-labeling stereo with the Potts model [4], [8], stereo with occlusions [23], and multicamera scene reconstruction [24]. Note that the last method is designed

TABLE 3

method	**input:** Diamond, 100x100 (pix)				**input:** Bell Quad, 125x125 (pix)			
	$\mathcal{L}$=27	$\mathcal{L}$=54	$\mathcal{L}$=108	$\mathcal{L}$=215	$\mathcal{L}$=32	$\mathcal{L}$=63	$\mathcal{L}$=125	$\mathcal{L}$=250
DINIC	6.89	18.16	50.81	166.27	6.91	17.69	46.64	102.74
H_PRF	3.05	7.38	15.50	47.49	7.47	19.30	58.14	192.39
Q_PRF	2.36	6.41	17.22	43.47	2.39	7.95	15.83	45.64
Our	0.55	1.39	4.34	16.81	1.13	2.95	10.44	41.11

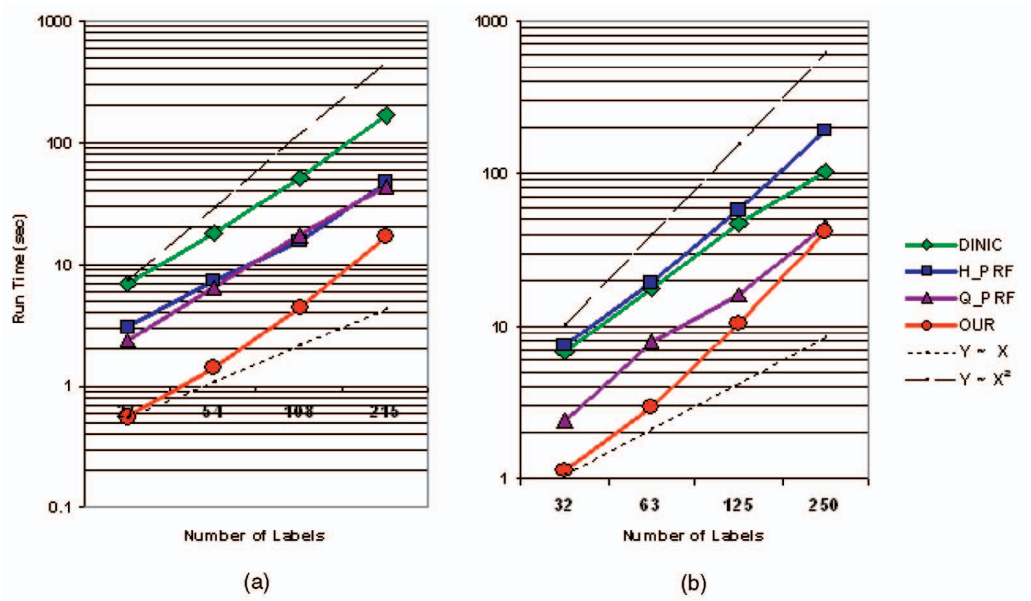


Fig. 8. Running times for "multilayered" graphs (e.g., [31], [19]) in Section 4.2.2. Here, we fixed the size of each image and tested running times with respect to growth in the number of allowed labels (graph layers). In this case, all algorithms were closer to quadratic complexity. (a) *Diamond*, 100×100 pix. (b) *Bell Quad*, 125×125 pix.

for a generalization of the stereo problem to the case of more than two cameras.

4.3.1 Pixel-Labeling Stereo with the Potts Model

First, we consider a formulation of stereo problem given in [4], [8] which is practically identical to our formulation of the restoration problem in Section 4.2.1. We seek a disparity labeling $d = \{d_p | p \in \mathcal{P}\}$ which minimizes the energy

$$E(d) = \sum_{p \in \mathcal{P}} D(p, d_p) + \sum_{(p,q) \in \mathcal{N}} K_{(p,q)} \cdot T(d_p \neq d_q), \qquad (4)$$

where d_p is a disparity label of pixel p in the left image, and $D(p, d)$ is a penalty for assigning a label d to a pixel p (the squared difference in intensities between corresponding pixels in the left and in the right images). We use the same iterative α-expansion method from [8] as in the restoration section above.

The tests were done on three stereo examples shown in Fig. 9. We used the *Head* pair from the University of Tsukuba and the well-known *Tree* pair from SRI. To diversify our tests, we compared the speed of algorithms on a *Random* pair where the left and the right images did not correspond to the same scene (they were taken from the *Head* and the *Tree* pairs, respectively).

Running times for the stereo examples in Fig. 9 are shown in seconds (450 MHz UltraSPARC II Processor) in Table 4. As in the restoration section, the running times correspond to the first cycle of the algorithm. The relative performance of different max-flow/min-cut algorithms is very similar when the energy minimization is run to convergence, while the number of cycles it takes to converge varies between three and five for different data sets. We performed two sets of experiments: one with a four-neighborhood system and the other with an eight-neighborhood system. The corresponding running times are marked by "N4" and "N8." The disparity maps at convergence are shown in Figs. 9b, 9e, and 9h. The convergence results are slightly better than the results after the first cycle of iterations. We obtained very similar disparity maps in the N4 and N8 cases.

4.3.2 Stereo with Occlusions

Any stereo images of multidepth objects contain occluded pixels. The presence of occlusions adds significant technical difficulties to the problem of stereo as the space of solutions needs to be constrained in a very intricate fashion. Most stereo techniques ignore the issue to make the problem tractable. Inevitably, such simplification can generate errors that range from minor inconsistencies to major misinterpretation of the scene geometry. Recently, [1] reported some progress in solving stereo with occlusions. Ishikawa and Geiger [17] were first to suggest a graph-cut-based solution for stereo that elegantly handles occlusions assuming monotonicity constraint.

Here, we consider a more recent graph-based formulation of stereo [23] that takes occlusions into consideration

Fig. 9. (a) Left image of *Head* pair. (b) Potts model stereo. (c) Stereo with occlusions. Disparity maps obtained for the *Head* pair. (d) Left image of *Tree* pair. (e) Potts model stereo. (f) Stereo with occlusions. Disparity maps obtained for the *Tree* pair. (g) *Random* pair. (h) Potts model stereo. (i) Stereo with occlusions. Disparity maps obtained for the *Random* pair. Stereo results. The sizes of images are 384×288 in (a), (b), and (c). 256×233 in (d), (e), and (f). 384×288 in (g), (h), and (i). The results in (c), (f), and (i) show occluded pixels in red.

TABLE 4

method	*Head*, 384x288 (pix)		*Tree*, 256x233 (pix)		*Random*, 384x288 (pix)	
	N4	N8	N4	N8	N4	N8
DINIC	104.18	151.32	9.53	19.80	105.93	167.16
H_PRF	12.00	18.03	1.65	2.86	14.25	18.22
Q_PRF	10.40	14.69	2.13	3.33	12.05	15.64
Our	3.41	6.47	0.68	1.42	3.50	6.87

TABLE 5

method	*Head*, 384x288 (pix)		*Tree*, 256x233 (pix)		*Random*, 384x288 (pix)	
	N4	N8	N4	N8	N4	N8
DINIC	376.70	370.94	66.19	102.60	81.70	115.58
H_PRF	35.65	49.81	9.07	15.41	5.48	8.32
Q_PRF	33.12	44.86	8.55	13.64	9.36	14.02
Our	10.64	19.14	2.73	5.51	3.61	6.42

Fig. 10. Multicamera reconstruction results. There are five images of size 384×288 in (a), eight images of size 352×240 in (c), and five images of size 384×256 in (e). (a) Middle image of *Head* data set. (b) Scene reconstruction for *Head* data set. (c) Middle image of *Garden* sequence. (d) Scene reconstruction for *Garden* sequence. (e) Middle image of *Dayton* sequence. (f) Scene reconstruction for *Dayton* sequence.

without making extra assumptions about scene geometry. The problem is formulated as a labeling problem. We want to assign a binary label (0 or 1) to each pair $\langle p, q \rangle$, where p is a pixel in the left image and q is a pixel in the right image that can potentially correspond to p. The set of pairs with the label 1 describes the correspondence between the images.

The energy of configuration f is given by

$$E(f) = \sum_{f_{\langle p,q \rangle}=1} D_{\langle p,q \rangle} + \sum_{p \in \mathcal{P}} C_p \cdot T(p \text{ is occluded in the configuration } f) + \sum_{\{\langle p,q \rangle, \langle p,q' \rangle\} \in \mathcal{N}} K_{\{\langle p,q \rangle, \langle p,q' \rangle\}} \cdot T(f_{\langle p,q \rangle} \neq f_{\langle p,q' \rangle}).$$

The first term is the data term, the second is the occlusion penalty, and the third is the smoothness term. $\mathcal{P}$ is the set of pixels in both images and $\mathcal{N}$ is the neighboring system consisting of tuples of neighboring pairs $\{\langle p, q \rangle, \langle p, q' \rangle\}$ having the same disparity (parallel pairs). Kolmogorov and Zabih [23] give an approximate algorithm minimizing this energy among all feasible configurations f. In contrast to other energy minimization methods, nodes of the graph constructed in [23] represent *pairs* rather than pixels or voxels.

We used the same three data sets as in the previous section. Running times for these stereo examples in Fig. 9 are shown in seconds (450 MHz UltraSPARC II Processor) in Table 5. The times are for the first cycle of the algorithm. Algorithm results after convergence are shown in Figs. 9c, 9f, and 9i.

4.3.3 Multicamera Scene Reconstruction

In this section, we consider a graph cuts-based algorithm for reconstructing a shape of an object taken by several cameras [24].

Suppose we are given n calibrated images of the same scene taken from different viewpoints (or at different moments of time). Let $\mathcal{P}_i$ be the set of pixels in the camera i and let $\mathcal{P} = \mathcal{P}_1 \cup \ldots \cup \mathcal{P}_n$ be the set of all pixels. A pixel $p \in \mathcal{P}$ corresponds to a ray in 3D-space. Consider the point of the first intersection of this ray with an object in the scene. Our goal is to find the depth of this point for all pixels in all images. Thus, we want to find a labeling $f : \mathcal{P} \rightarrow \mathcal{L}$, where $\mathcal{L}$ is a discrete set of labels corresponding to different depths. We tested the algorithm for image sequences with labels corresponding to parallel planes in 3D-space.

A pair $\langle (p, l) \rangle$, where $p \in \mathcal{P}$, $l \in \mathcal{L}$, corresponds to some point in 3D-space. We will refer to such pairs as *3D-points*. The set of interactions I will consist of (unordered) pairs of 3D-points with the same label $\langle (p_1, l) \rangle$, $\langle (p_2, l) \rangle$ "close" to each other in 3D-space.

We minimize the energy function consisting of three terms:

$$E(f) = E_{data}(f) + E_{smoothness}(f) + E_{visibility}(f). \quad (5)$$

The data term imposes photoconsistency. It is

$$E_{data}(f) = \sum_{\langle (p,f(p)) \rangle, \langle (q,f(q)) \rangle \in I} D(p, q),$$

where $D(p, q)$ is a nonpositive value depending on intensities of pixels p and q (for example, $D(p, q) = min\{0, (Intensity(p) - Intensity(q))^2 - K\}$ for some constant $K > 0$).

The smoothness term is the sum of Potts energy terms over all cameras. The visibility term is infinity if a

TABLE 6

method	**input sequence**		
	Head, 5 views 384x288	*Garden*, 8 views 352x240	*Dayton*, 5 views 384x256
DINIC	2793.48	2894.85	2680.91
H_PRF	282.35	308.52	349.60
Q_PRF	292.93	296.48	266.08
Our	104.33	81.30	85.56

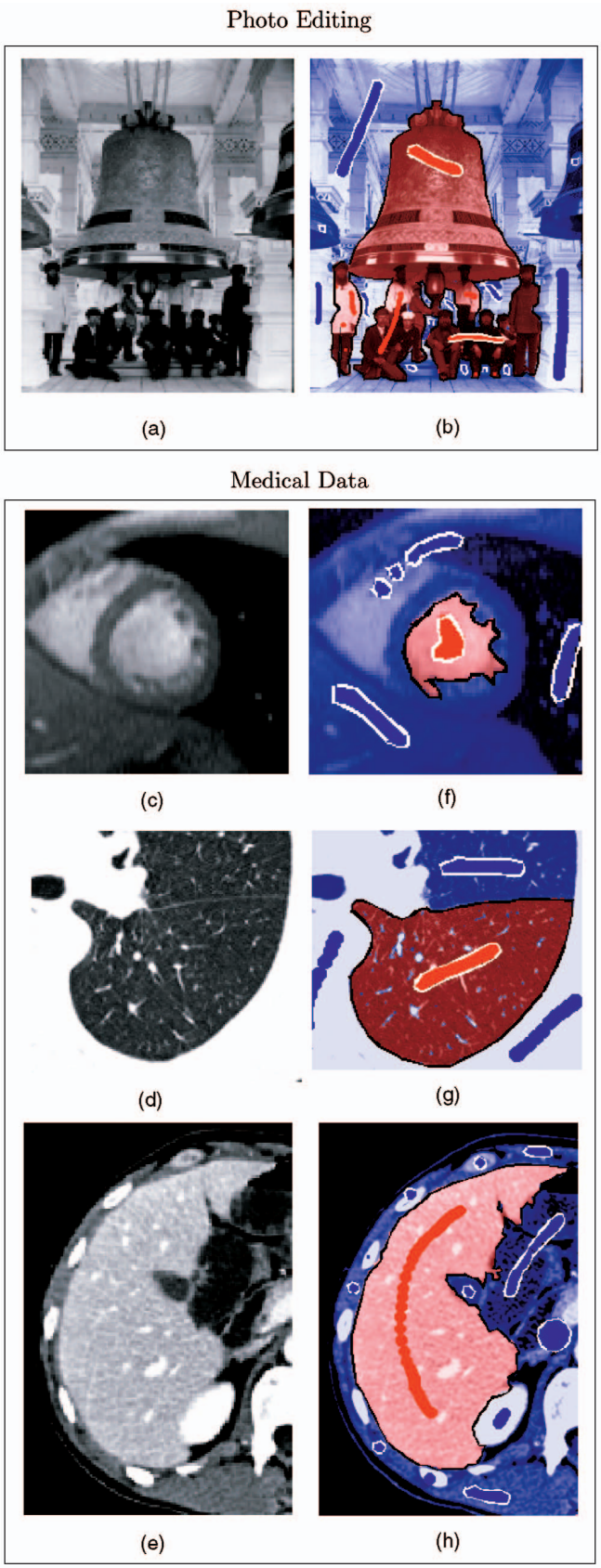


Fig. 11. Segmentation experiments. (a) Bell photo. (b) Bell segmentation. (c) Cardiac MR. (d) Lung CT. (e) Liver MR. (f) LV segment. (g) Lobe segment. (h) Liver segment.

TABLE 7

method	2D examples					
	Bell photo (255x313)		Lung CT (409x314)		Liver MR (511x511)	
	N4	N8	N4	N8	N4	N8
DINIC	2.73	3.99	2.91	3.45	6.33	22.86
H_PRF	1.27	1.86	1.00	1.22	1.94	2.59
Q_PRF	1.34	0.83	1.17	0.77	1.72	3.45
Our	0.09	0.17	0.22	0.33	0.20	0.45

3D examples					
Heart MR (127x127x12)		Heart US (76x339x38)		Kidney MR (99x66x31)	
N6	N26	N6	N26	N6	N26
20.16	39.13	172.41	443.88	3.39	8.20
1.38	2.44	18.19	47.99	0.19	0.50
1.30	3.52	23.03	45.08	0.19	0.53
0.70	2.44	13.22	90.64	0.20	0.58

configuration f violates the visibility constraint and zero otherwise. More details can be found in [24].

The tests were done for three data sets: the *Head* sequence from the University of Tsukuba, the *Garden* sequence, and the *Dayton* sequence. The middle images of these data sets are shown in Fig. 10. Table 6 gives running times (in seconds, 450 MHz UltraSPARC II Processor) for these three data sets. The times are for the first cycle of the algorithm. Algorithm results after three cycles are shown in Figs. 10b, 10d, 10f.

4.4 Segmentation

In this section, we compare the running times of the selected min-cut/max-flow algorithms in case of an object extraction technique [5] using appropriately constrained N-D grid-graphs.[6] The method in [5] can be applied to objects of interest in images or volumes of any dimension. This technique generalizes the MAP-MRF method of Greig et al. [15] by incorporating additional contextual constraints into minimization of the Potts energy

$$E(L) = \sum_{p \in \mathcal{P}} D_p(L_p) + \sum_{(p,q) \in \mathcal{N}} K_{(p,q)} \cdot T(L_p \neq L_q)$$

over binary (object/background) labelings. High-level contextual information is used to properly constrain the search space of possible solutions. In particular, some hard constraints may come directly from a user (object and background seeds). As shown in [3], graph construction in [5] can be generalized to find geodesics and minimum surfaces in Riemannian metric spaces. This result links graph-cut segmentation methods with popular geometric techniques based on level-sets [35], [29], [33], [28].

The technique in [5] finds a globally optimal binary segmentation of N-dimensional image under appropriate constraints. The computation is done in one pass of a max-flow/min-cut algorithm on a certain graph. In case of 2D images, the structure of the underlying graph is exactly the same as shown in Fig. 2. In 3D cases, [5] build a regular 3D grid graph.

We tested min-cut/max-flow algorithms on 2D and 3D segmentation examples illustrated in Fig. 11. This figure demonstrates original data and our segmentation results corresponding to some sets of seeds. Note that the user can place seeds interactively. New seeds can be added to correct segmentation imperfections. The technique in [5] efficiently

6. An earlier version of this work appeared in [6].

recomputes the optimal solution starting at the previous segmentation result.

Figs. 11a and 11b shows one of our experiments where a group of people around a bell were segmented on a real photo image (255×313 pixels). Other segmentation examples in Figs. 11c, 11d, 11e, 11f, 11g, and 11h are for 2D and 3D medical data. In Figs. 11c and 11d, we segmented a left ventricle in 3D cardiac MR data ($127 \times 127 \times 12$ voxels). In our 3D experiments, the seeds were placed in one slice in the middle of the volume. Often, this is enough to segment the whole volume correctly. The tests with lung CT data (Figs. 11e and 11f) were made in the 2D (409×314 pixels) case. The goal was to segment out a lower lung lobe. In Figs. 11g and 11h, we tested the algorithms on the 2D liver MR data (511×511 pixels). Additional 3D experiments were performed on heart ultrasound and kidney MR volumes.

Table 7 compares running times (in seconds, 1.4 GHz Pentium IV) of the selected min-cut/max-flow algorithms for a number of segmentation examples. Note that these times include only min-cut/max-flow computation.[7] In each column, we show running times of max-flow/min-cut algorithms corresponding to exactly the same set of seeds. The running times were obtained for the "6" and "26" neighborhood systems (N6 and N26). Switching from N6 to N26 increases the complexity of graphs but does not affect the quality of segmentation results much.

5 Conclusions

We tested a reasonable sample of typical vision graphs. In most examples, our new min-cut/max-flow algorithm worked 2-5 times faster than any of the other methods, including the push-relabel and the Dinic algorithms (which are known to outperform other min-cut/max-flow techniques). In some cases, the new algorithm made possible near real-time performance of the corresponding applications.

More specifically, we can conclude that our algorithm is consistently several times faster (than the second best method) in all applications where graphs are 2D grids. However, our algorithm is not a clear outperformer when the complexity of underlying graphs is increased. For example, linear interaction energy graphs (Section 4.2.2) with a large number of grid-layers (labels) is one example where Q_PRF performance was comparable to our algorithm. Similarly, experiments in Section 4.4 show that push-relabel methods (H_PRF and Q_PRF) are comparable to our algorithm in 3D segmentation tests even though it was several times faster in all 2D segmentation examples. Going from the "6" neighborhood system to the "26" system further decreased relative performance of our method in 3D segmentation.

Note that we do not have a polynomial bound for our algorithm.[8] Interestingly, in all our practical tests on 2D and 3D graphs that occur in real computer vision applications, our algorithm significantly outperformed a polynomial method of DINIC. Our results suggest that grid graphs in vision are a very specific application for min-cut/max-flow algorithms. In fact, Q_PRF outperformed H_PRF in many of our tests (especially in Section 4.2.2) despite the fact that H_PRF is generally regarded as the fastest algorithm in the combinatorial optimization community.

7. Time for entering seeds may vary between different users. For the experiments in Fig. 10, all seeds were placed within 10 to 20 seconds.

8. The trivial bound given in Section 3 involves the cost of a minimum cut and, theoretically, it is not a polynomial bound. In fact, additional experiments showed that our algorithm is, by several orders of magnitude, slower than Q_PRF, H_PRF, and DINIC on several standard (outside computer vision) types of graphs commonly used for tests in the combinatorial optimization community.

Acknowledgments

A portion of this work was done while the authors were at Siemens Research, New Jersey, and it would not have been possible without the strong support from Alok Gupta and Gareth Funka-Lea. The authors would like to thank Olga Veksler (University of Western Ontario, Canada) who provided implementations for Section 4.2. They would also like to thank Ramin Zabih (Cornell University, New York) for a number of discussions that helped to improve the paper. The anonymous reviewers gave numerous suggestions that significantly clarified presentation.

References

[1] A.F. Bobick and S.S. Intille, "Large Occlusion Stereo," *Int'l J. Computer Vision,* vol. 33, no. 3, pp. 181-200, Sept. 1999.

[2] Y. Boykov and D. Huttenlocher, "A New Bayesian Framework for Object Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. II, pp. 517-523, 1999.

[3] Y. Boykov and V. Kolmogorov, "Computing Geodesics and Minimal Surfaces via Graph Cuts," *Proc. Int'l Conf. Computer Vision,* vol. I, pp. 26-33, 2003.

[4] Y. Boykov, O. Veksler, and R. Zabih, "Markov Random Fields with Efficient Approximations," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 648-655, 1998.

[5] Y. Boykov and G. Funka-Lea, "Optimal Object Extraction via Constrained Graph-Cuts," *Int'l J. Computer Vision,* 2004. to appear.

[6] Y. Boykov and M.-P. Jolly, "*Interactive Graph Cuts* for Optimal Boundary & Region Segmentation of Objects in N-D Images," *Proc. Int'l Conf. Computer Vision,* vol. I, pp. 105-112, July 2001.

[7] Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision," *Proc. Int'l Workshop Energy Minimization Methods in Computer Vision and Pattern Recognition,* pp. 359-374, Sept. 2001.

[8] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 11, pp. 1222-1239, Nov. 2001.

[9] C. Buehler, S.J. Gortler, M.F. Cohen, and L. McMillan, "Minimal Surfaces for Stereo," *Proc. Seventh European Conf. Computer Vision,* vol. III, pp. 885-899, May 2002.

[10] B.V. Cherkassky and A.V. Goldberg, "On Implementing Push-Relabel Method for the Maximum Flow Problem," *Algorithmica,* vol. 19, pp. 390-410, 1997.

[11] W.J. Cook, W.H. Cunningham, W.R. Pulleyblank, and A. Schrijver, *Combinatorial Optimization.* John Wiley & Sons, 1998.

[12] E.A. Dinic, "Algorithm for Solution of a Problem of Maximum Flow in Networks with Power Estimation," *Soviet Math. Dokl.,* vol. 11, pp. 1277-1280, 1970.

[13] L. Ford and D. Fulkerson, *Flows in Networks.* Princeton Univ. Press, 1962.

[14] A.V. Goldberg and R.E. Tarjan, "A New Approach to the Maximum-Flow Problem," *J. ACM,* vol. 35, no. 4, pp. 921-940, Oct. 1988.

[15] D. Greig, B. Porteous, and A. Seheult, "Exact Maximum A Posteriori Estimation for Binary Images," *J. Royal Statistical Soc., Series B,* vol. 51, no. 2, pp. 271-279, 1989.

[16] M. R. Henzinger, P. Klein, S. Rao, and S. Subramanian, "Faster Shortest-Path Algorithms for Planar Graphs," *J. Computer and System Sciences,* vol. 55, pp. 3-23, 1997.

[17] H. Ishikawa and D. Geiger, "Occlusions, Discontinuities, and Epipolar Lines in Stereo," *Proc. Fifth European Conf. Computer Vision,* pp. 232-248, 1998.

[18] H. Ishikawa and D. Geiger, "Segmentation by Grouping Junctions," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 125-131, 1998.

[19] H. Ishikawa, "Exact Optimization for Markov Random Fields with Convex Priors," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 10, pp. 1333-1336, Oct. 2003.
[20] D.R. Karger, "Random Sampling in Cut, Flow, and Network Design Problems," *Math. Operations Research,* vol. 24, no. 2, pp. 383-413, May 1999.
[21] J. Kim, J.W. Fisher III, A. Tsai, C. Wible, A.S. Willsky, and W.M. Wells III, "Incorporating Spatial Priors into an Information Theoretic Approach for *f* MRI Data Analysis," *Medical Image Computing and Computer-Assisted Intervention,* pp. 62-71, 2000.
[22] V. Kolmogorov, " Graph-Based Algorithms for Multi-Camera Reconstruction Problem," PhD thesis, Computer Science Dept., Cornell Univ., 2003.
[23] V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions via Graph Cuts," *Proc. Int'l Conf. Computer Vision,* July 2001.
[24] V. Kolmogorov and R. Zabih, "Multi-Camera Scene Reconstruction via Graph Cuts," *Proc. Seventh European Conf. Computer Vision,* vol. III, pp. 82-96, May 2002.
[25] V. Kolmogorov and R. Zabih, "What Energy Functions Can Be Minimized via Graph Cuts?" *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 2, pp. 147-159, Feb. 2004.
[26] V. Kwatra, A. Schodl, I. Essa, and A. Bobick, "Graphcut Textures: Image and Video Synthesis Using Graph Cuts," *Proc. SIGGRAPH,* July 2003.
[27] G. Miller and J. Naor, "Flows in Planar Graphs with Multiple Sources and Sinks," *Proc. 30th IEEE Symp. Foundations of Computer Science,* pp. 112-117, 1991.
[28] S. Osher and N. Paragios, *Geometric Level Set Methods in Imaging, Vision, and Graphics.* Springer Verlag 2003.
[29] S. J. Osher and R.P. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces.* Springer Verlag, 2002.
[30] S. Roy and V. Govindu, "MRF Solutions for Probabilistic Optical Flow Formulations," *Proc. Int'l Conf. Pattern Recognition,* Sept. 2000.
[31] S. Roy, "Stereo without Epipolar Lines: A Maximum-Flow Formulation," *Int'l J. Computer Vision,* vol. 34, nos. 2/3, pp. 147-162, Aug. 1999.
[32] S. Roy and I. Cox, "A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem," *IEEE Proc. Int'l Conf. Computer Vision,* pp. 492-499, 1998.
[33] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis.* Cambridge Univ. Press, 2001.
[34] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision,* 2002.
[35] J.A. Sethian, *Level Set Methods and Fast Marching Methods.* Cambridge Univ. Press, 1999.
[36] D. Snow, P. Viola, and R. Zabih, "Exact Voxel Occupancy with Graph Cuts," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 345-352, 2000.
[37] R. Szeliski and R. Zabih, "An Experimental Comparison of Stereo Algorithms," *Proc. Vision Algorithms: Theory and Practice,* pp. 1-19, Sept. 1999.
[38] B. Thirion, B. Bascle, V. Ramesh, and N. Navab, "Fusion of Color, Shading and Boundary Information for Factory Pipe Segmentation," *IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 349-356, 2000.
[39] O. Veksler, "Image Segmentation by Nested Cuts," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 339-344, 2000.
[40] Z. Wu and R. Leahy, "An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 11, pp. 1101-1113, Nov. 1993.

Yuri Boykov received the "Diploma of High Education" with honors from the Moscow Institute of Physics and Technology (Department of Radio Engineering and Cybernetics) in 1992 and completed the PhD degree in the Department of Operations Research at Cornell University, Ithaca, New York, in 1996. He first became interested in combinatorial approach to generic problems in low-level vision while he was a postdoctoral researcher in the Computer Science Department at Cornell. As a scientist at Siemens Research, Princeton, New Jersey, he developed a powerful graph-cuts methodology for context extraction in volumetric imagery that, in particular, works well in many medical applications. Currently, he is an assistant professor in the Department of Computer Science at the University of Western Ontario, Canada. He is interested in problems of segmentation, restoration, registration, stereo, feature-based object recognition, tracking, photovideo editing, learning graph-based representation models, graph-cuts geometry, and others. He is a member of the IEEE and the IEEE Computer Society.

Vladimir Kolmogorov received the MS degree from the Moscow Institue of Physics and Technology in applied mathematics and physics in 1999 and the PhD degree in computer science from Cornell University in January 2004. He is currently a postdoctoral researcher at Microsoft Research, Cambridge, United Kingdom. His research interests are graph algorithms, stereo correspondence, image segmentation, parameter estimation, and mutual information. Two of his papers (written with Ramin Zabih) received a best paper award at the European Conference on Computer Vision, 2002. He is a member of the IEEE and the IEEE Computer Society.