Efficient Energy Minimization for Enforcing Label Statistics

Yongsub Lim, Kyomin Jung, and Pushmeet Kohli

Abstract—Energy minimization algorithms, such as graph cuts, enable the computation of the MAP solution under certain probabilistic models such as Markov random fields. However, for many computer vision problems, the MAP solution under the model is not the ground truth solution. In many problem scenarios, the system has access to certain statistics of the ground truth. For instance, in image segmentation, the area and boundary length of the object may be known. In these cases, we want to estimate the most probable solution that is consistent with such statistics, i.e., satisfies certain equality or inequality constraints. The above constrained energy minimization problem is NP-hard in general, and is usually solved using Linear Programming formulations, which relax the integrality constraints. This paper proposes a novel method that directly finds the discrete approximate solution of such problems by maximizing the corresponding Lagrangian dual. This method can be applied to any constrained energy minimization problem whose unconstrained version is polynomial time solvable, and can handle multiple, equality or inequality, and linear or non-linear constraints. One important advantage of our method is the ability to handle second order constraints with both-side inequalities with a weak restriction, not trivial in the relaxation based methods, and show that the restriction does not affect the accuracy in our cases.We demonstrate the efficacy of our method on the foreground/background image segmentation problem, and show that it produces impressive segmentation results with less error, and runs more than 20 times faster than the state-of-the-art LP relaxation based approaches.

Index Terms—Computer vision, image segmentation, energy minimization, Markov random fields

1 Introduction

Energy minimization has become a key tool for various computer vision problems such as image segmentation [1], [2], 3D reconstruction [3], [4], and stereo [5]. It is equivalent to performing Maximum A Posteriori (MAP) inference in Markov Random Fields (MRF), and in general, is NP-hard [6]. However, there are subclasses of energy functions such as submodular functions for which inference can be solved in polynomial time [6], [7], [8], [9]. Also many approximation algorithms based on belief propagation [10], tree reweighted message passing [11], and graph cuts [12], [13] have been studied.

Although sophisticated energy minimization algorithms have enabled researchers to compute the exact MAP solution under many probabilistic models, they do not provide the desired solution of image labelling problems. The reason for this is that most probabilistic models used in computer vision such as MRFs are mis-specified, i.e., the most probable solution under the model is not the ground truth (desired solution) [5], [14]. This has led researchers to consider more sophisticated models [8], [15], but these too have not been shown to be free from mis-specification.

For many vision problems, the system may have access to certain statistics of the ground truth. For instance, in the case of the foreground/background image segmentation problem, the area and boundary length of the object to be segmented may be known. Similarly, for 3D voxel segmentation, we may know the volume and surface area of the object. In these scenarios, researchers require computation of the most probable solution under the model that is consistent with such statistics, or in other words, the solution satisfies certain equality or inequality constraints. Note that the inequality constraint allows researchers, for instance, to find the most probable segmentation that is bigger or smaller than a specified value in the image segmentation problem. While imposing such constraints in the optimization problem results in an improved solution for the image labelling problem, the resulting optimization problem becomes NP-hard in general.

The constrained optimization problem (MAP inference problem) is usually solved using Linear Programming formulations that relax the integrality constraints [16], [17]. However, this approach suffers from a number of drawbacks. First, the size of the resulting Linear Program (LP) can be very large, making the algorithm very slow. Second, rounding error introduced while transforming the fractional solution of the LP to a discrete solution can be arbitrarily large and may result in solutions having a high energy value. Lastly, although LP relaxation based methods can deal with linear constraints, they cannot handle powerful second order or higher order constraints such as the boundary length constraint (Br) described in Section 5.

In this paper, we propose a novel method to solve constrained energy minimization problems that can handle linear and higher order constraints. In contrast to previous methods that employ a continuous relaxation of the problem, we directly exploit its discrete structure. We consider the Lagrangian dual of the primal constrained problem, and directly compute a discrete solution corresponding to the dual maximum. It can be shown that there exist some values of the parameters of the constraints for which the returned solution is optimal (see Lemma 1). Hence, due to such a minimal optimality guarantee, solutions returned by our method can be more accurate than those by LP relaxation based methods with rounding. Through extensive experiments on the image segmentation problem, we demonstrate that the proposed method produces solutions with less error, and runs more than 20 times faster compared with state-of-the-art continuous relaxation based methods. Although we show the efficacy of our method via the image segmentation problem, the method is not limited to a particular problem and can be applied to various constrained energy minimization problems in computer vision and machine learning.

2 Related Work

Our work has been inspired by a number of recent studies on enforcing statistics during inference [18], [19], [20], [21], [22], [23]. These methods can be divided into two broad categories on the basis of how they enforce constraints: methods that enforce statistics softly by incorporating global potentials in probabilistic models, and methods that use statistics as hard constraints during inference.

Many studies have considered the problem of performing inference in probabilistic models that contain global potentials that encourage solutions to have a particular distribution of labels [21], [22], [23]. Werner [21] and Kohli et al. [8] proposed inference algorithms that can handle a global potential that encourages a specific number of variables to be assigned a particular label. This potential can be used for encouraging foreground segmentations to be of a particular size, i.e., to cover a specific number of pixels. Woodford et al. [22] considered histogram distribution preserving potentials for image labelling problems such as image denoising and texture synthesis and proposed a number of sophisticated methods for performing inference in models containing such potentials.

The methods discussed above enforce statistics softly, and there is no guarantee that their solution would have statistics that match the desired statistics. In fact the statistics can be very different from the desired statistics. A number of inference algorithms that ensure

- *Y. Lim is with the Department of Computer Science, KAIST, Daejeon, Korea. E-mail: yongsub@kaist.ac.kr.*
- *K. Jung is with the Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea. E-mail: kjung@snu.ac.kr.*
- *P. Kohli is with the Machine Learning and Perception Department, Microsoft Research, Cambridge, United Kingdom. E-mail: pkohli@microsoft.com.*

Manuscript received 28 May 2013; revised 20 Jan. 2014; accepted 28 Jan. 2014. Date of publication 13 Feb. 2014; date of current version 29 July 2014. Recommended for acceptance by C.H. Lampert.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2014.2306415

0162-8828 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

that solutions have the desired statistics have been proposed. One of the first constraints used for image labelling problems was the silhouette constraint, used for the problem of 3D reconstruction [3], [24]. This constraint ensures that a ray emanating from any silhouette pixel must pass through at least one voxel that belongs to the 'object.' Algorithms that ensure topological constraints such as connectivity of the object segmentation [25] or that the boundary of the object segmentation is close to the sides of a bounding box [16] have also been proposed.

Convex relaxation has been widely used to solve discrete optimization problems. Ravikumar and Lafferty [26] proposed a convex quadratic relaxation (QR) for MAP inference, and it was shown to be more accurate than LP relaxation and propagation based methods. Another approach that is closely related to the present work is the method of Klodt and Cremers [18], which produces impressive results for labelling problems such as image segmentation. This method works by solving a continuous relaxation of the constrained energy minimization problem and rounding its solution to obtain a discrete solution. Although their methods can handle multiple constraints, it is not trivial to deal with constraints on second order statistics, such as the boundary length constraint introduced in Section 5. This is because their approach requires convexity of the final problem. Even applying our method on top of theirs would result in an uncompetitive algorithm as our method would require multiple executions of their algorithm and even a single execution of their algorithm takes much more time than our full method. Furthermore, the method proposed in [18] also differs from our method in the fact that it needs a explicit rounding step to arrive at a valid discrete solution, whereas our method uses the oracle to generate the discrete solution.

We now discuss the most relevant methods to the present work. Kolmogorov et al. [19] and Lim et al. [20] considered the problem of minimizing a submodular energy function under the 'label count' constraint. For the foreground/background segmentation problem, the label count constraint ensures that a specific number of pixels take the foreground label. Both methods use the parametric mincut algorithm for guiding their search for a lowest energy solution that satisfies the constraint. The results of these methods have been shown to improve the accuracy under Hamming error with respect to the ground truth [19], [20]. However, both methods suffer from the limitation that they can handle only one constraint at a time. In contrast, our method can handle multiple, equality or inequality, and linear or non-linear constraints. Especially, our method can be understood as a generalization of [19] to multiple constraints. This extension also allows our method to be used for sampling techniques, e.g., to obtain various segmentation samples in image segmentation [27].

3 Setup and Preliminaries

This section provides the notations and definitions used in the paper. It also provides our formulation for the constrained energy minimization problem and explains the optimality certificate associated with the solutions of our method.

3.1 Energy Minimization

Many computer vision problems are formulated using a random field model defined over a graph $G = (V, E)$. The energy of the random field is the negative log of the posterior probability distribution. For image labelling problems with two labels, such as image segmentation and volumetric segmentation, the energy has the following form:

$$E(\mathbf{x}) = \sum_{c \in C_G} \phi_c(\mathbf{x}_c), \tag{1}$$

where $\mathbf{x} \in \{0,1\}^n$, $\mathbf{x}_c$ is a vector having elements of $\mathbf{x}$ corresponding to a clique c, C_G is the set of cliques in G, and ϕ_c is called a potential and is defined over a clique c. Although minimizing E is NP-hard in general, if E is submodular, it becomes solvable in polynomial time. Moreover, if an energy function is defined over cliques of size up to 2 as

$$E(\mathbf{x}) = \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{ij}(x_i, x_j), \tag{2}$$

the problem can be efficiently solved by a st-mincut algorithm. Owing to this property, submodular second order energy functions are widely used in computer vision for problems such as image segmentation.

3.2 Constrained Energy Minimization

In this paper, we tackle the problem of minimizing an energy function under certain pre-specified constraints. This problem is formulated as

$$\min_{\mathbf{x} \in \{0,1\}^n} \left\{ E(\mathbf{x}) : b_i^- \le h_i(\mathbf{x}) \le b_i^+, 1 \le i \le m \right\}, \tag{3}$$

where $h_i : \{0,1\}^n \to \mathbb{R}$ and $b_i^-, b_i^+ \in \mathbb{R}$. We denote $(h_1(x), \ldots, h_m(x))$ by $H(x)$. In the context of image segmentation, h_i may correspond to prior knowledge aboutstatistics such as the size and boundary length of an object. For instance, $h_i = \sum_i x_i$ is the number of pixels that are assigned the label 1 (foreground).

Before explaining how to solve the inequality problem, we first examine the equality case as follows:

$$\min_{\mathbf{x} \in \{0,1\}^n} \{ E(\mathbf{x}) : H(\mathbf{x}) = \mathbf{b} \}. \tag{4}$$

To solve (4), we exploit the Lagrangian dual $D(\boldsymbol{\lambda})$ of (4).

$$D(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in \{0,1\}^n} L(\mathbf{x}, \boldsymbol{\lambda}), \tag{5}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^m$, and

$$L(\mathbf{x}, \boldsymbol{\lambda}) = E(\mathbf{x}) + \boldsymbol{\lambda}^T (H(\mathbf{x}) - \mathbf{b}). \tag{6}$$

Note that for any $\mathbf{b}$, maximizing $D(\boldsymbol{\lambda})$ gives a lower bound of the minimum for (4). This leads us to the following known lemma [28], which, as will become apparent later, provides us with an optimality certificate that guarantees that our solution is the lowest energy solution with its statistics.

Lemma 1. *Let $\mathbf{x}^*$ be such that $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = D(\boldsymbol{\lambda}^*)$ for some $\boldsymbol{\lambda}^*$, and $\mathbf{b}^* = H(\mathbf{x}^*)$. Then, $E(\mathbf{x}^*) = \min_{\mathbf{x} \in \{0,1\}^n} \{ E(\mathbf{x}) : H(\mathbf{x}) = \mathbf{b}^* \}$.*

Proof. Let $\mathbf{x}$ satisfy that $H(\mathbf{x}) = \mathbf{b}^*$. Then, $\boldsymbol{\lambda}^T (H(\mathbf{x}) - \mathbf{b}^*) = \boldsymbol{\lambda}^T (H(\mathbf{x}^*) - \mathbf{b}^*) = 0$ for all $\boldsymbol{\lambda} \in \mathbb{R}^m$. This implies $L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \le L(\mathbf{x}, \boldsymbol{\lambda}^*)$. Thus, from (6), $E(\mathbf{x}^*) \le E(\mathbf{x})$. □

Note that for a fixed $\mathbf{x}$, $L(\mathbf{x}, \boldsymbol{\lambda}) = z$ corresponds to an $(m+1)$-dimensional hyperplane on $(\boldsymbol{\lambda}, z) \in \mathbb{R}^{m+1}$ with $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $z \in \mathbb{R}$, and when $H(\mathbf{x}) = \mathbf{b}$, the slope of the hyperplane becomes 0. Also note that $D(\boldsymbol{\lambda})$ is defined by the pointwise minimum of a finite number of hyperplanes, that is, $D(\boldsymbol{\lambda})$ is piecewise linear concave. From these facts, if a hyperplane with slope 0 contributes to the shape of the dual $D(\boldsymbol{\lambda})$, that hyperplane corresponds to a primal optimal solution and we can find it by maximizing $D(\boldsymbol{\lambda})$. Even if there is no hyperplane of slope 0 that contributes to the shape of $D(\boldsymbol{\lambda})$, we can still compute a (primal) solution at the dual maximum. In our experimental analysis, we have observed that very often such a (primal) solution $\hat{\mathbf{x}}$ has a slope close to 0, i.e., $H(\hat{\mathbf{x}}) \approx \mathbf{b}$[1], and by

1. Recently, it was shown that the dual consists of quite a large number of hyperplanes [29]. This indicates that $H(\hat{\mathbf{x}})$ is not significantly different from $\mathbf{b}$. Hence, $H(\hat{\mathbf{x}}) - \mathbf{b} \approx 0$.

Lemma 1, it is optimal among $\mathbf{x} \in \{0,1\}^n$ such that $H(\mathbf{x}) = H(\hat{\mathbf{x}}) \approx \mathbf{b}$. Thus, it is a good approximate solution for (4). The algorithm in Section 4 essentially computes this solution when $H(\mathbf{x}) = b$ ensures that the dual is polynomial time computable for $\boldsymbol{\lambda} \in \mathbb{R}^m$.

Now we return to the inequality constrained problem. An inequality constraint can be handled as an equality one with the insertion of a *slack variable* $y_i \in \mathbb{R}$. As a first step, we rewrite problem (3) as follows:

$$\min_{\mathbf{x}}\{E(\mathbf{x}) : \mathbf{b} - \mathbf{s} \leq H(\mathbf{x}) \leq \mathbf{b}\}, \tag{7}$$

where $\mathbf{b} = \mathbf{b}^+$, and $\mathbf{s} = \mathbf{b}^+ - \mathbf{b}^-$. The corresponding equality problem then becomes

$$\min_{\mathbf{x},\mathbf{y}}\{\hat{E}(\mathbf{x},\mathbf{y}) : H(\mathbf{x}) + \mathbf{y} = \mathbf{b}\}, \tag{8}$$

where $y_i \in [0, s_i]$, and $\hat{E}(\mathbf{x},\mathbf{y}) = E(\mathbf{x})$. The following is the corresponding Lagrangian:

$$\hat{L}(\mathbf{x},\mathbf{y},\boldsymbol{\lambda}) = \hat{E}(\mathbf{x},\mathbf{y}) + \boldsymbol{\lambda}^T(H(\mathbf{x}) + \mathbf{y} - \mathbf{b}). \tag{9}$$

Let $(\mathbf{x}^*, \mathbf{y}^*)$ be a minimizer of $\hat{L}$ for some $\boldsymbol{\lambda}$; then $y_i^* = 0$ if $\lambda_i > 0$, $y_i^* = s_i$ if $\lambda_i < 0$, and y_i^* can be any number in $[0, s_i]$ if $\lambda_i = 0$. Hence, $\mathbf{y}^*$ only depends on $\boldsymbol{\lambda}$, and the dual is as follows:

$$\hat{D}(\boldsymbol{\lambda}) = \min_{\mathbf{x}}\{E(\mathbf{x}) + \boldsymbol{\lambda}^T(H(\mathbf{x}) + \mathbf{y}^*(\boldsymbol{\lambda}) - \mathbf{b})\}. \tag{10}$$

This form is the same as (5). Thus, we can handle an inequality constrained problem as the corresponding equality constrained problem. We obtain the following Lemma, which shows the optimality of our solution.

Lemma 2. *Let $(\mathbf{x}^*, \mathbf{y}^*)$ be such that $\hat{L}(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\lambda}^*) = \hat{D}(\boldsymbol{\lambda}^*)$ for some $\boldsymbol{\lambda}^*$, and $\mathbf{b}^* = H(\mathbf{x}^*) + \mathbf{y}^*$. Then $E(\mathbf{x}^*) = \min_{\mathbf{x}}\{E(\mathbf{x}) : \mathbf{b}^* - \mathbf{s} \leq H(\mathbf{x}) \leq \mathbf{b}^*\}$.*

Proof. Let $(\mathbf{x},\mathbf{y})$ satisfy $H(\mathbf{x}) + \mathbf{y} = \mathbf{b}^*$. Then, $\boldsymbol{\lambda}^T(H(\mathbf{x}) + \mathbf{y} - \mathbf{b}^*) = \boldsymbol{\lambda}^T(H(\mathbf{x}^*) + \mathbf{y}^* - \mathbf{b}^*)$ for any $\boldsymbol{\lambda} \in \mathbb{R}^m$. This implies $\hat{L}(\mathbf{x}^*, \mathbf{y}^*, \boldsymbol{\lambda}^*) \leq \hat{L}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}^*)$. Then $\hat{E}(\mathbf{x}^*, \mathbf{y}^*) \leq \hat{E}(\mathbf{x}, \mathbf{y})$, and finally we get $E(\mathbf{x}^*) \leq E(\mathbf{x})$ from the definition of $\hat{E}$. □

4 Algorithm to Maximize the Dual

We have seen that an energy minimization problem with inequalities can be solved by considering the corresponding problem with equalities. Thus, in this section we focus on solving the problem with equalities only. To this end, we use the standard cutting plane algorithm [28] for maximizing the Lagrangian dual. The cutting plane algorithm runs iteratively by computing an $(m+1)$-dimensional hyperplane consisting of the dual D one by one until finding the maximum. In the following, we assume the existence of a polynomial time oracle to compute $D(\boldsymbol{\lambda})$ for any $\boldsymbol{\lambda} \in S$, where $S = \prod_{i=1}^m [M_i^-, M_i^+]$ is a large enough set. Note that when (2) is submodular and constraints are linear, $D(\boldsymbol{\lambda})$ can be efficiently computed by the graph cut algorithm for any $\boldsymbol{\lambda} \in S$ with $M_i^- = -\infty$ and $M_i^+ = \infty$. In Section 5, we show that this holds for many useful constraints. We denote the oracle call by the function:

$$\mathcal{O}(\boldsymbol{\lambda}) = \operatorname*{argmin}_{\mathbf{x}\in\{0,1\}^n} L(\mathbf{x}, \boldsymbol{\lambda}). \tag{11}$$

As we mentioned, each $\mathbf{x} \in \{0,1\}^n$ corresponds to a hyperplane $\{(\boldsymbol{\lambda}, z) \in \mathbb{R}^{m+1} : \boldsymbol{\lambda} \in \mathbb{R}^m, z = L(\mathbf{x}, \boldsymbol{\lambda})\}$. The cutting plane algorithm works by iteratively finding $\boldsymbol{\lambda} \in S$, for which the oracle call needs to be made. Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be a set of computed solutions by the oracle call until iteration k; we then compute the maximum $(\boldsymbol{\lambda}^*, z^*) \in \mathbb{R}^{m+1}$ of the following linear programming over the variables $(\boldsymbol{\lambda}, z) \in \mathbb{R}^{m+1}$.

$$\begin{aligned} \max_{z\in\mathbb{R}} \quad & z \\ \text{subject to} \quad & z \leq L(\mathbf{x}, \boldsymbol{\lambda}), \\ & \mathbf{x} \in X \text{ and } \boldsymbol{\lambda} \in S. \end{aligned} \tag{12}$$

We let $(\boldsymbol{\lambda}^*, z^*)$ be the optimal solution of (12). We subsequently call the oracle to obtain an optimal solution $\mathbf{x}_{k+1}$ for $\boldsymbol{\lambda}^*$, that is, $\mathbf{x}_{k+1} = \mathcal{O}(\boldsymbol{\lambda}^*)$. If $z^* = L(\mathbf{x}_{k+1}, \boldsymbol{\lambda}^*)$, the algorithm terminates with the output $\mathbf{x}_{k+1}$; otherwise, it updates $X = X \cup \{\mathbf{x}_{k+1}\}$, and repeats the procedure. Note that as new hyperplanes are added over the iterations, the optimal value of (12) becomes smaller, and the condition $z^* = L(\mathbf{x}_{k+1}, \boldsymbol{\lambda}^*)$ guarantees the optimality. Algorithm 1 summarizes the whole procedure.

Algorithm 1: Cutting Plane Algorithm

Input: Oracle $\mathcal{O}$
Output: $\mathbf{x}$

$X \leftarrow \emptyset$
do
  Compute the maximum $(\boldsymbol{\lambda}^*, z^*)$ by the linear program (12).
  $\mathbf{x} \leftarrow \mathcal{O}(\boldsymbol{\lambda}^*)$, and $X \leftarrow X \cup \{\mathbf{x}\}$
while $L(\mathbf{x}, \boldsymbol{\lambda}^*) < z^*$

The running time of the algorithm is dominated by the oracle call and solving the LP in (12). Note that the number of variables, m, of (12) is small, regardless of that of the primal problem. Also, the number of constraints is at most $2m + r$, where r is the number of iterations. Note that r is bounded by the number of possible b values, which is $poly(n)$ for each constraint explained in Section 5. For example, $r \leq n$ for the size constraint (Sz), and if we consider $m > 1$ constraints, $r \leq (poly(n))^m$. Hence, the running time of our algorithm is polynomial in n for fixed m. Although the running time is exponential in the number of constraints m, this is not a problem in practice since its value is typically a small constant, and thus we treat it as bounded in our complexity analysis. We also observed that in practice, the number of iterations is at most 50 for almost all cases, thus inducing very fast running time (see Section 6).

5 Constraints for Matching Statistics

In this section, we explain how our method can be applied to labelling problems in vision such as image segmentation. The energy function for the pairwise random field model for image segmentation can be written as

$$E(\mathbf{x}) = \sum_{i\in V} \phi_i(x_i) + \sum_{(i,j)\in E} C_{ij}|x_i - x_j|, \tag{13}$$

where V is the set of all pixels, E is the set of all neighboring edges, $C_{ij} \geq 0$, and ϕ_i encodes the likelihood that each pixel i belongs to the object or background, and C_{ij} encodes the smoothness of the boundary of the object. Note that (13) is submodular since all the second order terms are submodular. Functions such as (13) can be efficiently minimized by solving an equivalent st-mincut problem [2]. In the following, we introduce some useful constraints for image segmentation, which can be handled by our method.

5.1 Size Constraint [20]

The most natural constraint for segmentation is the size (area for image segmentation, volume for 3D segmentation) of the object being segmented. This can be represented as

$$b_1^- \leq \sum_{i\in V} x_i \leq b_1^+. \tag{Sz}$$

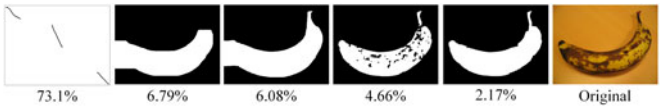

Fig. 1. Example of a sequence of segmentations computed by the cutting plane algorithm to maximize the dual. The size and boundary length constraints are considered in this figure. The segmentations shown in the figure are labeled with the percentage of erroneous pixels (inconsistent with the ground truth labels) in the segmentation. The image has a size of 640×480.

5.2 Boundary Length Constraint

The number of discontinuities in the labelling is a measure of the length of the object boundary in image segmentation, and the surface area in 3D reconstruction. These constraints can ensure that the segmentation boundary is smooth. Hence, we suggest the following boundary length constraint:

$$b_2^- \leq \sum_{(i,j)\in E} |x_i - x_j| \leq b_2^+. \tag{Br}$$

Note that $|x_i - x_j| = 1$ only when pixels i and j contribute to the boundary of an object. When we use the boundary length constraint, the search space S for the cutting plane algorithm may be restricted to ensure submodularity. Note that for some negative $\lambda_\ell \in \mathbb{R}$, where λ_ℓ is the Lagrangian multiplier for this constraint, the Lagrangian $L(\mathbf{x}, \lambda_\ell)$ is no longer submodular over $\mathbf{x}$. We can ensure submodularity by truncating the search space S such that λ_ℓ is always larger than $-\min_{i,j} C_{ij}$. This restriction appears to cause a problem in accuracy, but we observed that the solution for the parameter $\lambda_\ell = 0$ (the case of no constraint) resulted in a speckled segmentation whose boundary length was much larger than that of the ground truth in all cases. To obtain a shorter boundary length, we need to search for $\lambda_\ell \geq 0$. In other words, the maximum of the dual is found on a restricted search range, $\lambda_\ell \in [-\min_{ij} C_{ij}, \infty]$. Hence, this restriction does not affect the accuracy in our case.

5.3 Mean Constraint (Mn) [18]

The center of the object to be segmented can be easily specified by the user by drawing a circle roughly containing the object. This information can be used to define constraints on the mean horizonal and vertical coordinates of the pixels belonging to the object to be segmented. Enforcing the mean statistic involves the following constraint:

$$b_3^- \leq \frac{\sum_{i\in V} c_i x_i}{\sum_{i\in V} x_i} \leq b_3^+, \tag{Mn}$$

where $b_3^-, b_3^+ \in \mathbb{R}^2$, $c_i = (h_i, v_i)^T$, and h_i and v_i are the horizontal and vertical coordinates of a pixel i, respectively. Note that (Mn) can be rewritten using a slack variable $y \in \mathbb{R}^2$ as follows:

$$\sum_{i\in V} c_i x_i - y \sum_{i\in V} x_i = b_3^- \sum_{i\in V} x_i, \tag{14}$$

$$\sum_{i\in V} \left(c_i - b_3^- - y\right) x_i = 0, \tag{15}$$

where $y \in [0, b_{31}^+ - b_{31}^-] \times [0, b_{32}^+ - b_{32}^-]$. Note that if we consider the dual $D(\boldsymbol{\lambda})$ with the mean constraint (15), y is determined only by $\boldsymbol{\lambda}$. That is, for $i \in \{1, 2\}$, $y_i = b_{3i}^+ - b_{3i}^-$ for $\lambda_i > 0$, $y_i = 0$ for $\lambda_i < 0$, and y_i can be any number for $\lambda_i = 0$. Consequently, (15) can be handled as a linear equality constraint for our method.

5.4 Variance Constraint (Vr) [18]

If we have knowledge of the center of an object, we can also impose a variance constraint on the distance of the object pixels from the center as follows:

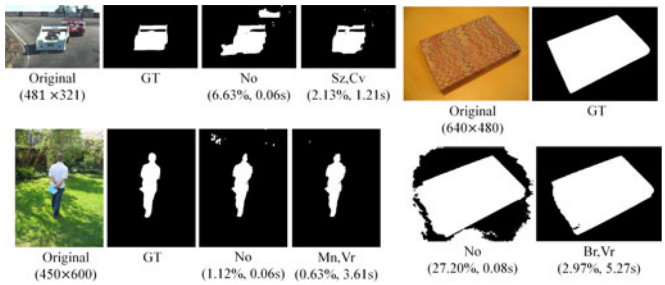

Fig. 2. Effect of constraints on the segmentation results. Each segmentation is labelled by the percentage of erroneous pixels with respect to the ground truth, and the computation time.

$$b_4^- \leq \frac{\sum_{i\in V} (c_i - \mu)^T (c_i - \mu) x_i}{\sum_{i\in V} x_i} \leq b_4^+, \tag{Vr}$$

where $b_4^-, b_4^+ \in \mathbb{R}^2$, $c_i = (h_i, v_i)^T$, h_i and v_i are horizontal and vertical coordinates of a pixel i, and $\mu \in \mathbb{R}^2$ denotes the coordinate mean of the object. This constraint can be also handled in the same manner as done for Mn.

5.5 Covariance Constraint (Cv)

Similarly, a covariance constraint can also be enforced as

$$b_5^- \leq \frac{\sum_{i\in V} (h_i - \mu_h)(v_i - \mu_v) x_i}{\sum_{i\in V} x_i} \leq b_5^+. \tag{Cv}$$

5.6 Local Size Constraint (Lsz)

This is a generalization of the size constraint. In contrast to the size constraint, in this constraint, the size of the object being segmented is locally fixed. For instance, we divide the image into 2×2 subimages, and impose the size constraint for each subimage. If we could obtain the size information locally, it would be more powerful than the size constraint. For instance, we can decompose an object having a complex boundary into parts having simpler boundaries and impose the size constraint for each part, respectively. In general, subimages are allowed to overlap each other or may not cover the whole image. When we enforce the size constraint for c number of subimages, the local size constraint can be represented as follows. For $1 \leq j \leq c$,

$$b_{6j}^- \leq \sum_{i\in V_j} x_i \leq b_{6j}^+. \tag{Lsz}$$

All the constraints above, except Br, are linear in $\mathbf{x}$. Note that for a linear constraint, we can adopt any search range S in the cutting plane algorithm since it does not break submodularity. Now, we can use any combination of these constraints with an appropriate search range. Fig. 1 shows a sequence of segmentations produced by our method for maximizing the dual. Fig. 2 shows some examples of segmentations improved by imposing constraints.

6 Experiments

We now evaluate the performance of our algorithm on the image segmentation problem using images from the GrabCut data set [30]. The first experiment checks the accuracy of the segmentation results obtained with different combinations of constraints. We also compare the results of our method with those obtained by the continuous relaxation based methods in [16], [26]. To measure the quality of segmentations, we compute

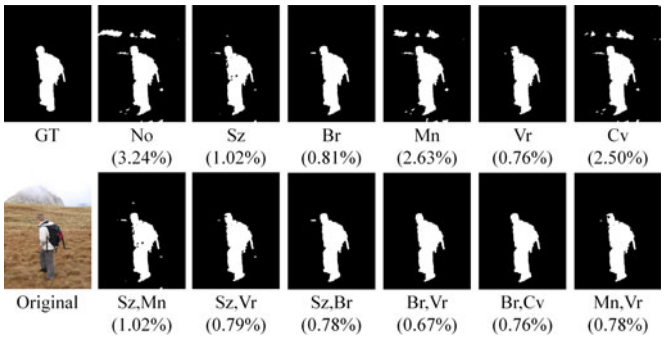

Fig. 3. Comparison of segmentations computed with different combination of constraints. The reported number is the percentage error (ER) with respect to the ground truth segmentations. The size of the image is 450×600.

the percentage of erroneously labelled pixels (with respect to the ground truth (GT)). We use ER to denote this percentage error. The experiment environment is a server with 2.8 GHZ Quad-Core Intel i7 930 and 24 G memory.

6.1 Effect of Imposing Constraints

Fig. 3 shows segmentation results corresponding to various combinations of constraints. We find that the variance constraint is quite effective: the results with constraint combinations including the variance constraint have less error compared with the other results. Imposing the boundary length constraint affected the smoothness of segmentation. For instance, the ER values of segmentations with boundary length/covariance constraints and variance constraint are nearly the same, but the former is much smoother than the latter. Table 1 shows quantitative results of our method. We assumed that statistics of the ground truth such as area, mean, and variance are known, and set inequality gaps in intervals of -10-10 percent and -5-5 percent from the ground truth values. We considered various combinations of constraints. As expected, imposition of any constraint improves the segmentation accuracy on average. From the table, we can see that the size and variance constraints are especially powerful for image segmentation. Note that the variance constraint alone results in a highly accurate segmentation. We speculate that this is because the variance constraint causes the segmentation result to be rounder, and reduces noise fragmentations. We provide more segmentation results by our method with various constraint combinations in the supplementary material.

6.2 Comparison with Continuous Relaxation

We compared the proposed method with two continuous relaxation based methods.

6.2.1 Linear Relaxation (LR) [16]

The first one is to relax the problem to a linear programming using additional variables $z \in \mathbb{R}^{|E|}$, as follows:

$$\begin{aligned} \min_{\mathbf{x}\in[0,1]^n, \mathbf{z}\in\mathbb{R}^{|E|}} \overline{E}(\mathbf{x},\mathbf{z}) &:= \boldsymbol{\varphi}'\mathbf{x} + \sum_{(i,j)\in E} C_{ij} z_{ij} \\ \text{subject to} \quad & z_{ij} \geq x_i - x_j, \text{ and} \\ & z_{ij} \geq x_j - x_i, \text{ for every } (i,j) \in E, \end{aligned} \tag{16}$$

where $\varphi_i = \phi_i(1) - \phi_i(0)$. This LP relaxation can handle all constraints described in Section 5, but cannot deal with both-side inequality or an equality boundary length constraint, because each of these breaks the convexity of the feasible region. Also, it is not reasonable to use $\sum_{(i,j)\in E} z_{ij}$ as the value of the boundary length constraint because this value is not exactly the same as $\sum_{(i,j)\in E} |x_i - x_j|$. Hence, the boundary length constraint cannot be properly handled by the LR.

6.2.2 Quadratic Relaxation [26]

For this relaxation, we first write our energy function as follows:

$$E(\mathbf{x}) = \boldsymbol{\varphi}' x_i + \sum_{(i,j)\in E} C_{ij}(x_i - x_j)^2. \tag{17}$$

Note that $(x_i - x_j)^2 = |x_i - x_j|$ since $\mathbf{x} \in \{0,1\}^n$. The relaxed problem can then be formulated as a convex quadratic programming.

$$\min_{\mathbf{x}\in[0,1]^n} \tilde{E}(x) = \mathbf{x}' H \mathbf{x} + \boldsymbol{\varphi}'\mathbf{x}, \tag{18}$$

where $H_{ij} = -C_{ij}$ for $i \neq j$, and $H_{ii} = \sum_{j:(i,j)\in E} C_{ij}$. This quadratic relaxation can handle all constraints described in Section 5, except the boundary length constraint. As in LR, both-side inequality or an equality boundary length constraint breaks the convexity of the feasible region.

6.2.3 Results

We implemented the LR and QR based methods using CPLEX Optimizer with MATLAB interface, and implemented our method by combining the CPLEX Optimizer and C++ codes. Table 2 summarizes the comparison between the proposed approach and the relaxation based methods outlined above.

For all constraint combinations, our method produces segmentations with less percentage pixel error (ER) compared to those obtained by LR and QR. Further, our method is extremely fast: 200-300 times faster than QR and 20-100 times faster than LR. Fig. 4 shows some examples of segmentation by the three methods. We provide more segmentation results to compare our method with LR and QR in the supplementary material.

TABLE 1
Percentage Pixel Error (ER) and the Running Time for Results Obtained with Our Method

		$-10 \sim 10\%$		$-5 \sim 5\%$				$-10 \sim 10\%$		$-5 \sim 5\%$	
		ER(%)	Time(s)	ER(%)	Time(s)			ER(%)	Time(s)	ER(%)	Time(s)
1	Sz,Br,Vr	2.13	6.42	1.83	7.04	12	Sz,Mn,Cv	3.27	2.81	2.73	2.95
2	Sz,Vr	2.30	2.05	2.00	2.22	13	Sz,Mn	3.29	1.64	2.72	1.84
3	Vr,Cv	2.30	2.41	2.12	2.40	14	Sz	3.49	0.65	3.10	0.69
4	Br,Vr	2.33	3.92	2.16	4.87	15	Br,Mn,Cv	6.39	10.43	6.32	11.51
5	Vr	2.41	1.27	2.28	1.37	16	Br,Cv	6.50	6.11	6.61	6.11
6	Br,Vr,Cv	2.41	7.24	2.13	6.67	17	Br,Mn	6.57	6.12	6.22	7.34
7	Mn,Vr	2.53	2.27	2.28	2.60	18	Mn,Cv	6.70	1.85	6.43	2.09
8	Sz,Br,Mn	3.04	6.46	2.52	9.28	19	Cv	6.76	0.81	6.76	0.81
9	Sz,Br,Cv	3.05	5.53	2.60	6.08	20	Mn	6.79	0.93	6.12	1.07
10	Sz,Br	3.19	3.57	2.40	3.39	21	Br	6.96	4.49	7.04	4.57
11	Sz,Cv	3.32	1.54	3.00	1.60	22	No	7.05	0.06	7.05	0.06

Each number was averaged over 14 images. Rows are sorted in ER for an inequality gap interval of -10-10 percent.

TABLE 2
Comparison of the Segmentation Errors of Our Method with Those Obtained from the Continuous Relaxation and Rounding Based Methods

	±10%						±5%					
	Our Method		LR		QR		Our Method		LR		QR	
	ER(%)	Time(s)	ER(%)	Time(s)	ER(%)	Time(s)	ER(%)	Time(s)	ER(%)	Time(s)	ER(%)	Time(s)
Sz	**3.49**	**0.65**	4.22	39.7	4.70	280	**3.10**	**0.69**	3.84	68.4	4.48	316
Mn	6.79	**0.93**	**5.90**	21.7	6.61	320	6.12	**1.07**	**5.67**	26.8	6.23	336
Vr	**2.41**	**1.27**	3.04	51.5	3.10	430	**2.28**	**1.37**	2.90	50.7	2.86	425
Sz,Mn	**3.29**	**1.64**	4.09	57.7	4.50	307	**2.72**	**1.84**	3.46	83.6	3.95	349
Sz,Vr	**2.30**	**2.05**	3.02	64.9	3.06	445	**2.00**	**2.22**	2.73	77.2	2.64	461
Mn,Vr	**2.53**	**2.27**	3.04	62.3	3.10	487	**2.28**	**2.60**	2.90	59.4	2.86	477

The inequality gaps used for this experiment were -10-10 *and* -5-5 *percent. The percentage error is averaged over six images of size* 321×481. *Bold indicates the best error and time in each row.*

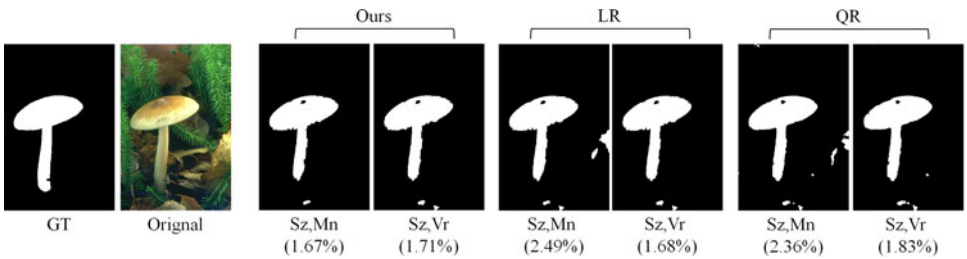

Fig. 4. Examples of segmentations computed by LR, QR, and our method. Inequality gaps for each constraint combination are -10-10 percent. The size of the image is 321×481.

Recall that our method runs by calling the oracle and solving the LP in (12) at each iteration. Since we consider the pixel grid graph, the oracle call, which involves minimizing a submodular pseudo-boolean function using the graph cut algorithm, takes very little time. The number of variables constituting the LP in (12) for our approach is small. Furthermore, for all constraint combinations in Table 2, the total number of iterations does not exceed 50 (average: 27, std.: 11.7), making the number of constraints of the LP small. The small size of the LP enables our method to run much faster than LR, which has to solves a relatively large linear programming problem.

The results of our experiments comparing the computational complexity of the three methods (ours, LR, QR) are shown in Fig. 6. We used four $1,024 \times 1,024$ images, and each image was scaled down to eight small images of size $2^{i/2} \times 2^{i/2}$, $12 \leq i \leq 19$. Each data point denotes the average running time of the specified method over four images having the same size specified on the x-axis values. It can be seen that our methods is substantially faster than the state-of-the-art LR and QR based approaches. Fig. 5 shows segmentation results yielded by our method for four original images of size $1,024 \times 1,024$.

Table 3 shows how solutions by the three methods satisfy given constraints. Here, we used two measures: $\mathrm{ER_a}$ and $\mathrm{ER_b}$. $\mathrm{ER_a}$ denotes the percentage of constraints that are unsatisfied. For instance, if the size and variance constraints are considered and one of them is not satisfied, then $\mathrm{ER_a}$ is counted as 0.5. $\mathrm{ER_b}$ represents the deviation of the constraint values from the inequality gap, thus providing a measure of the degree of satisfaction. For instance, let the inequality gap for Sz be $[a, b]$ and the size of a computed solution be c; $\mathrm{ER_b}$ is then counted by $\min(|c-a|, |c-b|)/((a+b)/2)$ if $c \notin [a, b]$ and counted by 0 if $c \in [a, b]$. We observed that solutions computed by three methods satisfy the constraints or are very close to the constraints gap in general.

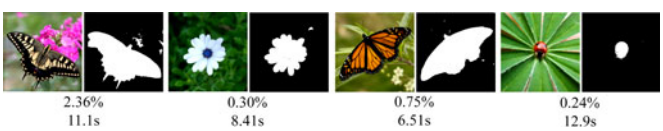

Fig. 5. Example of segmentations by our method. Each original image has a size of $1,024 \times 1,024$.

7 Discussion and Conclusions

We have proposed a novel approach to constrained energy minimization. Our method is efficient and can handle a very large class of constraints including those that enforce second order statistics, which cannot be achieved by using previously proposed methods since they usually requires the convexity of the problem [16], [18], [26]. To handle such second order constraints, the boundary length constraint in our paper, we use a restricted search region in which the dual is computable. In our cases, the dual maximum is always found in that restricted search region which means that it does not affect the accuracy of our method. Although our approach can be used with continuous relaxations based methods [16], [26], it would require repeated solutions of the relaxed problem to find

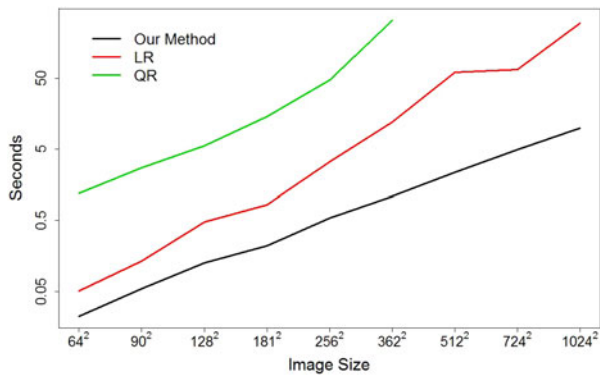

Fig. 6. Comparison of scalability between our method and two relaxation based methods.

TABLE 3
Comparison between Our Method and Two Relaxation Based Methods on Whether Computed Solutions Satisfy Constraints or Not

	Our Method		LR		QR	
	$\mathrm{ER_a}$	$\mathrm{ER_b}$	$\mathrm{ER_a}$	$\mathrm{ER_b}$	$\mathrm{ER_a}$	$\mathrm{ER_b}$
±10%	14.35	0.88	22.69	0.40	7.2	0.06
±5%	29.63	1.12	23.84	0.54	26.00	0.42

the dual maximum and would be much slower than our method as shown in Section 6. We evaluated our method on image segmentation with constraints described in Section 5. Our method produced segmentations, with less error, in much faster time compared to state-of-the-art continuous relaxation based methods.

ACKNOWLEDGMENTS

This research was in part funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2012R1A1A1014965).

REFERENCES

[1] A. Blake, C. Rother, M. Brown, P. Pérez, and P. Torr, "Interactive Image Segmentation Using an Adaptive GMMRF Model," *Proc. Eighth European Conf. Computer Vision (ECCV)*, 2004.
[2] Y. Boykov and M. Jolly, "Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images," *Proc. IEEE Eighth Int'l Conf. Computer Vision (ICCV)*, 2001.
[3] S. Sinha and M. Pollefeys, "Multi-View Reconstruction Using Photo-Consistency and Exact Silhouette Constraints: A Maximum-Flow Formulation," *Proc. IEEE 10th Int'l Conf. Computer Vision (ICCV)*, 2005.
[4] G. Vogiatzis, P. Torr, and R. Cipolla, "Multi-View Stereo via Volumetric Graph-Cuts," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005.
[5] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A Comparative Study of Energy Minimization Methods for Markov Random Fields," *Proc. European Conf. Computer Vision (ECCV)*, 2006.
[6] E. Boros and P. Hammer, "Pseudo-Boolean Optimization," *Discrete Applied Math.*, vol. 123, pp. 155-225, 2002.
[7] P. Kohli, M. Kumar, and P. Torr, "p^3 and Beyond: Solving Energies with Higher Order Cliques," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2007.
[8] P. Kohli, L. Ladicky, and P. Torr, "Robust Higher Order Potentials for Enforcing Label Consistency," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2008.
[9] V. Kolmogorov and R. Zabih, "What Energy Functions Can Be Minimized Using Graph Cuts?" *Proc. European Conf. Computer Vision (ECCV)*, 2002.
[10] J. Yedidia, W. Freeman, and Y. Weiss, "Generalized Belief Propagation," *Proc. Neural Information Processing Systems (NIPS)*, 2001.
[11] V. Kolmogorov, "Convergent Tree-Reweighted Message Passing for Energy Minimization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568-1583, Oct. 2006.
[12] 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.
[13] N. Komodakis, N. Paragios, and G. Tziritas, "MRF Optimization via Dual Decomposition: Message-Passing Revisited," *Proc. IEEE 11th Int'l Conf. Computer Vision (ICCV)*, 2007.
[14] P. Pletscher, S. Nowozin, P. Kohli, and C. Rother, "Putting Map Back on the Map," *Proc. 33rd Int'l Conf. Pattern Recognition (DAGM-Symp.)*, pp. 111-121, 2011.
[15] S. Roth and M. Black, "Fields of Experts: A Framework for Learning Image Priors," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2005.
[16] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp, "Image Segmentation with a Bounding Box Prior," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2009.
[17] S. Nowozin and C. Lampert, "Global Connectivity Potentials for Random Field Models," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2009.
[18] A. Klodt and D. Cremers, "A Convex Framework for Image Segmentation with Moment Constraints," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2011.
[19] V. Kolmogorov, Y. Boykov, and C. Rother, "Application of Parametric Maxflow in Computer Vision," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2007.
[20] Y. Lim, K. Jung, and P. Kohli, "Energy Minimization under Constraints on Label Counts," *Proc. 11th European Conf. Computer Vision (ECCV)*, 2010.
[21] T. Werner, "High-Arity Interactions, Polyhedral Relaxations, and Cutting Plane Algorithm for Soft Contraint Optimisation (MAP-MRF)," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2008.
[22] O. Woodford, C. Rother, and V. Kolmogorov, "A Global Perspective on MAP Inference for Low-Level Vision," *Proc. IEEE 12th Int'l Conf. Computer Vision (ICCV)*, 2009.
[23] P. Kohli and M.P. Kumar, "Energy Minimization for Linear Envelope MRFs," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1863-1870, 2010.
[24] K. Kolev and D. Cremers, "Integration of Multiview Stereo and Silhouettes via Convex Functionals on Convex Domains," *Proc. European Conf. Computer Vision (ECCV)*, 2008.
[25] S. Vicente, V. Kolmogorov, and C. Rother, "Graph Cut Based Image Segmentation with Connectivity Priors," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2008.
[26] P.D. Ravikumar and J.D. Lafferty, "Quadratic Programming Relaxations for Metric Labeling and Markov Random Field Map Estimation," *Proc. 23rd Int'l Conf. Machine Learning (ICML)*, 2006.
[27] J. Carreira and C. Sminchisescu, "CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1312-1328, July 2012.
[28] M. Guignard, "Lagrangean Relaxation," *TOP*, vol. 11, pp. 151-200, 2003.
[29] Y. Lim, K. Jung, and P. Kohli, "Multi-Dimensional Parametric Mincuts for Constrained Map Inference," arXiv:1307.7793, 2013..
[30] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts," *ACM Trans. Graphics*, vol. 23, pp. 309-314, 2004.