

A Compact Linear Programming Relaxation for Binary Sub-modular MRF

Junyan Wang and Sai-Kit Yeung

Singapore University of Technology and Design
{junyan_wang, saikit}@sutd.edu.sg

Abstract. Direct linear programming (LP) solution to binary sub-modular MRF energy has recently been promoted because i) the solution is identical to the solution by graph cuts, ii) LP is naturally parallelizable and iii) it is flexible in incorporation of constraints. Nevertheless, the conventional LP relaxation for MRF incurs a large number of auxiliary variables and constraints, resulting in expensive consumption in memory and computation. In this work, we propose to approximate the solution of the conventional LP at a significantly smaller complexity by solving a novel compact LP model. We further establish a tightenable approximation bound between our LP model and the conventional LP model. Our LP model is obtained by linearizing a novel l_1 -norm energy derived from the Cholesky factorization of the quadratic form of the MRF energy, and it contains significantly fewer variables and constraints compared to the conventional LP relaxation. We also show that our model is closely related to the total-variation minimization problem, and it can therefore preserve the discontinuities in the labels. The latter property is very desirable in most of the imaging and vision applications. In the experiments, our method achieves similarly satisfactory results compared to the conventional LP, yet it requires significantly smaller computation cost.

1 Introduction

Markov Random Field (MRF) has become one of the most popular models for fundamental computer vision tasks. In an MRF model, an MRF energy is minimized in order to find an optimal solution to the task. Minimizing general MRF energies is NP-hard [1], while certain types of the MRF energies can be minimized efficiently and exactly by using, for example, graph cuts [2].

Conventional LP relaxation. Recently, Bhusnurmath and Taylor [6] promoted the direct continuous linear programming (LP) solution to the binary sub-modular MRF. The LP model was obtained by linearizing the l_1 -norm pairwise potential in the binary sub-modular MRF using auxiliary variables. Bhusnurmath and Taylor proved that the solution to the continuous LP model is identical to the graph-cuts solution given the same binary MRF energy. Their work was motivated by the fact that LP algorithms, e.g. the interior point method, can be easily parallelized. This is natural, since the interior point method is based on elementary matrix operations. The GPU version of all common matrix operations can easily be found in many toolboxes, such as MATLAB and

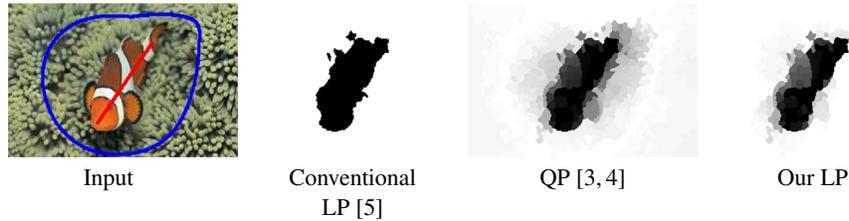


Fig. 1: Conventional LP is computationally demanding but it preserves discontinuity of the labels at the true boundary. QP [3, 4] has much lower computational complexity but often produces over-smooth labels at the boundary. Our method provides a solution sharper at object boundary at an affordable computational cost.

CULA¹. On the contrary, the parallel implementation of graph cuts is very challenging, on which consensus has not yet been reached [7, 8]. Furthermore, incorporating linear constraints into an LP model is straightforward, while this is not the case for graph cuts. Lempitsky et al. [9] also showed linear constraints can be useful to segmentation.

Motivations. As reported in [5, 6], the conventional LP relaxation contains a large number of auxiliary variables and constraints, which would cause large consumption in memory and computation. Consequently, the computation upon shared with multiple computing units may still remain expensive.

In contrast to the LP model, the computational complexity of the quadratic programming (QP) relaxation for the binary sub-modular MRF energy proposed in [3, 4] is much smaller than that required by the conventional LP model. This is largely because no auxiliary variables or constraints are required in the model. However, the QP model may produce over-smooth ambiguous labels at the desired discontinuities in the solution. For instance in object segmentation in images, this may cause incorrect segmentation. As shown in Fig. 1, the solution by conventional LP is clean and more desirable than that from QP.

Our contributions. To gain high quality solution similar to that from LP, at a computational cost similar to that of QP models, we propose a novel LP relaxation for binary sub-modular MRF to leverage both the compactness of the QP relaxation and the edge preservability of LP relaxation. Our LP relaxation is obtained by linearizing a novel l_1 -norm minimization problem that is derived from the Cholesky factorization of the QP relaxation model. We further establish a tightenable approximation bound between our LP relaxation and the conventional LP relaxation. The complexity of the resultant algorithm for solving the proposed LP problem is of the same order of the corresponding QP model, and it is significantly smaller than that of the conventional LP. In addition, the derived novel l_1 -norm minimization is strongly related to the total-variation minimization problem according to our theoretical analysis. Thus, it is able to preserve discontinuities in labels.

¹ <http://www.culatools.com/>

2 Background

2.1 The binary submodular MRF model in computer vision

In the generic MRF model for the labeling problems in computer vision, the labels in the image are formulated as an Markov random field, and the corresponding distribution is in the form of Gibbs distribution according to the HammersleyClifford theorem. The labeling task is therefore cast into an Maximum *a posterior* (MAP) problem. Due to the Gibbs distribution form, the MAP problem becomes an energy minimization problem, and the energy is often written in the following standard form:

$$E(\mathbf{x}) = \sum_{p \in \mathcal{P}} D_p(x_p) + \sum_{\{p,q\} \in \mathcal{N}} V_{pq}(x_p, x_q),$$

where \mathbf{x} is a label vector corresponding to all elements in the image, $D_p(\cdot)$ is known as the unary term, or data fidelity term, and $V_{pq}(\cdot, \cdot)$ is a pairwise potential.

Due to the fundamental works by Boykov, Olga and Zabih [2] and Komogorov and Zabih [10], it is well-known that the above energy, especially for binary label, can be solved exactly by graph cuts, as long as E is submodular. One of the most successful applications of this formulation is object segmentation [11].

More recently, approximate solution to general MRF models attracts much attention from the energy minimization community [12, 13]. We argue that a more generalizable approach for solving the binary submodular problem can make approximations to general problems easier.

2.2 Conventional LP relaxation for binary submodular MRF

In the binary submodular MRF energy, the unary term is often formulated as a term linear in the label vector. The complexity of the optimization for the MRF model only lies in the pairwise potential. The pairwise potential can be written as:

$$V_{pq}(x_p, x_q) = w_{pq}|x_p - x_q|^o, \quad (1)$$

where p and q are the indices of image elements, \mathcal{E} is a neighborhood system and o is either 1 or 2 in this paper. We will elaborate on the choice of the value of o in this paper. In the context of segmentation, w_{pq} can be defined as $w_{pq} = \frac{1}{1 + \{\|I_p - I_q\|^2\}} + c$, where I_p, I_q are the image values at the p - and q -th pixel/superpixel in the image. The first component in w_{pq} encourages discontinuous labeling at image edges, and the constant c that imposes smoothness to the resultant boundary. The constant weight in the latter part is related to the curve-shortening flow in the active contour models [14, 15].

It has been pointed out that when $o = 1$, the minimization of the binary submodular MRF energy with the above pairwise potential term in Eq. (1) is equivalent to an l_1 -norm minimization problem in [6].

Formally, we may rewrite the total pairwise potential as

$$\sum_{\{p,q\} \in \mathcal{E}} w_{pq}|x_p - x_q| = \sum_{i,j} w_{ij}^e |x_i - x_j| = \|\text{diag}(\mathbf{w}^e) \mathbf{D}\mathbf{x}\|_{l_1}, \quad (2)$$

where $w_{ij}^e = w_{pq}$ if $i = p, j = q$, and $w_{ij}^e = 0$ if $\{i, j\} \notin \mathcal{E}$, $[\text{diag}(\mathbf{w}^e)]^{N^2 \times N^2}$ is the diagonal matrix composed of \mathbf{w}^e , where \mathbf{w}^e is the vectorized $\{w_{ij}^e\}$. \mathbf{D} is an incidence matrix defined as follow:

$$[D]_{ij}^{N^2 \times N} = \begin{cases} 1, & \text{if } j = (i \bmod N) \\ -1, & \text{if } (i \bmod N, j) \in \mathcal{E} \end{cases} \quad (3)$$

The LP model of the full MRF energy can be rewritten as follows:

$$\begin{aligned} \min_x \quad & \mathbf{v}^T \mathbf{x} + \mathbf{1}^T \mathbf{y} \\ \text{s.t.} \quad & -\mathbf{y} \leq \text{diag}(\mathbf{w}^e) \mathbf{D} \mathbf{x} \leq \mathbf{y} \\ & 0 \leq \mathbf{x} \leq 1, 0 \leq \mathbf{y}. \end{aligned} \quad (4)$$

where \mathbf{v} is the weights in unary term, and the variable x_{pq} is an auxiliary variable induced by the linearization process. It is further shown in [5] and [6] that the l_1 -norm minimization problem can be solved by LP, and it is proven in [6] that the solution to the LP problem in [6] converges to either 0 or 1 without any external prodding.

A drawback of this LP formulation is that it requires a large number of auxiliary variables and constraints. Suppose that there are N elements to be labeled, then there can be *as many as* $N + N \times N$ variables and $N + 2N \times N$ linear constraints, which is the worst case. The computational complexity of LP is known as $O(n^3)$ [16] where n is the number of variables, and when n is fixed the complexity is $O(m)$ [17] where m is the number of constraints. As a result, the computational complexity for solving the above LP problem is $O(N^6)$, and the computational cost can be high, which has been witnessed in [5].

2.3 Comparing l_1 -norm minimization with l_2 -norm minimization

Two decades ago, it was observed that the minimization of square of image gradients will result in blurry edges. This leads to the invention of the celebrated ROF total-variation minimization model for denoising [18]. It has already been pointed out that the l_1 -norm minimization in our context corresponds to total variation minimization [19]. Likewise, the l_2 -norm minimization corresponds to the problem of minimization of square of gradients in the context of denoising.

In segmentation, the solution from l_2 -norm minimization may also become over-smooth and therefore ambiguous at the boundaries. This can affect the accuracy of boundary locating in the segmentation, as shown in Fig. 1. Accordingly, we also expect the solution of our model to contain sharp discontinuities, and the l_1 -norm minimization seems promising.

3 A compact LP relaxation for binary submodular MRF

3.1 Deriving a compact LP relaxation via Cholesky factorization of l_2 -norm

Since the conventional l_1 -norm minimization is computationally expensive, we propose to seek alternatives to it. In the following, we will show that a new l_1 -norm, which is

induced by factorizing the l_2 -norm form of the boundary term in Eq.(1), can lead to a more compact LP problem with significantly less computational complexity compared to the original LP problem.

First, we rewrite the l_2 -norm in quadratic form:

$$\sum_{i,j} w_{ij}^e (\mathbf{x}_i - \mathbf{x}_j)^2 = \mathbf{x}^T \widetilde{\mathbf{W}} \mathbf{x} \quad (5)$$

where $\widetilde{\mathbf{W}} = \text{diag}(\widehat{\mathbf{w}}) + \text{diag}(\widehat{\mathbf{w}}) - 2\mathbf{W}$, $\widehat{w}_i = \sum_j w_{ij}^e$, $\widehat{w}_j = \sum_i w_{ij}^e$ and $\mathbf{W} = [w_{ij}^e]$. The full derivation of the above is included in the Appendix.

A quadratic continuous optimization problem is NP-hard if the matrix in the quadratic term is non-definite, i.e. the optimization is non-convex. In fact, having even single negative eigenvalue leads to NP-hard problem [20]. Regarding the convexity of the formulation, we have the following proposition.

Proposition 1. *The matrix $\widetilde{\mathbf{W}}$ in Eq.(5) is positive semi-definite.*

The proof is included in the Appendix. Since $\widetilde{\mathbf{W}}$ is positive semi-definite, the formulation is convex. It is also possible to ensure the matrix to be positive definite by adding a small positive value to the diagonals. In addition to the well-posedness of this formulation, we show that positive definiteness of the matrix $\widetilde{\mathbf{W}}$ allows the problem to be linearized.

Our linear relaxation is based on the following facts:

$$\mathbf{x}^T \widetilde{\mathbf{W}} \mathbf{x} = \mathbf{x}^T \mathbf{U}^T \mathbf{U} \mathbf{x} = \|\mathbf{U} \mathbf{x}\|_{l_2}^2,$$

where \mathbf{U} is an upper triangular matrix of the same dimension of $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{W}} = \mathbf{U}^T \mathbf{U}$ is known as the Cholesky factorization/decomposition. The squared matrix \mathbf{U} is *unique* for symmetric positive definite matrix $\widetilde{\mathbf{W}}$. The Cholesky factorization of it generally uses $n^3/3$ FLOPs, where n is the rank of the matrix, and it is instantaneous for very large matrix on modern processors.

We observe that the matrix $[\text{diag}(\mathbf{w}^e)\mathbf{D}]$ operating on \mathbf{x} in the conventional l_1 -norm can also be thought of as being factorized from the matrix $\widetilde{\mathbf{W}}$. To see this, we can rewrite Eq. (1) as follows:

$$\|\text{diag}(\mathbf{w}^e)\mathbf{D}\mathbf{x}\|_{l_2}^2 = \mathbf{x}^T [\text{diag}(\mathbf{w}^e)\mathbf{D}]^T [\text{diag}(\mathbf{w}^e)\mathbf{D}] \mathbf{x} = \mathbf{x}^T \widetilde{\mathbf{W}} \mathbf{x}.$$

This motivates us to have the following new reformulation of the pairwise potential as:

$$E_{l_1^+}^2(\mathbf{x}) = \|\mathbf{U}\mathbf{x}\|_{l_1} \quad (6)$$

Here, we call the above norm to be minimized as the Cholesky l_1 -norm.

A major difference between the conventional l_1 -norm and our Cholesky l_1 -norm is that the linear operator \mathbf{U} has much smaller dimension than $[\text{diag}(\mathbf{w})\mathbf{D}]$, giving rise to a LP relaxation with significantly fewer variables and constraints.

$$\begin{aligned} \min_{\mathbf{x}, \delta^+} \quad & \mathbf{v}^T \mathbf{x} + \mathbf{1}^T \delta^+ \\ \text{s. t. :} \quad & -\delta^+ \preceq \mathbf{U}\mathbf{x} \preceq \delta^+ \\ & 0 \leq \mathbf{x}_i \leq 1, \delta_i^+ \geq 0, \end{aligned} \quad (7)$$

where the first term is the same as in Eq. (4) and δ^+ is an additional vector of auxiliary variables used for the linear relaxation and its dimension is N , as the same as \mathbf{x} . Essentially, Eq.(7) tries to reduce the bounding values of $\mathbf{U}\mathbf{x}$. The above LP is obtained by applying the equivalence between l_1 -norm minimization and linear programming.

Compared with the conventional LP model in Eq.(4), our model in Eq.(7) has a significantly smaller number of variables and constraints. Specifically, for the image containing N superpixels, there are $N + N \times N$ variables and $N + 2N \times N$ linear constraints for the worst case in the original model [6, 5], whereas there are only $2N$ variables and $2N$ linear constraints in our model. The complexity of our model is therefore $O(N^3)$ which is the same as QP according to Eq.(5). The number of variables and constraints does not change when increasing the number of edges in the graph. We will compare the performance of the two formulations experimentally. The matrices were all set to sparse mode in the implementation.

3.2 Mathematical relationship between l_1 -norm and Cholesky l_1 -norm

In this subsection, we are particularly interested in how tightly the proposed Cholesky l_1 -norm can be related to the conventional l_1 -norm energy, and we are interested in the relationship between the Cholesky l_1^+ -norm and total variation.

Let us consider the reduced QR factorization of the rectangular matrix $[\text{diag}(\mathbf{w})\mathbf{D}]$ in the l_1 -norm boundary term, i.e. $[\text{diag}(\mathbf{w})\mathbf{D}] = \mathbf{Q}^{N^2 \times N} \mathbf{R}^{N \times N}$, where \mathbf{Q} is an orthogonal matrix, such that $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}^{N \times N}$, and \mathbf{R} is an upper triangular matrix. The following fact will relate our Cholesky l_1 relaxation to the original l_1 -norm minimization.

Theorem 1. *The upper triangular matrix \mathbf{U} in the Cholesky l_1 -norm minimization model in Eq.(6) is identical to the upper triangular matrix \mathbf{R} in the QR factorization of $[\text{diag}(\mathbf{w})\mathbf{D}]$ in the l_1 -norm minimization model in Eq.(2)*

The proof of this theorem is presented in the Appendix. This theorem implies several additional relationships between the l_1 -norm and the Cholesky l_1 -norm.

Corollary 1. $\mathbf{U}\mathbf{x} = \mathbf{Q}^T \mathbf{Q}\mathbf{U}\mathbf{x} = \mathbf{Q}^T [\text{diag}(\mathbf{w})\mathbf{D}\mathbf{x}]$.

The above equality implies that the Cholesky l_1 -norm is the l_1 -norm of the linearly transformed weighted gradients, and the transformation matrix is \mathbf{Q} . The weighted variations in \mathbf{x} are projected on the subspace of \mathbf{Q} before calculating the total. Hence, we may also view our Cholesky l_1 -norm as a total subspace-variation. This observation implies that the quasi-total variation minimization may share the discontinuity preservability of the total variation minimization.

Besides, Theorem 1 offers us a stronger relationship between the two formulations in terms of a tight equivalence-of-norm bound.

Theorem 2. *The difference between Cholesky l_1 -norm and l_1 -norm satisfies the following inequalities:*

$$(\|\text{diag}(\mathbf{w})\mathbf{D}\mathbf{x}\|/\|\mathbf{Q}\|) \leq \|\mathbf{U}\mathbf{x}\| \leq \|\mathbf{Q}^T\| \|\text{diag}(\mathbf{w})\mathbf{D}\mathbf{x}\|$$

where the norms are all l_1 -norm, and they are either the l_1 -norm of vector or the induced l_1 -norm of matrix.

The proof of this theorem is included in the appendix.

Remarks. From the above, we can observe that the difference between the Cholesky l_1 -norm and the l_1 -norm is determined by $\|\mathbf{Q}\|_{l_1}$ and $\|\mathbf{Q}^T\|_{l_1}$ which are variable and hopefully reducible by selecting a proper $\widetilde{\mathbf{W}}$ at the beginning. For example, the weight matrix $\widetilde{\mathbf{W}}$ may be chosen such that its unique Cholesky factor \mathbf{Q} gives $\|\mathbf{Q}\|_{l_1} \approx \|\mathbf{Q}^T\|_{l_1} \approx 1$, without any loss of accuracy in modeling. This means the above bound is *tightenable* in principle. This result encourages us to further explore the useful subspaces in the Cholesky l_1 -norm to approximate the total variation norm.

4 Experiments

In the experiment, we will evaluate our method in the context of interactive object segmentation, in which the unary term encodes the seeding information [11] and the pairwise potential is defined as under Eq. (1). We compare our method with the original graph cuts (GC) [2], the l_1 -norm minimization via LP [5, 6], and the l_2 -norm minimization by QP [3, 4].

4.1 Experimental settings

Data and performance measure To evaluate the performance gain in terms of computation. We perform the conventional LP and our proposed LP on GPU for synthetic data. In this experiment, we randomly generate the model parameters and apply the interior point method to solving the LP.

To evaluate the effectiveness of our method, we evaluate on a clownfish dataset and the Oxford interactive segmentation benchmark dataset [21]. Ground truth results and user input seeds on objects and backgrounds are provided in both datasets. The performances of the methods measured by the overlapping ratio between the labeled region and the ground truth object region: $\Gamma = \frac{\text{size}(\text{Result Region} \cap \text{True Region})}{\text{size}(\text{Result Region} \cup \text{True Region})}$.

Implementation issues We adopt superpixelization [22] as a preprocessing to reduce the computational cost. The number of superpixels is around 800 for all test images. We choose the average color of each superpixel to represent the superpixel. We implement all the methods in MATLAB. We used the `linprog` function and `quadprog` function. We use default option settings of the functions. The graph cuts is based on Michael Rubinstein’s implementation². There are some parameters in the model for segmentation. We used $c = 0.00001$, $\lambda = 10$ in all the experiments. The threshold value for converting the continuous labels to binary labels is empirically chosen as 0.08. We also experiment on the effect of different threshold values. We perform the experiments on a PC with Intel Core i5-450M (2.4GHz) processor and 4GB memory.

² <http://www.mathworks.com/matlabcentral/fileexchange/21310-maxflow>

4.2 Results

The clownfish dataset. We first present and analyze the experimental results for the clownfish dataset which contains 62 images. See Fig. 2 for example segmentation results and input seeds. In addition to the manually drawn background seeds, we include the points at the image border as the background seeds in this experiment. As expected, we can see that the results of the conventional LP is very similar to those by graph cuts. A characteristic of them is that they suffer from the small-cut problem. In contrast, QP may produce larger regions due to the possible diffusion of labels at the boundaries. Thus, the resultant regions can be larger than the desired region. Our method compromises the two types of methods and the overall results may outperform the others, e.g., when LP suffers from small-cut problem and/or QP suffers from large-cut problem. We also visualize the continuous labels of conventional LP, QP and our method in Fig. 3. The solutions of LP are binary without thresholding, and the solutions of QP can be over-smooth. The boundaries in the solutions of our LP are clearer than QP, and the solutions are smoother than LP. Quantitative segmentation results of the clownfish dataset are shown in Fig. 4. The results show that QP slightly outperforms the conventional LP on this dataset, and our method slightly outperforms the others. From Table. 1, we can see that the computational cost of our compact LP model is comparable to QP and requires significantly less computational expenses compared to conventional LP. We also note that there is some minor difference between the results by graph cuts and those by conventional LP. We conjecture that the difference is a result of early termination of the interior point method for solving the LP.

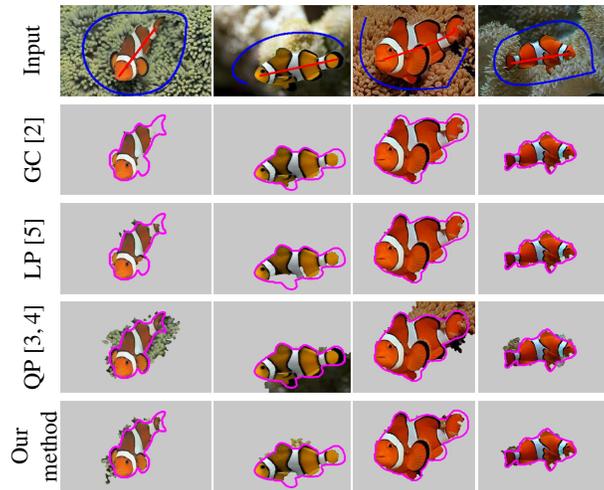


Fig. 2: Example results of the seed-initialized interactive segmentation on clownfish dataset. The results are shown as extracted image regions against the ground truth shape contours in purple.

The Oxford dataset. We mainly evaluate our method on the Oxford dataset which contains 151 images. The user input seeds provided in this dataset are generally insufficient

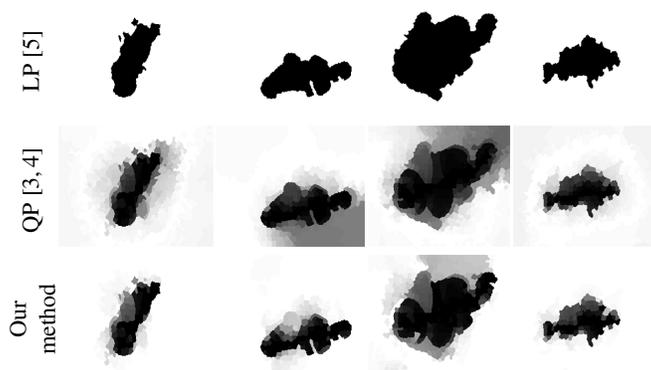


Fig. 3: Example labels of segmentation results in Fig. 2.

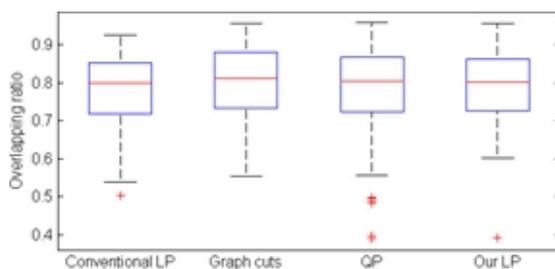


Fig. 4: Quantitative results on clownfish dataset.

for producing a satisfactory segmentation. We adopt the robotuser [21] to simulate the additional user interactions. By increasing the number of interactions, the segmentation results can finally become satisfactory. The maximum number of user interactions is set to 20 in our experiments. See Fig. 6 for example results. We can observe that GC and LP performs quite alike, while QP may produce larger regions. In most of the situations our methods produce more accurate segmentation results than QP. We present the solutions of QP and our method before thresholding in Fig. 7. The LP produces binary labels as expected, the QP produces smooth labels near the object boundaries and our method produces piecewise smooth labels with relatively clear discontinuities at the boundaries. The quantitative results are shown as red boxes in Fig. 8(a).

To quantitatively reveal the effect of the discontinuity preservability of our method, we further consider the robustness of the segmentation to threshold values. We hypothesize that the continuous labels with clear discontinuities at the boundaries will be robust to different threshold values. Therefore, we generate a vector of 100 threshold values equally spaced in $[0, 1]$ for the evaluation. We apply all these threshold values to the continuous labels of QP and our method. Surprisingly, we observe that our method overwhelmingly outperforms the QP for almost all the threshold values in the sense of average overlapping ratio. See Fig. 8(b) for the plots of mean performance with standard deviation.

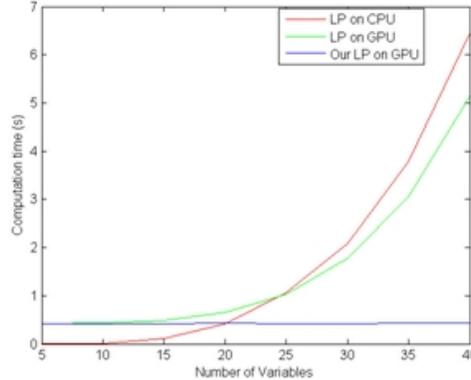


Fig. 5: Comparison of computational times on GPU. As a reference, in [6], the average computation time was 0.66 sec. for GC on CPU and the 0.76 sec. for LP on GPU.

Computational costs. We propose to compare the computational costs for solving conventional LP and our method using the same implementation of interior point method on CPU and GPU. The GPU implementation is realized by simply using `gpuarray` in MATLAB. We used small number of variables because MATLAB does not support sparse matrix in GPU. The results are shown in Fig. 5. From the plot we can observe that the computational cost of our method is almost unchanged but slightly oscillating when increasing the number of variables.

The statistics of the computational costs for our experiment on Oxford dataset are shown in Table 1. Very recently, a fast optimization approach has been proposed for solving a similar segmentation model [23]. However, the computational cost of their approach for 760 superpixels is 23.7 sec. on a machine with 2.7GHz Intel CPU.

Table 1: Comparison of computational costs.

	CPU-LP[6, 5]	CPU-QP [3, 4]	CPU-Our method
Worst-case complexity	$O(N^6)$	$O(N^3)$	$O(N^3)$
Average time (s)	72.35	1.13	12.9

5 Conclusion and future work

In this paper, we proposed a novel LP relaxation for the binary sub-modular MRF model. Our LP relaxation contains significantly fewer variables and constraints compared to the conventional LP. We also showed that our l_1 -norm minimization is tightly related to the total variation minimization through mathematical analysis. Experimental results show that our method is significantly faster than the conventional LP, and it uniformly outperforms QP when converting the continuous labels to binary labels. Our model may be of use to other MRF models, e.g. the TV-MRF [24], as well as many applications, such as shape estimation [25–27].



Fig. 6: Comparison of segmentation performance on Oxford dataset. The upper images in each row show the input images overlaid with input seeds. The lower images in each row show extracted image regions against the ground truth shape contours in purple.

Acknowledgement

Junyan Wang is supported by SUTD-ZJU Collaboration Research Grant 2012 SUTD-ZJU/RES/03/2012. Sai-Kit Yeung is supported by SUTD Start-Up Grant ISTD 2011 016, SUTD-MIT International Design Center Grant IDG31300106, and Singapore MOE Academic Research Fund MOE2013-T2-1-159.

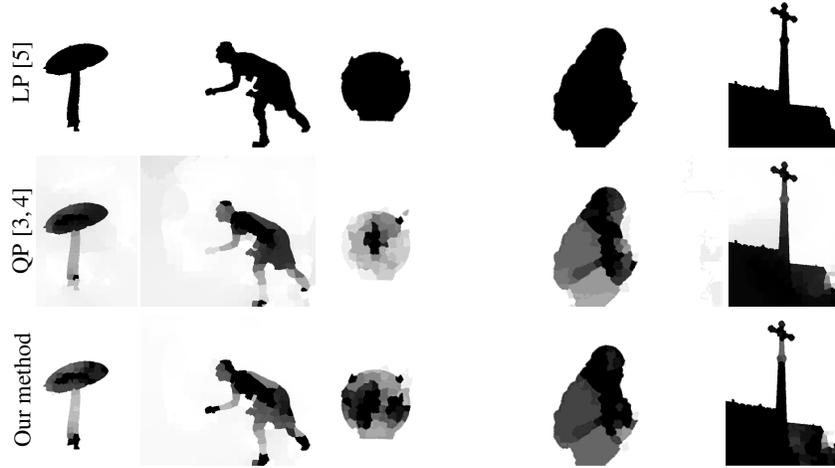


Fig. 7: Continuous labels before thresholding from LP, QP and our method on example inputs in Fig. 6.

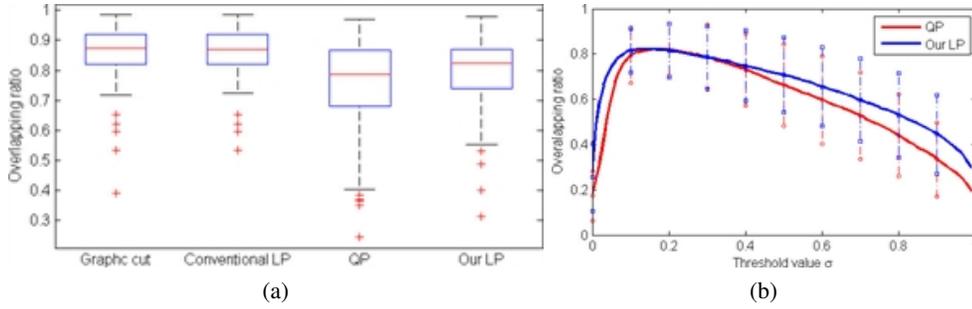


Fig. 8: Quantitative results of the experiments on Oxford dataset. a) Comparison of segmentation accuracy. b) Comparison of QP and our LP for all threshold values.

A Appendix

A.1 Derivation of Eq. (5)

$$\begin{aligned} \sum_{ij} w_{ij}^e (x_i^2 + x_j^2 - 2x_i x_j) &= \sum_i x_i^2 \sum_j w_{ij}^e + \sum_j x_j^2 \sum_i w_{ij}^e - 2 \sum_{ij} w_{ij}^e x_i x_j \\ &= \sum_i x_i^2 \bar{w}_i + \sum_j x_j^2 \hat{w}_j - 2 \sum_{ij} w_{ij}^e x_i x_j = \mathbf{x}^T \tilde{\mathbf{W}} \mathbf{x} \end{aligned}$$

where $\tilde{\mathbf{W}} = \text{diag}(\bar{w}) + \text{diag}(\hat{w}) - 2\mathbf{W}$, $\mathbf{W} = [w_{ij}^e]$.

A.2 Proof of Proposition 1

Proof. The definition of $\tilde{\mathbf{W}}$ is as follows.

$$\tilde{\mathbf{W}} = \text{diag}(\bar{w}) + \text{diag}(\hat{w}) - 2\mathbf{W}$$

where $\bar{w}_j = \sum_k w_{jl} = \sum_k w_{lj'} = \hat{w}_{j'}$, if $j = j'$. In short $\text{diag}(\bar{w}) = \text{diag}(\hat{w})$. Note that $w_{jj'} = 0$ for $j = j'$. Hence, we have the following.

$$\widetilde{\mathbf{W}}_{jj'} = \begin{cases} 2\bar{w}_j, & \text{for } j = j' \\ -2w_{jj'}, & \text{otherwise} \end{cases}$$

Therefore, matrix $\widetilde{\mathbf{W}}$ is a symmetric diagonal dominant matrix, and the diagonal elements are nonnegative. Such matrix is a positive semi-definite matrix. \square

A.3 Proof of Theorem 1

Proof. Substituting $[\text{diag}(\mathbf{w})\mathbf{D}] = \mathbf{Q}^{N^2 \times N} \mathbf{R}^{N \times N}$ into Eq. (2), we obtain the following form of the boundary term.

$$\mathbf{B}_{l_1}(\mathbf{x}) = \|\mathbf{QRx}\|_{l_1}$$

where we applied the QR factorization. The l_2 relaxation of this form will lead to

$$\mathbf{B}_{l_2}(\mathbf{x}) = (\mathbf{x}^T \mathbf{R}^T \mathbf{Q}^T \mathbf{QRx})^{1/2} = (\mathbf{x}^T \mathbf{R}^T \mathbf{Rx})^{1/2} = \|\mathbf{Rx}\|_{l_2}$$

The corresponding l_1^+ -norm minimization is therefore the following

$$\mathbf{B}_{l_1^+}(\mathbf{x}) = \|\mathbf{Rx}\|_{l_1} = \|\mathbf{QRx}\|_{l_1}$$

Note that the Cholesky decomposition is unique and \mathbf{R} is upper-triangular. We can conclude that $\mathbf{U} = \mathbf{R}$. \square

A.4 Proof of Theorem 2

Proof. We prove the left hand side first.

$$\begin{aligned} \|\text{diag}(\mathbf{w}^e)\mathbf{Dx}\|_{l_1} &= \|\mathbf{QUx}\|_{l_1} \leq \|\mathbf{Q}\|_{l_1} \|\mathbf{Ux}\|_{l_1} \\ &\Leftrightarrow \frac{1}{\|\mathbf{Q}\|_{l_1}} \|\text{diag}(\mathbf{w}^e)\mathbf{Dx}\|_{l_1} \leq \|\mathbf{Ux}\|_{l_1} \end{aligned}$$

where we have replaced \mathbf{R} with \mathbf{U} . The right hand side is likewise.

$$\|\mathbf{Ux}\|_{l_1} = \|\mathbf{Q}^T \mathbf{QUx}\|_{l_1} \leq \|\mathbf{Q}^T\|_{l_1} \|\mathbf{QUx}\|_{l_1} = \|\mathbf{Q}^T\|_{l_1} \|\text{diag}(\mathbf{w}^e)\mathbf{Dx}\|_{l_1},$$

which completes the proof. \square

References

1. Kolmogorov, V., Rother, C.: Minimizing nonsubmodular functions with graph cuts—a review. TPAMI **29** (2007) 1274–1279
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. TPAMI **23** (2001) 1222–1239

3. Grady, L.: Random walks for image segmentation. *TPAMI* **28** (2006) 1768–1783
4. Sinop, A.K., Grady, L.: A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In: *CVPR, IEEE* (2007)
5. Li, H., Shen, C.: Interactive color image segmentation with linear programming. *Machine Vision and Applications* **21** (2010) 403–412
6. Bhusnurmath, A., Taylor, C.J.: Graph cuts via l_1 norm minimization. *TPAMI* **30** (2008) 1866–1871
7. Derigs, U., Meier, W.: Implementing goldberg’s max-flow-algorithm a computational investigation. *Zeitschrift fr Operations Research* **33** (1989) 383–403
8. Jamriska, O., Sykora, D., Hornung, A.: Cache-efficient graph cuts on structured grids. In: *CVPR, IEEE* (2012) 3673–3680
9. Lempitsky, V.S., Kohli, P., Rother, C., Sharp, T.: Image segmentation with a bounding box prior. In: *ICCV*. (2009)
10. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *TPAMI* **26** (2004) 147–159
11. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: *ICCV*. (2001)
12. Komodakis, N., Paragios, N., Tziritas, G.: Mrf energy minimization and beyond via dual decomposition. *TPAMI* **33** (2011) 531–552
13. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnorr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Lellmann, J., Komodakis, N., et al.: A comparative study of modern inference techniques for discrete energy minimization problems. In: *CVPR*. (2013) 1328–1335
14. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *IJCV* **1** (1988) 321–331
15. Chan, T., Vese, L.: Active contours without edges. *TIP* **10** (2001) 266–277
16. Ye, Y.: An $o(n^{3l})$ potential reduction algorithm for linear programming. *Mathematical programming* **50** (1991) 239–258
17. Megiddo, N.: Linear programming in linear time when the dimension is fixed. *Journal of the ACM (JACM)* **31** (1984) 114–127
18. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60** (1992) 259 – 268
19. Chambolle, A.: Total variation minimization and a class of binary mrf models. In: *Energy minimization methods in computer vision and pattern recognition*, Springer (2005) 136–152
20. Pardalos, P.M., Vavasis, S.A.: Quadratic programming with one negative eigenvalue is NP-hard. *Journal of Global Optimization* **1** (1991) 15–22
21. Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A.: Geodesic star convexity for interactive image segmentation. In: *CVPR*. (2010)
22. Levinshtein, A., Stere, A., Kutulakos, K.N., Fleet, D.J., Dickinson, S.J., Siddiqi, K.: Turbopixels: Fast superpixels using geometric flows. *TPAMI* **31** (2009) 2290–2297
23. Wang, P., Shen, C., van den Hengel, A.: A fast semidefinite approach to solving binary quadratic problems. In: *CVPR*. (2013)
24. Wu, T.P., Yeung, S.K., Jia, J., Tang, C.K., Medioni, G.G.: A closed-form solution to tensor voting: Theory and applications. *TPAMI* **34** (2012) 1482–1495
25. Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F., Osher, S.J.: Normal estimation of a transparent object using a video. *TPAMI* (2014)
26. Yeung, S.K., Wu, T.P., Tang, C.K.: Extracting smooth and transparent layers from a single image. In: *CVPR*. (2008)
27. Yeung, S.K., Wu, T.P., Tang, C.K., Chan, T.F., Osher, S.: Adequate reconstruction of transparent objects on a shoestring budget. In: *CVPR*. (2011)