Graph-without-Cut: An Ideal Graph Learning for Image Segmentation

Lianli Gao[1]*, Jingkuan Song[2], Feiping Nie[3], Fuhao Zou[4], Nicu Sebe[2], Heng Tao Shen[1,5]
[1]University of Electronic Science and Technology of China, Chengdu 611731, China. [2]University of Trento, Trento 38122, Italy.
[3]Northwestern Polytechnical University, Xi'an 710072, China. [4]Huazhong University of Science and Technology, China.
[5]The University of Queensland, Brisbane 4067, Australia. lianli.gao@uestc.edu.cn, {jingkuan.song, niculae.sebe}@unitn.it,
feipingnie@gmail.com, fuhao.zou@hust.edu.cn, shenht@itee.uq.edu.au

Abstract

Graph-based image segmentation organizes the image elements into graphs and partitions an image based on the graph. It has been widely used and many promising results are obtained. Since the segmentation performance highly depends on the graph, most of existing methods focus on obtaining a precise similarity graph or on designing efficient cutting/merging strategies. However, these two components are often conducted in two separated steps, and thus the obtained graph similarity may not be the optimal one for segmentation and this may lead to suboptimal results. In this paper, we propose a novel framework, Graph-Without-Cut (GWC), for learning the similarity graph and image segmentations simultaneously. GWC learns the similarity graph by assigning adaptive and optimal neighbors to each vertex based on the spatial and visual information. Meanwhile, the new rank constraint is imposed to the Laplacian matrix of the similarity graph, such that the connected components in the resulted similarity graph are exactly equal to the region number. Extensive empirical results on three public data sets (i.e, BSDS300, BSDS500 and MSRC) show that our unsupervised GWC achieves state-of-the-art performance compared with supervised and unsupervised image segmentation approaches.

Introduction

Image segmentation i.e., partitioning an image into several disjoint subsets such that each subset has similar color, intensity or texture achieved an extraordinary success and has become popular in a wide range of applications, such as objects recognition, tracking and image analysis (Liu, Seyedhosseini, and Tasdizen 2015; Arbelaez et al. 2011; Ren and Bo 2012; Maitin-Shepard et al. 2015; Yi and Moon 2012; Kim et al. 2014; Falcao, Udupa, and Miyazawa 2000; Sundaramoorthi, Yezzi, and Mennucci 2008; Willett and Nowak 2007; Boykov and Funka-Lea 2006). The simplest approach of segmentation is to find an appropriate threshold for separating the image into foreground and background group (Yi and Moon 2012). More recently, tremendous efforts have been devoted to designing more efficient and effective segmentation methods, such as edge-based algorithms (Arbelaez et al. 2011), region-based algorithms (Ren and Shakhnarovich 2013), watershed-based segmentation (Donoser and Schmalstieg 2014) and graph-based methods (Peng, Zhang, and Zhang 2013; Arbeláez et al. 2014; Delong and Boykov 2008; Falcao, Udupa, and Miyazawa 2000; Grady 2005; Pavan and Pelillo 2003).

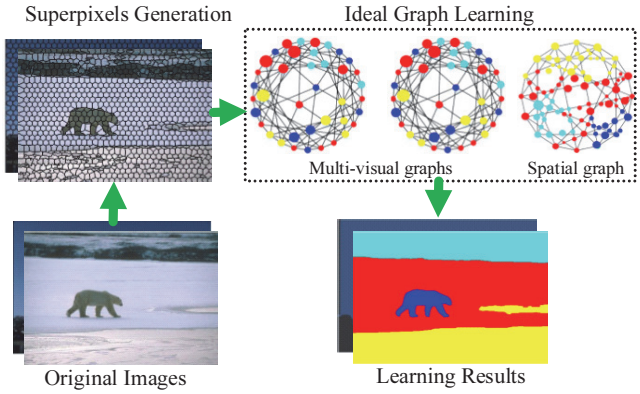

Figure 1: The overview of GWC.

Among the existing image segmentation techniques, many successful ones benefit from mapping the image elements onto a graph $\mathbf{G} = (\mathbb{V}, \mathbb{E})$ which pixels/superpixels are nodes and the weights of edges measure the similarity between nodes. Cutting/merging is then applied on this graph to generate the image segments. Most of the existing graph-based methods focus on two research problems: 1) how to define a precise similarity graph; and 2) how to cut/merge the nodes effectively. The Euclidean distance on the visual features is the simplest way to calculate the similarity graph. However, the segmentation performance is usually inferior to a learned similarity. In (Ren and Shakhnarovich 2013; Liu, Seyedhosseini, and Tasdizen 2015), the authors start with a fine superpixel over-segmentation and train a logistic regression model to determine the similarity between two regions. Based on the similarity graph, they greedily merge regions at each stage until the whole region merging process finishes. By contrast, another pipeline of methods work on

*This work was done when she was a visiting scholar at University of Trento. Jingkuan Song is the Corresponding author.

cutting approaches, which explicitly organize the image elements into mathematically sound structures based on the optimization of the predefined cutting loss function. One representative criterion is normalized cut (Shi and Malik 2000). By minimizing a cutting cost objective function, the best segmentation can be obtained. This objective function is further proved to be equivalent to the generalized eigenvalue decomposition problem and lots of follow-ups proposed efficient solutions for this problem (Arbeláez et al. 2014; Peng, Zhang, and Zhang 2013). In (Comaniciu and Meer 2002), pixels are represented in the joint spatial-range domain by concatenating their spatial and color values into a single vector. Applying mean shift filtering in this domain yields a convergence point for each pixel. Regions are formed by grouping together all pixels whose convergence points are closer than a threshold in the spatial domain and in the range domain.

Graph cut methods provide a well-defined relationship between the segments, while the problem of finding a cut in an arbitrary graph may be NP-hard. More importantly, because the graph similarity learning and the graph cut are conducted in two separated steps, the learned graph similarity matrix may not be the optimal one for cutting leading to suboptimal results. To tackle this problem, in this paper we propose a novel image segmentation framework: Graph-Without-Cut (GWC) approach, which learns the similarity graph and cutting structure simultaneously. It is worthwhile to highlight three aspects of GWC:

- Our image segmentation framework learns the similarity graph and cutting structure simultaneously to achieve the optimal segmentation results. We derive a novel and efficient algorithm to solve this challenging problem.
- We impose the rank constraint on the Laplacian matrix of the learned similarity graph to achieve the ideal neighbors assignment, such that the connected superpixels in the graph are exactly the number of cuts and each connected subgraph corresponds to one region.
- Extensive empirical results on three publicly data sets (i.e, BSDS300, BSDS500 amd MSRC) show that our unsupervised GWC method achieves competitive performance to the related state-of-the-art supervised and unsupervised image segmentation approaches (e.g., gPB, ISCRA, HOCC, MCG-S, MCG and DC).

Related Work

Image segmentation is one of the most fundamental computer vision problems that has been well-studied for a long time yet remains challenging (Liu, Seyedhosseini, and Tasdizen 2015; Arbelaez et al. 2011; Ren and Bo 2012; Maitin-Shepard et al. 2015; Yi and Moon 2012; Kim et al. 2014; Falcao, Udupa, and Miyazawa 2000; Sundaramoorthi, Yezzi, and Mennucci 2008; Willett and Nowak 2007; Boykov and Funka-Lea 2006). Conventionally, segmentation can be grouped into two categories: 1) non-graph based segmentation, which is solely based on boundary or regional information, and 2) graph-based segmentation, which is based on both regional and boundary information and can generally achieve globally optimal results. Note that we leave out of our discussion here the extensive recent literatures on "object-dependent segmentation" in the image. In this paper we are only concerned with methods that are agnostic about any semantic segmentation.

First, we briefly summarize some related non-graph works. A novel Watershed-based segmentation approach (discrete-continuous (DC)) (Donoser and Schmalstieg 2014) is proposed to locally predict oriented gradient signals by analyzing mid-level patches in a discrete-continuous setup. The predicted gradients are passed to an oriented watershed transform and the results are analyzed to obtain the hierarchical segments. In addition, Ren and Shakhnarovich proposed ISCRA (Ren and Shakhnarovich 2013), which adaptively divides the whole region merging process into different cascaded stages and trains a respective logistic regression model at each stage to determine the greedy merging, starting with a fine superpixel oversegmentation. It achieves state-of-the-art performance for object-independent image segmentation.

Compared with non-graph based segmentation, graph-based segmentation have been shown to produce more consistent segmentation and it can be further categorized into five categories: (1) Minimal spanning tree based methods (Peng, Zhang, and Zhang 2013), which provide a mechanism for converting any over-segmentation into the high-level counterparts without loss of the cluster features; (2) Graph cut with cost functions (Arbelaez et al. 2011; Arbeláez et al. 2014), which include minimal cuts illustrating the idea of gestalt principle and normalized cut aiming at alleviating the bias problem toward finding small component; (3) Graph cut on Markov random field (MRF) models (Delong and Boykov 2008), in which the mutual influences among pixels can be formulated into conditional MRF distributions; (4) The shortest path based methods (Falcao, Udupa, and Miyazawa 2000), which describe a certain nature of the object boundaries in the image; and (5) Other methods, which refer to several efficient graph theoretic methods that do not belong to any of the above categories, such as random walker (Grady 2005) and the dominant set based method (Pavan and Pelillo 2003).

Following the graph cut segmentation, object independent segmentation methods seek to partition an image only based on its appearance and does not utilize the underlying semantics about the scene or specific information about target objects. One of the most recent competitive object-independent graph cut frameworks is the higher-order correlation clustering (HOCC), which is proposed in (Kim et al. 2014), by generalizing the correlation clustering over a hypergraph. Specifically, the hypergraph representation is defined first and then the LP relaxation for hypergraphs is generalized, followed by the presentation of a feature vector which consists of pairwise and higher-order feature vectors to characterize the relationship among superpixels over a hypergraph. In addition, a Multiscale Combinatorial Grouping (MCG) (Arbeláez et al. 2014) is proposed by developing a fast normalized cuts algorithm, a high-performance hierarchical segmenter, and a grouping strategy that combines multiscale regions into highly-accurate object candidates. They report the best results on the BSDS dataset.

A substantial difference between our approach and the previous work is that, instead of finding a cut in an arbitrary graph that may be NP-hard, we propose a graph-without-cut method by firstly generating a set of over-segmenting superpixels and then learning an similarity graph between superpixels which has exactly K connected components.

Graph-Without-Cut for Image Segmentation

In this section, we first introduce the preliminary concepts for image segmentation, and then propose our Graph-Without-Cut (GWC) method.

The framework of GWC

Given an image I consisting of pixels V, a segmentation is a partition of V, denoted as $G = \{G_i\}, 1 \leq i \leq K$. Each G_i, which is a connected set of pixels in V, is called a segment or region, and K is the number of total segmentations. All possible partitions form a segmentation space G_V. A ground truth segmentation $G_g \in G_V$ is usually generated by humans and considered as the gold standard. The accuracy of a segmentation G is measured based on its agreement with G_g.

Instead of following the conventional pipeline for image segmentation, in our GWC framework (see Fig. 1), we propose a novel perspective to solve the graph-based image segmentation problem. In conventional graph-based image segmentation methods, the similarity weights w_{ij} are usually computed based on the features, and all the data points are connected as just one connected component. Ideally, if the data points are connected as exactly K components, the pixels are naturally divided into K segmentations and this segmentation is the optimal solution for different criteria, e.g., normalized cut (Shi and Malik 2000) and minimal spanning tree (Felzenszwalb and Huttenlocher 2004). In GWC, we first generate a set of over-segmenting superpixels as a preprocessing. Then we learn a similarity graph between superpixels which has exactly K connected components. A superpixel is an image segment consisting of pixels that have similar visual characteristics. A number of algorithms (Achanta et al. 2012; Levinshtein et al. 2009) can be used to generate superpixels. Next we introduce our graph learning method.

Ideal graph learning

Suppose that for each superpixel, we extracted T features. Let $X^t = \{x_i^t\}_{i=1}^N$ denote the feature matrix of the t-th feature of a set of N superpixels, where $t \in \{1, ..., T\}$. $Y = \{y_1, y_2, ..., y_N\}$ is the average location information for the superpixels. The goal is to learn the similarity matrix S between each superpixel based on different features as well as the spatial information, and all the superpixels have exactly K connected components.

An optimal graph S should be smooth on different features as well as the spatial information, which can be formulated as:

$$\min_{S,\alpha} g(Y,S) + \mu \sum\nolimits_{t=1}^{T} \alpha^t h\left(X^t, S\right) + \beta r(S,\alpha) \qquad (1)$$

where $g(Y,S)$ is the penalty function to measure the smoothness of S on the spatial information Y and $h(X^t, S)$ is the loss function to measure the smoothness of S on the feature X^t. $r(S,\alpha)$ is regularizer defined on the target S and α. μ and β are balancing parameters, and α^t determines the importance of each feature.

The penalty function $g(F,S)$ should be defined in such a way that close superpixels should have high similarity and vice versa. In this paper, we define it as follows:

$$g(Y,S) = \sum\nolimits_{ij} \|y_i - y_j\|_2^2 s_{ij} \qquad (2)$$

where y_i and y_j are the locations of superpixels x_i and x_j. Similarly, $h(X^t, S)$ can be defined as:

$$h\left(X^t, S\right) = \sum\nolimits_{ij} \left\|x_i^t - x_j^t\right\|_2^2 s_{ij} \qquad (3)$$

Note that for simplicity, we use a distance based method to build the similarity graph here. Other options based on the reconstruction coefficients methods can be utilized (Nie, Wang, and Huang 2014; Cai et al. 2011; Gao et al. 2015a; Li et al. 2015; Song et al. 2011; Nie et al. 2011; Song et al. 2013; 2015; Gao et al. 2015b). The regularizer term $r(S,\alpha)$ is defined as:

$$r(S,\alpha) = \|S\|_F^2 + \gamma \|\alpha\|_2^2 \qquad (4)$$

We further constrain that $S \geq 0$,$S1 = 1$,$\alpha \geq 0$ and $\alpha^T 1 = 1$. Then we can obtain the objective function for learning the optimal graph by replacing $g(F,S)$, $h(X^t, S)$ and $r(S,\alpha)$ in (1) using (2), (3) and (4), as follows:

$$\begin{aligned} &\min_{S,F,\alpha} \sum_{ij} \|y_i - y_j\|_2^2 s_{ij} + \mu \sum_{tij} \left(\alpha_t \left\|x_i^t - x_j^t\right\|_2^2 s_{ij}\right) \\ &+\beta \|S\|_F^2 + \beta\gamma \|\alpha\|_2^2 \\ &s.t. S \geq 0, S1 = 1, \alpha \geq 0, \alpha^T 1 = 1 \end{aligned} \qquad (5)$$

The matrix $S \in \mathbb{R}^{N \times N}$ obtained in the neighbors assignment can be seen as a similarity matrix of the graph with the N data points as the nodes. For a nonnegative similarity matrix S, there is a Laplacian matrix L associated with it. According to the definition of Laplacian matrix, suppose each superpixel x_i is assigned a random value as $f_i \in \mathbb{R}^{K \times 1}$, then L can be calculated as:

$$\sum\nolimits_{ij} \|f_i - f_j\|_2^2 s_{ij} = 2tr\left(F^T L F\right) \qquad (6)$$

where $F \in \mathbb{R}^{N \times K}$ with the i-th row formed by f_i, $L = D - \frac{S^t + S}{2}$ is called the Laplacian matrix in graph theory, the degree matrix $D \in \mathbb{R}^{N \times N}$ is defined as a diagonal matrix where the i-th diagonal element is $\sum_j (s_{ji} + s_{ij})/2$. The Laplacian matrix L has the following property.

Theorem 1 *The number K of the eigenvalue 0 of the Laplacian matrix L is equal to the number of connected components in the graph with the similarity matrix S if S is nonnegative.*

Theorem 1 indicates that if $rank(L) = N - K$, then the superpixels have K connected components based on S. Motivated by Theorem 1, we add an additional constraint

$rank(L) = N - K$ into (5) to achieve the ideal similarity matrix. Thus, our new graph learning model is to solve:

$$\begin{aligned} &\min_{S,\alpha} \sum_{ij} \|y_i - y_j\|_2^2 s_{ij} + \mu \sum_{tij} \left(\alpha_t \left\| x_i^t - x_j^t \right\|_2^2 s_{ij} \right) \\ &+\beta \|S\|_F^2 + \beta\gamma \|\alpha\|_2^2 \\ &s.t. \begin{cases} S \geq 0, S1 = 1, \alpha \geq 0, \alpha^T 1 = 1 \\ rank\,(L) = N - K,\ L = f(S) \end{cases} \end{aligned} \tag{7}$$

where $L = f(S) = D - \frac{S^t + S}{2}$. It is difficult to solve the problem (7). Because $L = D - \left(S^T + S\right)/2$ and D also depends on S, the constraint $rank(L) = N - K$ is not easy to tackle. In the next subsection, we will propose a novel and efficient algorithm to solve this challenging problem.

Iterative optimization

Suppose e_i is the i-th smallest eigenvalue of L, we know $e_i \geq 0$ since L is positive semi-definite. It can be seen that the problem (7) is equivalent to the following problem for a large enough value of ρ:

$$\begin{aligned} &\min_{S,\alpha} \sum_{ij} \|y_i - y_j\|_2^2 s_{ij} + \mu \sum_{tij} \left(\alpha_t \left\| x_i^t - x_j^t \right\|_2^2 s_{ij} \right) \\ &+\beta \|S\|_F^2 + \beta\gamma \|\alpha\|_2^2 + 2\rho \sum_{i=1}^{K} e_i \\ &s.t. S \geq 0, S1 = 1, \alpha \geq 0, \alpha^T 1 = 1 \end{aligned} \tag{8}$$

When ρ is set to a large enough value, $\sum_i^K e_i$ will be imposed to be close to 0, which results in $rank\,(L) = N - K$.

According to the Ky Fans Theorem (Fan 1949), we have:

$$\sum_{i=1}^{K} e_i = \min_{F \in R^{N \times K}, F^T F = I} tr\left(F^T L F\right) \tag{9}$$

Therefore, the problem (8) is further equivalent to the following problem:

$$\begin{aligned} &\min_{S,F,\alpha} \sum_{ij} \|y_i - y_j\|_2^2 s_{ij} + \mu \sum_{tij} \left(\alpha_t \left\| x_i^t - x_j^t \right\|_2^2 s_{ij} \right) \\ &+\beta \|S\|_F^2 + \beta\gamma \|\alpha\|_2^2 + 2\rho tr\left(F^T L F\right) \\ &s.t. \begin{cases} S \geq 0, S1 = 1, \alpha \geq 0, \alpha^T 1 = 1 \\ F \in R^{N \times K}, F^T F = I \end{cases} \end{aligned} \tag{10}$$

Compared with the original problem (7), the problem (10) is much easier to solve. We propose an iterative method to minimize the above objective function (10).

Firstly, we initialize $\alpha^t = 1/T$ and we initialize S by the optimal solution to the problem (5). Once these initial values are given, in each iteration, we first update F given S and α, and then update S and α by fixing the other parameters. These steps are describe as below:

Update F By fixing S and α, the problem (10) is equivalent to optimize the following objective function:

$$\min_{F \in R^{N \times K}, F^T F = I} tr\left(F^T L F\right) \tag{11}$$

The optimal solution F to the problem (11) is formed by the K eigenvectors of L corresponding to the K smallest eigenvalues.

Update S By fixing F and α, we can obtain S by optimizing (5). It is equivalent to optimize the following objective function:

$$\begin{aligned} &\min_{S \geq 0, S1 = 1} \sum_{ij} \|y_i - y_j\|_2^2 s_{ij} + \rho \sum_{ij} \|f_i - f_j\|_2^2 s_{ij} \\ &+\beta \|S\|_F^2 + \mu \sum_{tij} \left(\alpha_t \left\| x_i^t - x_j^t \right\|_2^2 s_{ij} \right) \end{aligned} \tag{12}$$

It can be reformulated as:

$$\begin{aligned} &\min_{S \geq 0, S1 = 1} \sum_{i} \left(\beta s_i s_i^T + (a_i + \mu b_i + \rho c_i)\, s_i^T \right) \\ &\Rightarrow \min_{S \geq 0, S1 = 1} \sum_{i} \left(s_i s_i^T + \tfrac{a_i + \mu b_i + \rho c_i}{\beta} s_i^T \right) \end{aligned} \tag{13}$$

where $a_i = \{a_{ij}, 1 \leq j \leq n\}$ with $a_{ij} = \|y_i - y_j\|_2^2$, $b_i = \{b_{ij}, 1 \leq j \leq n\}$ with $b_{ij} = \sum_t \alpha_t \left\| x_i^t - x_j^t \right\|_2^2$ and $c_i = \{c_{ij}, 1 \leq j \leq n\} \in \mathbb{R}^{1 \times n}$ with $c_{ij} = \|f_i - f_j\|_2^2$. It is further equivalent to:

$$\min_{S \geq 0, S1 = 1} \sum_i \left(s_i + \tfrac{a_i + \mu b_i + \rho c_i}{2\beta} \right)_2^2 \tag{14}$$

Then each s_i can be efficiently solved by using a quadratic programming solver, which will be introduced in the next subsection (solution for problem (14)).

Update α By fixing F and S, we can obtain α by optimizing (5). It is equivalent to optimize the following objective function:

$$\begin{aligned} &\min_{\alpha \geq 0, \alpha^T 1 = 1} \mu \sum_t \alpha_t \left(\sum_{ij} \left\| x_i^t - x_j^t \right\|_2^2 s_{ij} \right) + \beta \|\alpha\|_2^2 \\ &\Rightarrow \min_{\alpha \geq 0, \alpha^T 1 = 1} \mu d\alpha + \beta \|\alpha\|_2^2 \end{aligned} \tag{15}$$

where $d = \{d_t, 1 \leq t \leq T\}$ with $d_t = \sum_{ij} \left\| x_i^t - x_j^t \right\|_2^2 s_{ij}$. Then we can be use a quadratic programming solver to obtain α.

We update F, S and α iteratively until $\mathbf{S}$ has K connected components, as shown in Algorithm 1.

Algorithm 1 Solution for ideal graph learning

Input: Initialized α, $\mathbf{S}$, segmentation number K, parameters β, γ, μ, a large enough ρ;
Output: $\mathbf{F}$, $\mathbf{S}$, α;

1: **repeat**
2: Fix S and α, calculate F according to the solution of problem (11);
3: Fix F and α, update $\mathbf{S}$ by solving the problem (14);
4: Fix F and S, update α by solving the problem (15);
5: **until** $\mathbf{S}$ has K connected components or max iteration is reached.
6: **return** $\mathbf{F}$, $\mathbf{S}$, α;

Solution for problem (14) In this subsection, we introduce an efficient solution for problem (14). In the meantime, we present an effective method to determine the regularization parameter β in the problem (14), so that we have

fewer parameters to tune. The Lagrangian function of problem (14) is:

$$\ell(s_i,\eta,\varepsilon_i)=\frac{1}{2}\sum_i\left\|s_i+\frac{a_i+\mu b_i+\rho c_i}{2\beta}\right\|_2^2-\eta(s_i^T1-1)-s_i^T\varepsilon_i \quad (16)$$

where η, $\varepsilon_i \geq 0$ are the Lagrangian multipliers, and β is the regularization parameter for each s_i. According to the KKT condition, it can be verified that the optimal solution s_i should be:

$$s_{ij}=\left(-\frac{a_{ij}+\mu b_{ij}+\rho c_{ij}}{2\beta}+\eta\right)_+ \quad (17)$$

According to Eq.(17) and the constraint $s_i^T1=1$, we have

$$s_i^T1=\sum_{j=1}^{N}\left(-\frac{d_{ij}}{2\beta}+\eta\right)=1\Rightarrow\eta=\frac{1}{N}+\frac{1}{2N\beta}\sum_{j=1}^{N}d_{ij} \quad (18)$$

where $d_{ij}=a_{ij}+\mu b_{ij}+\rho c_{ij}$. Without loss of generality, suppose $d_{i1},d_{i2},...,d_{iN}$ in are ordered from small to large. If the optimal s_i has only P nonzero elements, then according to (17), we know $s_{iP}>0$ and $s_{i,P+1}=0$. Therefore, we have:

$$-\frac{d_{iP}}{2\beta_P}+\eta>0,\ -\frac{d_{i,P+1}}{2\beta_{P+1}}+\eta\leq 0 \quad (19)$$

and

$$s_i^T1=\sum_{j=1}^{P}(-\frac{d_{ij}}{2\beta_i}+\eta)=1\Rightarrow\eta=\frac{1}{P}+\frac{1}{2P\beta_i}\sum_{j=1}^{P}d_{ij} \quad (20)$$

By replacing η in (19) using (20), we have the following inequality for β_i:

$$\frac{P}{2}d_{iP}-\frac{1}{2}\sum_{j=1}^{P}d_{ij}<\beta_i\leq\frac{P}{2}d_{i,P+1}-\frac{1}{2}\sum_{j=1}^{P}d_{ij} \quad (21)$$

Therefore, in order to obtain an optimal solution s_i to the problem (14) that has exactly P nonzero values, we set

$$\beta_i=\frac{P}{2}d_{i,P+1}-\frac{1}{2}\sum\nolimits_{j=1}^{P}d_{ij} \quad (22)$$

The overall β could be set to the mean of $\beta_1,\beta_2,...,\beta_N$. The number of neighbors P is much easier to tune than the regularization parameter β since P is an integer and has explicit meaning.

Experiments

We conduct experiments with two validation goals. First, we study the influence of parameters in our algorithm. Second, we compare our results with other state-of-the-art algorithms on three public datasets.

Datasets We experiment with three publicly available data sets for image segmentation: (1) Berkeley Segmentation Data Set 300 (BSDS300) (Martin et al. 2001) consists of all of the grayscale and color segmentations for 300 images. Specifically, it collects 12,000 hand-labeled segmentations with half of them obtained from presenting the subject with a color image and the other half from presenting a grayscale image. (2) Berkeley Segmentation Data Set 500 (BSDS500) (Arbelaez et al. 2011) is an extension of the BSDS300 with 200 fresh images of the same size. This dataset consists of 500 natural images, ground-truth human annotations and benchmarking code. (3) Microsoft Research Cambridge (MSRC) Object Recognition Data Set cleaned version (Malisiewicz and Efros 2007) contains 591 natural images with dense labeling and a large number of object categories and 591 320 $\times$ 213 natural images with single ground truth. We utilize gPb (Arbelaez et al. 2011) to generate the over-segmentations, and we further extract color features (32D LAB, 32D HSV, 3D LAB, 3D HSV) and texture features (64D Texton) from each superpixel.

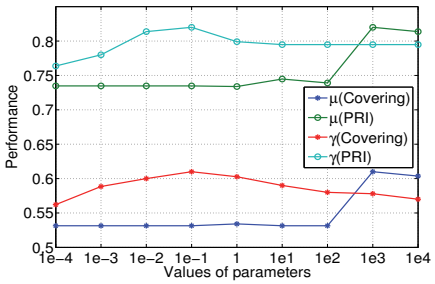

Figure 2: Effect of parameters on BSDS500 dataset

Baselines To evaluate the performance of the GWC approach, we compare our methods with several state-of-the-art image segmentation methods: a) non-graph-based methods (i.e., DC (Donoser and Schmalstieg 2014) and ISCRA (Ren and Shakhnarovich 2013)); and b) graph based approaches (i.e., gPb (Arbelaez et al. 2011), HOCC (Kim et al. 2014), MCG (single-scale hierarchy) and MCG-S (multi-scale hierarchy) (Arbeláez et al. 2014)) on the above three publicly available dataset.

Evaluation metrics Following (Arbelaez et al. 2011), we use three evaluation metrics: 1) segmentation covering (Everingham et al.) that measures averaged matching between proposed segments with a ground truth labeling; 2) probabilistic rand index (PRI) (Unnikrishnan, Pantofaru, and Hebert 2007) that measures pairwise similarity between two multi-label clusterings; and 3) variation of information (VI) (Meila 2005) that measures the relative entropy between a proposed segmentation and a ground truth labeling. For a family of machine segmentations associated with different scales of a hierarchical algorithm or different sets of

Table 1: Results of different methods on the BSDS300 dataset

	BSDS300					
	Covering		PRI		VI	
Method	ODS	OIS	ODS	OIS	ODS	OIS
gPb	0.59	0.65	0.81	0.85	1.65	1.47
ISCRA	0.60	0.67	0.81	**0.86**	1.61	**1.40**
HOCC	0.60	-	0.81	-	1.74	-
GWC	**0.61**	**0.68**	**0.82**	**0.86**	1.60	1.42

Table 2: Results of different methods on the BSDS500 dataset

	BSDS500					
	Covering		PRI		VI	
Method	ODS	OIS	ODS	OIS	ODS	OIS
gPb	0.59	0.65	0.83	0.86	1.69	1.48
ISCRA	0.59	0.66	0.82	0.86	1.60	1.42
HOCC	0.60	-	**0.83**	-	1.79	-
DC	0.59	0.64	0.82	0.85	1.68	1.54
MCG-S	0.60	0.66	0.81	0.86	1.62	1.41
MCG	**0.61**	**0.67**	0.81	0.86	**1.55**	**1.37**
GWC	**0.61**	0.66	**0.83**	**0.87**	1.62	1.41

Table 3: Results of different methods on the MSRC dataset

	MSRC					
	Covering		PRI		VI	
Method	ODS	OIS	ODS	OIS	ODS	OIS
gPb	0.65	0.75	**0.78**	0.85	1.28	0.99
ISCRA	0.67	0.75	0.77	0.85	**1.18**	1.02
GWC	**0.68**	**0.76**	**0.78**	**0.85**	1.24	**0.98**

parameters (e.g., K for GWC), we report the segmentation results at optimal dataset scale (ODS) and optimal image scale (OIS).

Parameters study

There are two balancing parameters (i.e., μ and γ) affecting the performance of our algorithm. In this subsection, we study the performance variation with different parameters. Due to the space limit, we only report the results at ODS on the BSDS500 dataset in Fig. 2. We can see that the parameter μ is important to the performance. In general, when $\mu < 10^2$, an unsatisfactory performance is achieved, probably due to that visual features are assigned with a small weight. By contrast, the performance is not very sensitive to γ. When γ is set to a large constant, similar weights α_v will be assigned to different features. On the other hand, when γ is small, GWC tends to depend on the best performance of a single feature.

Results

We show the comparison of our GWC with the baselines in Table 1,2 and 3 and also in the quantitative examples in Fig. 3. From these results, we have several observations.

- Compared with other state-of-the-art graph-based methods, our method outperforms most of them, and can achieve comparable performance with the current best result (MCG). Specifically, the results for HOCC are not available on the MSRC dataset, but on the BSDS500 datasets GWC performs better than HOCC in terms of all metrics at ODS. In addition, the results for MCG-S and MCG are both only available on the BSDS500 dataset, the results indicate that GWC is comparable with MCG in terms of Covering and PRI, and better than MCG-S.
- We can see that by learning an ideal graph, the performance is improved compared with gPb around (0.02 in covering ODS, 0.013 in covering OIS, 0.01 PRI OIS, 0.05 in vi ODS and 0.04 in OIS) and ISCRA about (0.013 for covering at ODS, 0.01 for PRI at ODS and 0.01 for VI at ODS). The result for the DC is available on the BSDS500 dataset and our experimental results show that GWC performs better than DC in terms of all evaluation metrics.
- The quantitative examples show that our GWC can achieve promising performance in practice. Some of the segmentation results are even more reasonable than the groundtruth. However, our GWC fails to segment the regions when the adjacent components have similar visual features (e.g., the hairs of the girl and the trees in the background are very similar in the fourth line).

Figure 3: The qualitative results of GWC on BSDS500. first column: input images, second: results of gPb (Arbelaez et al. 2011), third: our results, fourth: ground truth.

Conclusions

This paper proposed the GWC framework by simultaneously learning the graph similarity matrix and image segmentation instead of first organizing the image elements into graphs and then cutting the generated graph for segmentation. Based on the spatial and visual information, each vertex is assigned with adaptive and optimal neighbors for graph similarity learning. By using the Laplacian matrix, the connected components in the generated similarity graph are exactly equal to the segmentation numbers. Experimental results show that the proposed unsupervised GWC achieved comparable results or even outperformed other image segmentation algorithms on various datasets.

Acknowledgments

The work of Jingkuan Song and Nicu Sebe has been partially supported by the FP7 EC project xLiMe. The work of Lianli Gao is partially supported by the Fundamental Research Funds for the Central Universities-Grant No.ZYGX2014J063) and the National Natural Science Foundation of China (Grant No. 61502080).

References

Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; and Süsstrunk, S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(11):2274–2282.

Arbelaez, P.; Maire, M.; Fowlkes, C.; and Malik, J. 2011. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(5):898–916.

Arbeláez, P. A.; Pont-Tuset, J.; Barron, J. T.; Marqués, F.; and Malik, J. 2014. Multiscale combinatorial grouping. In *CVPR*, 328–335.

Boykov, Y., and Funka-Lea, G. 2006. Graph cuts and efficient N-D image segmentation. *IJCV* 70(2):109–131.

Cai, X.; Nie, F.; Huang, H.; and Kamangar, F. 2011. Heterogeneous image feature integration via multi-modal spectral clustering. In *CVPR*, 1977–1984.

Comaniciu, D., and Meer, P. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(5):603–619.

Delong, A., and Boykov, Y. 2008. A scalable graph-cut algorithm for n-d grids. In *CVPR*, 1–8.

Donoser, M., and Schmalstieg, D. 2014. Discrete-continuous gradient orientation estimation for faster image segmentation. In *CVPR*, 3158–3165.

Everingham, M.; Van Gool, L.; Williams, C. K. I.; Winn, J.; and Zisserman, A. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

Falcao, A.; Udupa, J.; and Miyazawa, F. 2000. An ultra-fast user-steered image segmentation paradigm: live wire on the fly. *Medical Imaging, IEEE Transactions on* 19(1):55–62.

Fan, K. 1949. On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations. I. *Proceedings of the National Academy of Science* 35:652–655.

Felzenszwalb, P. F., and Huttenlocher, D. P. 2004. Efficient graph-based image segmentation. *IJCV* 59(2):167–181.

Gao, L.; Song, J.; Nie, F.; Yan, Y.; Sebe, N.; and Shen, H. T. 2015a. Optimal graph learning with partial tags and multiple features for image and video annotation. In *CVPR*, 4371–4379.

Gao, L.; Song, J.; Zou, F.; Zhang, D.; and Shao, J. 2015b. Scalable multimedia retrieval by deep learning hashing with relative similarity learning. In *ACM Multimedia*.

Grady, L. 2005. Multilabel random walker image segmentation using prior models. In *CVPR*, 763–770.

Kim, S.; Yoo, C.; Nowozin, S.; and Kohli, P. 2014. Image segmentation using higher-order correlation clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 36(9):1761–1774.

Levinshtein, A.; Stere, A.; Kutulakos, K. N.; Fleet, D. J.; Dickinson, S. J.; and Siddiqi, K. 2009. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.* 31(12):2290–2297.

Li, Y.; Nie, F.; Huang, H.; and Huang, J. 2015. Large-scale multi-view spectral clustering via bipartite graph. In *AAAI*, 2750–2756.

Liu, T.; Seyedhosseini, M.; and Tasdizen, T. 2015. Image segmentation using hierarchical merge tree. *CoRR* abs/1505.06389.

Maitin-Shepard, J.; Jain, V.; Januszewski, M.; Li, P.; Kornfeld, J.; Buhmann, J.; and Abbeel, P. 2015. Combinatorial energy learning for image segmentation. *CoRR* abs/1506.04304.

Malisiewicz, T., and Efros, A. A. 2007. Improving spatial support for objects via multiple segmentations. In *BMVC*.

Martin, D.; Fowlkes, C.; Tal, D.; and Malik, J. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, 416–423.

Meila, M. 2005. Comparing clusterings: an axiomatic view. In Raedt, L. D., and Wrobel, S., eds., *ICML*, 577–584.

Nie, F.; Zeng, Z.; Tsang, I. W.; Xu, D.; and Zhang, C. 2011. Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering. *IEEE Transactions on Neural Networks* 22(11):1796–1808.

Nie, F.; Wang, X.; and Huang, H. 2014. Clustering and projected clustering with adaptive neighbors. In *SIGKDD*, 977–986.

Pavan, M., and Pelillo, M. 2003. Graph-theoretic approach to clustering and segmentation. In *CVPR*, 145–152.

Peng, B.; Zhang, L.; and Zhang, D. 2013. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition* 46(3):1020–1038.

Ren, X., and Bo, L. 2012. Discriminatively trained sparse code gradients for contour detection. In *NIPS*, 593–601.

Ren, Z., and Shakhnarovich, G. 2013. Image segmentation by cascaded region agglomeration. In *CVPR*, 2011–2018.

Shi, J., and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(8):888–905.

Song, J.; Yang, Y.; Huang, Z.; Shen, H. T.; and Hong, R. 2011. Multiple feature hashing for real-time large scale near-duplicate video retrieval. In *ACM Multimedia*, 423–432.

Song, J.; Yang, Y.; Yang, Y.; Huang, Z.; and Shen, H. T. 2013. Inter-media hashing for large-scale retrieval from heterogeneous data sources. In *SIGMOD*, 785–796.

Song, J.; Gao, L.; Yan, Y.; Zhang, D.; and Sebe, N. 2015. Supervised hashing with pseudo labels for scalable multimedia retrieval. In *ACM Multimedia*.

Sundaramoorthi, G.; Yezzi, A. J.; and Mennucci, A. 2008. Coarse-to-fine segmentation and tracking using sobolev active contours. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(5):851–864.

Unnikrishnan, R.; Pantofaru, C.; and Hebert, M. 2007. Toward objective evaluation of image segmentation algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29(6):929–944.

Willett, R., and Nowak, R. D. 2007. Minimax optimal level-set estimation. *TIP* 16(12):2965–2979.

Yi, F., and Moon, I. 2012. Image segmentation: A survey of graph-cut methods. In *ICSAI*, 1936–1941.