J. Vis. Commun. Image R. 33 (2015) 10–19

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Image segmentation based on weighting boundary information via graph cut ☆

Tao Wang [a], Zexuan Ji [a], Quansen Sun [a,*], Qiang Chen [a], Shoudong Han [b]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 24 November 2014
Accepted 27 August 2015
Available online 6 September 2015

Keywords:
Image segmentation
Graph cut
Weighting parameter
Image patch
Color image
Region information
Boundary information
Parameter learning

ABSTRACT

The graph cut model has been widely used in image segmentation, in which both the region and boundary information play important roles for accurate segmentation. However, how to effectively model and combine these two information is still a challenge. In this paper, we improve the conventional graph cut methods by combining the region and boundary information with an effective and straightforward way. When modeling the region information, the component-wise expectation–maximization for Gaussian mixtures algorithm is used to learn the parameters of the prior knowledge. When modeling the boundary information, the weighting patch is used to represent the similarities of the neighboring pixels. Then the region and boundary information are combined by a weighting parameter, where the weight is small for boundary pixels and is large for non-boundary pixels. Finally, experiments on various images from the Berkeley and MSRC data sets were conducted to demonstrate the effectiveness of the proposed method.

1. Introduction

Extracting a foreground object in a complex environment is of great practical importance in computer vision [1], the process of which is known as image segmentation. It is useful in various applications such as image analyzing, image editing and recognition. The existing image segmentation methods can be classified into unsupervised and supervised approaches. Due to the variety and complexity of the images, the supervised method is more practical than unsupervised method [2]. As a supervised method, the graph cut approach is very popular due to many excellent characteristics, such as multiple feature fusion ability, globally optimal framework, and highly effectiveness of the performance. For quadratic binary submodular functions, the graph cut model can obtain the global optimal solutions [3]. Even for non-submodular functions, the extended graph cut algorithm such as the roof duality [4] can also provide a reliable optimization. Based on this theoretical character of the graph cut model, in this paper, we mainly focus on the graph cut model for the two-label segmentation.

Recently, many graph cut based methods have been proposed, such as Graph Cut [2,5], Grabcut [6], Lazy Snapping [7], and Texture-color [8,9]. These methods treat an image as a weighted graph, and then minimize a certain energy function based on the graph to produce a segmentation. Compared with the region-based [10,11] and boundary-based [12,13] segmentation methods, the graph cut based methods can often obtain more satisfactory results due to the combination of region and boundary information together. So how to effectively model and combine this two information is the key problem in the graph cut based methods.

To model the region information, many histogram [2,5] and clustering [6,7] based methods have been proposed. In histogram based methods, the seeds of foreground and background are counted by a histogram. Due to the impracticability to construct adequate color space histogram, these methods are hard to segment color images. In clustering based methods, the features in the image are analyzed by clustering algorithms, and then the obtained clusters are used to represent the prior label information. The commonly used clustering algorithms include the K-means algorithm and the Gaussian mixture model (GMM). Compared with K-means algorithm, the GMM is more suitable for graph cut model by providing the maximum likelihood estimation. The Expectation–maximization (EM) algorithm is usually utilized to update the parameters of GMM. However, for certain types of mixtures, the EM algorithm may converge to the boundary of the parameter space (where the likelihood is unbounded) leading to meaningless estimates [14]. Furthermore, the above clustering algorithms cannot automatically select the cluster number. The

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author.
E-mail addresses: wangtaoatnjust@163.com (T. Wang), jizexuan@njust.edu.cn (Z. Ji), sunquansen@njust.edu.cn (Q. Sun).

http://dx.doi.org/10.1016/j.jvcir.2015.08.013
1047-3203/

mixture model may over-fit the data with too many components, while with too few components it may not be flexible enough to approximate the true underlying model. Figueiredo and Jain [14] proposed the component-wise expectation–maximization for Gaussian mixtures (CEMGM) algorithm, in which the number of components can be selected automatically and the boundary of the parameters space problem is avoided.

The boundary information is usually denoted by the penalties of neighborhood pixels. Many methods [2,6,15] compute the penalties of pixels based on their intensities. However, it is hard to accurately describe the relationships of pixels based on their intensities when the image contains noise and texture. Recently, many methods [7,16–18] replace pixels with image patches to conduct the segmentation. The Lazy Snapping algorithm [7] pre-segments the image into superpixels which can be regarded as irregular image patches. Though the speed of Lazy Snapping is fast, the precision of the superpixel is related to the pre-segmentation algorithm and it is easy to over-segment the image. The rectangular image patch is also widely used to describe the similarities of pixels [16–18]. All the pixels in the patch are considered equally. Compared with the pixel level feature, the patch level feature is more robust due to the consideration of the neighborhood information. However, though the texture details can be effectively preserved by the patch, the image edges cannot be preserved well.

The graph cut model amalgamates the region and boundary information together. Most of the existing methods [2,6,19,20] balance this two information by a regularization parameter which is usually set by experiments. The segmentation is very sensitive to this parameter, which may lead to over-segmentation or under-segmentation results. Peng and Veksler [21] extends the conventional fixed parameter to an adaptive one, in which each image is assigned an optimal parameter based on the evaluations of the segmentations. Although this method can obtain a suitable parameter for each image, the cost is very expensive and the corresponding estimation relies on the ground true images. Candemir and Akgül [22] point out that different image parts may need different parameters due to the complexity of the natural images. In their method, each pixel is automatically set a value based on its edge probability. If the pixel is likely to be an edge point, then its parameter value is small, vice versa. Compared with the above method, this method is simpler and more practical. However, though the details in boundary can be well preserved, the segmentation results in non-boundary regions are sometimes unsatisfactory. Because not all edge points should be assigned small parameters. For example, the edge points within the foreground, background and texture regions are not true boundary points, and they should be set large parameters due to the fact that they belong to the same category with their neighborhoods.

To overcome the aforementioned problems, this paper extends the graph cut based methods in three aspects. Firstly, the CEMGM algorithm is used to model the region information and the number of Gaussian components can be selected automatically. Secondly, the weighting rectangular patch is used to model the boundary information. Unlike other patch based methods, pixels in the patch are considered unequally in the proposed method. Each pixel in the patch is allocated a weight which is based on the structural similarity with the central pixel. In this way, the texture and edge can be both well preserved. Thirdly, a proposed weighting parameter is used to combine the region and boundary information. The weight of each pixel not only depends on its edge probability but also the difference with its neighborhoods. So the edge pixels within the foreground, background or texture regions can also obtain large weights. Only true boundary pixels are assigned small weights. In this case, more details in the image are preserved and the segmentation accuracy can be improved.

2. Graph cut model

The segmentation problem can be viewed as a binary labeling problem in graph cut framework. An image is a set of pixels $V = \{v_{ij}:(i,j) \in \Omega\}$ where (i,j) is the position of the pixel v_{ij} and $\Omega \subset R^2$ is the domain of the image. Users label some pixels to construct two subsets of feature vectors as foreground seeds $F \subset V$ and background seeds $B \subset V$. The goal of the segmentation is to find a labeling f that assigns each pixel $v_{ij} \in V$ a label $f_{ij} \in \{0,1\}$ with 0 for background and 1 for foreground.

The region and boundary information in the image are combined together in the graph cut model. The labeling f should be both piecewise smooth and consistent with the observed data. The labeling problem can be naturally formulated in terms of energy minimization. In the graph cut framework, one seeks the labeling f that minimizes the energy

$$E(f) = E_r(f) + E_b(f) \tag{1}$$

where the energy function consists of a region term E_r and a boundary term E_b. The regional term $E_r(f)$ represents the region information of an image, which assumes that the individual penalties for assigning pixels to "foreground" and "background" and reflects how the intensity of pixels fit into given intensity models of the foreground and background:

$$E_r(f) = \sum_{(i,j)\in\Omega} -\ln(\Pr(v_{ij}|f_{ij})) \tag{2}$$

where $\Pr(v_{ij}|f_{ij})$ denotes the probability of a pixel v_{ij} fitting the model of foreground ($f_{ij} = 1$) or background ($f_{ij} = 0$). The boundary term is defined as a penalty for a discontinuity between neighboring pixels:

$$E_b(f) = \sum_{(ij,pq)\in\aleph} \lambda e^{-\beta\|v_{ij}-v_{pq}\|^2} \frac{1}{\mathrm{dist}(v_{ij}, v_{pq})} \cdot \delta(f_{ij} \neq f_{pq}) \tag{3}$$

where constant $\lambda \geqslant 0$ specifies a relative importance of the region term E_r versus the boundary properties term E_b. $\delta(f_{ij} \neq f_{pq}) = \begin{cases} 1 & \text{if } f_{ij} \neq f_{pq} \\ 0 & \text{otherwise} \end{cases}$, $\aleph$ is a set of all pairs of neighboring pixels, dist() is the Euclidean distance of neighboring pixels, and $\|\cdot\|$ is the L^2-*norm*. The parameter β controls the smoothness and preciseness of the segmentation boundary. In general, β is chosen to be [2]:

$$\beta = \frac{1}{2\mathrm{EP}(\|v_{ij} - v_{pq}\|^2)} \tag{4}$$

where EP() is the expectation over an image sample. Normally, the penalty is large when pixels v_{ij} and v_{pq} are similar and the penalty is close to zero when the two are very different.

The minimization of Eq. (1) can be solved by graph cut which can provide a global optimal solution for binary submodular functions [2] by using a min-cut/max-flow algorithm [23]. For nonsubmodular functions, many methods [24–26] have also been developed, which makes the graph cut model adapt to larger class of energy functions. In graph cut model, the image is viewed as a graph and two additional terminals are added to represent the foreground and background, respectively. Edges between nodes and terminals represent the region information, and edges between pairwise nodes represent the boundary information. The minimum cost cut in the graph gives an optimal solution for the energy in Eq. (1) [5]. Referring to Fig. 1 for example, the left is the construction of the graph where S and T are two terminals representing two labels. The right shows a cut partitioning the graph into two sets. Then pixels linking to S are assigned the label S and pixels linking to T are assigned the label T.