Information Fusion 20 (2014) 292–304

Contents lists available at ScienceDirect

Information Fusion

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

Interactive color image segmentation via iterative evidential labeling

Yin Chen [a], Armin B. Cremers [b], Zhiguo Cao [a,*]

[a] National Key Laboratory of Science and Technology on Multi-spectral Information Processing, School of Automation, Huazhong University of Science and Technology, Luo Yu Road, No. 1037, Hongshan District, 430074 Wuhan, China

[b] Institute of Computer Science III, Rheinische Friedrich-Wilhelms-Universität Bonn, Römerstr. 164, 53117 Bonn, Germany

ARTICLE INFO

Article history:
Received 9 November 2013
Received in revised form 21 February 2014
Accepted 24 March 2014
Available online 16 April 2014

Keywords:
Interactive image segmentation
Markov random fields (MRFs)
Dempster–Shafer's (DS) theory of evidence
Bayesian information criterion (BIC)
Information fusion

ABSTRACT

We develop an interactive color image segmentation method in this paper. This method makes use of the conception of Markov random fields (MRFs) and D–S evidence theory to obtain segmentation results by considering both likelihood information and priori information under Bayesian framework. The method first uses expectation maximization (EM) algorithm to estimate the parameter of the user input regions, and the Bayesian information criterion (BIC) is used for model selection. Then the beliefs of each pixel are assigned by a predefined scheme. The result is obtained by iteratively fusion of the pixel likelihood information and the pixel contextual information until convergence. The method is initially designed for two-label segmentation, however it can be easily generalized to multi-label segmentation. Experimental results show that the proposed method is comparable to other prevalent interactive image segmentation algorithms in most cases of two-label segmentation task, both qualitatively and quantitatively.

1. Introduction

Image segmentation is the first step of many computer vision tasks. It involves partitioning an image into several homogeneous parts, which are spatially connected clusters of pixels, while the union of any two neighboring parts is heterogeneous. In general, image segmentation methods can be categorized as: fully automatic ones, semi-automatic ones, and manual ones. It is time consuming and tedious as well as lacking in precision to manually segment images. Such methods are impractical for images with a large size or long image sequences. On the other hand, fully automatic methods can segment images without human intervention, which greatly simplifies the operation. These methods can achieve high accuracy in many uncomplicated image scenes. However, fully automatic methods often fail when the image scene is complex. In these situations, semi-automatic methods can be the best choice. The segmentation is obtained after a few interactions (usually scribbles or strokes) are provided, which indicate the region of interest. This kind of user interaction can help to segment difficult scenes. In the pattern classification view, such user inputs can be viewed as supervised information which provides visual hints to model and group visual patterns. Many existing machine learning algorithms can be employed to segment images with such supervised information [1].

Practically, image segmentation is not an easy task due to all sorts of difficulties, such as noise pollution, illumination variation and background clutter. Color images contain more information, which makes it more difficult to segment color ones [2]. Thus, there has been much research on color image segmentation, and it has received much attention for visual surveillance, intelligent transportation, special film effects, and so on. Many different color image segmentation methods have been reported. For example, the methods based on mathematical morphology [3], MRFs [4,5], neural networks [6], support vector machines (SVMs) [7], and so on.

MRFs consider the spatial–contextual information contained in images in the framework of Bayesian decision theory. In this framework, the labels representing segmentation results are decided by considering both likelihood information given by pixel values and priori information given by the labels of neighborhoods [8].

Data fusion has gained a lot of research interests in the last decade [9,10]. There are many data fusion techniques. They are fusion by Bayesian inference [11], fusion by probabilistic [12], fuzzy fusion [13] and evidence theory, also known as D–S theory [14,15], which is the base of this work. The D–S theory has been used for MRI segmentation and classification [16–20]. In [16], an unsupervised algorithm based on D–S evidence theory is proposed to segmenting and visualizing left heart ventricles. In [17], some key features of D–S evidence theory are pointed out, as well as with examples of brain tissue classification in pathological dual echo MR images. In [18], a segmentation scheme of multi-echo MR image is proposed. The scheme combines spatial information by the fusion

* Corresponding author. Tel: +86 27 87558918; fax: +86 27 87540131.
E-mail address: zgcao@mail.hust.edu.cn (Z. Cao).

http://dx.doi.org/10.1016/j.inffus.2014.03.007
1566-2535/

of the information of spatial neighborhoods. In [19], the information of spatial neighborhood is introduced in Evidential C-Means to deal with the problem of multi-source image segmentation, with applications to prostate multi-parametric MRI. In [20], a pixel labeling method is proposed based on evidence theory. In [21], the problem of color image segmentation is tackled by considering tristimulus *R*, *G* and *B* as three independent information sources and fusing the information provided by different sources. However, in some cases, the information of different color channel may conflict, which will lead to non-sense fusion results [22].

The main contribution of this paper is that we present a new interactive color image segmentation method by making use of the conception of Markov random fields (MRFs) and D–S evidence theory. In [18], the spatial information is introduced by fusion of basic belief assignments of neighboring pixels. But such treatment is intuitive. Since segmentation results are decided by considering both likelihood information and priori information under Bayesian framework, we here consider spatial information by generalizing it under the evidential framework. The proposed method has only one parameter. Starting with two-label segmentation, the method can be generalized to multi-label segmentation. Experimental results have demonstrated the effectiveness of our method. In [20], although the pixel labeling method is based on evidence theory, which is also the basis of our method, there are essential differences between the methods. In particular the contribution of our method includes: (1) the spatial contextual information is introduced in a Bayesian framework; (2) the assignment of non-singleton belief is inspired by [23], which considers how large the difference among the involved singletons; (3) class variances are considered in assigning the belief of an unlabeled pixel.

The rest of the paper is organized as follows. First, related works are summarized in Section 2. Then the basic conception of MRF and D–S evidence theory are introduced in Section 3. We describe in Section 4 our segmentation scheme. Experimental results are given in Section 5, and discussions and conclusions are given in Sections 6 and 7.

2. Related works

In [24], image segmentation is formulated as a labeling problem in which image pixels or features are assigned with labels. In particular, a set of *sites* and a set of *labels* are defined, with a neighborhood system representing the interrelationship between sites. The contextual constraints are integrated into energy functions under Bayesian decision rule [24]. The labeling result is obtained by different optimization methods.

Another interactive image segmentation method related to MRF is graph cut, proposed in [25,26]. It determines a globally optimal solution using a fast min-cut/max flow algorithm. The graph cut method boasts high speed, high stability and strong mathematical foundation, and has become popular. In [27], the "GrabCut" algorithm based on graph cuts is proposed. It iteratively uses the min-cut/max flow algorithm to minimize energy, instead of the one-shot algorithm in [26].

In [20], the formulation of image segmentation is also a pixel labeling problem. However unlike the above mentioned methods, the strategy adopted in that work is to firstly label those pixels with low degrees of doubt. Those with high doubt degrees are labeled progressively by using iterative regularization, which step-by-step achieve more accurate results. Although the strategy adopted in that work is similar to our proposed method, there are several differences between the two, including the assignment of belief and regularization scheme.

In [28], a region merging strategy is proposed which uses maximal-similarity mechanism to guide the process of merging. In the mechanism, the input image is firstly segmented into regions by any kind of general methods, such as mean shift or watershed. Then, a region is merged with one of its neighboring regions if the similarity between the two neighboring regions is the highest among all neighboring region pairs. The region merging process neither requires similarity threshold, nor does it depend on the image content.

Besides the above mentioned methods, there are still other different classifiers like "linear discriminant analysis (LDA) + K-nearest neighbor (KNN) classifier" [29,30], support vector machine (SVM) [7], random walks (RW) [31], lazy snapping [32], paint selection [33], etc. These methods are effective in many cases. However they may generate unsatisfactory results in complex natural scenes.

In graph cuts [25,26] or Grabcut [27], the energy function is optimized by finding the minimal cut of a graph. However the result of graph cut may be inaccurate due to the complexity of scenes and inaccurate parameter estimation. In [20] the pixel labeling process is by utilizing D–S evidence theory. But the regularization method in [20] is not in the sense of maximum a posteriori (MAP). Intuitively, data fusion has the potential to improve the performance of image segmentation. In our method we combine the D–S evidence theory with the MRF framework, yielding a new method which is in the sense of MAP. In the following sections, we present our interactive color image segmentation method based on MRF and D–S evidence theory as well as experiments and discussions.

3. Basics of MRF and D–S evidence theory

3.1. MRF image segmentation

MRF formulates image segmentation as a maximum a posterior (MAP) problem. It iteratively optimizes the class labels by maximizing the global posterior probability. Let S be the pixel set, $X = \{x_1, x_2, \cdots, x_S\}$ the observation field, and $\Omega = \{\omega_1, \omega_2, \cdots \omega_S\}$ the corresponding label field, where $\omega_i \in \{1, 2, \cdots, C\}$ can take any label of the class set $\{1, 2, \cdots, C\}$. The optimization of global posterior probability $p(\Omega|X)$ under Bayesian rule is given by [8]:

$$\begin{aligned}\hat{\Omega} &= \arg\max_{\Omega} p(\Omega|X) \\ &= \arg\max_{\Omega} \{p(X|\Omega)p(\Omega)\}\end{aligned} \tag{1}$$

Here in (1), $p(X|\Omega)$ is the likelihood probability of the observation field X conditioned by the label field Ω, while $p(\Omega)$ is the prior probability of the full scene label field. If Gaussian distribution is assumed for the observation field, we have:

$$p(X|\Omega) = \prod_{i=1}^{S} \frac{1}{\sqrt{2\pi}\sigma_i}\left[-\exp\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right] \tag{2}$$

where μ_i and σ_i are the average value and standard deviation of the Gaussian distribution corresponding to the class ω_i, respectively.

The prior probability of the label field, $p(\Omega)$, is computationally intractable. Because for an image with $I \times J$ pixels and each pixel having C possible class labels, Ω has C^{IJ} different configurations. We consider the posterior probability at the individual pixel level, rather than for the whole scene. Making use of the conception of neighborhood systems, we have:

$$\hat{\omega}_c = \arg\max_{c} p(\omega_c|x_i, \omega_{\partial i}) \tag{3}$$

where ∂i is the neighborhood of pixel i. An example of neighborhood system is shown in Fig. 1, and the neighborhood size chosen in this study is 3×3. Assuming that the central observation x_i is independent of neighboring labels, it can be shown that [34]:

$$p(\omega_i|x_i, \omega_{\partial i}) \propto p(x_i|\omega_i)p(\omega_i|\omega_{\partial i}) \tag{4}$$