

Evidential based Expectation Maximization for Image Segmentation

Yin Chen ^{1, a}, Siyu Hu ^{1, b} and Armin Cremers ^{2, c}

¹Department of mathematics, Taiyuan University of Technology, Taiyuan 030024, China

²Institute of Computer Science III, University of Bonn, Bonn 53117, Germany

^a314581243@qq.com, ^b1260857053@qq.com, ^cabc@iai.uni-bonn.de

Keywords: Expectation Maximization, Spatial Contextual Information, Dempster-Shafer's Theory (DST) of Evidence, Maximum A Posteriori (MAP).

Abstract. In statistics, expectation–maximization (EM) algorithm is an iterative method which finds the maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models depending on unobserved latent variables. In image segmentation, EM is widely used to determine the unknown parameters of different visual objects existing in an image. However, the main drawback of the EM method is that it does not consider spatial contextual information, which may entail rather noisy segmentation results. To remedy this, we develop an evidence theory based EM method (EEM) which incorporates spatial contextual information in EM by iteratively fusing the belief assignments of neighboring pixels to the central pixel. A simulated image set is used to evaluate the proposed method. Experimental results show that the new evidential method can achieve relative high accuracy.

1. Introduction

Image segmentation is the first step of many pattern recognition and computer vision tasks. The general problem of image segmentation involves the partitioning of a given image into a number of homogeneous segments, which are spatially connected groups of pixels. Meanwhile, the union of any two neighboring segments yields a heterogeneous segment. Image segmentation has been studied for many decades, and received much attention for visual surveillance, intelligent transportation, special film effects, etc. Many different techniques for improving image segmentation have been reported. For example, the methods, Markov random fields [1], [2], neural networks [3], support vector machines [4], and so on.

Although these methods produce relatively satisfactory results in many cases, they usually need parameter estimation for initialization and their results usually depend on the accuracy of parameter estimation. The mostly used parameter estimation method for incomplete data is EM [5], which, however, is inappropriate due to ignoring spatial contextual information. The inadequacy of EM often leads to the local optima of ensuing algorithms such as MRFs or wavelet domain analysis.

Data fusion has gained a lot of research interests in the last decade, e.g., fusion by Bayesian inference, fusion by probabilistic, fuzzy fusion and evidence theory, also known as DST [6,7], which is the basis of this work. Intuitively, data fusion has the potential to improve the performance of parameter estimation. To deal with the above-mentioned problem of EM, we propose an evidence theory based EM (EEM) method which incorporates spatial contextual information in each iteration of EM. Then, we test the proposed EEM method with a simulated image set.

The rest of this paper is organized as follows. Section 2 introduces the basic theory of DST and EM, then detailedly demonstrates the proposed EEM method. Experimental results are shown in Section 3. Finally, the conclusion is drawn in Section 4.

2. Method

2.1 Basics of DST.

Under DST, a frame of discernment Θ is defined, which is composed of N hypotheses H_n . The problem's solution set is $\Theta = \{H_1, H_2, \dots, H_N\}$. The power set of all 2^N propositions is defined on Θ as:

$$2^\Theta = \{\emptyset, H_1, H_2, \dots, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \dots, \Theta\} \quad (1)$$

where \emptyset represents an empty set.

A proposition A can be either a singleton H_n or a union of hypotheses. Define the basic belief assignment (bba) m assigned by a source as

$$m: 2^\Theta \rightarrow [0,1] \\ \sum_{A \subseteq \Theta} m(A) = 1 \quad (2)$$

If we consider the closed-world case, the bba satisfies $m(\emptyset) = 0$, which means that the solution belongs to the discernment frame. If A is a union of hypotheses, redistributing a part of the belief mass $m(A)$ to any subset of A is not possible without additional information. An element $A \subseteq \Theta$ is called a focal element if $m(A) \neq 0$.

The credibility (Bel) and the plausibility (Pl) functions are defined as

$$\text{Bel}(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad (3)$$

$$\text{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (4)$$

We can interpret $\text{Bel}(A)$ and $\text{Pl}(A)$, respectively, as the total amount of belief in the proposition A , and the maximum amount of belief potentially assigned to A . Functions m , $\text{Bel}(A)$ and $\text{Pl}(A)$ can be viewed as three different representations of the same information. By using Möbius transformation, these functions are transformed to each other.

According to Dempster's combination rule, the combination of J masses of belief from J distinct sources is defined as

$$m_\oplus = m_1 \oplus m_2 \oplus \dots \oplus m_J \quad (5)$$

where

$$m_\oplus(A) = \frac{m_\cap(A)}{1 - \kappa} \quad \forall A \subseteq \Theta, \quad (6)$$

$$m_\cap(A) = \sum_{A_1 \cap \dots \cap A_J = A} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_J(A_J)$$

where κ is the normalization term and is given by

$$\kappa = \sum_{A_1 \cap \dots \cap A_J = \emptyset} m_1(A_1) \cdot m_2(A_2) \cdot \dots \cdot m_J(A_J) \quad (7)$$

Term κ , with $0 \leq \kappa \leq 1$, is a conflict measure between the J information sources. When it is high, combining the sources is incoherent. Moreover, when $\kappa = 1$, the sources completely contradict each other and data fusion is not possible. There are different solutions dealing with the conflict. For example, it is proposed in [7] to avoid the normalization, and in [8] a general framework unifying different classical combination rules is developed.

2.2 Standardized EM in Image Segmentation.

For simplicity, we take two-class problem as an example to demonstrate the application of EM method in image segmentation. Within the Bayesian framework, an image is parameterized by both the probability density functions $p(X/\omega_1)$ and $p(X/\omega_2)$, and the a priori probabilities $P(\omega_1)$ and $P(\omega_2)$ of the two classes, respectively. Furthermore, the difference image is assumed to have a Gaussian mixture density distribution, which consists of two density components associated with both classes, i.e.

$$p(X) = p(X / \omega_1)P(\omega_1) + p(X / \omega_2)P(\omega_2) \quad (8)$$

where the parameters are estimated by EM, which is an unsupervised method.

The EM algorithm iteratively execute two steps until convergence, namely an expectation step (E-step) and a maximization step (M-step). The E-step uses current estimates of the parameters to compute the unknown underlying variables conditioned by observations, while new parameter estimations are obtained in the maximization step.

Updating the *a priori* probability according to current class parameters:

$$P^{t+1}(\omega_c) = \frac{\sum_{i,j} \frac{P^t(\omega_c) p^t(X(i,j)/\omega_c)}{p^t(X(i,j))}}{MN} \quad (9)$$

Updating the class parameters according to current class assignment:

$$\mu_l^{t+1}(\omega_c) = \frac{\sum_{i,j} \frac{P^t(\omega_c) p^t(X(i,j)/\omega_c)}{p^t(X(i,j))} X(i,j)}{\sum_{i,j} \frac{P^t(\omega_c) p^t(X(i,j)/\omega_c)}{p^t(X(i,j))}} \quad (10)$$

$$(\sigma_c^2)^{t+1} = \frac{\sum_{i,j} \frac{P^t(\omega_c) p^t(X(i,j)/\omega_c)}{p^t(X(i,j))} [X(i,j) - \mu_c^t]^2}{\sum_{i,j} \frac{P^t(\omega_c) p^t(X(i,j)/\omega_c)}{p^t(X(i,j))}} \quad (11)$$

where t is the number of iteration steps, and $l \in \{1,2\}$ represents the two classes, respectively. The parameter values are initialized by selecting a threshold for the two class to generate two initial pseudo-training sets for each class.

2.3 Incorporating DST in EM.

The EM parameter estimation for the difference image mainly has two drawbacks. First, it does not consider the spatial contextual information in difference images. The effect of the noise existing in the difference image is typically tackled by considering spatial contextual information. However, the EM algorithm just takes the pixel intensity into account while ignoring spatial contextual information, which may lead to inaccurate decision of pixel classes. Second, in each iteration of the EM algorithm, the parameter updating considers a pixel either to be unchanged or changed. However, the decision of a pixel is not certain in real situations. That is, we cannot assign the quantities in (10) and (11) by only considering the hard decision, instead, we should assign soft values to them.

We solve the above mentioned problems by incorporating DST in EM to utilize spatial contextual information, resulting in an evidential EM (EEM) algorithm. In order to make use of spatial contextual information in EM, we shall modify the two iteration steps in EM. Additionally, a new step is inserted between the two steps of the original EM.

In order to deal with uncertainty, the label set is augmented to $\Omega^+ = \{\omega_1, \omega_2, \omega_{12}\}$, with ω_{12} representing that at this time, we do not know the exact label of the current pixel. Different from the class conditional probability in (9)-(11) which quantifies the probability that a pixel belongs to a given class, the belief assignment of a pixel has to deal with the uncertainty in judging the pixel class. We use the model in [9] to calculate the bba, which is given by:

$$\begin{aligned} m_{ij}^t(\omega_1) &= \alpha \cdot \frac{p^t(X(i,j)/\omega_1)}{p^t(X(i,j)/\omega_1) + p^t(X(i,j)/\omega_2)} \\ m_{ij}^t(\omega_2) &= \alpha \cdot \frac{p^t(X(i,j)/\omega_2)}{p^t(X(i,j)/\omega_1) + p^t(X(i,j)/\omega_2)} \\ m_{ij}^t(\omega_{12}) &= 1 - \alpha \end{aligned} \quad (12)$$

where t represents iteration step, and $0 < \alpha \leq 1$ is a belief discounting parameter [10] whose value controls the uncertainty degree in the belief assignment. The bigger value the of it is, the less uncertainty

is in the assignment. When it equals to 1, there is no uncertainty in the assignment. It is easy to show the belief assignment in (12) satisfies the requirements in (2).

With the augmented label set and the belief assignment method defined, we can modify the updating equations for *a priori* probabilities and class parameters. The modification is straightforward, i.e., the class conditional probabilities in (9)-(11) are replaced by belief assignments. In particular, the updating of *a priori* probabilities is given by

$$P^{t+1}(\omega_1) = \frac{\sum_{i,j} \frac{m_{ij}(\omega_1)P^t(\omega_1)}{\sum_c m_{ij}(\omega_c)P^t(\omega_c)}}{MN}$$

$$P^{t+1}(\omega_2) = \frac{\sum_{i,j} \frac{m_{ij}(\omega_2)P^t(\omega_2)}{\sum_c m_{ij}(\omega_c)P^t(\omega_c)}}{MN} \quad (13)$$

$$P^{t+1}(\omega_{12}) = 1 - P^{t+1}(\omega_1) - P^{t+1}(\omega_2)$$

while the class parameters are updated by:

$$\mu_c^{t+1}(\omega_c) = \frac{\sum_{i,j} \frac{m_{ij}(\omega_c)P^t(\omega_c)}{\sum_c m_{ij}(\omega_c)P^t(\omega_c)} X(i,j)}{P^t(\omega_c) \cdot MN} \quad (14)$$

$$(\sigma_c^2)^{t+1} = \frac{\sum_{i,j} \frac{m_{ij}(\omega_c)P^t(\omega_c)}{\sum_c m_{ij}(\omega_c)P^t(\omega_c)} [X(i,j) - \mu_c^t]^2}{P^t(\omega_c) \cdot MN} \quad (15)$$

where $c \in \{1,2,12\}$, and the initial values of *a priori* probabilities and class parameters are assigned by using the same strategy in [11]. Note that the initial value of $P^0(\omega_{12})$ is not defined in [11], so for simplicity, it is computed as $P^0(\omega_{12}) = 1 - P^0(\omega_1) - P^0(\omega_2)$.

Between the iterations for updating *a priori* probabilities and class parameters, spatial contextual information can be utilized by fusing the belief assignment of pixels in the neighborhood of the current pixel. That is,

$$m_{ij}^t(\omega) = \bigoplus_{\forall (g,h) \in N(i,j)} m_{gh}^t(\omega) \quad (16)$$

where $N(i,j)$ is the neighborhood of pixel (i,j) . In this study, the neighborhood size is set to 3×3 . The algorithm of EEM is summarized as follows.

- (i) Select two thresholds for both classes and calculate the *a priori* probability and parameters for each class using the method in [7];
- (ii) Assign the bba for each pixel using (12);
- (iii) Update *a priori* probability for each class using (13);
- (iv) Fuse the bba to use spatial contextual information by using (16);
- (v) Update class parameters using (14) and (15);
- (vi) Iterate Step (ii) to (v) until convergence;
- (vii) The final pixel class is assigned by considering the bba for each pixel at convergence, i.e.

$$C_l(i,j) = \begin{cases} \omega_n, & m_{ij}(\omega_n) \geq m_{ij}(\omega_c) \\ \omega_c, & \text{otherwise} \end{cases} \quad (17)$$

The only parameter in EEM is the belief discounting parameter. Its effect on the accuracy of image segmentation is shown and discussed in Section 3. In the proposed EEM algorithm, the spatial contextual information is utilized by fusing the bba of pixels in the neighborhood of the current pixel, which is an intuitive treatment.

3. Experimental Results

To evaluate the performance of the proposed method, we perform experiments on a set of simulated images. The simulated image set is made up of an original image and three noisy versions of it. The original image is composed of a dark background with an average gray level of 80 and a bright foreground with an average gray level of 180, and polluted by Gaussian noise with zero mean and the a standard deviation of 10. Three different probability levels of salt and pepper noise, i.e. 0.1, 0.2, and 0.4, is added to the original image, resulting in three noisy versions of the original image. The simulated image set and the histograms of corresponding images are shown in Fig. 1, with the first row showing the images in the set, and the second row showing their corresponding histograms.

The final classification results of the standard EM and the proposed EEM are shown in Fig. 2, with the first row representing the results of the standard EM and the second row for the proposed EEM. From the first column to the fourth column of Fig. 2, the images correspond to the results for noise level of 0, 0.1, 0.2 and 0.4, respectively. It is clear in Fig. 2 that the results of EEM is much more accurate than those of standard EM.

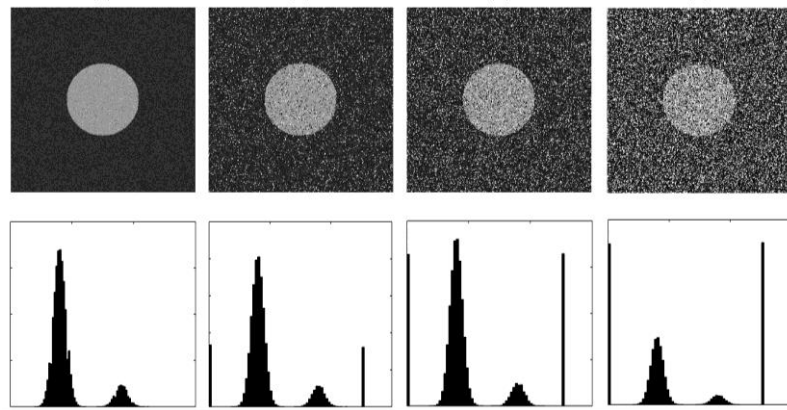


Fig. 1 Simulated image set.

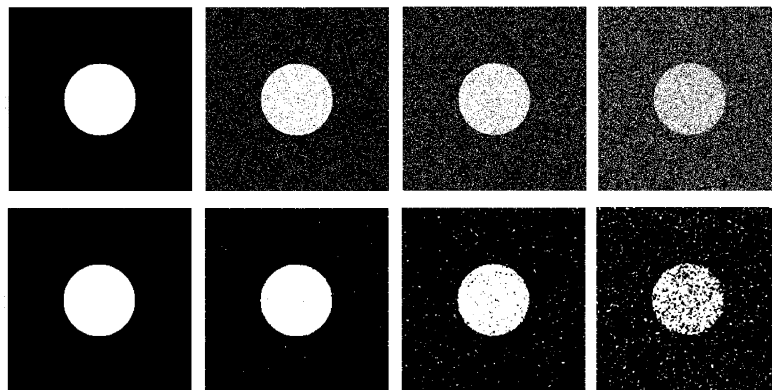


Fig. 2 Classification results for EM and EEM.

4. Conclusion

We develop two methods for unsupervised CD in multitemporal remote sensing images based on DST. We first develop an EEM method which incorporates spatial contextual information in EM method by considering the belief assignments of neighboring pixels. In each iteration of EEM method, the belief assignments of neighboring pixels are fused to the central pixel so that the belief assignment of the central pixel contains spatial contextual information. We test our proposed methods with a simulated image set. Experimental results show that the proposed EEM method can overcome the drawback of ordinary EM in parameter estimation.

References

- [1]. D. K. Panjwani, G. E. Healey. Unsupervised segmentation of textured color images using Markov random field models. in Proceedings of 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1993, p. 776-777.
- [2]. S. Wesolkowski, P. W. Fieguth. Adaptive color image segmentation using Markov random fields. in Proceedings of 2002 International Conference on Image Processing. 2002, vol. 3, pp. 769-772.
- [3]. G. Dong, M. Xie. Color clustering and learning for image segmentation based on neural networks. IEEE Transactions on Neural Networks, Vol. 16 (2005) No. 4, p. 925-936.
- [4]. J. Wang, M. Cohen. An iterative optimization approach for unified image segmentation and matting, in Proceedings of International Conference on Computer Vision. Beijing, China, 2005, p. 936-943.
- [5]. A.P. Dempster, N.M. Laird, D.B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". Journal of the Royal Statistical Society, Series B. Vol. 39 (1977) No. 1, p. 1-38.
- [6]. Dempster, Arthur P. "Upper and lower probabilities induced by a multivalued mapping." The annals of mathematical statistics. Vol. 38 (1967) No. 2, p. 325-339.
- [7]. Shafer, Glenn. A mathematical theory of evidence. Vol. 1: Princeton University press Princeton, 1976.
- [8]. Lefevre Eric, Olivier Colot, and Patrick Vannoorenberghe. Belief function combination and conflict management. Information fusion. Vol. 3 (2002) No. 2, p. 149-162.
- [9]. Appriou, Alain. 1999. Multisensor signal processing in the framework of the theory of evidence. Lecture series 216 on Application of Mathematical Signal Processing Techniques to Mission Systems: 5-31.
- [10]. Capelle, A-S, Olivier Colot, Christine Fernandez-Maloigne. "Evidential segmentation scheme of multi-echo MR images for the detection of brain tumors using neighborhood information. Information fusion Vol. 5 (2004) No. 3, p. 203-216.
- [11]. Bruzzone, Lorenzo, Diego F Prieto, Automatic analysis of the difference image for unsupervised change detection, IEEE Transactions on Geoscience and Remote Sensing, Vol. 38 (2000) No. 3, p. 1171-1182.