

# AN EXPECTATION-MAXIMIZATION APPROACH ASSISTED BY DEMPSTER-SHAFER THEORY AND ITS APPLICATION TO SONAR IMAGE SEGMENTATION

\*Tai Fei<sup>1,2</sup> and Dieter Kraus<sup>1</sup>

<sup>1</sup> IWSS, University of Applied Sciences Bremen, Neustadtswall 30, 28199 Bremen, Germany

<sup>2</sup> SPG at the Institute of Telecommunications, TU Darmstadt, Merckstraße 25,  
64283 Darmstadt, Germany

{Tai.Fei, dieter.Kraus}@hs-bremen.de

## ABSTRACT

In this paper we deal with an unsupervised segmentation approach for images given by a synthetic aperture sonar (SAS). The images with objects are segmented into highlight, background and shadow. Since the shape features are extracted from these segmented images, correctness and precision of the segmentation are highly required. We improve the expectation-maximization (EM) methods of Sanjay-Gopal *et al.* by using the gamma mixture model. Moreover an intermediate step (I-step) based on Dempster-Shafer theory (DST) is introduced between the E- and M-steps of the EM to consider the pixel spatial dependency. Finally, numerical tests are carried out on both synthetic images and SAS images. The results are compared to iterative conditional mode (ICM) and diffused EM (DEM). Our approach provides segmentations with less false alarms and better shape preservation.

**Index Terms**— Expectation-maximization algorithms, Dempster-Shafer theory, Theory of evidence, image segmentation, Clustering methods

## 1. INTRODUCTION

Most recently, with the help of synthetic aperture sonars (SAS) much attention has been attracted to process the SAS images for the purpose of underwater mine countermeasures, since SAS can provide images with very high resolution of around a few centimeters. The mine countermeasure procedure is in general divided into 3 phases: mine-like object detection, object feature extraction and mine type classification. The image segmentation takes place during the first phase. An imperfect segmentation leads to low classification accuracy since the features used for classification distort the key information in regard to the objects' shape. Moreover, a poor segmentation can also bring into many false alarms which burden the work of classifiers.

---

\* The first author is a registered PhD student at the TU Darmstadt and works for his PhD project in the University of Applied Sciences Bremen. The second author is his project supervisor.

In order to obtain an accurate segmentation, there are many algorithms proposed in the literature such as iterative conditional mode (ICM) [1] as well as expectation maximization (EM) based methods, i.e. an MAP estimator of Zhang *et al.* in [2] and the diffused EM (DEM) [3] of Boccignone *et al.* There is spatial dependency among neighboring image pixels. Thus in the energy function of ICM, *a priori* distribution is incorporated, while Zhang *et al.* replaces the pixel class probability provided by the M-step of previous iteration with an MRF based estimate. Boccignone *et al.* introduce an anisotropic diffusion [4] step between each E- and M-step to realize the spatial dependency. ICM assumes that the probability of a pixel belonging to a specific class is given by one distribution (i.e. Gaussian or gamma distribution), but not a finite mixture model as in EM based methods. Moreover, although it is assumed in the classical finite mixture model that the pixels are independently distributed, our study shows that there is still some coupling due to the sharing of the same class probability by all the pixels, which controls the probability of a certain distribution within the finite distribution mixture model.

We adopt the structure of DEM and generalize the diffusion step to an intermediate step (I-step). The distribution mixture model proposed by Sanjay-Gopal *et al.* [5] is chosen. We substitute the Gaussian distribution with the gamma distribution, since a gamma mixture model is proven to be more suitable to approximate nonnegative distributions [6] and it also fits the SAS data in our study better. In addition, we apply the Dempster-Shafer theory (DST) [7] to the pixel clustering. The assignment of class index to a pixel of interest relies on its neighborhood. The neighboring pixel provides some support that the pixel of interest belongs to the same class of this neighboring pixel. The algorithm we propose is applied to both synthetic as well as SAS images. ICM and DEM are also implemented and applied to the above images. Their results are used as benchmarks for comparison.

## 2. EM BASED SEGMENTATION ALGORITHM

### 2.1. Spatially Independent EM with Gamma Mixture Model

Let  $u_i$  be the observed intensity of pixel  $i$ ,  $1 \leq i \leq N$ , and  $\{f(u_i|\boldsymbol{\theta}_j)\}$  is a set of probability density functions (pdf),  $\boldsymbol{\theta}_j$  is the parameter vector of function  $f$ , with  $1 \leq j \leq M$ .  $M$  denotes the number of classes. The binary value vector  $\mathbf{r}_i = (r_{i,1}, \dots, r_{i,M})^T$  contains  $M$  elements. If  $r_{i,j} = 1$ , then  $r_{i,j'} = 0, j' \neq j, 1 \leq j' \leq M$ , which means that pixel  $i$  belongs to class  $j$ . Its corresponding label is then given as  $l_i = j$ . The probability of that the  $i$ -th pixel belongs to  $j$ -th class is given by  $p_{i,j} = \text{Prob}(r_{i,j} = 1)$ .

$\mathbf{D} = (u_1, \dots, u_i, \dots, u_N, \mathbf{r}_1^T, \dots, \mathbf{r}_i^T, \dots, \mathbf{r}_N^T)^T$  is the complete data, and  $\Psi = (\mathbf{p}_1^T, \dots, \mathbf{p}_N^T, \boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_M^T)^T$  is the parameter vector with  $\mathbf{p}_i = (p_{i,1}, \dots, p_{i,M})^T$ . The conditional probability density function  $f_{\mathbf{D}}(\mathbf{D}|\Psi)$  is given as

$$f_{\mathbf{D}}(\mathbf{D}|\Psi) = \prod_{i=1}^N \prod_{j=1}^M [p_{i,j} f(u_i|\boldsymbol{\theta}_j)]^{r_{i,j}}, \quad (1)$$

where the gamma pdf,  $f$ , is given as

$$f(u_i|\boldsymbol{\theta}_j) = \frac{\left(\frac{\alpha_j}{\mu_j}\right)^{\alpha_j} u_i^{\alpha_j-1} \exp\left(-\frac{\alpha_j u_i}{\mu_j}\right)}{\Gamma(\alpha_j)}, \quad (2)$$

where  $\boldsymbol{\theta}_j = (\alpha_j, \mu_j)^T$ . The E-step of iteration  $k$  is derived similarly as in [5]

$$\omega_{i,j}^{(k)} = \frac{p_{i,j}^{(k)} f(u_i|\boldsymbol{\theta}_j^{(k)})}{\sum_{m=1}^M p_{i,m}^{(k)} f(u_i|\boldsymbol{\theta}_m^{(k)})}, \quad (3)$$

where  $\omega_{i,j}$  is the expectation of  $r_{i,j}$ . We then maximize (1) and obtain the M-step,

$$p_{i,j}^{(k+1)} = \frac{\omega_{i,j}^{(k)}}{\sum_{m=1}^M \omega_{i,m}^{(k)}} = \omega_{i,j}^{(k)}, \quad (4)$$

$$\mu_j^{(k+1)} = \frac{\sum_{i=1}^N u_i \omega_{i,j}^{(k)}}{\sum_{i=1}^N \omega_{i,j}^{(k)}}. \quad (5)$$

Finally,  $\alpha_j$  can be obtained by solving the equation

$$\ln(\alpha_j) - \psi(\alpha_j) = -\frac{\sum_{i=1}^N \omega_{i,j} \ln\left(\frac{u_i}{\mu_j}\right)}{\sum_{i=1}^N \omega_{i,j}}, \quad (6)$$

where  $\psi$  is the digamma function and the left side of (6) can be substituted by an approximation given in [8].

## 2.2. Dempster-Shafer Theory assisted EM segmentation

The DST can be applied to pixel clustering. It is very similar to Bayesian probability theory, but the counterpart of pdf in DST is called basic probability assignment (bpa),  $b(A)$ . In contrast the argument of  $b$  can be either an element or a set. In general, a finite set of class indexes  $\mathcal{L} = \{1, 2, 3\}$  is called the frame of discernment, which contains all the poss-

ible states of class label  $l_i$ . Its power set is  $2^{\mathcal{L}} = \{A|A \subseteq \mathcal{L}\}$ . The bpa of certain evidence fulfills the conditions:

$$b(\emptyset) = 0, \quad (7)$$

$$\sum_{A \subseteq \mathcal{L}} b(A) = 1. \quad (8)$$

When more than one evidence is available, the information can be fused sequentially with the Dempster's rule, e.g. 2 evidences,

$$b_{1 \oplus 2}(a) = \frac{\sum_{A \cap B = a} b_1(A) b_2(B)}{1 - \sum_{A \cap B = \emptyset} b_1(A) b_2(B)}. \quad (9)$$

In our application, the neighborhood  $\mathcal{N}_i$  of pixel  $i$  can be viewed as a pool of evidences as visualized in Fig. 1. The information of the neighbors is fused through (9) one by one. And the result is independent on the fusion ordering due to the commutativity of (9). If  $u_\eta \in \mathcal{N}_i$  is of class  $j$ , then it provides some support that pixel  $i$  is also of class  $j$ . Its bpa is defined as

$$\begin{cases} b_\eta(l_\eta) = \vartheta_\eta w_\eta \\ b_\eta(\mathcal{L}) = 1 - \vartheta_\eta w_\eta \end{cases}, \quad (10)$$

where  $\vartheta_\eta$  and  $w_\eta$  can be determined by

$$\vartheta_\eta = \frac{\exp(-\gamma_1 |u_\eta - \text{med}_i|)}{\max_{\xi \in \mathcal{N}_i} \exp(-\gamma_1 |u_\xi - \text{med}_i|)}, \quad (11)$$

$$w_\eta = \exp\left(-\gamma_2 \frac{|u_i - \mu_{l_\eta}|}{\sigma_{l_\eta}}\right). \quad (12)$$

$\mu_{l_\eta}$  and  $\sigma_{l_\eta}$  are the mean value and standard deviation of class  $l_\eta$ ,  $\text{med}_i$  is the median of the pixel intensity of  $\mathcal{N}_i$ , and  $\gamma_1, \gamma_2$  are positive constants. The choice of these two free parameters will be given in the section of numerical study. The  $w_\eta$  gives the total belief portion which can be given by pixel  $\eta$ , and  $\vartheta_\eta$  counts the perfectness of the evidence itself, since the information supplied by an outlier is normally less plausible. Furthermore, since  $w_\eta$  is distance dependent, it is necessary to normalize all the distance measures into the same scale by dividing the measures with  $\sigma_{l_\eta}$  as in (12).

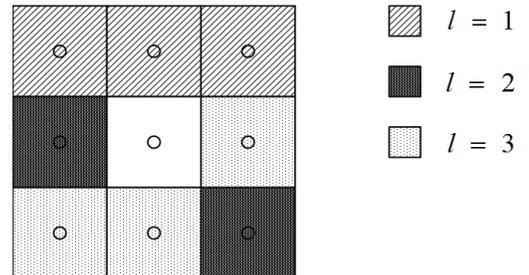


Fig. 1. The pool of evidences for the center pixel  $i$ .

The decision-making of DST is still open, and there are many proposals in the literature, such as *pignistic* probabili-

ty [9] and the expected cost function [10]. Since the simple belief structure is chosen in this paper as shown by (10), the results obtained from the *pignistic* level is identical to those from the bpa function. Hence, the decision-making at location  $i$  is given by

$$l_i = \arg \max_{j \in \mathcal{L}} b_{total}(j), j \in \mathcal{L}, \quad (13)$$

where  $b_{total}$  is the bpa induced by fusing eight neighboring evidences in  $\mathcal{N}_i$ . During the DST step, we should finally turn the label information into  $\bar{\omega}_{i,j}$  by

$$\begin{cases} \bar{\omega}_{i,j} = 1, & l_i = j \\ \bar{\omega}_{i,j'} = 0, & j' \neq j \end{cases} \quad (14)$$

Hence, the proposed method called E-DS-M can be summarized as follows:

- i. Initialization for E-DS-M
- ii. Run E-Step, and obtain  $\{\omega_{i,j}^{(k)}\}$
- iii. Perform a hard decision for  $\{\omega_{i,j}^{(k)}\}$ , then get  $\{l_i^{(k)}\}$
- iv. Run DST clustering on  $\{l_i^{(k)}\}$ , and with the help of (14) get  $\{\bar{\omega}_{i,j}^{(k)}\}$
- v. Forward  $\{\bar{\omega}_{i,j}^{(k)}\}$  to the M-step, substitute  $\omega_{i,j}^{(k)}$  with  $\bar{\omega}_{i,j}^{(k)}$  in (4), (5) and (6). And then return to step ii until the results converge.

### 3. NUMERICAL STUDY

In this section we present the experimental results on both synthetic and real SAS images. Segmentation methods are firstly tested on synthetic images, since the ground truth is available. The accuracy can be solidly compared and verified. Secondly, they are applied to the real SAS images to evaluate their generalization ability to real data. EM based methods are very sensitive to the initial inputs. Therefore we choose the initialization scheme in [11] proposed by Fandos *et al.* The pixels of the input images are clustered by  $k$ -means into a number of classes, which are more than required. The initial mean values of E-DS-M are the averages of the pixel mean values of the neighboring classes, which are obtained by  $k$ -means.

#### 3.1. Experiments on Synthetic Images

There are two synthetic images whose dimensions are  $300 \times 300$  pixels, and each contains three objects, i.e. either cylinder or truncated cone mines. The background is ripple sediment. The object regions and background are initially synthesized separately. The highlights and shadows are gamma distributed and the ripple background is simulated as proposed in [12]. The backgrounds are corrupted by multiplicative noise. Finally, we fuse the object region and background as follows

$$u_{syn} = 0.8u_{object} + 0.2u_{ripple}. \quad (15)$$

The free parameters of DST are set to  $\gamma_1 = 1.5$ , and  $\gamma_2 = 0.15$ . And the shape parameter  $\alpha_j$  in (2) is initially set to 10 as the E-DS-M inputs.

In Fig. 2 and Fig. 3, on the first row are ground truth and the synthetic images. Considering the comparison on the second row, it is obvious that the E-DS-M can suppress the ripple background much better than ICM and DEM. Moreover, we find that E-DS-M is also reliable when the object region is relatively small, e.g. when a truncated cone mine is considered.

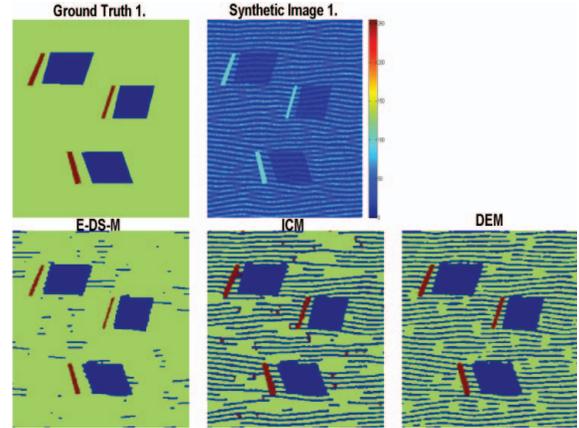


Fig. 2. Tests on Synthetic Image 1, cylinder mines.

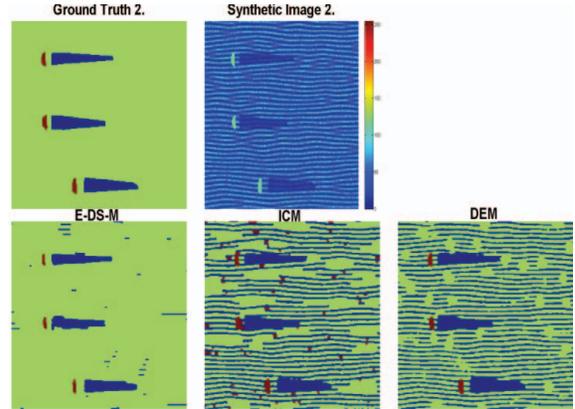


Fig. 3. Tests on Synthetic Image 2, truncated cone mines.

#### 3.2. Experiments on Real SAS Images

At last, the segmentation methods are applied on real SAS images to verify their segmentation ability. All the images are of dimension  $100 \times 100$ , and the pixel intensity lies in the interval of  $[0, 255]$ .

In Fig. 4 and Fig. 5, totally 10 images are tested by E-DS-M, ICM and DEM. Although the results of the image No. 8 and No. 9 given by all the three methods are very poor due to the low image quality, E-DS-M outperforms the other two segmentation methods in most of the cases. It preserves the shape of object region better than ICM and DEM, while

ICM is more likely to enlarge the highlight area. And DEM is inclined to introduce more noisy segmentations.

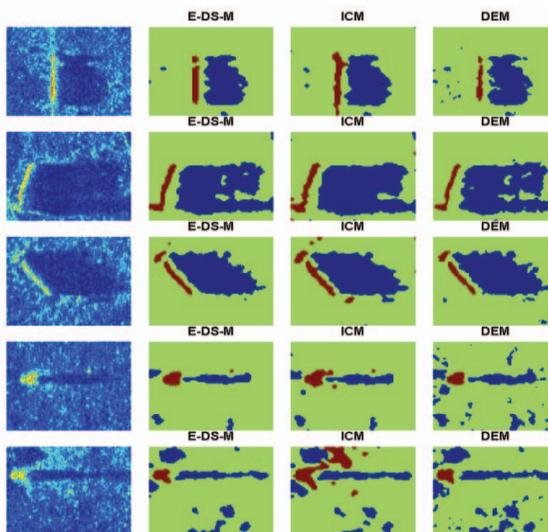


Fig. 4. Tests on real SAS images, No. 1-5.

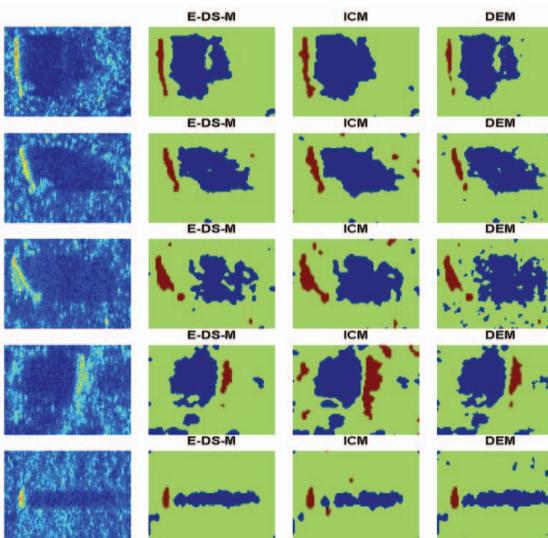


Fig. 5. Tests on real SAS images, No. 6-10.

#### 4. CONCLUSION AND FUTURE WORK

In this paper, DST pixel clustering method is incorporated into the classical EM structure to suppress the noisy segmentations when the image is corrupted. A simple belief structure is proposed to catch the belief portion provided by the evidence in the neighborhood. It considers not only the amount of belief the evidence can supply but also the perfectness of the evidence itself. And we also extend the spatial independent model of Sanjay-Gopal *et al.* to a gamma mixture model. According to our tests, the E-DS-M outperforms the classical segmentation algorithms such as ICM and DEM.

Some future work on spatial correlation of pixel labels is foreseen. DST clustering method adopting more complicated belief structures should be considered, since the current clustering depends highly on the choice of  $\gamma_1$  and  $\gamma_2$ . The manner, in which the proposed E-DS-M reacts to deviations in the initial shape parameter of the gamma distribution, should be investigated. Furthermore, the E-DS-M should also be applied to real SAS with ripple background to verify its advantages on ripple sediments.

#### 5. ACKNOWLEDGEMENT

The authors would like to thank ATLAS ELEKTRONIK for providing high quality SAS data collected with the VISION 600 system.

#### 6. REFERENCES

- [1] Besag, J., "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. Series B*, Vol. 48, No. 3, pp. 259-302, 1986.
- [2] Zhang, J., Modestino, J. W., and Langan, D. A., "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Process.*, Vol. 3, No.4, pp. 404-420, Jul. 1994.
- [3] Boccignone, G., Ferraro, M., and Napoletano, P., "Diffused expectation maximization for image segmentation," *Elec. Lett.*, Vol. 40, No. 18, pp. 1107-1108, Sep. 2004.
- [4] Weickert, J., "Applications of nonlinear diffusion in image processing and computer vision," *Acta Math. Univ. Comenianae*, Vol. 70, pp. 33-50, 2001.
- [5] Sanjay-Gopal, S., and Hebert, T. J., "Bayesian pixel classification using spatial variant finite mixtures and the generalized EM algorithm," *IEEE Trans. Image Process.*, Vol. 7, No. 7, pp. 1014-1028, Jul. 1998.
- [6] Almhana, J., Liu, Z., Choulakian, V., and McGorman, R., "A Recursive Algorithm for Gamma Mixture Models," *IEEE Int. Conf. Comm., ICC '06*, Vol. 1, pp. 197-202, Istanbul, Jun. 2006.
- [7] Yager, R., and Liu, L., "Chapter 1: Classic Works of the Dempster-Shafer Theory of Belief Functions: An introduction," *Classic Works of the Dempster-Shafer Theory of Belief Functions*, Springer-Verlag, Berlin Heidelberg, 2008.
- [8] Webb, A. R., "Gamma mixture models for target recognition," *Pattern Recognition*, Vol. 33, No. 12, pp. 2045-2054, 2000.
- [9] Smets, P., "Constructiong the pignistic probability function in a context of uncertainty," *Proc. 5. Ann. Conf. Uncertainty in Artificial Intel.*, pp. 29-39, North-Holland, Amsterdam, 1989.
- [10] Donooux, T., "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE Trans. Syst. Man Cyb.*, Vol. 25, No. 5, pp. 804-813, 1995.
- [11] Fandos, R., and Zoubir, A. M., "Enhanced Initialization Scheme for a Three-Region Markovian Segmentation Algorithm and its Application to SAS images," *Proc. European Conf. Underwater Acous.*, 2010.
- [12] Tang, D., Henyey, F. S., Hefner, B. T., and Traykovski, P. A., "Simulating Realistic-Looking Sediment Ripple Fields," *IEEE J. Oceanic Eng.*, Vol. 34, No. 4, pp. 444-450, Oct. 2009.