SAR IMAGE CLASSIFICATION WITH NON-STATIONARY MULTINOMIAL LOGISTIC MIXTURE OF AMPLITUDE AND TEXTURE DENSITIES

Koray Kayabol[1], Aurélie Voisin[1] and Josiane Zerubia [1]

[1]Ariana, INRIA Sophia Antipolis Mediterranee, 2004 route des Lucioles, BP93,
06902 Sophia Antipolis Cedex, France,
{koray.kayabol, aurelie.voisin, josiane.zerubia }@inria.fr

ABSTRACT

We combine both amplitude and texture statistics of the Synthetic Aperture Radar (SAR) images using Products of Experts (PoE) approach for classification purpose. We use Nakagami density to model the class amplitudes. To model the textures of the classes, we exploit a non-Gaussian Markov Random Field (MRF) texture model with t-distributed regression error. Non-stationary Multinomial Logistic (MnL) latent class label model is used as a mixture density to obtain spatially smooth class segments. We perform the Classification Expectation-Maximization (CEM) algorithm to estimate the class parameters and classify the pixels. We obtained some classification results of water, land and urban areas in both supervised and semi-supervised cases on TerraSAR-X data.

Index Terms— High resolution SAR, TerraSAR-X, classification, texture, multinomial logistic, Classification EM, Products of Experts

1. INTRODUCTION

The aim of SAR image classification is to assign each pixel to a class with regard to a feature space. Some of the basic features used in SAR image classification are the intensity or the amplitude and the phase of the observed image [1]. To model the statistics of such data, both empirical and theoretical probability density functions (pdfs) have been proposed [1]. Finite mixture model is a suitable statistical model to represent SAR image histogram and to perform classification. A combination of the pdfs into a finite mixture model has been used in [2].

The texture which represents the context of the image can also be used as a feature. Correlated K-distributed noise is used to capture the texture information of the SAR images in [3]. In [4], Gray Level Co-occurrence Matrix (GLCM) [5] and semivariogram [6] textural features are resorted to classify very high resolution SAR images including urban areas. In this study, we use a non-Gaussian MRF model, so-

called t-MRF, for texture representation [7]. In this autoregressive model, we express a pixel as a linear combination of its neighboring pixels. We assume that the regression error is an independent and identically distributed (iid) Student's t-distribution. t-distribution is a convenient model for robust regression and it has been used in image processing as a robust statistical model [7].

There are some methods to combine the results of different classifiers [8]. Rather than this, we construct a single classifier using the Products of Experts (PoE) approach [9] to combine both SAR amplitude and texture features into a finite mixture model. In our mixture model, we assume that each latent class label is a categorical random variable which is a special version of the multinomial random variable where each pixel belongs to only one class. For each class, we have a binary map that indicates the pixels belonging to that class. We introduce the spatial interaction of each binary map using multinomial logistic model [10] to obtain a smooth segmentation map. Note that the edge preserving segmentation is out of the scope of this paper. In this logistic regression model, the probability of the class label is proportional to a linear combination of surrounding binary pixels. In contrast to Potts-Markov image model [11], we have K different probability density functions for random fields of each class, instead of a single Gibbs distribution. The final density of the class label is constituted by combining K probability densities into a multinomial density. In this way, we obtain a non-stationary multinomial class density function which incorporates both class mixture probabilities and spatial smoothness into a single density. A non-stationary finite mixtures model been has introduced for image classification in [12]. A single model and algorithm are preferred to avoid the propagation of the error between different models and algorithms.

The EM algorithm and its stochastic versions have been used for parameter estimation in latent variable models. We use a computationally less expensive version of EM algorithm, namely Classification EM (CEM) [13], for both parameter estimation and classification, using the advantage of categorical random variables. After the classification step of CEM, we estimate the parameters of the class densities using

only the pixels which belong to that class.

In Section 2 and 3, the MnL mixture model and CEM algorithm are given. The simulation results are presented in Section 4. Section 5 presents the conclusion and future work.

2. MULTINOMIAL LOGISTIC MIXTURE OF AMPLITUDE AND TEXTURE DENSITIES

We assume that the observed amplitude at the nth pixel, s_n, where $n \in \{1, 2, \ldots, N\}$ represents the lexicographically ordered pixel indices, is free from any noise and instrumental degradation. Every pixel in the image has a latent class label. Denoting K the number of classes, we encode the class label as a K dimensional categorical vector $\mathbf{z}_n$ whose elements $z_{n,k}$, $k \in \{1, 2, \ldots, K\}$ have the following properties: 1) $z_{n,k} \in \{0, 1\}$ and 2) $\sum_{k=1}^{K} z_{n,k} = 1$. We may write the probability of s_n as the marginalization of the joint probability of $p(s_n, \mathbf{z}_n | \Theta) = p(s_n | \mathbf{z}_n, \Theta) p(\mathbf{z}_n | \theta_z)$

$$p(s_n | \Theta) = \sum_{\mathbf{z}_n} \prod_{k=1}^{K} [p(s_n | \theta_k) \pi_{n,k}]^{z_{n,k}} \qquad (1)$$

where $\pi_{n,k} = p(z_{n,k} = 1 | \theta_z)$, θ_k and θ_z are the parameters of the class and mixture densities respectively and $\Theta = \{\theta_{1:K}, \theta_z\}$ is the set of all parameters. We give the details of the class and the mixture densities in the following two sections.

2.1. Class Amplitude and Texture Densities

Our aim is to use the amplitude and the texture statistics together to classify the SAR images. For this purpose, we combine both statistics by using the idea of PoE [9]. We model the class amplitudes using Nakagami density, which is a basic theoretical multi-look amplitude model for SAR images [1]. We express the class amplitude density as

$$p_A(s_n | \mu_k, \nu_k) = \frac{2}{\Gamma(\nu_k)} \left(\frac{\nu_k}{\mu_k}\right)^{\nu_k} s_n^{2\nu_k - 1} e^{\left(-\nu_k \frac{s_n^2}{\mu_k}\right)}. \qquad (2)$$

We introduce a t-MRF texture model to use the contextual information for classification. We write the t-MRF texture model using the neighbors of the pixel in $\mathcal{N}(n)$:

$$s_n = \sum_{n' \in \mathcal{N}(n)} \alpha_{k,n'} s_{n'} + t_{k,n} \qquad (3)$$

where $\alpha_{k,n'}$ is the regression coefficient and the regression error $t_{k,n}$ is an iid t-distributed zero-mean random variable with degree of freedom parameter β_k and scale parameters δ_k. In this way, we write the class texture density as a t-distribution such that

$$p_T(s_n | \alpha_{1:D,k}, \beta_k, \delta_k) = \frac{\Gamma((1 + \beta_k)/2)}{\Gamma(\beta_k/2)(\pi \beta_k \delta_k)^{1/2}} \\ \times \left[1 + \frac{(s_n - \phi_n^T \boldsymbol{\alpha}_k)^2}{\beta_k \delta_k}\right]^{-\frac{\beta_k + 1}{2}} \qquad (4)$$

where the $D = |\mathcal{N}(n)|$ dimensional vectors ϕ_n and $\boldsymbol{\alpha}_k$ contain the neighboring pixels $s_{n'}$ and regression coefficients $\alpha_{k,n'}$, respectively.

We constitute the class density by multiplying the amplitude and texture densities: $p(s_n | \theta_k) = p_A(s_n | \mu_k, \nu_k) \times p_T(s_n | \alpha_{1:D,k}, \beta_k, \delta_k)$ where $\theta_k = \{\alpha_{1:D,k}, \beta_k, \delta_k, \mu_k, \nu_k\}$.

2.2. Mixture Density - Class Prior

The prior density $p(z_{n,k} | \theta_z)$ of the categorical random variable is naturally an iid multinomial density, but we are not able to obtain a smooth class label map in case that we use an iid multinomial. We need to use a density which models the spatial smoothness of the class labels as well. Multinomial logistic model, [10], allows us to introduce the spatial interaction of the categorical random field. We can write the MnL density for our problem as

$$p(\mathbf{z}_n | \theta_z) = \prod_{k=1}^{K} \left(\frac{\exp(\theta_z v_k(z_{n,k}))}{\sum_{j=1}^{K} \exp(\theta_z v_j(z_{n,j}))}\right)^{z_{n,k}} \qquad (5)$$

where $v_k(z_{n,k}) = 1 + \sum_{m \in \mathcal{M}(n)} z_{m,k}$. The function $v_k(z_{n,k})$ returns the number of labels which belong to class k in a given window. The mixture density in (5) is spatially-varying with given function $v_k(z_{n,k})$.

3. CLASSIFICATION EM ALGORITHM

Since our purpose is to cluster the observed image pixels by maximizing the marginal likelihood given in (1), we suggest to use EM type algorithm to deal with the summation. The EM log-likelihood function is written as

$$Q(\Theta | \Theta^{t-1}) = \sum_{k=1}^{K} z_{n,k} \log\{p(s_n | \theta_k) \pi_{n,k}\} p(z_{n,k} | s_n, \Theta^{t-1}) \qquad (6)$$

If we used the exact EM algorithm to find the maximum of $Q(\Theta | \Theta^{t-1})$ with respect to Θ, we would need to maximize the parameters for each class given the expected value of the class labels. Instead of this, we use the advantage of working with categorical random variables and resort to Classification EM algorithm [13]. The CEM algorithm incorporates a classification step between the E-step and the M-step which performs a Maximum-a-Posteriori (MAP) estimation. Since the posterior $p(z_{n,k} | s_n, \Theta^{t-1})$ is a discrete probability density function of a finite number of classes, we can perform the MAP estimation by choosing the maximum class probability. We summarize the CEM algorithm for our problem as follows:

E-step: For $k = 1, \ldots, K$ and $n = 1, \ldots, N$, calculate the posterior probabilities

$$p(z_{n,k} | s_n, \Theta^{t-1}) = p(s_n | \theta_k^{t-1}) \frac{\exp(\theta_z^{t-1} v_k(z_{n,k}))}{\sum_{j=1}^{K} \exp(\theta_z v_j(z_{n,j}))} \qquad (7)$$

given the previously estimated parameter set Θ^{t-1}.

C-step: For $n = 1, \ldots, N$, classify the nth pixel into class j as $z_{n,j} = 1$ by choosing j which maximizes the posterior $p(z_{n,k}|s_n, \Theta^{t-1})$ over $k = 1, \ldots, K$ as

$$j = \arg\max_k p(z_{n,k}|s_n, \Theta^{t-1}) \tag{8}$$

M-step: Maximize the EM log-likelihood function in (6) with respect to Θ. To maximize this function, we alternate among the variables μ_k, ν_k, $\boldsymbol{\alpha}_k$, β_k and δ_k. The functions of the amplitude parameters over all pixels are written as follows

$$Q(\mu_k; \Theta^{t-1}) = -N\nu_k \log \mu_k - \frac{\nu_k}{\mu_k} \sum_{n=1}^{N} s_n^2 \tag{9}$$

$$
\begin{aligned}
Q(\nu_k; \Theta^{t-1}) \;=\; & N\nu_k \log \frac{\nu_k}{\mu_k} - N \log \Gamma(\nu_k) + \\
& (2v - 1) \sum_{n=1}^{N} \log s_n - \frac{\nu_k}{\mu_k} \sum_{n=1}^{N} s_n^2
\end{aligned}
\tag{10}
$$

The maximum of (9) is found analytically. For (10), we use a zero finding method to determine its maximum [14].

We estimate the texture parameters using another sub-EM algorithm nested within CEM. The nested EM algorithm has already been studied in [15]. We can express the t-distribution as a Gaussian scale mixture of gamma distributed latent variables $\tau_{n,k}$. Thereby, the EM log-likelihood functions of the t-distribution in (4) are written as [16], [7]:

$$Q(\boldsymbol{\alpha}_k; \Theta^{t-1}) = -\sum_{n=1}^{N} \frac{(s_n - \phi_n^T \boldsymbol{\alpha}_k)^2}{2\delta_k} \langle \tau_{n,k} \rangle \tag{11}$$

$$Q(\delta_k; \Theta^{t-1}) = -\frac{N}{2} \log \delta_k - \sum_{n=1}^{N} \frac{(s_n - \phi_n^T \boldsymbol{\alpha}_k)^2}{2\delta_k} \langle \tau_{n,k} \rangle \tag{12}$$

$$
\begin{aligned}
Q(\beta_k; \Theta^{t-1}) \;=\; & -N \log \Gamma(\tfrac{\beta_k}{2}) + \frac{N\beta_k}{2} \log \frac{\beta_k}{2} \\
& + \sum_{n=1}^{N} \left(\frac{\beta_k}{2} \right) \langle \log \tau_{n,k} \rangle \\
& - \sum_{n=1}^{N} \frac{\langle \tau_{n,k} \rangle \beta_k}{2} \left(1 + \frac{(s_n - \phi_n^T \boldsymbol{\alpha}_k)^2}{2\delta_k \beta_k} \right)
\end{aligned}
\tag{13}
$$

where $\langle \tau_{n,k} \rangle$ is the posterior expectation of the gamma distributed latent variable.

4. SIMULATION RESULTS

This section presents the high resolution SAR image classification results of the proposed method, compared to the corresponding results obtained with other methods. We use a 900 $\times$ 600 pixels, HH polarized, TerraSAR-X SpotLight (8.2 m ground resolution) 4-look image which was acquired over the city of Rosenheim in Germany (see Fig. 1(a)).

The proposed method is denoted as ATML-CEM (Amplitude and Texture density mixtures of MnL with CEM) and is compared to three other methods. The competitors are DSEM [2], CoDSEM-GLCM [4] and K-NN-MRF. We have

Table 1. Accuracy of the classification in water, urban and land areas and overall.

	water	urban	land	overall
CoDSEM-GLCM (Sup.)	91.28	**98.82**	93.53	94.54
DSEM (Sup.)	92.95	98.32	81.33	90.87
K-NN-MRF (Sup.)	90.56	98.49	94.99	94.68
ATML-CEM (Sup.)	**98.05**	98.30	**95.87**	**97.41**
ATML-CEM (Semi-sup.)	94.53	97.85	86.09	92.82

Table 2. The confusion matrix to show the mean percentages of the true and the false classified water, urban and land areas.

	water	urban	land
water	84.40	5.51	10.09
urban	0.03	99.97	0
land	0.03	17.26	82.72

tested the ATML-CEM method in both supervised and semi-supervised cases. For supervised case, we manually determine a ground-truth map which covers 20% of the whole image and we divide it into two sets, for training and for testing. In semi-supervised case, we initialize the algorithm by manually selecting some regions (totaly about 26%) regarding to each class. We use 5×5 and 3×3 windows for autoregressive texture model in supervised and semi-supervised cases, respectively. For MnL label model, a 13×13 window is used with parameter $\theta_z = 7/13^2$ which is found by trial and error. Fig. 1(b) and 1(c) show the supervised and semi-supervised classification maps of ATML-CEM method.

Table 1 lists the accuracy of the classification in water, urban and land areas and overall according to a groundtruth map. In overall, ATML-CEM provides significantly better results than the others. Although the performance of semi-supervised ATML-CEM is not good in land areas, it surpasses the supervised DSEM in overall. Table 2 shows the confusion matrix obtained by running supervised ATML-CEM 28 times using some disjoint training and test set pairs selected over groundtruth map. From the table, we can see that the urban areas are classified better than the water and the land areas, but some of the land areas are classified as urban areas. Missclassification in water areas is caused by the dark shadowed regions.

5. CONCLUSION

We have shown that the use of amplitude and textural features in the same finite mixture model has improved the performance of classification for our data set. Different features, like edges and corners, can be added to improve the performance and classify the different areas. The performance of the algorithm depends on the initialization of the classes. As a future work, we plan to derive a hierarchical method to estimate the number of classes and to avoid the effect of initialization.
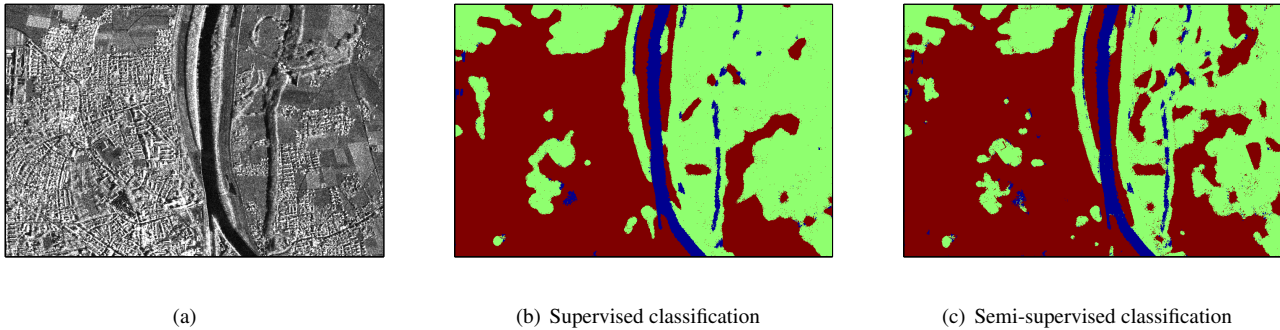
(a)　　　　　　　　(b) Supervised classification　　　　　　　　(c) Semi-supervised classification

Fig. 1. (a) 900×600 pixels, HH polarized TerraSAR-X SpotLight (8.2 m ground resolution) 4-look image which is acquired over Rosenheim, Germany. ©Infoterra, 2008. (b) and (c) Classification maps obtained by ATML-CEM method. Blue (dark), red (medium) and green (light) colors represent water, urban and land areas, respectively.

6. ACKNOWLEDGMENT

7. REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech House, Norwood, 1998.

[2] V.A. Krylov, G. Moser, S.B. Serpico, and J. Zerubia, "Dictionary-based probability density function estimation for high resolution SAR data," in *IS&T/SPIE Electronic Imaging*, 2009, vol. 7246, p. 72460S.

[3] C.J. Oliver, "The interpretation and simulation of clutter textures in coherent images," *Inverse Problems*, vol. 2, no. 4, pp. 481–518, 1986.

[4] A. Voisin, G. Moser, V.A. Krylov, S.B. Serpico, and J. Zerubia, "Classification of very high resolution SAR images of urban areas by disctionary-based mixture models, copulas and Markov random fields using textural features," in *SPIE Symposium Remote Sensing*, 2010, vol. 7830, p. 78300O.

[5] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.

[6] Q. Chen and P. Gong, "Automatic variogram parameter extraction for textural classification of the panchromatic IKINOS imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1106–1115, 2004.

[7] K. Kayabol, E.E. Kuruoglu, J.L. Sanz, B. Sankur, E. Salerno, and D. Herranz, "Adaptive Langevin sampler for separation of t-distribution modelled astrophys-

ical maps," *IEEE Trans. Image Processing*, vol. 19, no. 9, pp. 2357–2368, 2010.

[8] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 20, no. 3, pp. 226–239, 1998.

[9] G. E. Hinton, "Products of experts," in *Int. Conf. on Artificial Neural Net. ICANN'99*, 1999, vol. 1, pp. 1–6.

[10] B. Krishnapuram, L. Carin, M.A.T. Figueiredo, and A.J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 27, no. 6, pp. 957–968, 2005.

[11] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721–741, 1984.

[12] S. Sanjay-Gopal and T.J. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1014–1028, 1998.

[13] G. Celeux and G. Govaert, "A classification EM algorithm for clustering and two stochastic versions," *Comput. Statist. Data Anal.*, vol. 14, pp. 315–332, 1992.

[14] G.E. Forsythe, M.A. Malcolm, and C.B. Moler, *Computer Methods for Mathematical Computations*, Prentice-Hall, 1976.

[15] D. A. van Dyk, "Nesting EM algorithms for computational efficiency," *Statistica Sinica*, vol. 10, pp. 203–225, 2000.

[16] C. Liu and D.B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, pp. 19–39, 1995.