A Study on Classification of Polarimetric SAR Image by Target Decomposition and Support Vector Machines

JIANG Yong, ZHANG Xiao-ling, SHI Jun
(*School of Electronic Engineering University of Electronic Science and Technology of China, Chengdu* 610054)

Abstract This paper presents a new method for unsupervised classification of terrain types using polarimetric synthetic aperture radar data. This unsupervised classification combines the target decomposition theory and the support vector machines. The initial cluster centers are firstly determined by target decomposition advanced by Cloude and Pottier. Then the pixels near to the cluster centers are selected to train the support vector machines using Wishart distribution. The classified results are then used to define training sets for the next iteration if necessary. Finally, by the optimal separating hyperplanes and the kernel method this method obtains extraordinary classification results and needn't much iteration. And the effects of feature vectors consisted of several polarimetric parameters are discussed in detail.

Keywords synthetic aperture radar, radar polarimetry, image classification, support vector machine
中图法分类号:TP751.1 **文献标识码**:A **文章编号**:1006-8961(2008)08-1511-06

基于目标分解与支持向量机的极化 SAR 图像分类研究

江 勇 张晓玲 师 君
(电子科技大学电子工程学院, 成都 610054)

摘 要 为了有效地对极化 SAR 图像进行分类,基于目标分解和支持向量机,提出了一种极化 SAR 图像非监督分类法。该方法首先利用目标分解理论获得极化熵和平均散射角,并在熵-平均散射角平面对图像进行初分类,以确定类中心;然后利用 Wishart 分布定义的距离函数寻找训练样本,同时选择一定的极化参数组成特征矢量,并利用训练样本和特征矢量训练支持向量机;最后用训练好的分类器对极化 SAR 图像进行分类。通过对 ESAR 图像进行分类,比较了多种参数组合的分类结果,并与 Wishart 方法进行了比较,结果表明,该方法特征选择非常灵活,不仅结果类内离散度更小,且不需要太多的迭代次数。

关键词 合成孔径雷达 极化 图像分类 支持向量机

1 Introduction

Polarimetric SAR receives more and more attention on account of the large amount of information they convey. One of the important applications of polarimetric SAR is the classification of the terrain type. For unsupervised classification polarimetric SAR data, Cloude and Pottier proposed an algorithm based on their target decomposition theory[1]. The data are classified in the H/α plane into eight zones when the entropy H measures of randomness of the scattering and the angle α characterizes the scattering mechanism. Nevertheless, only two parameters are used,

收稿日期:2006-10-12;**改回日期**: 2007-02-27

第一作者简介:江 勇(1979~),男。2007 年 6 月获电子科技大学工学硕士学位。主要从事极化雷达图像处理及合成孔径雷达成像方面的研究。E-mail:largefield@163.com

polarimetric information is not used completely. Support vector machines (SVMs)[2] have much attention recently as promising approach to pattern recognition. SVMs are first introduced to the supervised polarimetric SAR classification[3] and then unsupervised classification[4] by S. Fukuda and H. Hirosawa. The results are successful, but the programme of selecting training sets is so complex, and they cannot expose the targets' physical scattering characteristic. For the sake of more effectively use the polarimetric parameters and select training sets, the target decomposition theory and the SVMs are combined in this paper. This combined method begins with the target decomposition, then classifies the data by the H/α plane and calculates each cluster center. The pixels near to cluster center are selected for training sets. At last we classify the polarimetric synthetic aperture radar data by SVMs iteratively. DLR E-SAR L-band images of Oberpfaffenhofen as showed in Fig. 1 ~ Fig. 3 are used for comparison, Fig. 1 ~ Fig. 3 is the S_{HH}, S_{HV}, S_{VV} data respectively. Results

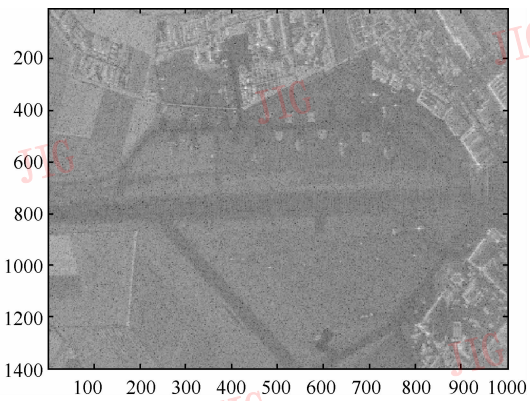

Fig. 1 S_{HH}

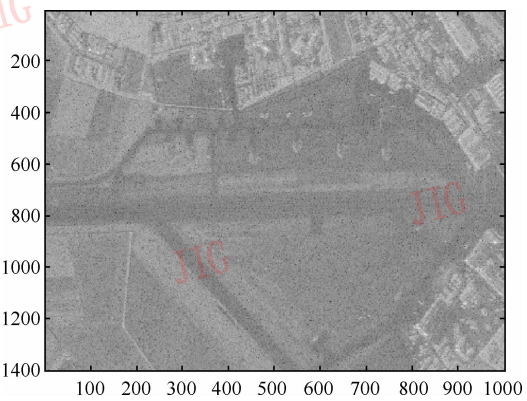

Fig. 2 S_{HV}

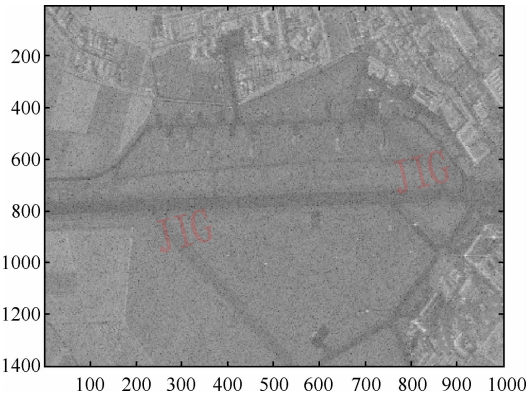

Fig. 3 S_{VV}

come from various feature vectors which consist of several polarimetric parameters and the result of H/α / Wishart method advanced by Lee[5] are compared at the last.

2 Initial Classification

The target decomposition advanced by Cloude and Pottier are introduced to obtain initial cluster center. This decomposition based on the eigenvalue analysis of the multilook coherency matrix

$$\langle \boldsymbol{T} \rangle = \frac{1}{n}\sum_{i}^{n} \boldsymbol{k}_i (\boldsymbol{k}_i^*)^{\mathrm{T}} \tag{1}$$

and $$\boldsymbol{k} = [S_{HH} + S_{VV} \quad S_{HH} - S_{VV} \quad 2S_{HV}]^{\mathrm{T}} \tag{2}$$

where the superscript "$*$" and "$\boldsymbol{T}$" denote the complex conjugate and the matrix transpose, respectively.

Then the coherency matrix $\langle \boldsymbol{T} \rangle$ can be decomposed as

$$\langle \boldsymbol{T} \rangle = \lambda_1 \boldsymbol{e}_1 (\boldsymbol{e}_1^*)^{\mathrm{T}} + \lambda_2 \boldsymbol{e}_2 (\boldsymbol{e}_2^*)^{\mathrm{T}} + \lambda_3 \boldsymbol{e}_3 (\boldsymbol{e}_3^*)^{\mathrm{T}} \tag{3}$$

where λ_i and $\boldsymbol{e}_i$ are eigenvalues and eigenvectors, respectively. The eigenvectors can be written as

$$\boldsymbol{e}_i = \mathrm{e}^{\mathrm{j}\theta i}[\cos\alpha_i \quad \sin\alpha_i\cos\beta_i \mathrm{e}^{\mathrm{j}\delta i} \quad \sin\alpha_i\cos\beta_i \mathrm{e}^{\mathrm{j}\gamma i}] \tag{4}$$

Then entropy H and the average of α are defined as

$$H = \sum_{i=1}^{3} (-p_i \log_3 p_i) \tag{5}$$

$$\alpha = \sum_{i=1}^{3} p_i \alpha_i \tag{6}$$

where

$$p_i = \frac{\lambda_i}{\left(\sum_{j=1}^{3}\lambda_j\right)} \tag{7}$$

The entropy H represents the randomness of the scattering and α stands for the mean scattering mechanism. The polarimetric SAR data are then projecting onto the H/α plane to be segmented into eight regions. The eight zones are defined as: Low entropy surface scattering; Low entropy dipole scattering; Low entropy mutiple scattering; Medium entropy surface scattering; Medium entropy vegetation scattering; Medium entropy mutiple scattering; High entropy vegetation scattering; High entropy multiple scattering.

3　Cluster Centers and Training Sets

The cluster centers of coherency matrices are defined as

$$\boldsymbol{V}_m = E\left[\frac{\langle \boldsymbol{T}\rangle}{\langle \boldsymbol{T}\rangle \in \boldsymbol{\omega}_m}\right] \tag{8}$$

where $\boldsymbol{\omega}_m$ is the set of pixels belonging to the mth class. One pixel (i,j) is selected to be class m's training set only if

$$\begin{cases} d(\langle \boldsymbol{T}\rangle_{i,j}, \boldsymbol{V}_m) < d(\langle \boldsymbol{T}\rangle_{i,j}, \boldsymbol{V}_p) & (p=1,2,\cdots,N, p\neq m) \\ d(\langle \boldsymbol{T}\rangle_{i,j}, \boldsymbol{V}_m) < d(\langle \boldsymbol{T}\rangle_{k,l}, \boldsymbol{V}_m) & ((k,l)\neq \boldsymbol{\Omega}_m) \end{cases} \tag{9}$$

where $\boldsymbol{\Omega}_m$ is class m's training sets, and $d(\langle \boldsymbol{T}\rangle, \boldsymbol{V}_m)$ is the distance between a sample coherency matrix $\langle \boldsymbol{T}\rangle$ and the cluster center $\boldsymbol{V}_m$ based on Wishart distribution proposed by Lee *et al*[6] as

$$d(\langle \boldsymbol{T}\rangle, \boldsymbol{V}_m) = \log|\boldsymbol{V}_m| + \mathrm{tr}(\boldsymbol{V}_m^{-1}\langle \boldsymbol{T}\rangle) \tag{10}$$

4　Feature Vector

To compare the polarimetric parameters, various combinations of these polarimetric parameters are presented in Tab. 1.

Tab. 1　Feature vectors

	components of feature vectors
A	$T_{1,1}\ T_{2,2}\ T_{3,3}$
B	$T_{1,1}\ T_{2,2}\ T_{3,3}\ span$
C	$T_{1,1}\ T_{1,2}T_{1,3}T_{2,2}T_{2,3}T_{3,3}$
D	$T_{1,1}\ T_{1,2}T_{1,3}T_{2,2}T_{2,3}T_{3,3}\ span$
E	$T_{1,1}\ T_{2,2}\ T_{3,3}\ \ H\ \ \alpha$
F	$T_{1,1}\ T_{2,2}\ T_{3,3}\ \ H\ \ \alpha\ \ span$

where $T_{1,1}$, $T_{1,2}$, $T_{1,3}$, $T_{2,2}$, $T_{2,3}$, $T_{3,3}$ are the components of matrix $\langle \boldsymbol{T}\rangle$, and

$$span = |S_{\mathrm{HH}}|^2 + 2|S_{\mathrm{HV}}|^2 + |S_{\mathrm{VV}}|^2 \tag{11}$$

The classification results of these feature vectors are showed in Tab. 2 ~ Tab. 3 and analyzed later.

Tab. 2　Mean distance without iterate

	Cluste1	Cluste2	Cluste3	Cluste4	Cluste5	Cluste6	Cluste7	Cluste8
A	21.249 0	15.319 6	15.607 9	12.050 8	10.915 5	11.746 9	9.630 6	6.824 3
B	20.538 3	15.594 0	15.662 7	11.505 5	10.949 4	11.383 1	9.196 2	6.201 7
C	20.635 5	15.071 2	15.075 9	11.799 1	10.807 3	11.506 3	9.824 1	6.855 5
D	20.222 3	15.075 0	15.140 7	11.449 4	10.822 8	11.245 5	9.331 1	6.234 3
E	21.106 2	15.323 2	15.609 0	11.666 3	10.743 4	11.625 6	9.773 1	7.165 3
F	20.091 3	15.595 2	15.662 6	11.143 4	10.573 8	11.342 9	9.384 1	6.781 5
G	18.911 2	18.172 4	17.856 7	13.978 3	13.635 9	13.864 7	13.303 1	13.069 0

Tab. 3　Mean distance with two iterate

	Cluste1	Cluste2	Cluste3	Cluste4	Cluste5	Cluste6	Cluste7	Cluste8
A	21.338 6	13.299 0	12.540 7	11.339 9	9.118 7	10.269 5	7.709 7	4.754 3
B	21.091 1	13.102 9	12.291 9	10.411 5	9.610 0	10.027 6	7.584 2	4.480 3
C	21.123 5	12.990 1	12.627 1	10.750 8	9.059 7	10.096 9	7.857 2	4.843 8
D	20.921 0	12.764 1	12.330 0	10.225 9	9.310 3	9.969 8	7.646 9	4.368 0
E	21.195 7	13.031 8	12.309 9	10.620 1	8.667 6	10.006 9	7.735 8	5.262 9
F	21.023 7	12.981 5	12.100 8	9.762 9	8.514 7	9.914 2	7.535 1	5.029 7
G	24.522 3	31.248 3	18.946 5	15.338 9	11.451 2	15.025 5	11.353 4	7.723 2

5 Classification Based on SVM

When training sets and feature vectors are prepared, the SVM can be conducted. For binary linear classifiers such as SVM, feature vector $\boldsymbol{x}$ is correctly segmented into both classes by a hyperplane

$$f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b, (\boldsymbol{w} \in \mathbf{R}^n, b \in \mathbf{R}) \quad (12)$$

SVM tries to maximize the margin (the distance between the separating hyperplane and training samples) to find out the optimal $\boldsymbol{w}$ and b. Given $|f(\boldsymbol{x})| |\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b| \geqslant 1$, for any training samples, there is no sample in the gap from $\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b = 1$ to $\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b = -1$, so the margin is calculated as $\frac{1}{\|\boldsymbol{w}\|}$, the maximal margin hyperplane is obtained by minimizing $\|\boldsymbol{w}\|^2$ under the constraints

$$y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geqslant 1 \quad (13)$$

where y_i is the class label. Introducing Lagrange multipliers $\alpha_i (1 \leqslant i \leqslant l, \alpha_i \geqslant 0)$, the problem turns into looking for the solutions α_i that maximize the objective function

$$L = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \quad (14)$$

Subject to

$$\sum_{i=1}^{l} \alpha_i y_i = 0$$
$$\alpha_i \geqslant 0 \quad (15)$$

After solving the equation, the decision function becomes

$$\operatorname{sgn}(f(\boldsymbol{x})) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i \langle \boldsymbol{x}_i, \boldsymbol{x} \rangle + b \right) \quad (16)$$

However, since the polarimetric SAR data are not linearly separable, the soft margin is introduced to relax the constraints in the learning phase and Gaussian kernel, which maps the input data into a high-dimensional space, is applied to instead of the inner product in Eq. 16.

Since the SVM is a kind of binary classifiers as mentioned above, we have to cope with the multiclass some modified schemes. In the following experiments one class against one method is used to deal with multiclass problem.

6 Summary of this Method

The proposed unsupervised classification method begins with the target decomposition, then classifies the data by the H/α plane and calculates each cluster center. The pixels near to cluster center are selected for training sets. At last classify the polarimetric synthetic aperture radar data by SVM iteratively. The procedure is presented as follows

(1) Calculate the entropy H, α and cluster the image by H/α classifier;

(2) Compute the cluster mean $\boldsymbol{V}_m$;

(3) Select training sets by Eq. 9;

(4) Train the SVM with the training sets;

(5) Classify the whole image by the trained SVM classifier;

(6) Repeat step 2-5 until good classification result is obtained.

Note that the class number is not fixed to eight, if the whole imaging is complex, the zones in the H/α plane can be divided into more classes.

7 Experiment Results and Analysis

To illustrate the effectiveness, the presented method is applied to classify a segment of the DLR E-SAR L-band image of Oberpfaffenhofen as displayed in Fig. 1 ~ Fig. 4. Fig. 4 is span image, and span image is much clearer compared to Fig. 1 ~ Fig. 3. Fig. 5 is the result of the H/α method, the result cannot be satisfactory by the reason of arbitrary linear boundaries and only two parameters are used. The mean distances between cluster center and the points belong to this cluster calculated by Eq. 10 are used to show the classification performance. These mean distances of the classification results without iteration and with two iterations are given in Tab. 2 ~ Tab. 3, respectively. According to the two tables, the mean distances of each cluster using the feature vector $\boldsymbol{A}$ in table1 are larger than that using the other feature vectors. It is shown that when only $T_{1,1}$, $T_{2,2}$, $T_{3,3}$

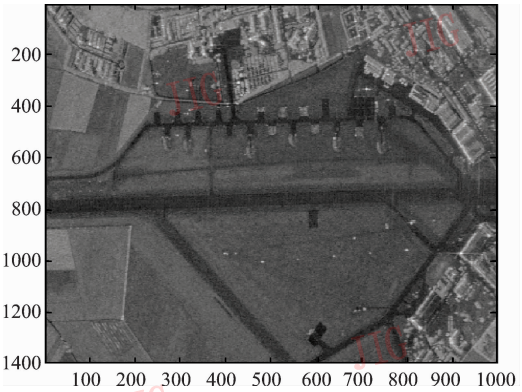

Fig. 4　Span image

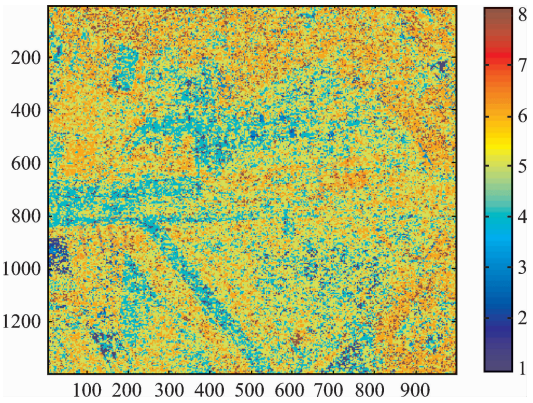

Fig. 5　Classification result of H/α method

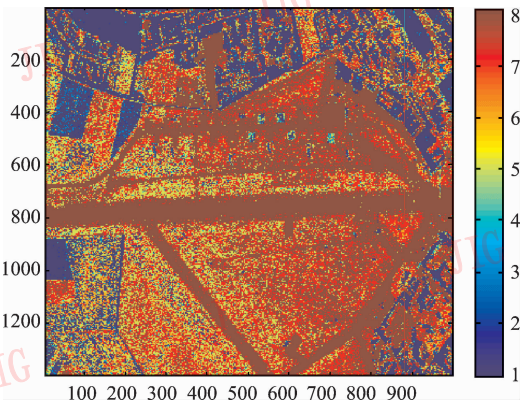

Fig. 6　Classification result of feature vector $\boldsymbol{D}$ without iteration

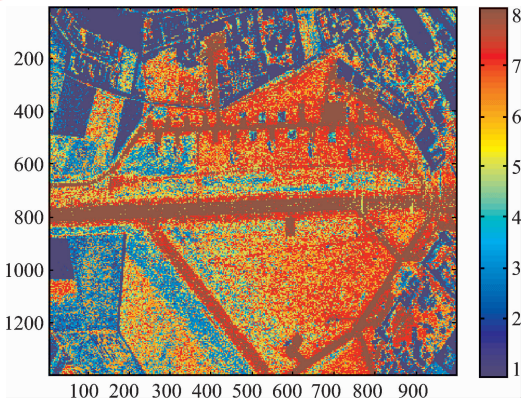

Fig. 7　Classification result of feature vector $\boldsymbol{D}$ with two iteration

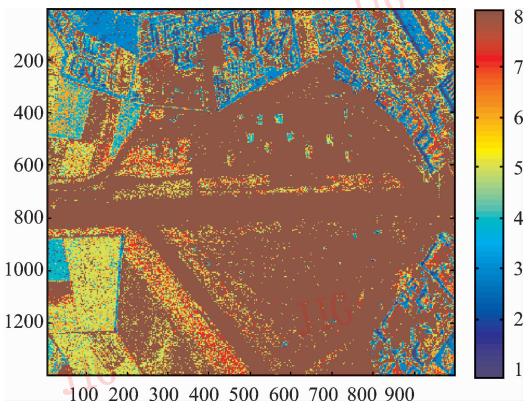

Fig. 8　Classification result of H/α /Wishart method with two iteration

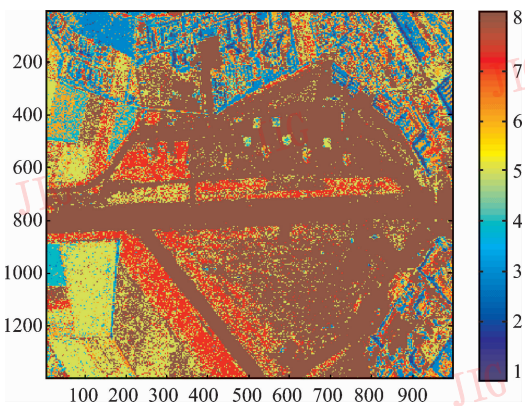

Fig. 9　Classification result of H/α /Wishart method with four iteration

are applied, the results are not as good as expected since the polarimetric information is not utilized completely. When H and α are appended as feature vector $\boldsymbol{E}$ in table1, the mean distances of each cluster don not improve obviously as shown in Tab. 2 ~ Tab. 3, since H and α are tiny relative to $T_{1,1}$, $T_{2,2}$ and $T_{3,3}$. The improvement is obvious when $T_{1,2}$, $T_{1,3}$ and $T_{2,3}$, which are comparable to $T_{1,1}$, $T_{2,2}$ and $T_{3,3}$, are added as feature vector $\boldsymbol{C}$ in Tab. 1. While span added, the results using the feature vector $\boldsymbol{B}$, $\boldsymbol{D}$, $\boldsymbol{F}$ are improved greatly compare to that using feature vector $\boldsymbol{A}$, $\boldsymbol{C}$, $\boldsymbol{E}$ respectively, it means that span is a very important parameter for classification. The mean distances of cluster 1 are large in both table, since there are more

than one class in this cluster and the results of SVM depend on training sets greatly owe to the excellent learning ability, when divide more class in the H/α plane or in iteration, these problems can be resolved. Fig. 6 ~ Fig. 7 are the corresponding classification results of feature vector $\boldsymbol{D}$ without iteration and with two iteration respectively. Fig. 6 shows that even without iteration the classification result is acceptable, and the result with two iterations become more accurate as shown in Fig. 7. The Fig. 8 ~ Fig. 9 and the rows $\boldsymbol{G}$ in the Tab. 2 ~ Tab. 3 are the results of H/α /Wishart method presented by Lee[5]. The mean distances in rows $\boldsymbol{G}$ in the Tab. 2 ~ Tab. 3 are larger than any rows in the two tables. Fig. 8 is the result of H/α /Wishart method with same iteration as Fig. 7. From Fig. 7, Fig. 8 and Tab. 2 ~ Tab. 3, the advantage of the method presented in this paper is the evident. Fig. 9 is the result of H/α / Wishart method with four iterations, although the result is acceptable, the iterative time is rather long.

8 Conclusion

In this paper, a new method for unsupervised classification of terrain types based on target decomposition theory and support vector machines is presented, which can cluster the fully polarimetric SAR data rather good and has very good flexibility. The effects of each feature vector consisted of several polarimetric parameter are discussed in detail, and it is shown that this method generates more accurate results with less iterations, comparing to the H/α /Wishart method.

References

1 Cloude S R, Pottier E. An entropy based classification scheme for land applications of polarimetric SAR[J]. IEEE Transactions on Geoscience and Remote Sensing, 1997, **35**(1), 68 ~ 78.

2 Christianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge, UK: Cambridge University Press, 2000.

3 Fukuda S, Hirosawa H. Support vector machine classification of land cover: application to polarimetric SAR data[A]. In: Proceedings of IEEE 2001 Geoscience and Remote Sensing Symposium [C], Sydney, Australia, 2001:187 ~ 189.

4 Fukuda S, Hirosawa H. Unsupervised approach for polarimetric SAR image classification using support vector machines [A]. In: Proceedings of IEEE Geoscience and Remote Sensing Symposium [C], Toronto, Canda, 2002:2599 ~ 2601.

5 Lee Jong Sen, Grunes Mitchell R. Unsupervised classification using polarimetric decomposition and the complex wishart classifier[J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, **37**(5): 2249 ~ 2258.

6 Lee Jong Sen, Grunes M R, Kwok R. Classification of multi-look polarimetric SAR imagery based on complex wishart distribution[J]. International Journal of Remote Sensing, 1994, **15**(11):2299 ~ 2311.