Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Multispectral Remote Sensing Image Classification Algorithm Based on Rough Set Theory

Ying Wang , Xiaoyun Liu , Zhensong Wang , Wufan Chen
School of Automation Engineering
University of Electronic Science and Technology of China
Chengdu, Sichuan Province, China
gfwy_01@163.com

***Abstract*—Rough set theory is a relatively new mathematical tool to deal with imprecise, incomplete and inconsistent data. A method of multispectral image classification using rough set theory is proposed. First, to decrease computational time and complexity, band reduction of multispectral image using attribute reduct concept in rough set theory and information entropy is performed. Then, mixture model initial parameters of remote sensing image are mapped from crude classes, which are generated using equivalent relation. Finally image cluster is obtained unsupervised with Gaussian mixture model whose parameters are refined by Expectation Maximization algorithm. The proposed method is performed on a multispectral image, and the experimental results show the feasibility and effectiveness of the algorithm by means of comparison and analysis.**

***Keywords*—rough set, attribute reduction, multispectral image, classification**

I. INTRODUCTION

With the fast development of spatial, digital image processing and computer technology, remote sense technology has made great progress. The appearance of multispectral and hyperspectral technology is the focus of development of remote sense technology in 21st century. Because of the large dimension of multispectral remote sensing image, correlation and data redundancy among bands; how to obtain effective reduction of multi-spectrum is the important problem to decrease computational complexity. Remote sense technology forms images for spatial objects of earth surface using satellite sensor, thus the remote sensing data had the characteristic of complicacy and uncertainty, the general mathematical statistic method can not deal with the uncertainty effectively.

Pawlak introduced rough set theory in the early 1980s [1] as a tool for representing and reasoning about imprecise or uncertain information. It is one of the most challenging areas of modern computer applications nowadays. Rough set theory is an effective tool to analyze remote sense image. There are many literatures about image classification based on rough set theory home and abroad, but all of them are almost used for gray or color images. Dr. Pal used rough set to make reduction of multispectral image [2], but there is a problem in discretization of the feature space which performed by gray-level thresholds of the individual band images, for they do not exist in such image with single-peak histogram.

This research is supported by National Natural Science Foundation of China (No.70871084)

In this paper, we focus on multispectral image dimension reduction unsupervised and image classification with statistical means. The experimental result of band reduction is compared with common bands selection methods which overcome the disadvantage of changing image characteristic [3], and the result of image cluster is shown in section V.

II. ROUGH SET THEORY

We present some preliminaries of rough-set theory that are relevant to this paper. Details have been nicely characterized in [4] and [5]. An information system is $S=(U,A,F)$, where $U=\{x_1,x_2,\cdots,x_n\}$ is a nonempty finite set called the universe, and $A=\{a_1,a_2,\cdots,a_m\}$ is a nonempty finite set of attributes, each element $a_l(l\le m)$ in A is one of the attributes. F is a relative set between U and A, $F=\{f_l:U\to V_l\ (l\le m)\}$, V_l is the value domain of $a_l(l\le m)$.

With every subset of attributes $B\subseteq A$, one can easily associate an equivalence relation I_B on U:

$$I_B=\{(x_i,x_j)\mid f_l(x_i)=f_l(x_j)\ \textit{for every}\ a_l\in B\}. \quad (1)$$

Then $I_B=\bigcap_{a\in B} I_a$.

A. Attribute Reduct

One way to increase computation efficiency is to reduce the size of data by reducing attributes. Attributes that do not contribute to the classification results can be omitted such that the indiscernibility relation remains intact, and the attributes left constitute reducts of the system.

Consider $U=\{x_1,x_2,\cdots x_n\}$ and $A=\{a_1,a_2,\cdots,a_m\}$ in the information system. By the discernibility matrix, $M(S)$ of S means an $n\times n$ matrix such that

$$c_{ij}=\{a_l\in A: f_l(x_i)\ne f_l(x_j)\}. \quad (2)$$

A discernibility function f_S is a function of m Boolean variables $\overline{a}_1, \overline{a}_2, \cdots, \overline{a}_m$ corresponding to the attributes $a_1, a_2, \cdots, a_m$, respectively, and defined as follows:

$f_S(\overline{a}_1, \overline{a}_2, \cdots, \overline{a}_m) = \wedge\{\vee(c_{ij}) : 1 \le i, j \le n, j < i, c_{ij} \ne \varnothing\}$. (3)

Where $\vee(c_{ij})$ is the disjunction of all variables $\overline{a}$ with $a \in c_{ij}$. $\{a_{i_1}, a_{i_2}, \cdots, a_{i_p}\}$ is a reduct in S if and only if $a_{i_1} \wedge a_{i_2} \wedge \cdots \wedge a_{i_p}$ is a prime implicant (constituent of the disjunctive normal form) of f_S.

A decision system $S = (U, A, F, d)$ is similar with an information system, where A called condition attributes, d is decision attribute, V_d is the value domain of d.

B. Description of Attributes Discretization Based on Rough Set Theory

Decision system $S = (U, A, F, d)$, V_l is the value domain of attribute a_l. For $a_l \in A$, on $V_l = [b_l, t_l] \subset R$, any cuts set $\{(a_l, c_1^l), (a_l, c_2^l), \cdots, (a_l, c_k^l)\}$ is defined as the following partition: $P_l = \{[c_0^l, c_1^l), [c_1^l, c_2^l), \cdots [c_k^l, c_{k+1}^l]\}$, where $b_l = c_0^l < c_1^l < \cdots < c_{k+1}^l = t_l$, $V_l = [c_0^l, c_1^l) \bigcup [c_1^l, c_2^l) \cdots \bigcup [c_k^l, c_{k+1}^l]$. Therefore, any set $P = \bigcup P_l$ defines a new decision table $S^P = (U, A^P, F, d)$, where $A^P = \{a_l^P : f_l^P(x) = i \Leftrightarrow f_l(x) \in [c_i^l, c_{i+1}^l]\}$, for $x \in U$, $i \in \{0, 1, \cdots, k\}$. The original system S has been replaced by a new one after discretization [5].

III. GAUSSIAN MIXTURE MODEL AND EM ALGORITHM

Statistical methods are widely used in unsupervised pixel classification framework because of their capability of handling uncertainties arising from both measurement error and the presence of mixed pixels [2].

The Gaussian mixture model approximates the data distribution by fitting g component density functions $p_j, j = 1, 2, \cdots, g$ to a dataset $\boldsymbol{D}$. For $\boldsymbol{x} \in \boldsymbol{D}$, the mixture model probability density function evaluated at $\boldsymbol{x}$ is

$$p(\boldsymbol{x}) = \sum_{j=1}^{g} \pi_j p_j(\boldsymbol{x}; \boldsymbol{\theta}_j) = \sum_{j=1}^{g} \pi_j \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j)\right\}}{(2\pi)^{n/2} \left|\Sigma_j\right|^{1/2}}. \quad (4)$$

The weights π_j denote the fraction of data points belonging to model j, and they sum to one $\left(\sum_{j=1}^{g} \pi_j = 1\right)$. The functions $p_j(\boldsymbol{x}; \boldsymbol{\theta}_j), j = 1, 2, \cdots, g$ are the component density functions modeling the points of the jth cluster. $\boldsymbol{\theta}_j = \left(\boldsymbol{\mu}_j, \Sigma_j\right)$ represents the specific parameters used to compute the value of p_j [6].

The Expectation Maximization (EM) algorithm is an effective and popular technique for estimating the mixture model parameters. It iteratively refines an initial cluster model to better fit the data and terminates at a solution that is locally optimal for the underlying clustering criterion. An advantage of EM is that it is capable for handling uncertainties due to mixed pixels, but one disadvantage is that the convergence speed and solution of EM algorithm depend strongly on initial condition. Dr. Pal initialized the mixture model parameters using equivalence class which partitioned by equivalence relation in rough set [2].

IV. PROPOSED ALGORITHM FOR BAND REDUCTION OF MULTISPECTRAL IMAGE

Remote sensing images can be analyzed using rough set theory. However, since the value domains of attributes are too wide to decrease computational time, the attributes discretization process of multispectral image is required. In literature [2], Dr. Pal used gray-level threshold for discretization of 4-band multispectral images, which is not suitable for image with single peak histogram.

A. Attributes Discretization of Remote Sensing Image

Different requirements of attribute discretization in rough set are needed in different application fields. In this paper, we choose information entropy to discretize the spectral vector of multispectral image, which keeps the compatibility of decision table. Documents about discrete threshold in rough set theory, one may refer to [7] and [8].

Decision system $S = (U, A, F, d)$, subset $X \subseteq U$, its cardinality is $|X|$, the number of examples with decision attribute j are k_j, $(j = 1, 2, \cdots, r(d))$. Information entropy of subset X is:

$$H(X) = -\sum_{j=1}^{r(d)} p_j \log_2 p_j, \quad p_j = k_j / |X|. \quad (5)$$

Obviously, $H(X) \ge 0$, the smaller the $H(X)$ is, the lower the chaos is. This property ensures that the algorithm can not change the compatibility of decision table.

For any $a_l \in A$, finite attribute values are sorted as: $b_l = v_0^l < v_1^l < \cdots < v_m^l = t_l$, the candidate cuts set can be chosen as

$$c_i^l = (v_{i-1}^l + v_i^l)/2 \; (i = 1, 2, \cdots, m). \quad (6)$$

For breakpoint c_i^l, in all examples which decision attribute value be $j (j = 1, 2, \cdots, r(d))$, the number of examples belonged to X which attribute a_l's value lower than c_i^l is $b_j^X(c_i^l)$, otherwise, $t_j^X(c_i^l)$. Then, we have:

$$b^X(c_i^l) = \sum_{j=1}^{r(d)} b_j^X(c_i^l), \quad t^X(c_i^l) = \sum_{j=1}^{r(d)} t_j^X(c_i^l). \quad (7)$$

Obviously, breakpoint c_i^l divided X into two parts X_b and X_t,

$$H(X_b) = -\sum_{i=1}^{r(d)} p_j \log_2 p_j, p_j = b_j^X(c_i^l) / b^X(c_i^l)\,. \quad (8)$$

$$H(X_t) = -\sum_{i=1}^{r(d)} q_j \log_2 q_j, q_j = t_j^X(c_i^l) / t^X(c_i^l)\,. \quad (9)$$

So, define information entropy for breakpoint c_i^l on X :

$$H^X(c_i^l) = \frac{|X_b|}{|X|} H(X_b) + \frac{|X_t|}{|X|} H(X_t)\,. \quad (10)$$

The breakpoint c^l for which $H^U(c^l)$ is minimal among all the candidate breakpoints is taken to be the best breakpoint, c^l partitions U into two sets U_1 and U_2. Let c_1^l be the best break point with $H^{U_1}(c_1^l)$, and c_2^l be the best breakpoint with $H^{U_2}(c_2^l)$. If

$$H^{U_1}(c_1^l) > H^{U_2}(c_2^l)\,, \quad (11)$$

we partition U_1 , otherwise we partition U_2 . The worse one is divided. Now we can get cuts set C using (10) and (11).

B. *Band Reduct Algorithm Based on Rough Set Theory*

Information entropy based discretization does not change the compatibility of decision table, but it has the limitation that the decision table generated from remote sensing image supervised is needed, which not only costs lots of time but also brings human effect. The proposed band reduct algorithm can obtain decision table and band reduction unsupervised. The block diagram of band reduct algorithm is shown in Fig. 1.

Algorithm steps are as follows:

1) Map the multispectral image to an information system, and each band is an attribute column.

2) Divide the universe into equivalence classes using equal interval discretization and indiscernibility relation, then append different class labels as decision values to each equivalence class.

3) A decision table is consisted of the original information system and the decision attribute column appended, which is unsupervised rather than supervised sampling.

4) Obtain candidate break points using (6) for each attribute of the decision table, repeat (10) and (11) to get final cuts set.

5) Reduce the attributes according to discernibility matrix (2) and discernibility function (3), now we get the final band reduct.

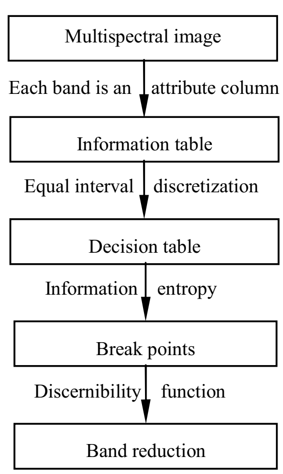
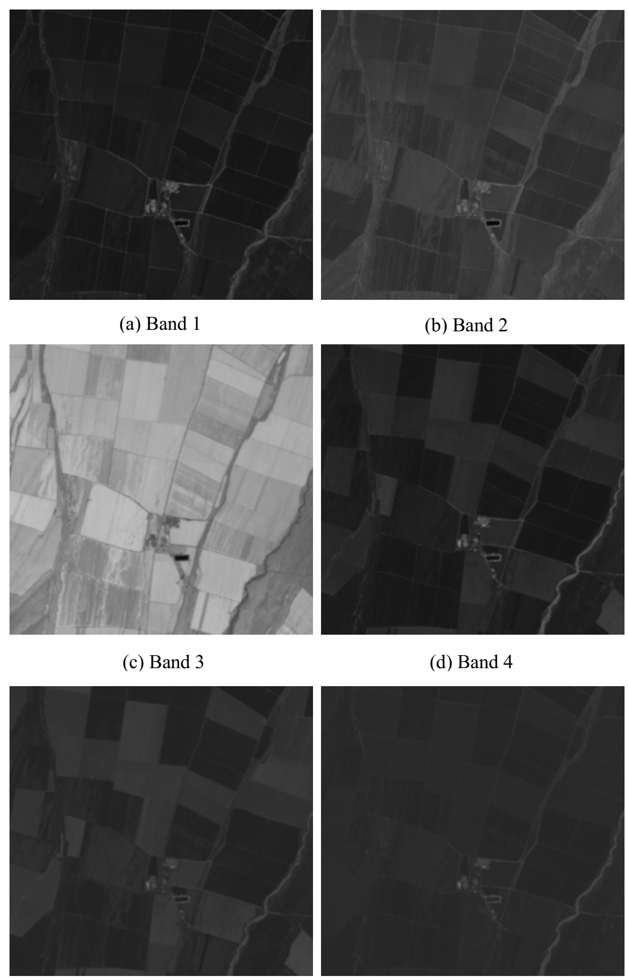


Figure 1. Block diagram of band reduct algorithm.

V. Classification of Multispectral Image and Experimental Results

After the band reduct in section IV, to obtain classification of multispectral image using Gaussian mixture model which parameters are refined by EM algorithm. Results are presented on 6-band TM image which size is 263×275. Fig. 2 shows each band gray image of the multispectral image.

For the 6-band multispectral image, the information system is mapped and a small part of the data is chosen as seen in Table I. All the attributes are discretized using equal interval discretization with the width 30, Table II shows the data after



(a) Band 1 (b) Band 2

(c) Band 3 (d) Band 4

(e) Band 5 (f) Band 6

Figure 2. 6-band remote sensing image.

attribute discretization. Different class labels are given to each class after divided the information table into different equivalent classes by means of indiscernibility relation; append the class labels to the information table as decision attribute, as given in Table III. Now we change an information system to a decision table which is combined with original image data and the appended decision attribute column, which is presented in Table IV. Breakpoints are obtained from attributes discretization using information entropy. After discretization, we get the information system Table V. Then reduce attributes according to (2) and (3), here the reduct are bands {1,3,5}. Finally, map the reduct to the initial parameters of Gaussian mixture model and refine them to convergence using EM algorithm, and classify the image according to Bayesian classification rule; for details, one may refer to [2].

TABLE I. ORIGINAL REMOTE SENSING DATA

U	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
1	23	46	163	34	49	43
2	24	46	164	34	49	44
3	23	46	164	34	50	44
4	29	54	146	38	49	43
5	34	66	127	29	39	40
6	29	61	128	29	38	38
7	29	62	117	28	36	38
8	34	65	114	29	38	39
9	34	62	126	32	40	43
10	31	59	125	29	38	42
	…	…	…	…	…	…

TABLE II. EQUAL INTERVAL DISCRETIZATION DATA

U	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
1	1	2	6	2	2	2
2	1	2	6	2	2	2
3	1	2	6	2	2	2
4	1	2	5	2	2	2
5	2	3	5	1	2	2
6	1	3	5	1	2	2
7	1	3	4	1	2	2
8	2	3	4	1	2	2
9	2	3	5	2	2	2
10	2	2	5	1	2	2
	…	…	…	…	…	…

TABLE III. DIFFERENT CLASS LABELS APPENDED

U	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Class label
1	1	2	6	2	2	2	1
2	1	2	6	2	2	2	1
3	1	2	6	2	2	2	1
4	1	2	5	2	2	2	12
5	2	3	5	1	2	2	13
6	1	3	5	1	2	2	18
7	1	3	4	1	2	2	22
8	2	3	4	1	2	2	19
9	2	3	5	2	2	2	2
10	2	2	5	1	2	2	11
	…	…	…	…	…	…	…

TABLE IV. DECISION TABLE

U	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Class label
1	23	46	163	34	49	43	1
2	24	46	164	34	49	44	1
3	23	46	164	34	50	44	1
4	29	54	146	38	49	43	12
5	34	66	127	29	39	40	13
6	29	61	128	29	38	38	18
7	29	62	117	28	36	38	22
8	34	65	114	29	38	39	19
9	34	62	126	32	40	43	2
10	31	59	125	29	38	42	11
	…	…	…	…	…	…	…

TABLE V. INFORMATION ENTROPY DISCRETIZATION DATA

U	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
1	1	1	3	4	5	3
2	1	1	4	4	5	3
3	1	1	2	1	1	2
4	3	5	7	2	4	3
5	1	1	2	4	5	3
6	1	1	1	1	1	1
7	1	1	3	1	1	1
8	1	1	4	1	1	1
9	1	1	4	1	1	2
10	1	1	1	1	1	2
	…	…	…	…	…	…

The performance of the proposed band reduct method is compared with the common band selection means [3] as in Table VI - Table VIII. The order of band selection according to standard deviation is 3>4>2>5>1>6, band3 will be chosen if only one band is requested. The order of Optimum Index Factor (OIF) with bands combined is {2,3,5}>{2,3,4}>{1,3,5}>…, but the correlation of bands{1,3,5} is better than {2,3,4}. So, as we have seen, the reduct bands {1, 3, 5} obtained form proposed approach is valid.

Figs. 3-6 are demonstrated the validity and applicability of the proposed method for multispectral image cluster. Fig. 4-Fig. 6 are classified image of 6 clusters using different methods. Compared with false-color image in visual effect, misclassification of the river right side of the image mixed with its right land exists in both Fig.3 and Fig.4. Therefore, not all the bands are essential to classify image with comparison above; choosing suitable bands can not only improve accuracy of classification, but also decrease computational time and complexity.

TABLE VI. STANDARD DEVIATION OF DIFFERENT BANDS

Band	1	2	3	4	5	6
SD	7.757	8.508	21.014	9.533	8.279	3.705

TABLE VII. CORRELATION MATRIX OF DIFFERENT BANDS

	Band1	Band2	Band3	Band4	Band5	Band6
Band1	1	0.807	0.197	0.428	0.152	0.688
Band2		1	0.056	0.305	0.144	0.463
Band3			1	0.0005	0.118	0.051
Band4				1	0.822	0.633
Band5					1	0.387
Band6						1

TABLE VIII. OPTIMUM INDEX FACTOR

Bands	123	124	125	126	134
OIF	35.172	16.749	22.256	10.2	61.214
Bands	135	136	145	146	156
OIF	79.319	34.703	18.236	12.006	16.1
Bands	234	235	236	245	246
OIF	108.16	118.94	58.367	20.699	15.523
Bands	256	345	346	356	456
OIF	20.624	41.249	50.055	59.363	11.68

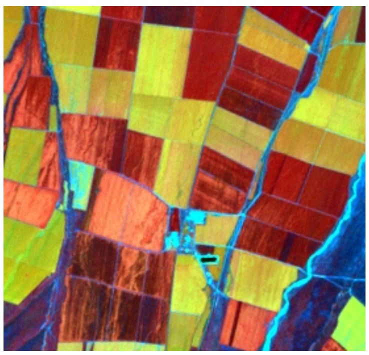


Figure 3. Original false-color image with bands1,3,5.

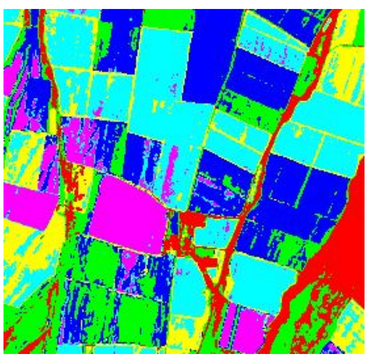


Figure 4. Classification using k-means method with bands1,3,5.

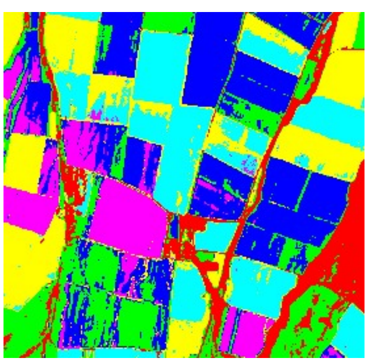


Figure 5. Classification using k-means method with 6 bands.

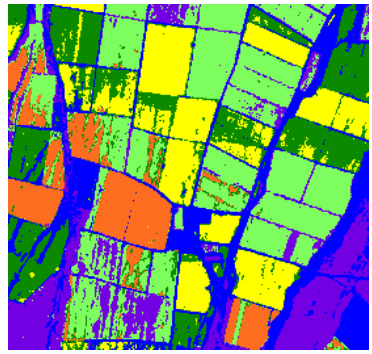


Figure 6. Classification using proposed method.

VI. CONCLUSION AND DISCUSSION

The content of this paper is twofold. First, to decrease computational time, band reduct is performed on multispectral image using rough set and information entropy. Second, image classification is obtained after band reduction, Gaussian mixture model and EM algorithm are considered. The algorithm designed in this paper can make bands reduction and multispectral image classification unsupervised. Experimental results show that the proposed method did have effective and valid performance. But there are still some miss clustering problems to be solved.

REFERENCES

[1] Z. Pawlak, "Rough sets," Int. J. Inform. Comput. Sci., vol. 11, no. 5, pp. 341-356, 1982.

[2] S. K. Pal and P. Mitra, "Multispectral image segmentation using the rough-set-initialized EM algorithm," IEEE Trans. Geosci. Remote Sensing, vol. 40, pp. 2495-2501, November 2002.

[3] C. Zhao, W. Chen and L. Yang, "Research advances and analysis of hyperspectral remote sensing image band selection," J. Natural Sci. Heilongjiang Univers., vol. 24, no. 5, pp. 592-602, October 2007.

[4] W. Zhang and G. Qiu, Uncertain decision making based on rough sets, Beijing: Tsinghua University Press, 2005.

[5] S. Hu, Y. He, Rough decision making theory and application, Beijing: Beihang University Press, 2006.

[6] A. R. Webb, Statistical Pattern Recognition (Second Edition), Beijing: Publishing House of Electronics Industry, 2004. pp. 27-40.

[7] U. M. Fayyad and K. B. Irani, "On the handling of continuous-valued attributes in decision tree generation," Machine Learning, vol.8, pp. 87-102, 1992.

[8] H. Xie, H. Cheng and D. Niu, "Discretization of continuous attributes in rough set theory based on information entropy," Chinese J. Comput., vol. 28, no. 9, pp. 1570-1574, September 2005.