

Study on Increasing the Accuracy of Classification Based on Ant Colony algorithm

YU Ming ^{a,c*}, CHEN Da-Wei ^b, DAI Chen-Yan ^a, LI Zhi-Lin ^{c,d}

^a College of Geographical Science, Fujian Normal University, Shangsang Road, Fuzhou, Fujian, China
(gisyuming@163.com)

^b School of Sociology & Anthropology, Sun Yat-Sen University, Guangzhou, Guangdong, China
(suyouyancheng@gmail.com)

^c Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University

^d Faculty of Geosciences and Environmental Engineering, Southwest Jiao Tong University
(dean.ge.swjtu@gmail.com)

KEY WORDS: Remote Sensing Image; Increasing the Accuracy of Classification; Ant Colony Algorithm; LUCC

ABSTRACT:

The application for GIS advances the ability of data analysis on remote sensing image. The classification and distill of remote sensing image is the primary information source for GIS in LUCC application. How to increase the accuracy of classification is an important content of remote sensing research. Adding features and researching new classification methods are the ways to improve accuracy of classification. Ant colony algorithm based on mode framework defined, agents of the algorithms in nature-inspired computation field can show a kind of uniform intelligent computation mode. It is applied in remote sensing image classification is a new method of preliminary swarm intelligence. Studying the applicability of ant colony algorithm based on more features and exploring the advantages and performance of ant colony algorithm are provided with very important significance. The study takes the outskirts of Fuzhou with complicated land use in Fujian Province as study area. The multi-source database which contains the integration of spectral information (TM1-5、TM7、NDVI、NDBI) and topography characters (DEM、Slope、Aspect) and textural information(Mean、Variance、Homogeneity、Contrast、Dissimilarity、Entropy、Second Moment、Correlation) were built. Classification rules based different characters are discovered from the samples through ant colony algorithm and the classification test is performed based on these rules. At the same time, we compare with traditional maximum likelihood method, C4.5 algorithm and rough sets classifications for checking over the accuracies. The study showed that the accuracy of classification based on the ant colony algorithm is higher than other methods. In addition, the land use and cover changes in Fuzhou for the near term is studied and display the figures by using remote sensing technology based on ant colony algorithm. In addition, the land use and cover changes in Fuzhou for the near term is studied and display the figures by using remote sensing technology based on ant colony algorithm. The causes of LUCC have been analysed and some suggestions to the development of this region were proposed.

1. INTRODUCTION

The classification by extracting of remote sensing (RS) data is the primary information source for GIS in land resource application (Pei Tao, Zhou Chenghu, Han zhijun, Wang Min, Qin Chengzhi and Cai Qiang. 2001; Treitz P., Howarth P. 2000; Yu Ming, Ai Ting-Hua. 2007). Automatic and accurate mapping of region land use and coverage changing (LUCC) from high spatial resolution satellite image is still a challenge. Currently the most commonly used traditional image classification methods based on spectrum, but the spectral based classification cannot obtain good result due to the spectral dimension shortage in spite of remote sensing image classification is an important means of extracting information (Erika, Lepers, Eric F. Lambin et al. 2005; Chen Shu-Peng, Tong Qing-Xi and Guo Hua-Dong. 1998). We must improve the classifying method to solve the uncertainty of classification. The methods on spatial data mining and knowledge discovery have been applied in some fields(Li De-ren, Wang Shu-liang, Li De-yi and Wang Xin_zhou. 2002; Li Shuang, Ding Sheng-Yan, Hu Shu-ming. 2002; Wang Hai-qi, Wang

Jin-feng. 2005; Chen Zhong-xiang, Yue Chao-yuan. 2003). Author has studied on C4.5 algorithm and rough sets and the combination of C4.5 algorithm and rough sets(Yu Ming, Ai Ting-Hua. 2009).The paper discussed remote sensing image data classification techniques based on ant colony algorithm under the support different variable and compared with the other algorithm. The results have been shown that we improve the classification accuracy. It is applied to LUCC.

2. STUDY AREA AND DATA

2.1 Study Area

This study is on the basis of Fuzhou city, as well as the area connecting city and country of the size of 512×512 . The geographical location of Fuzhou is in $25^{\circ}15'N - 26^{\circ}29'N$, $118^{\circ}08'E - 120^{\circ}31'E$, the southeast edge of the Eurasian continent, in the eastern of Fujian province in China's southeastern coastal, the lower reaches of the Minjiang River, the East to the Donghai, the location picture of the study area is shown in Figure 1.

* Corresponding author. This is useful to know for communication with the appropriate person in cases with more than one author.

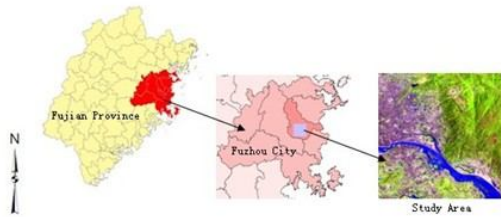


Figure1. Relative locations of study area

The moves to date have been better developing in Fuzhou. Our data come from early research item basic databases(Yu Ming, Ai Ting-Hua. 2009). We mainly consider the following principles are very important on classifications: the principle of unity, scientific principles, applicability principle, territorial principles, systemic principles, monitor principles of remote sensing technology. Based on the above principles and the eight categories demarcation methods of LUCC in conjunction with the characteristics of regional landuse, the types of LUCC connecting city and country are defined: water land, grassplot, wild green land, forest land, construction land, paddy fields, urban green land, and garden plot.

2.2 Data source

Remote sensing data the paper used is landsat7 ETM+ of Fuzhou after corrected at 10:30, on 4th May 2000; on 10th May 2003; on 8th May 2007; on 3th March 2009; the space number is 119/42. However the test area carried out classification is the size of 512×512 by bands with TM1-TM5、TM7 and panchromatic band with 15 meters resolution, which is intersecting region connecting city and country with the complicated landuse. And the assistant image is the SPOT5 image of Fuzhou already corrected in December fourteenth 2003, whose multi-spectral band is 10 meters resolving power and the panchromatic band is 2.5 meters. Non-RS data are regionalism map, landuse vector map, and contour vector data of Fuzhou whose scale is one to one hundred thousand.

3. METHOD

Ant Colony algorithm is a bionics technical method who M Dorigo and V Maniezzo put forward to look for optimization schemes(Colomi A, Dorigo M, Maniezzo V, *etal.*1999; Dorigo M. 1992; Parp Inell R S, LopesH S, FreitasA A. 2002). there are some Ant Colony algorithm application on RS image classification (Wang Shugen, Yang Yun, Lin Ying, Cao Chonghua,2009). But A Rare Few did it. The paper discussed remote sensing image data classification techniques based on Ant Colony algorithm. Work flow chart see figure 2. Steps are as follows:

Step1: Building classificatory system on LUCC.

According to RS images data and references(GB/T21010-2007), we got scheme, e.g. table 1 and table 2. The bases of our work have been confirmed.

I	II
cultivated land	paddy fields
garden plot	garden plot
forest land	forest land
grass land	urban green land
	Wild green land
Water body	Water body
construction land	construction land

Table1. Land use classification scheme of the test area

surface features	Distribution features	visual image (RGB543)	visual image (RGB432)
paddy fields	Rivers housing estate	Green	dark greenish blue
garden plot	Housing estate road	bright green	Pink
forest land	hilly area	Dark green, third dimension	Dark red, third dimension
urban green land	housing estate	green, regular	Formal red
Wild green land	hilly area	bro wnish red brownish red	yellow green
Water body	Rivers Lakes reservoir	Blue and dark blue Banding and irregular blocky	blue-green and dark green, Banding and irregular blocky
construction land	housing estate roads land for mining and industry	Purple	bluish white and white

Table2. Visual characteristics of the Land use of the test area

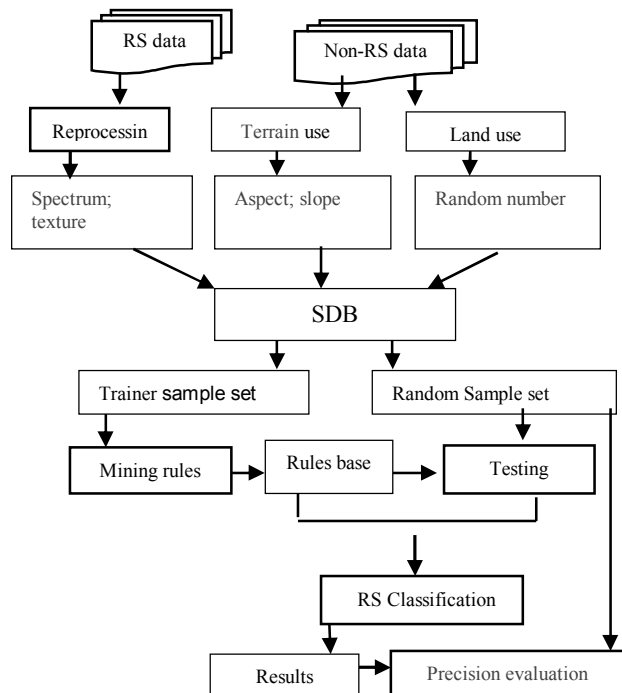


Figure2. Work flow chart

Step 2: Building spatial database.

We selected 19 correlative data from early research item basic databases as classificatory characters which included DEM, slope, aspect, NDVI, NDBI and 8 texture characters, 6 gray bands to form decision-making table about spatial information. The test integrates multi-source and multi-dimensional spatial data to build spatial database.

Step 3: Classify by ant colony algorithm under the support of different variables and compares their precision (see section 4)

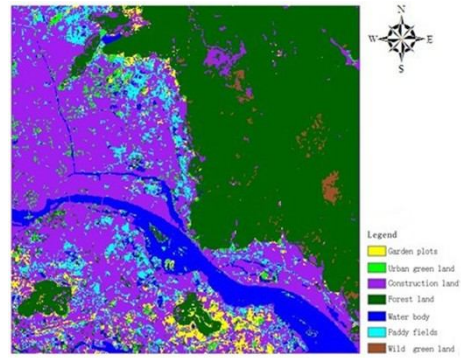
Step4: Four methods are compared and discussed under the support of 19 variable (see section 4)

4. COMPARISON AND DISCUSS

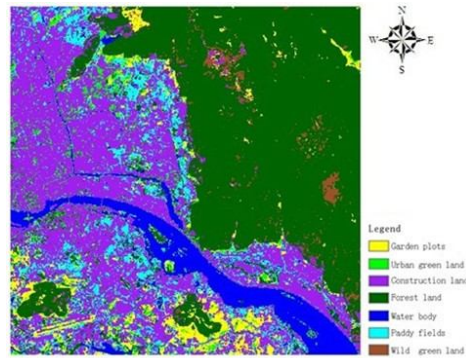
The classification accuracy refers to the extent of pixels correctly classified in the remote sensing image classification. The confusion matrix is the relatively common method on the evaluation of remote sensing image classification accuracy. The main parameters of classification accuracy are producer accuracy, user accuracy, overall accuracy, omission errors, commission errors and kappa coefficient. By contrasting the classification results with each other, we choice the best one.

4.1 Classify by ant colony algorithm under the support of different variables

In ant colony algorithm program (Dorigo M, Maniezzo V, Colorni A. 1996; Badr A, Fahlny A. 2003), decision trees are widely used of all machine learning methods. We applied ant colony algorithm to obtain some classification rules from spatial database under the support of different variables. Figure 3 are showed classification result from different variables. We have obtained confusion matrix and accuracy from ENVI. Total accuracy assessments are 89.12% and 90.18% and 92.22% based 8 and 11 and 19 factors. Kappa coefficients are 0.8651 and 0.8781 and 0.9034 based 8 and 11 and 19 factors. Figure 4 is comparison of producer's accuracy based different factors. Figure 5 is comparison of user's accuracy based different factors



(b) Remote sensing image classification results based on Ant Colony Algorithm Optimization under the support of 11 variables



(c) Remote sensing image classification results based on Ant Colony Algorithm Optimization under the support of 11 variables

Figure 3. Classification based on ant colony algorithm under the support of different variables (a to c)

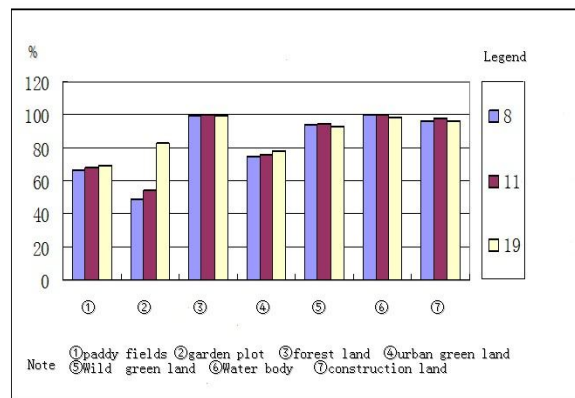
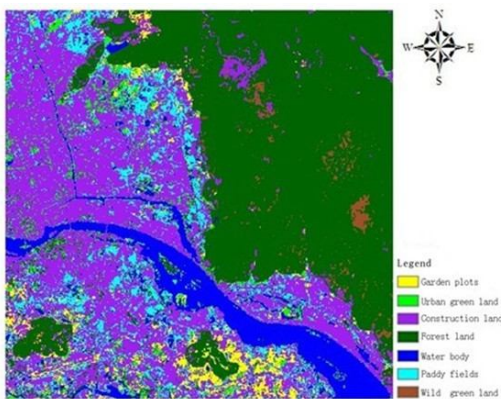


Figure4. Comparison of producer's accuracy based different factor



(a) Remote sensing image classification results based on Ant Colony Algorithm Optimization under the support of 8 variables

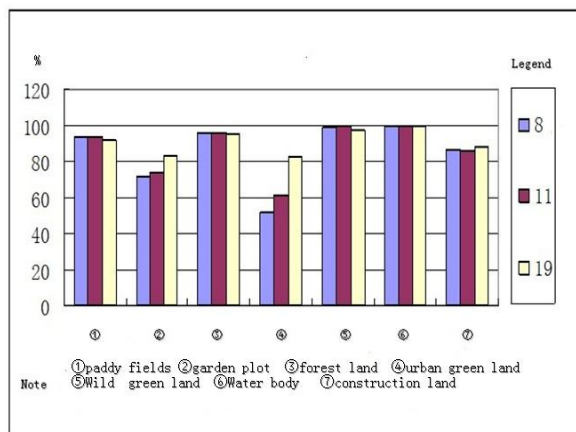


Figure5. Comparison of user's accuracy based different factors

4.2 Rules based on different variable / characteristics factors

Research variables and increasing classification according to Ant colony algorithm is an effective way to improve the classification accuracy, but is not classified for a particular classification algorithm based on the characteristics of the variable is the more the better. Characteristics factors are more and rules are less and complexity are higher. Validity checks on rules do not increase as the variable on the whole is added. We select different characteristics factors from variable set and Comparison is as follows. The table 3 shows comparison of the rules based different characteristics factors and the table 4 shows that is parts rules of landuse/cover classification in the test area. Of CF stand for validity check (confidence)

variables	rules	CF>0.8	0.5<CF<0.8	CF<0.5	complexity
8	27	7	7	13	simple
11	17	9	5	3	medium
19	17	7	7	3	complicated

Table 3. The comparison of the rules based different characteristics factors

IF-clause	Classification rules
19 variables:	
IF< B1 < 89.5 and B7 > 162.14 and B9 > 79.81 > THEN < class = forest land 2 >	(CF=1.000000)
IF< B9 > 79.81 and B12 < 39.5 > THEN < class = Wild green land >	(CF=0.998914)
IF< B4 < 54.5 and B12 < 39.5 > THEN < class = Water body >	(CF=0.997807)
IF< B4 < 54.5 and B8 > 148.5 and B12 < 39.5 > THEN < class = forest land 2 >	(CF=0.986413)
IF< B7 > 162.14 > THEN < class = garden plot >	(CF=0.870421)
IF< B8 < 118.5 > THEN < class = Water body >	(CF=0.844027)
11 variables	
IF< B2 < 77.5 and B10 < 3.06 > THEN < class = forest land 1 >	(CF=0.998873)
IF< B8 > 174.5 and 4.5 < B11 < 6.5 > THEN < class = Wild green land >	(CF=1.000000)
IF< B3 > 80.5 and B5 > 93.5 > THEN < class = construction land 1 >	(CF=0.947047)
IF< B4 < 54.5 and B9 < 48.81 > THEN < class = construction land 2 >	(CF=0.991424)
IF< 54.5 < B4 < 74.5 and 157.5 < B8 < 174.5 > THEN < class = construction land 2 >	(CF=0.982684)
IF< B10 < 3.06 > THEN < class = paddy fields >	(CF=0.895141)
IF< B9 > 84.33 and B10 > 3.06 > THEN < class = Wild green land >	(CF=0.836704)
8 variables	
IF< B8 > 176.5 > THEN < class = Wild green land >	(CF=0.986315)
IF< B4 < 54.5 and B5 < 63.5 > THEN < class = Water body >	(CF=0.997963)
IF< B7 < 99.89 and B8 > 176.5 > THEN < class = Water body >	(CF=0.947368)
IF< B3 > 79.5 and B6 > 81.5 > THEN < class = construction land 1 >	(CF=0.979235)
IF< 70.5 < B2 < 89.5 > THEN < class = garden plot >	(CF=0.863805)

Table 4. Parts from rules of landuse/cover classification in the test area

The experimental results show that is five points

(1) Increase as classification according to the number of variables, the classification results of overall accuracy and kappa coefficient increased.

(2) Ant colony algorithm under different variable support for remote sensing classification results are good, the overall accuracy is not less than 85%, the kappa coefficient were greater than 0.85, at 11 and 19 variable support under the overall accuracy is over 90%, under the support of 19 variables kappa coefficient over 0.9, shows that ant colony algorithm supports multiple characteristics of the remote sensing classification, and is a very effective method.

(3) Three cases producers accuracy is relatively high forest land, grassland, water, and land for construction; Paddy field, orchard, city green grass producers accuracy is low, because of the paddy field, garden land and urban green grass mixed distribution in remote sensing image and spectrum close to human-computer interaction to select for discovering classification rules of training sample accuracy is not high, directly affect the quality of the rules, then influence the classification accuracy.

(4) Most classes of users are high precision, 8 and 11 variables when user accuracy of gardens and urban green grass low, because the two similar spectral characteristics and terrain characteristics, influences the degree of differentiation.

(5) With the increase of variable precision of paddy field, orchard, city green grass are improved, especially when 19 variables of garden producers has greatly increased accuracy, city green grass users has greatly increased accuracy, show more support, the ant colony algorithm can reduce the error of the sample, found that higher degree of confidence of confusing rules to distinguish between classes.

4.3 Make comparison for evaluating precision

We tested 4 methods of data extraction by means of some examples for evaluating precision under the support 19 variable. Please see figure 6 (a-d). Comparison is as follows (see figure 7 to 9).

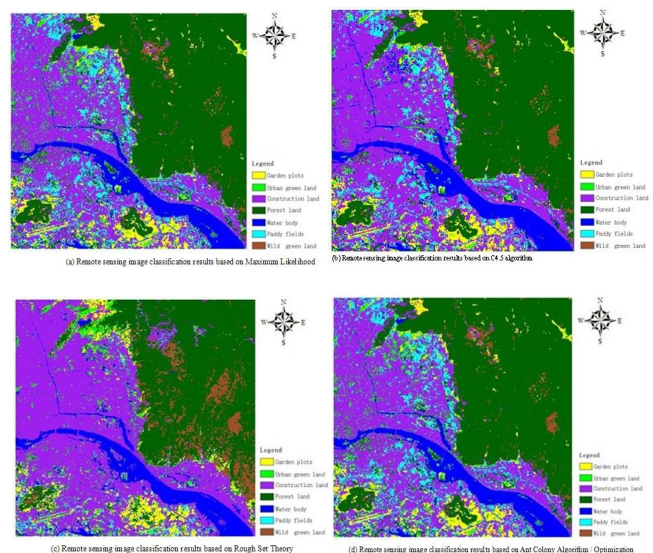
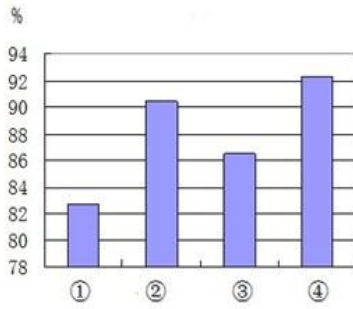


Figure 6. Different classification based remote sensing image (a-d)



Note ①Maximum Likelihood ②C4.5 Algorithm
 ③Rough Set Algorithm ④Ant Colony Algorithm

Figure7. Comparisons of total accuracy assessment based four classification methods

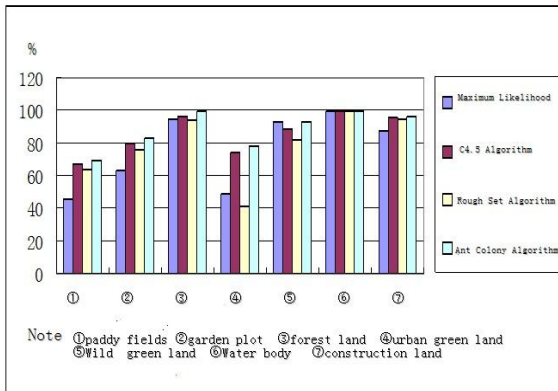


Figure8. Comparison of producer's accuracy based four classification methods

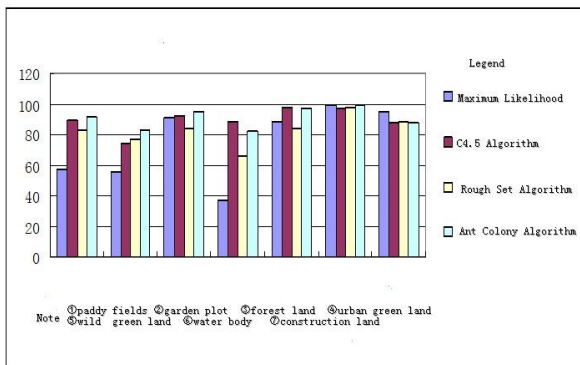


Figure9. Comparison of user's accuracy based four classification methods

In this section performances are compared. Studies have shown that the following points.

(1) Intelligent ant colony algorithm was applied to the precision of remote sensing classification is far higher than the traditional maximum likelihood classification based on rough set theory and the precision of classification, the classification accuracy is slightly higher than C4.5 algorithm. Intelligent ant colony algorithm for total classification accuracy of 92.22%, a maximum likelihood method, C4.5 algorithm and rough set method is increased by 9.46%, 1.84% and 5.67% respectively; the kappa coefficient was 0.9034, increased by 0.1154, 0.1154 and 0.0227, respectively.

(2) Four methods for classification of water effect is very good, all above 99%; Besides water, intelligent ant colony algorithm for producers of each land type classification accuracy is the highest, C4.5 algorithm for classification of land types except of weeds to producers for higher precision and rough set for paddy field, garden land, construction land use classification of producers for the third high precision, the maximum likelihood method of weeds to the producer of the classification accuracy is the same as the ant colony algorithm, the forest land, urban green grass high accuracy for the third classification of producer.

(3) Four methods of forest land, weeds land, water, and producers of construction land use classification precision is high, but relatively rough set approach to forest land and producers accuracy of classification of weeds to the minimum and maximum likelihood method to construction land use classification precision of producers. Four methods of paddy field, orchard and producers of city green grassland classification accuracy are relatively low, and the spectral characteristics, terrain characteristics are similar and mixed distribution, spectral similarity and mixed distribution caused by selection of training data accuracy is lower, affects the classification result, C4.5 algorithm and ant colony intelligence algorithm to reduce the impact of training data to make producers accuracy improved.

5. CASE

Above experimental research, found that the ant colony algorithm has obvious advantages for relative to C4.5 and rough set, the maximum likelihood.as the ant colony algorithm in remote sensing application in an example, this method was used to May 4, 2000, on May 10, 2003, May 8, 2007, and March 10, 2009, when four phase of Fuzhou Landsat ETM + image, obtain the land use/cover in Fuzhou and change information, detect the status of land use and change(lucc) in Fuzhou in 2000-2009 and the transfer mode. First use of ant colony algorithm intelligent mining classification rules, and then respectively in the remote sensing software based classifier, the classification results as shown in figure 10.

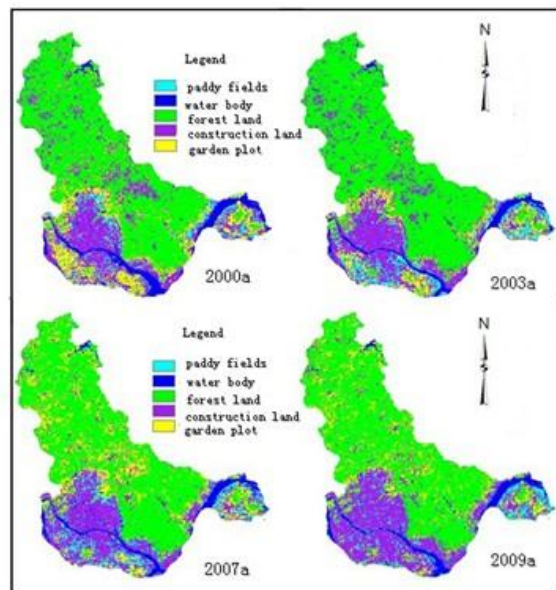


Figure10. Classified images during 2000 to 2003 to 2007 to 2009 of Fuzhou

Respectively, the results of each classification in remote sensing software diagram automatically in proportion randomly selected 500 validation samples, refer to the land use thematic map, secondary data of high resolution remote sensing image classification verifies the accuracy of the results, establish confusion matrix, the classification accuracy of the precision of the evaluation index selected producers, user accuracy and overall accuracy and Kappa coefficient. From the overall accuracy is more than 90% of all image, kappa coefficient greater than 0.85.

This part of the research results show that the ant colony algorithm is applied to a wide range area, intelligent ant colony algorithm was applied to Fuzhou multiple phase of remote sensing image classification, relying only on spectrum characteristics is achieved better classification accuracy of the results, for Fuzhou city land use/cover change research laid a good foundation, in addition, the ant colony algorithm is its characteristics of fast computing speed, high intellectualized degree also helped in the remote sensing application based data processing efficiency to obtain enhances greatly, intelligent ant colony algorithm can be effectively introduced into remote sensing field, is of great significance for remote sensing applications.

6. CONCLUSION

Some initial investigations are conducted to apply the Ant Colony Algorithm for classification of remotely sensed images. This paper explored the effectiveness of ant colony algorithm for multiple features of remote sensing classification. The availability of the ant colony algorithm is verified. As known from the results of classification, ant colony algorithm with higher precision is more fit to LUCC.

References from Journals:

Pei Tao, Zhou Chenghu, Han zhijun, Wang Min, Qin Chengzhi and Cai Qiang. 2001a. Progress Review on Spatial Data and Knowledge Discovery. *China image and graphics Journal*.6(9),854-860 .

Treitz P , Howarth P. 2000a.Integrating spectral, spatial and terrain variables for forest ecosystem classification. *Photogrammetric Engineering & Remote Sensing*, 66(3),305-317.

ERIKA LEPERS, ERIC F. LAMBIN etal. 2005a.A Synthesis of information on Rapid Land-cover Change for the Period 1981-2000,*BioScience*, 55(2),115-124.

Li De-ren, Wang Shu-liang, Li De_yi and Wang Xin-zhou. 2002a. The Theories and Methods on Spatial Data Mining and Knowledge Discovery. *Journal of Wuhan University on Science Edition*, 27 (3) ,221-233.

Li Shuang, Ding Shengyan, Hu Shuming. 2002a. Comparative Study on remote sensing image classification method ". *Journal of Henan University (Natural Science Edition)*, 32(2),70-73.

Wang Hai-qi, Wang Jin-feng. 2005a. Research progress on spatial data mining technology. *Geography and Geographic Information Science*, 21 (4) , 6-10.

Chen Zhongxiang, Yue Chaoyuan. 2003a. Research and Development on Spatial Data Mining, *Computer Engineering and Applications*, 5-7

Yu Ming, Ai Ting-Hua. 2009a. Data Mining and C4.5 Algorithm and its Classification Application. *Proc. Of SPIE*. Vol. 7492 74920B-1

Parp inell R S, LopesH S, FreitasA A. 2002a. Data mining with an ant colony optimization algorithm. *IEEE Transaction on Evolutionary Computation*, 6 (4): 321-332.

Badr A, Fahlny A. 2003a. A proof of convergence for ant algorithms. *International Journal of Intelligent Computing and Information*.3 (1) : 22-32.

DorigoM , Maniezzo V , Colorni A. 1996a. Ant system : Optimization by a colony of cooperating agents. *IEEE Transaction on Systems, Man, and Cybernetics 2 PartB*, 26 (1) : 29-41.

Wang Shugen, Yang Yun, Lin Ying, Cao Chonghua.2005a. Automatic Classification of Remotely Sensed Images Based on Artificial Ant Colony Algorithm, *COMPUTER ENGINEERING AND APPLICATIONS*.41(29):77-88,129

References from Books:

Yu Ming, Ai Ting-Hua. 2007a. *Study on Water Body Extraction and Wetland Sorts Based on SPOT5 Images. Proceeding of the Third International Symposium on Intelligent Earth Observation Satellites*. Science Press, Beijing China, pp.239-243

Chen Shu-Peng, Tong Qing-Xi and Guo Hua-Dong. 1998a. *Mechanism of Remote Sensing Information*. Science Press, Beijing, China.

References from Other Literature:

Colomi A, Dorigo M, Maniezzo V, et al. 1991a. *Distributed optimization by ant colonies* [C]. Proceedings of the 1st European Conference on Artificial Life.134-142.

Dorigo M. 1992a. *Optimization , Learning and natural algorithms[D]*. Ph. D. Thesis.Department of Electronics , Politecnico diMilano. Italy.

The People's Republic of China national standard (GB/T21010-2007), *Land-use Status Classification [S]*. 2007-08-10.

Yu Ming, Peng Yi-ru, Ji Qing.2011a. *Study on Urban Thermal Environment Based on RS and GIS Techniques - Taking as Example in Coastal Cities of Southeast Fujian Province*. IEEE International Conference on Electronics, Communications and Control (ICECC)

Acknowledgements

The author is extremely grateful to the headers of Fujian Normal University and Fujian Provincial Human Resources Development for support to carry out this work. This study was funded by exchange programs with Hong Kong, Macao and Taiwan.