RESEARCH ON IMAGE PROCESSING AND CLASSIFICATION OF MOUNTAIN SOILS IN THE SOUTHEAST CHINA

Li Xiaomei② Sha Jinming① Yang Wunian① Cai Qin③

Fujian Normal University, College of Geography , Fuzhou, P.R.C.350007 , Institute of Remote Sensing ,Chengdu University of Technology Chengdu 610059; College of Geography ,② Institute of Environmental Sciences, Fujian Normal University , Fuzhou , P.R.C.350007, ③Fuzhou Bureau Of Environment, Fuzhou , P.R.C.350007)

It is followed with interest by soil geographer that soils are surveyed in remote sensing materials .It is not only key points but also forward position of soil remote sensing to map and distinguish TM materials automatically by GIS .We did research on soil remote sensing in the southeast of China, which is supported by Sino-Europe cooperation project, and drew our inspiration from the research .The software used in the research are ARC/INFO, ENVI.

Basic information about the research area

The research area is located in Longyou county of Zhejiang province in the southeast of China(north latitude is 28°44’10”-29°17’15”,east longitude is 119°1’41”-119°19’52”).The total area is about 1726Km^2,and belongs to middle subtropical monsoon climate zone .Its four seasons are distinguished clearly ,full of sunshine and rainfall .No frost days are 247,accumulated temperature per year is 5502.9°C ,annual rainfall is 1671.6mm ,annual evaporation is 960mm,dry season is from July to September . Because of new structural movement and Qujiang river which flows from the west to the east , the county became to be a river basin with high mountains in the north and the south ,low plain in the middle .The bottom of the basin (or the middle) is about 50m high above sea-level .There is transition zone of hills from the edge of river to the mountain ,that elevation is between 60-150m.Hill mountains are distributed on the top of the basin ,the highest mountain is 1438m in the south ,while in the north the highest is 940m .Most of mountains are distributed in the south ,there are a little mountain and a great deal of hills in the north .There grew a lot of pine ,fir ,cryptomeria ,Huangshan

pine ,bamboo and etc .There are evergreen broadleaf tree ,evergreen coniferous-broadleaf forest ,fallen leaves coniferous-broadleaf forest .Vegetation rate is 70-80% in the south ,50-55% in the north .Most soil materials are alluvial material scattered at valley plain .On the slope mound are rocks of Cretaceous period and red clay of Quaternary period ;there are calcareous and no-calcareous brown sandstone on the north shore ;there are whetstone rock of Archean group and rocks of Jurassic periods .The area belongs to red soil zone and the mountain soil is yellow soil .

Remote sensing materials and their processing and analyzing

Materials includes :TM (05/06/1997),1:50000 topographic map,1:50000 soil map,1:50000 administrative map,1:50000 rock map,1:50000 land using map .The maps are digitized by ARC/INFO .After these maps were edited ,through CLEAN ,BUILD map's topology relation is created ,a coverage is formed .ARC/INFO interchange file is E00 ,ENVI(image processing software) can transform E00 file into EVF file ,so a vector map and image can be overlay together ,image can be cut and analyzed easily .

Image Processing

Because of high vegetation rate it is difficult to discern soils directly from image ,it is little possibility to survey soils .To the situation we design the technology ,which through many kinds of image processing ,composing model ,comparing method the whole research area is classified step by step supported by no-remote-sensing materials .Because research is for yellow soil ,texture analysis and image classification are only suited to yellow soils of the southeast.

2.1.1.Basic characters and processing

Using 1:50000 administrative map to cut the whole TM image ,Longyou's image was got .After statistics and analysis we got basic information of remote sensing material ,which will be used later.

Table 1 spectrum value and standard criterion of each band

Band	Min	Max	Mean	Stdev	Eigenval
1	0	194	34.1672	43.3603	10586.5434
2	0	95	14.5113	18.5915	192.7383
3	0	134	7.0163	12.7805	175.2471
4	0	178	29.4378	39.3536	23.9675
5	0	215	31.4123		

41.5100	2.9233						
6	0	172	56.2070	70.6815	2.2974		
7	0	153	12.6990	18.0974	0.5551		

Table 2 Covariance Matrix

Band	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
1	1880ÿ@15	803.449	430.619	1581.656	1736.944	3045.535	725.728
2	803.449	345.644	193.058	676.064	750.701	1295.677	317.234
3	430.619	193.058	163.341	314.633	424.489	644.392	210.077
4	1581.656	676.064	314.633	1548.706	1546.149	2608.249	604.528
5	1736.944	750.701	424.489	1546.149	1723.079	2808.429	726.235
6	3045.535	1295.677	644.392	2608.249	2808.429	4995.872	1146.018
7	725.728	317.234	210.077	604.528	726.235	1146.018	327.515

Table 3 Correlation Matrix

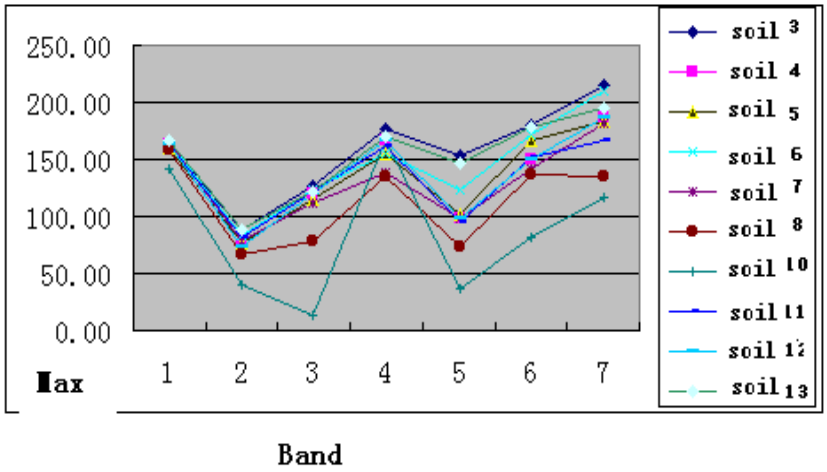
Band	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
1	1.000	0.997	0.777	0.927	0.965	0.994	0.925
2	0.997	1.000	0.813	0.924	0.973	0.986	0.943
3	0.777	0.813	1.000	0.626	0.800	0.713	0.908
4	0.927	0.924	0.626	1.000	0.946	0.938	0.849

5	0.965	0.973	0.800	0.946	1.000	0.957	0.967
6	0.994	0.986	0.713	0.938	0.957	1.000	0.896
7	0.925	0.943	0.908	0.849	0.967	0.896	1.000

Table 4 Eigenvectors Matrix

	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
1	0.419	-0.185	-0.229	0.380	-0.718	-0.173	0.217
2	0.179	-0.135	-0.035	0.217	-0.066	0.221	-0.921
3	0.093	-0.549	0.171	0.546	0.433	0.322	0.267
4	0.366	0.640	0.534	0.383	0.111	-0.108	0.013
5	0.395	-0.238	0.530	-0.546	-0.237	0.384	0.072
6	0.683	0.116	-0.530	-0.224	0.429	0.034	0.061
7	0.163	-0.408	0.264	-0.124	0.199	-0.811	-0.156

From the table ,relations between two bands arranged from low to high are 3-4,3-6,3-1,5-3,7-3.Shizhou's doctoral dissertation indicated that OIF value of 3-4-5 is the



largest, the relation is the

3.Yellow-red soil 4.Red soil 5.Young red soil 6. Calcareous purple soil 7. Dyskel soil 8. Mountain meadow soil 10.unknown soil 11. Percogenic paddy soil 12. Hydrogric paddy soil 13.yellow soil

Figure 1. Spectrum characters of forming environment of soils

least ,bands composing result is the best .Other composing models ,such as 3-4-2,1-4-7, are good .All these provided foundation of research.

The graph shows that law of forming environment of different soils is not clear ,difference between classes is little ,it is impossible to get good classification result directly from original image .To get ideal result ,different image processing ,texture analysis ,classification method have to be organized well to classify different soil types.

2.1.2.Research on image processing

Each eigenvectors has clear geographic meaning after tasseled cap transformation .Different aim has to chose different band of transformation to compose or to overlay or analyse .

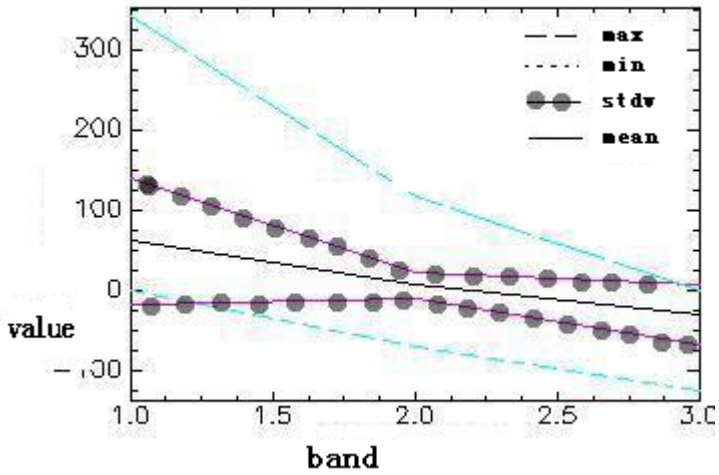


Figure 2. Image Characters after K-T transformation

Composing experiment (B,G,Third) indicates that K-T transformation erases effect of mountain shade ,strengthens difference between mountain soils and plain soils .It not only separates green part ,protrudes soil characters but also eliminates interrelated information .Algorithm of K-T transformation in the research is simple , but result is good and fits to survey soils in the mountain .It is regarded that seven TM bands can be divided into four groups :1-2-3 ,5-7 ,4 ,6 .Principal components analysis of the research takes two

group:1-2-3 and 5-7 to transform respectively .Bands of principal components are used to be basic bands of classification to solve the complicated problems .It is difficult to get good result only depending on spectrum value , so research takes attention on analysis of image texture , calculates texture characters of 3×3 windows and 8×8 windows ,such as information entropy skewness ,mean ,variance. Variance spreads out structure of surface features , 3×3 windows expressed vegetation community and shores of river ,while 8×8 windows express characters of large surface feature , such as direction of river . Different bands with different windows can resolve a certain problem .Information entropy expresses spectrum reflection of surface feature , while variance expresses information of texture intensity .Large windows means merging of small pixels , and is synthesis .

These algorithms are below:

Information entropy:

$$D(I,J)=\left(\sum_{i1=i-n/2}^{i+n/2}\sum_{j1=j-n/2}^{j+n/2} P(i_1,j_1)\log P(i_1,j_1)\right)/\log(N^2+1)$$

while $P_{ij}= g(i_1,j_1)/\left(\sum_{i1=i-n/2}^{i+n/2}\sum_{j1=j-n/2}^{j+n/2} g(i_1,j_1)\right)$

can be simplified into:

$$D(i,j)=\sum_{i1=i-n/2}^{i+n/2}\sum_{j1=j-n/2}^{j+n/2} P(i_1,j_1)\log P(i_1,j_1)$$

D（i,j）value of texture statistics

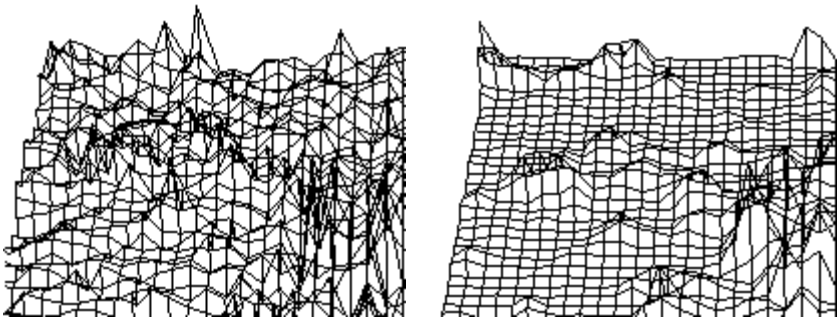
N :windows size

g(i,j) :original spectrum value

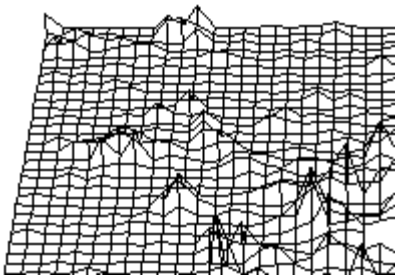
$$D(I,J)=\sum_{i1=i-n/2}^{i+n/2}\sum_{j1=j-n/2}^{j+n/2} [g(i_1,j_1)-X]^2$$

X :average spectrum value of N×N windows

These texture indexes show texture characters of surface feature and will be used in classification later .



Original Image　　8×8 Texture Transformation

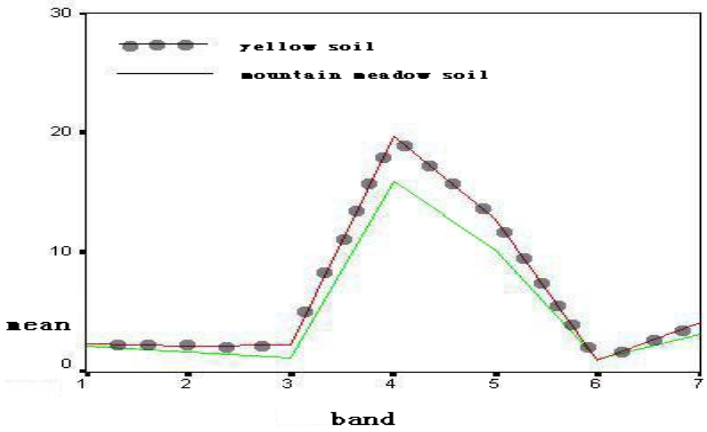


3×3 Texture Transformation

Figure 3. Different 3D maps with different processing (130×180)

The three figures indicated that topography of the original image without texture processing is fragmented and little regular patterns .Texture transformation of 8×8 windows can show large topography well ,while 3×3 windows can eliminate fragments and show regular patterns of topography .TM5/TM7 and (TM5/TM7)/(TM4/TM3) can show soil information well but both of them have disturbing information more and less , such as vegetation information ,water information or other information .These interfered information can be eliminated effectively with logical “and” method , it improves to survey soils of the southeast .

After division of bands composing image of (TM5/TM7) /(TM4/TM3) , (TM5/TM4) / (TM3/TM1) ,TM1 can eliminate plant effect ,through quick gaussian transformation especially density vegetation (arbors) is deep blue while naked soil is pale brown , farmland is grey brown ,river is yellow ,bamboo and forest of low vegetation is blue green ,other special object is purplish red .Because TM7 has an absorbing zone of hydroxy. .TM5/TM7 can reflect clay information ,while TM4/TM3 can reflect plant information.(TM5/TM7)/(TM4/TM3) not only protrudes soils information but also eliminates effect of vegetation and topography . (TM5/TM4) and (TM3/TM1) reflects Fe^{+3} well , their specific value can eliminate many kinds interfered information .



1.Yellow soil. 2. Mountain meadow soil

Figure 4.Spectrum characters of typical sample

2.2.2. Classification technology of image

In the research classification is to mountain soil-yellow soil , which is main soil on the LONGYOU mountain with 600-1200M , its subclasses are yellow soil and mountain meadow soil .There is coniferous-broadleaf forest with high vegetation .It can not get ideal result to monitor soils directly so we design many methods to classify and to improve accuracy .

Remote sensing classifications have two models: Supervised Classification and Unsupervised Classification . Supervised Classification has many ways : Parallelepiped ,Minimum Distance ,Mahalanobis Distance ,Maximum Likelihood ,Spectral Angle Mapper ,Binary Encoding . Unsupervised Classification has Isodata ,K-Means . Image is cut with different elevation to improve its speed and accuracy before classification .

The figure indicates clearly that the spectrum of environment of yellow soil and Mountain meadow soil is closed ,and the soils can be distinguished on TM4 or TM5 . Because TM5 can reflect soil information better it must to be selected to classify matching with other bands and models .Experiment classification results are below (number is classification area/standard area).

Table 5 classification reslut (yellow soil and Mountain meadow soil)

methods TM345 S3TM5Z1 W ZH S3TM5W
mahalanobis 20/18 89/85 76/70 130/143
spectral 150/128 32/35 87/84 132/112 129/142

minimum	little	75/80	23/15	149/158 12/19

Number (a/b) describe :a is yellow soil area ratio with the standard (%) b is mountain meadow soil area ratio with the standard(%);**TM345** is composing of original band 3 4 5 ; **S3TM5Z1** is composing of the third of K-T transformation ,original **TM5** and the first of principal components;w is composing of texture(3×3) processing bands; zh is composing of**(TM5/TM7)/(TM4/TM3) , (TM5/TM4)/(TM3/TM1) , TM1; S3TM5W** is composing of the third of K-T transformation , original **TM5** and texture(3×3) processing bands .

Parameters of classifications are respectively :Minimum Distance (maximum allowable distance error=3, Max Stdev From Mean=3),Mahalanobis Distance (max distance=2) ,Maximum Likelihood (Probability Threshold=0.25),Spectral Angle Mapper (max angle=0.1), Isodata(maximum allowable distance error=20, Max Stdev From Mean=20) ,K-Means (maximum allowable distance error=20, Max Stdev From Mean=20).

The actual situation showed that result of Mahalanobis Distance is more ideal and simple, the algorithm is below:

$$W(x)=(X-U)'V_2^{-1}(X-U_2)-(X-U_1)'V_1^{-1}(X-U_1)$$

$$X \begin{cases} X<A_1 \quad W(X)>=0 \\ X<A_2 \quad W(X)<0 \end{cases}$$

X,distinguished object， A_1,A_2:mather object， U_1,U_2:meanvector， V_1,V_2 **:**Covariance Matrix

Analysis of result

Image processing and classification experiment show a little laws:

When Spectral Angle Mapper is used classification effect of texture processing is better than that of no processing ,it is because that processing improves image information and eliminates disorder verbose information ,all these kinds consider relate information around more ,this fits the actual situation between soils and their forming environment . Because Spectral Angle Mapper distinguishes objects by spectrum types rather than by density of no-choosing absorbing it can not distinguish these objects with similar spectrum types and great spectrum difference .What is more , Spectral Angle Mapper is slow in spectrum change it results in that classification area is large than its reality .

In the mountain of southeast ,soils are distributed with elevation (certain contour line is boundary of some soils).Other distance methods have relation with their dimension

without considering relation between theirs . Mahalanobis Distance does not change with transformation ,is not affected by dimension ,considers relation between each other ,its result is good and can meet practice requirement .With complicating of landform and enlarging of area , Mahalanobis Distance can get good result too through texture processing and adjusting max distance error .It is clear that a good visual image must not be suited to classify ,but the image can be used as reference in comparing with result map , such as ,(TM5/TM7)/(TM4/TM3) ,(TM5/TM4)/(TM3/TM1),TM1.

It is law that all better methods have to draw support from vector map ,material map .Because high vegetation makes monitoring difficult soils distinguished should be regarded as typical section .Some contour ,material boundary or other boundary may be treated as temporary soil boundary , the final soil boundary can be confirmed with computer classification later.

Before research working author had ever been to Longyou county to survey field and built image symbols of naked fields ,treat them as typical sections ,try to obey gradually changing law of soils to find changing boundaries .The monitoring was finished finally by synthesizing technology supported by image and vector maps .

The research is only suited to the southeast of China ,which soils have obvious law changing with elevation ,material .Through effectively processing of image changing can be picked out exactly .With expending of research area and more complicated problems ,Artificial Intelligence and DEM model have to be added to improve speed and accuracy of monitoring ,it will be main direction of monitoring .

Supported by The State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing --wkl(02)0104

Reference Artilces

[1] Foody, G. M. Land cover classification using an artificial neural network with ancillary information. Int. J. Geographical Information Systems, 1995, 9: 527-542.

[2] Foody, G. M. and Arora, M. K. An evaluation of some factors affecting the accuracy of classification by an artificial neural network. Int. J. Remote Sensing, 1997, 18(4): 937-953.

[3] Foody, G. M. and Cox, D. P. Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership function. Int. J. Remote Sensing, 1994, 15: 619-

[4] Habbane, D'ubois, EL-Sabh. Empirical Algorithm using SeaWiFS Hyperspectral Bands: a Simple Test. Int. J. Remote Sensing, 1998, 19(11): 2161-2169.

[5]Jürgens, C. The modified normalized difference vegetation index (mNDVI) — a new index to determine forest damages in agriculture based on Landsat TM data. Int. J. Remote Sensing, 1997, 18(17): 3583-3594.

[6]Jeon, B. and Landgrebe, D. A. Decision fusion approach for multitemporal

classification. IEEE Trans. on Geoscience and Remote Sensing, 1999, 37(3): 1227-1233.

[7] Tassan, S. .Local Algorithms using SeaWiFS Data for the Retrieval of Phytoplankton Pigments, Suspended Sediment and Yellow Substance in Coastal Waters. Applied Optics, 1994, 33: 2369-2378.

[8] Townshend, J. R. G. and Justice, C. O. Spatial variability of images and the monitoring of changes in the Normalized Difference Vegetation Index. Int. J. Remote Sensing, 1995, 16(12): 2187-2195.

[9] Wilkinson G.G. A review of current issues in the integration of GIS and remote sensing data, 1996,10(1)：85-105.

[10] Wilkinson, G. G., Fierens, F. and Kanellopoulos, I. Integration of neural and statistical approaches in spatial data classification. Geographical Systems, 1995, 2: 1-20.

[11] Woodruff, Stumpf, Scope, etal. Remote Estimation of Water Clarity in Optically Complex Estuarine Waters. Remote Sensing of Env. , 1999, 68: 41-52.

[12] Xu Hanqiu, Wang Xiaoqin, Xiao Guirong. A remote sensing and GIS integrated study on urbanization with its impact on arable lands: Fuqing County, Fujian Province, China. Land Degrad & Develop, 2000, 11: 301-316.

[13] Yang X. and Lo, C. P. Relative radiometric normalization performance for change detection from multi-date satellite images. Photogrammetric Engineering & Remote Sensing, 2000, 66(8): 967-980.