Improving Classification Accuracy by Combining Spectral and Texture-based Feature Spaces in High Resolution Multispectral Images

H. Ashoori
K.N.Toosi University of Technology, Tehran, Iran
hamed_ashoori@yahoo.com

A. Alimohammadi
K.N.Toosi University of Technology, Tehran, Iran
Alimoh_abb@yahoo.com

M. J. Valadan Zoej
K.N.Toosi University of Technology, Tehran, Iran
valadanzouj@kntu.ac.ir

B. Mojarradi
K.N.Toosi University of Technology, Tehran, Iran
mojaradi@albors.kntu.ac.ir

Abstract : Although conventional classification methods based on the use of spectral information may provide acceptable results on low and medium resolution images, usually they have shown poor results when applied for high resolution images. Many researchers have shown that contextual information is a rich source of information for improvement of the classification accuracy. In this paper, texture quantization as an example of producing valuable features for the purpose of object discrimination from IKONOS pan-sharpened data of suburb areas has been investigated.
Several statistical features such as the mean, variance and median as well as the features originated from the gray level run-lengths matrix have been generated. In addition, autocorrelation and geo-statistical methods have been employed for production of the new features. Results of different tests have shown that by proper use and combination of texture-based features in the classification process, up to 20 percent improvement in the accuracy may be achieved
Keywords: Classification, Texture

1. Introduction

Classification is the most common method of extracting information from remotely sensed data. In conventional classification methods only spectral data are used. High resolution images have more spatial information but do not have a high spectral resolution, so using conventional classification methods seems to be ineffective. To improve the classification accuracy, spatial information, which is a reach source of useful information needs careful consideration. Texture quantization is an effective approach for utilization of the spatial information. There is no clear definition for image texture, but we can describe how the image texture look e.g. fine, coarse, smooth or irregular, homogeneous and so forth. [1]
Many authors have introduced methods to quantify spatial relations between pixels and have used them as an input data in the classification. There are a wide range of texture quantization methods that can be classified in three main groups, statistical, structural and spectral based methods [2]. Statistical methods produce statistical measures of gray level variation; Structural methods assume that the texture pattern is composed of spatial arrangement of texture primitives, so their task is to locate the primitives and quantify their spatial arrangement; and Spectral features are generated using the spectrum obtained through image transformations such as Fourier transform.
In this paper, four groups of features based on first order statistics, run-lengths matrix, autocorrelation and geostatistics which can be classified as the statistical methods have been generated. Then different classifications resulting from the combination of different texture features have been evaluated.

2. Generated Features

1) First Order Statistical Features

In this paper we generate mean, median and variance from the first order statistical measures.

If (I) is the random variable representing the gray levels in the region of interest, the first order histogram P (I) is defined as:

$$P(I) = \frac{number \quad of \quad pixels \quad whith \quad gray \quad level \quad I}{Total \quad number \quad of \quad pixels} \tag{1}$$

Now mean and variance are defined as:

$$mean = \sum_{i=0}^{N_g - 1} I * P(I) \tag{2}$$

$$Variance = \sum_{i=0}^{N_g - 1} (I - mean)^2 P(I) \tag{3}$$

Where N_g = number of gray levels.

Median is the middle value in a set of numbers arranged in increasing order. Because the kernel size always cover odd number of pixels, median can be extracted simply by choosing the mid member of an array which contains gray levels of pixels that covered by the mask and then is sorted.

2) Features Generated from Gray Level Run-lengths Matrix

A gray level run is a set of the consecutive pixels having the same gray level value. The length of the run is the number of pixel in the run. The gray level run-lengths matrix is generated in for directions (0°, 45°, 90°, 135°), it is a $N_g \times N_r$ matrix (N_g is the number of gray levels and N_r is the maximum run length that in the mask) , it's (i,j)th element ($Q_{RL}(i, j)$) is the number of runs with gray level (i), and length (j).

Run-length features encode textural information related to number of times each gray level appears in image by itself. After generating gray level run-length matrix, the five following features can be defined [1]:

1-Short Run Emphasis

$$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} (Q_{RL}(i, j) / j^2)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} Q_{RL}(i, j)} \tag{4}$$

2-Large Run Emphasis

$$LRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} (Q_{RL}(i, j) \times j^2)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} Q_{RL}(i, j)} \tag{5}$$

3- Gray Level Nonuniformity

$$GLN = \frac{\sum_{i=1}^{N_g} \left[\sum_{j=1}^{N_r} Q_{RL}(i, j) \right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} Q_{RL}(i, j)} \tag{6}$$

4- Run Length Nonuniformity

$$RLN = \frac{\sum_{j=1}^{N_r} \left[\sum_{i=1}^{N_g} Q_{RL}(i, j) \right]^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} Q_{RL}(i, j)} \tag{7}$$

5- Run Length Nonuniformity

$$RP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} Q_{RL}(i, j)}{L} \tag{8}$$

3) Geostatistical Features

Geostatistics is the statistical methods developed for and applied to geographic data. These statistical methods are required because geographical data do not usually conform to the requirements of standard statistical procedures, due to spatial autocorrelation and other problems associated with spatial data [3].

Semivariogram that represents half of the expectation of the quadratic increments of pixel pair values at the special distance can quantify both spatial and random correlation between the adjacent pixels. [4] I it is defined as:

$$\gamma(h) = \frac{1}{2} E[DN(x+h) - DN(x)]^2 \tag{9}$$

That is the classical expression of variogram (h) here represents a vectorial lag between pixels. In this study direct variogram, madogram, cross variogram and pseudo-cross variogram have been used. The first two operate separately for each image bands and the second two operate for pairs of image bands.

1- Direct Variogram

In this approach the following equation is used to estimate E.q. (9) :

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r - h_1} \sum_{j=1}^{N_c - h_2} [DN(i,j) - DN(i + h_1, j + h_2)]^2 \tag{10}$$

n(h) is the number of pairs that the are in mask filter.

2- Madogram

This is similar to direct variogram except squaring differences, but uses the absolute value of differences.

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r - h_1} \sum_{j=1}^{N_c - h_2} \left| DN(i,j) - DN(i + h_1, j + h_2)^2 \right| \tag{11}$$

3- Cross Variogram

Two image bands are used to quantify the joint spatial variability between bands.

$$\gamma_{m,n}(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r - h_1} \sum_{j=1}^{N_c - h_2} \{[DN_m(i,j) - DN_m(i + h_1, j + h_2)] * [DN_n(i,j) - DN_n(i + h_1, j + h_2)]\} \tag{12}$$

3- Pseudo-cross Variogram

It is similar to direct variogram, but uses pairs which are from two different bands (m,n).

$$\gamma_{m,n}(h) = \frac{1}{2n(h)} \sum_{i=1}^{N_r - h_1} \sum_{j=1}^{N_c - h_2} [DN_m(i,j) - DN_n(i + h_1, j + h_2)^2] \tag{13}$$

4) Autocorrelation Features

Autocorrelation refers to the degree of existing relationship between two or more spatial variables, such that when one changes, the other(s) also change. This change can either be in the same direction, which is a positive autocorrelation, or in the opposite direction, which is a negative autocorrelation [4].

Coarse texture consists of large primitives but fine texture has small primitives and autocorrelation can quantize this coarseness can be quantized through the autocorrelation function like: [5]

$$A.C.(m,n) = \frac{N_r N_c}{(N_r - m)(N_c - n)} \frac{\sum_{i=1}^{N_r - m} \sum_{j=1}^{N_c - n} DN(i,j) DN(i+m, j+n)}{\sum_{i=1}^{N_r} \sum_{j=1}^{N_c} DN^2(i,j)} \tag{14}$$

(m,n) is the lag vector.

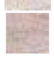
N_r & N_c are kernel size dimensions

3. Case Study

For evaluation of the above methods, we selected a 500*500 size subset from the IKONOS ortho image of part of Tehran suburb area (Fig. 1). This subset contains different textures. Ten classes (table 1) were defined in the area, and two images were produced to be used as training and testing sets. Some of these classes are spectrally similar and therefore are confused in the pixel-based conventional classification algorithms.

Figure 1. Selected Subset

Table 1. Classes and their sample

No.	Class Name	Sample
1	Ploughed Land	
2	Tree	
3	Row of Trees	
4	Bushed	
5	Cultivated Land	
6	Asphalt Road	
7	Bare Soil	
8	Bare Soil 2	
9	Building	
10	Building 2	

Different window sizes including 3 by 3, 5 by 5, 7 by 7 and 9 by 9 pixels were used to calculate each texture-based feature. Also, calculation of the first order statistical and run lengths matrix was based on varying number of pixels up to 25 pixels.
Because of the large number of features generated from the run lengths matrices and their correlation, the mean value of four directions was computed and therefore, omni directional features were produced.
The supervised maximum likelihood classification process was completed using each group of generated features plus the three spectral bands as input features. Independent test samples were used for evaluation of the accuracy of the classification. For this purpose, producer accuracy, overall accuracy, variance and mean of the producer and normalized overall accuracy were considered.

4. Results and Conclusions

Results of some tests are presented in tables 2, 3 and 4. Also three graphs representing maximum and minimum values of the mean, overall and variance of accuracy are shown in Figures 2-5. Considerable variation of the accuracy of classes as a function of the input features demonstrates the fact that selection of the proper feature for a given situation is an important task.
Overall accuracy is almost influenced by dominant classes and therefore, it is not so sensitive for change of the accuracy of the minor classes. So, consideration of the mean and variance of accuracy is as important as the overall accuracy.
Some feature display very specific roles for classification of a particular class. There are situations where, discrimination of a class with high accuracy is more important than obtaining a medium overall accuracy.
First order statistical features act as low pass filters, leading to significant reduction of the high frequency components in the data. It may lead to poor results when classes are spectrally similar with smooth and distinct textures. Also it may mix the adjacent classes.
Run lengths based features show better performance when the classes of interest such as the cultivated lands (see item 5 on table 3), are composed of strips, but generally they are weak. Autocorrelation-based features did not show good performance in most tests and needs more considerations.
Geostatistical-based features show a good performance in production of the homogeneous maps with high accuracies (table 2 & 3). Use of some features not only does not show an increase in the accuracy, it leads to lower accuracies (as an example, see item 7 on table 3). Therefore careful selection and use of different features is very important. It may be concluded that the first order statistical and geostatistical features work better for improvement of the accuracy of most classes and their performances is not specific. Whereas run lengths based features as mentioned above are most suited for discrimination of some special classes. These observations may be used as a guide for proper selection and use of different features.
The research will be continued towards testing various combinations of features and formulation of the possible influences of the window size in the results.

Table 2 Results considering all tests

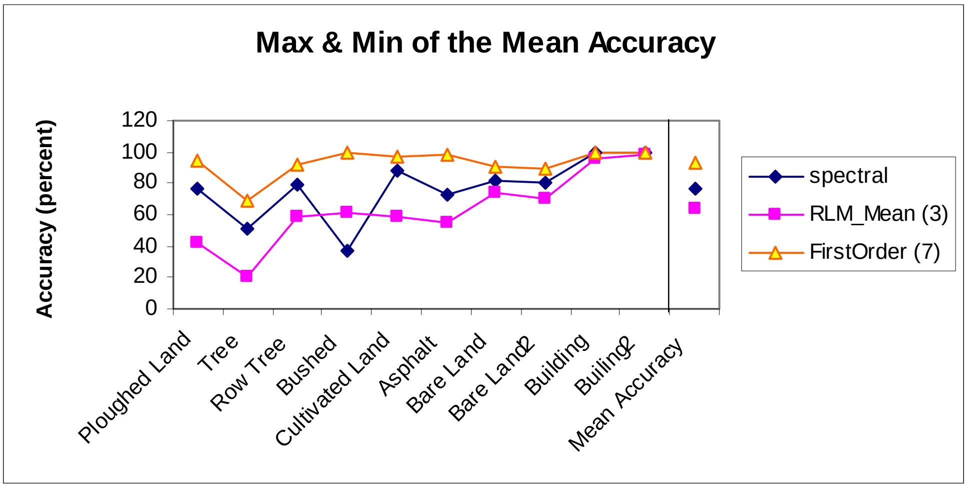
	Sorted by	Combined Features	Window Size	Lag Vector X	Lag Vector Y	Ploughed Land	Tree	Row of Trees	Bushed	Cultivated Land	Asphalt Road	Bare Land	Bare Land 2	Building	Buil2	Overall Accuracy	Mean Accuracy	Accuracies' Variance	Overall/ Variance
1	-	No feature	-	-	-	76.47	50.68	79.3	36.96	87.98	72.72	81.77	80.61	98.94	99.39	75.74	76.481	384.01	0.19
2	Overall Accuracy	First Order	11	-	-	95.34	43.38	94.08	100	95.61	100	94.22	91.54	100	100	95.27	91.417	294.4622	0.323539
		First Order	19	-	-	93.96	55.25	84.62	97.74	89.06	100	99.85	99.43	84.81	100	95.86	90.472	191.515	0.500535
		First Order	13	-	-	95.91	36.53	96.45	99.58	92.83	100	97.37	93.43	100	100	96.14	91.21	376.3108	0.25548
		First Order	17	-	-	94.61	52.51	91.72	98.44	90.31	100	99.83	96.34	91.7	100	96.2	91.546	201.8726	0.476538
		First Order	15	-	-	95.39	47.03	94.67	98.87	91.75	100	99.74	94.31	98.41	100	96.57	92.017	258.0567	0.37422
3	Mean Accuracy	First Order	17	-	-	94.61	52.51	91.72	98.44	90.31	100	99.83	96.34	91.7	100	96.2	91.546	201.8726	0.476538
		First Order	15	-	-	95.39	47.03	94.67	98.87	91.75	100	99.74	94.31	98.41	100	96.57	92.017	258.0567	0.37422
		First Order	5	-	-	90.81	71.69	90.53	97.31	97.31	95.32	91.44	86.73	100	100	92.54	92.114	71.04936	1.302475
		First Order	9	-	-	95.1	65.3	91.72	99.95	97.22	99.44	91.18	89.89	100	100	94.58	92.98	110.2115	0.858168
		First Order	7	-	-	94.66	68.49	92.31	99.48	97.58	98.11	90.49	88.95	100	100	94.13	93.007	90.41698	1.041066
4	Overall Accuracy/ Variance	Geostatistics	5	0	3	86.32	64.84	81.66	91.32	85.83	72.72	89.02	82	99.12	98.16	87.24	85.099	112.0417	0.778638
		First Order	9	-	-	95.1	65.3	91.72	99.95	97.22	99.44	91.18	89.89	100	100	94.58	92.98	110.2115	0.858168
		First Order	3	-	-	86.18	65.3	84.02	87.36	95.61	81.85	89.82	84.9	100	99.08	88.35	87.412	100.7576	0.876857
		First Order	7	-	-	94.66	68.49	92.31	99.48	97.58	98.11	90.49	88.95	100	100	94.13	93.007	90.41698	1.041066
		First Order	5	-	-	90.81	71.69	90.53	97.31	97.31	95.32	91.44	86.73	100	100	92.54	92.114	71.04936	1.302475

Table 3 Results omitting large kernel size features

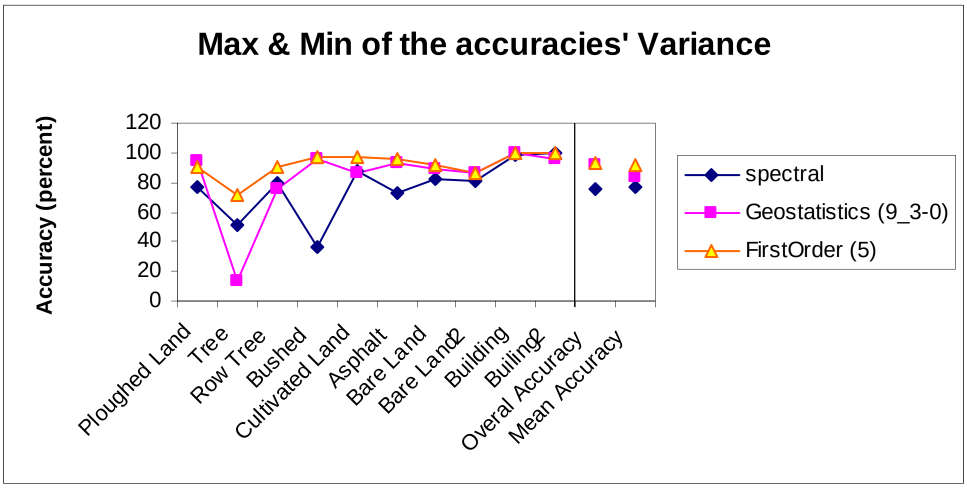
	Sorted By	Combined Features	Window Size	Lag Vector X	Lag Vector Y	Ploughed Land	Tree	Row of Trees	Bushed	Cultivated Land	Asphalt Road	Bare Land	Bare Land 2	Building	Buil2	Overall Accuracy	Mean Accuracy	Accuracies' Variance	Overall/ Variance
1	-	No feature	-	-	-	76.47	50.68	79.3	36.96	87.98	72.72	81.77	80.61	98.94	99.39	75.74	76.481	384.01	0.19
2	Overall Accuracy	Geostatistics	9	1	1	87.33	37.9	79.3	99.39	95.16	94.54	94.35	95.07	100	98.16	92.33	88.119	349.7204	0.264011
		Geostatistics	9	2	1	90.41	24.66	85.8	98.68	95.61	92.43	91.63	94.88	100	98.87	92.41	87.297	503.5477	0.183518
		FirstOrder	5	-	-	90.81	71.69	90.5	97.31	97.31	95.32	91.44	86.73	100	100	92.54	92.114	71.04936	1.302475
		FirstOrder	7	-	-	94.66	68.49	92.3	99.48	97.58	98.11	90.49	88.95	100	100	94.13	93.007	90.41698	1.041066
		FirstOrder	9	-	-	95.1	65.3	91.7	99.95	97.22	99.44	91.18	89.89	100	100	94.58	92.98	110.2115	0.858168
3	Mean Accuracy	Geostatistics	7	1	1	87.58	40.64	81.1	97.08	96.05	90.53	91.98	94.06	100	98.16	91.36	87.715	304.8665	0.299672
		Geostatistics	9	1	1	87.33	37.9	79.3	99.39	95.16	94.54	94.35	95.07	100	98.16	92.33	88.119	349.7204	0.264011
		First Order	5	-	-	90.81	71.69	90.5	97.31	97.31	95.32	91.44	86.73	100	100	92.54	92.114	71.04936	1.302475
		First Order	9	-	-	95.1	65.3	91.7	99.95	97.22	99.44	91.18	89.89	100	100	94.58	92.98	110.2115	0.858168
		First Order	7	-	-	94.66	68.49	92.3	99.48	97.58	98.11	90.49	88.95	100	100	94.13	93.007	90.41698	1.041066
4	Overall Accuracy/ Variance	Geostatistics	5	0	3	86.32	64.84	81.7	91.32	85.83	72.72	89.02	82	99.12	98.16	87.24	85.099	112.0417	0.778638
		First Order	9	-	-	95.1	65.3	91.7	99.95	97.22	99.44	91.18	89.89	100	100	94.58	92.98	110.2115	0.858168
		First Order	3	-	-	86.18	65.3	84	87.36	95.61	81.85	89.82	84.9	100	99.08	88.35	87.412	100.7576	0.876857
		First Order	7	-	-	94.66	68.49	92.3	99.48	97.58	98.11	90.49	88.95	100	100	94.13	93.007	90.41698	1.041066
		First Order	5	-	-	90.81	71.69	90.5	97.31	97.31	95.32	91.44	86.73	100	100	92.54	92.114	71.04936	1.302475
5	Cultivated Land	RLM_Mean	7	-	-	83.17	34.7	53.3	84.91	95.96	83.41	88.61	78.9	97.17	99.69	85.64	79.977	425.1112	0.201453
		Geostatistics	7	1	1	87.58	40.64	81.1	97.08	96.05	90.53	91.98	94.06	100	98.16	91.36	87.715	304.8665	0.299672
		First Order	9	-	-	95.1	65.3	91.7	99.95	97.22	99.44	91.18	89.89	100	100	94.58	92.98	110.2115	0.858168
		First Order	5	-	-	90.81	71.69	90.5	97.31	97.31	95.32	91.44	86.73	100	100	92.54	92.114	71.04936	1.302475
		First Order	7	-	-	94.66	68.49	92.3	99.48	97.58	98.11	90.49	88.95	100	100	94.13	93.007	90.41698	1.041066
6	Bare Land	Geostatistics	5	4	1	82.07	36.99	84.6	76.43	89.96	73.83	93.7	80.1	94.88	98.46	85.04	81.104	307.0225	82.07
		Geostatistics	5	3	2	84.25	42.47	80.5	76.19	85.2	71.05	93.77	79.15	95.94	98.77	85.3	80.726	259.9893	84.25
		Geostatistics	5	4	0	87.8	31.51	80.5	82.74	87.8	72.94	94.24	76.75	96.47	97.34	87.19	80.806	367.7014	87.8
		Geostatistics	9	1	1	87.33	37.9	79.3	99.39	95.16	94.54	94.35	95.07	100	98.16	92.33	88.119	349.7204	87.33
		Geostatistics	5	3	0	81.84	29.68	79.3	87.03	89.87	78.95	94.43	77.07	98.06	97.65	86.28	81.387	391.0212	81.84
7	Worst Mean Accuracy	RLM_Mean	3	-	-	41.68	21	59.17	61.62	59.28	55.46	73.88	70.81	95.23	97.85	61.06	63.598	526.0995	41.68
		RLM_Mean	25	-	-	73.96	11.87	37.87	96.89	89.51	84.86	95.47	91.6	55.12	94.47	84.55	73.162	843.0157	73.96
		Geostatistics	3	2	1	56.96	48.86	75.15	57.52	81.52	66.26	91.7	76.25	97.53	98.46	73.05	75.021	307.9592	56.96
		Geostatistics	3	2	2	72.72	55.25	76.33	45.59	78.03	59.35	89.5	78.96	96.82	98.98	76.02	75.153	310.3257	72.72
		RLM_Mean	5	-	-	69.97	39.27	55.03	68.08	90.31	75.84	83.65	77.76	97.35	99.59	77.37	75.685	351.7401	69.97

Table 4 Features grade considering better mean accuracy and normalized overall accuracy

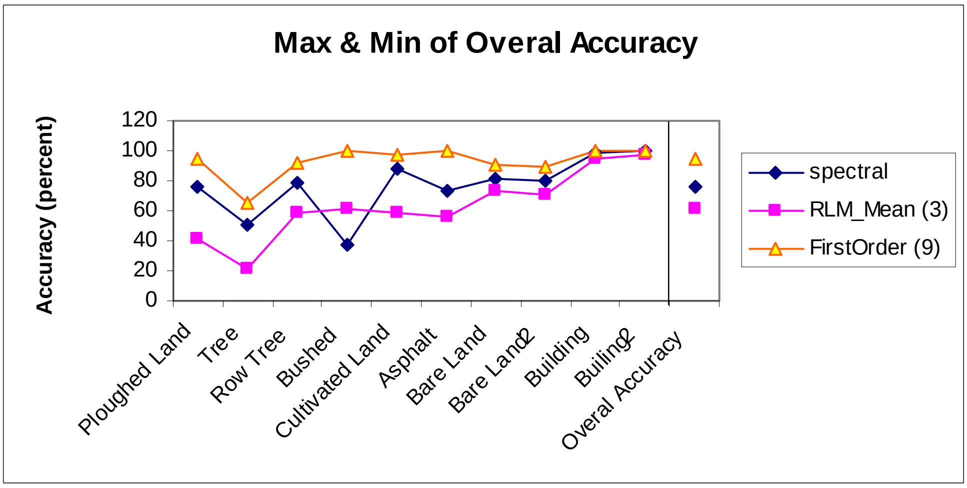
		ws	Lag		Mean Accuracy			ws	Lag		Normalized Overall
			x	y					x	Y	
1	First Order	7	-	-	93.007		First Order	5	-	-	1.302475
2	First Order	9	-	-	92.98		First Order	7	-	-	1.041066
3	First Order	5	-	-	92.114		First Order	3	-	-	0.876857
4	Geostatistics	9	1	1	92.017		First Order	9	-	-	0.858168
5	Geostatistics	7	1	1	91.546		Geostatistics	5	0	3	0.778638
6	Geostatistics	7	0	1	91.417		Geostatistics	7	4	5	0.664522
7	First Order	3	-	-	91.21		Geostatistics	7	3	5	0.663911
8	Geostatistics	9	2	1	90.472		Geostatistics	5	1	3	0.660369
9	Geostatistics	9	2	2	90.152		Geostatistics	7	6	3	0.623352
10	Geostatistics	9	1	2	89.468		Geostatistics	7	2	4	0.610159
11	Geostatistics	7	1	2	88.119		Autocorrelation	7	1	2	0.609127
12	Geostatistics	7	0	4	87.715		Geostatistics	5	0	2	0.606313
13	Geostatistics	7	0	2	87.441		Geostatistics	5	1	2	0.6062
14	Geostatistics	9	0	1	87.412		Geostatistics	7	0	4	0.596301
15	Geostatistics	7	2	1	87.297		Geostatistics	7	3	4	0.578512
16	Geostatistics	9	0	2	87.049		Geostatistics	7	2	5	0.567024
17	Geostatistics	9	0	4	87.019		Geostatistics	7	0	1	0.557598
18	Geostatistics	7	0	3	86.767		Geostatistics	7	3	6	0.550732
19	Geostatistics	9	1	4	86.587		Geostatistics	7	5	4	0.550048
20	Geostatistics	7	2	2	86.586		Autocorrelation	5	1	1	0.543256
21	Geostatistics	7	1	4	86.414		Geostatistics	5	2	3	0.539236
22	Geostatistics	9	3	3	86.265		Autocorrelation	5	0	2	0.530754
23	Geostatistics	5	1	1	86.196		Autocorrelation	7	0	3	0.529071
24	Geostatistics	7	1	3	86.104		Geostatistics	7	6	4	0.52864
25	Geostatistics	9	1	3	86.069		Geostatistics	7	2	6	0.52619
26	Geostatistics	9	3	1	85.959		Geostatistics	7	4	4	0.525452
27	Geostatistics	7	3	2	85.934		Autocorrelation	7	1	1	0.520172
28	Geostatistics	9	2	4	85.925		Geostatistics	7	1	4	0.519993
29	Geostatistics	7	3	1	85.763		Geostatistics	7	1	5	0.511445
30	Geostatistics	5	0	3	85.49		Autocorrelation	7	0	2	0.503506
31	Geostatistics	9	0	3	85.429		Geostatistics	7	5	5	0.495948
32	Geostatistics	9	3	2	85.398		Geostatistics	7	0	3	0.483315
33	Geostatistics	7	4	1	85.39		Geostatistics	7	1	3	0.481285
34	Geostatistics	5	0	2	85.299		Geostatistics	5	2	4	0.479202
35	Geostatistics	7	2	4	85.178		Geostatistics	5	0	4	0.473412
36	Geostatistics	7	2	0	85.157		Geostatistics	7	4	6	0.460928
37	Geostatistics	7	5	1	85.101		Autocorrelation	7	2	2	0.459864
38	Geostatistics	7	1	0	85.099		Geostatistics	7	1	6	0.457003
39	Geostatistics	9	2	3	84.996		Geostatistics	7	1	2	0.454169
40	Geostatistics	7	3	3	84.9		Geostatistics	7	2	3	0.452742



Figures 2-Minimum and Maximum of obtained mean accuracy



Figures 3 Minimum and Maximum of obtained accuracy variance



Figures 4 Minimum and Maximum of obtained overall accuracy

5. References

[1] Sergios Theodoridis, "Pattern Recognition", Academic Press,1999
[2] Kenneth R. Castleman, " Digital Image Processing”, Prentice-Hall, 1996
[3] http://www.geo.ed.ac.uk
[4] M. Chica-Olmo and F. Abarca-HernaÂ ndez, "Computing geostatistical image texture for remotely sensed data classification", Computers & Geosciences 26 (2000) 373-383
[5] Mona Sharma,Markos Markou, Sameer Singh "Evaluation of Texture Methods for Image Analysis", 2001