THE IMPORTANCE OF THE SELECTION OF THE TRAINING AREAS AT THE MAXIMUM-LIKELIHOOD CLASSIFICATION

Dr. Cankut ÖRMECİ

Technical University of Istanbul
Faculty of Civil Engineering, Dept. of Geodesy and Photogrammetry
80626 Ayazağa-Istanbul
Turkey

Commission VII

Abstract:

The result of a classification procedure is accurate only if the training samples are truely the representatives of the overall data set to be classified. Assuming that the classes have normal probability density functions, the distribution of a class response pattern can be described by the mean vector and the covariance matrix, which describes the variance and the correlation. Given these parameters, the statistical probability of a given pixel value can be computed. To see whether it is possible to use training areas,-chosen for a specific region-, for classifying neigbouring regions or not, a research is done on two neighbouring plains of Aegean region in Turkey

Introduction:

Maximum likelihood classification has found wide application in the field of remote sensing. The maximum likelihood classification is based on multivariate normal distribution theory. It provides a probabilistic method for recognizing similarities between individual measurements and predefined standards. Multispectral digital images of land areas make use of the maximum likelihood classification in producing thematic land-use maps of large areas.

The set of ground-cover classes being observed must be guessed from a digital image, or better, computed from the data source of a multispectral scanner. To classify the ground cover classes probability functions are useful, and these can be estimated from training patterns. The data which is the brightness value distribution in the training patterns can be displayed in the form of a histogram or a probability function. Generally the probability function for a class can be approximated by a normal (Gaussian) probability density function. Therefore, from the training samples, the statistical parameters, such as the mean or average value and the dispersion of the brightness values of the classes must be defined.

$$p(x/w_i) = (2\pi)^{-n/2} . |C_i|^{-1/2} . \exp(-2^{-1} . (x-m_i)^T . C_i^{-1} (x-m_i))$$

where m_i is the mean vector for class i,

C_i is the dispersion or the covariance matrix for class i,

$|C_i|$ is the determinant of the covariance matrix for class i.

This expression allows the calculation of the probability that an observation, a brightness value, is a member of each of i classes. In the maximum likelihood decision rule the observation is then assigned to the class for which the probability value is greatest.

In using this assumption adequate training samples must be available to estimate the mean vector and covariance matrix for each class. The success of the automatic spectral pattern recognition process relies on the quality of the training procedure and on the selection of the training samples. The result of the classification is accurate only if the training samples are truly the representatives of the overall data set to be classified. Otherwise, with training samples which are not representative for the classes, the classification would be inadequate, even wrong. In this paper the importance of the selection of training areas will be discussed with the aid of the land use classification of two neighbouring plains in the valleys of Gediz and Little Meander at the western part of Turkey. Further to see whether it is possible to use training areas chosen for a specific region, for classifying neighbouring regions or not, an analysis is done on the same neighbouring plains.

Description of the Study Area:

Western Anatolia, which lies west of longitude 30^0, comprises the Aegean coastland. The coast line is deeply indented and has many harbours. In this region the hills are well forested, and the valleys between them, running from east to west are extensively cultivated. The three east-west valleys of the Gediz, Little and Great Meander have the main importance in this region. These valleys are separated from each other by chains of mountains. A supervised classification for land-use is made at these neighbouring plains which are the Little Meander and Gediz. They are about 50 km apart from each other in north-west direction. The climate is very similar for both plains, the summers are hot and dry, the winters near the coast are mild and humid. Rainfall is moderate. These plains are very fertile and the principal crops are cotton, wheat, tobacco, olive, grape, fruit and vegetables.

Classification:

The areas selected for the application of the digital classification technique are two neighbouring plains at the Aegean region of Turkey. A wide variety of distinctive vegetation types such as cotton, tobacco, cropland, wineyard, forest are present. For the classification the CCT of Landsat satellite dated 31.07.1975 is used.

First the supervised land-use classification of the plain Little Meander has been done with the maximum likelihood method. Following an initial survey of the region, twelve cover classes were selected as being representative of the actual ground cover. These classes classify plant species, soil type, settlement and water, and are defined by a set of training sites ranging in size from approximately 20 to 130 pixels (Table 1). Then for each class the mean values and the covariance matrices are defined (Table 2).

Table 1. Land Use Classes Scheme Developed for Classification

Training Site No.	Class in the training site
1	Mixed forest
2	Wineyard
3	Cotton
4	Tobacco
5	Maquis
6	Olive grove
7	Vegetables
8	Cropland, corn
9	Water
10	Mixed class comprising sand-stone-and settlement
11	Soil with high content of water
12	Mud area

Table 2. The Mean Values and the Covariance Matrices of All the Classes

Class No.	Bands	Covariance matrix(symmetric) 4	5	6	7
1	Mean	28.9	32.1	59.8	69.9
	4	2.34	3.06	2.12	1.63
	5		6.93	1.38	-1.77
	6			53.84	68.96
	7				107.19
2	Mean	44.7	67.4	80.7	83.5
	4	10.82	8.14	9.42	8.65
	5		20.18	17.53	8.93
	6			29.69	19.33
	7				21.74
3	Mean	36.7	39.9	107.0	128.1
	4	4.00	5.60	-8.26	-15.83
	5		13.65	-20.86	-37.17
	6			92.69	134.91
	7				225.86

Class No.	Bands	Covariance matrix(symmetric) 4	5	6	7
4	Mean	42.3	55.9	83.1	89.6
	4	7.61	17.19	-15.70	-26.06
	5		50.73	-50.73	-83.10
	6			74.02	111.23
	7				182.44
5	Mean	39.4	52.9	72.6	80.7
	4	10.71	13.33	23.52	21.84
	5		21.67	32.75	31.31
	6			63.25	60.58
	7				64.28
6	Mean	45.8	67.8	78.9	79.8
	4	6.92	7.05	1.45	2.67
	5		11.99	4.02	5.42
	6			9.46	8.25
	7				16.76
7	Mean	52.6	75.1	80.8	74.6
	4	6.84	8.34	6.75	5.98
	5		17.84	14.89	11.20
	6			30.50	18.89
	7				15.41
8	Mean	62.2	105.4	111.6	111.5
	4	43.13	84.45	84.05	75.00
	5		187.92	174.50	148.68
	6			182.20	158.82
	7				161.59
9	Mean	42.1	38.3	22.4	9.9
	4	29.83	31.04	10.73	-3.26
	5		39.31	17.27	3.32
	6			14.74	9.97
	7				17.88
10	Mean	60.0	88.3	88.3	80.0
	4	38.27	59.13	49.40	38.40
	5		110.63	94.23	82.53
	6			87.96	77.00
	7				74.80

Class No.	Bands	Covariance matrix(symmetric) 4	5	6	7
11	Mean	42.9	63.5	67.3	64.7
	4	13.25	22.47	16.40	14.98
	5		47.67	34.23	31.40
	6			38.16	33.51
	7				39.86
12	Mean	28.8	34.1	33.4	33.0
	4	7.37	13.20	15.88	14.36
	5		28.64	32.55	28.98
	6			46.20	43.29
	7				49.09

In the Landsat image the plain of Little Meander consists of "2099 by 512" pixels. The classification is made with the bands 4,5 and 7. The results of the classification are presented in Table 3.

Table 3. Percent of Specific Classes after Max. Likelihood Classification

Class No.	Little Meander Plain	Gediz Plain with test sites of L.Meander	Gediz Plain with test sites of L.Meander-Gediz	Gediz Plain Class. with Eucl.Distance
1	16,0	8.0	6.7	6.2
2	13.1	16.5	16.0	14.1
3	9.4	6.0	5.6	5.7
4	13.5	15.5	13.6	15.7
5	24.6	19.3	18.1	18.3
6	3.7	7.7	6.8	7.7
7	0.6	2.9	2.5	2.4
8	18.0	18.1	17.9	17.6
9	0.04	0.3	1.6	1.6
10	1.1	5.7	2.7	2.8
11	-	-	7.0	6.4
12	-	-	1.5	1.5

Gediz plain, which consists of 2048 by 512 pixels in the Landsat image, is classified with the method of maximum likelihood with the training samples of the plain Little Meander. The results of the classification of Gediz plain using the test sites of the Little Meander plain are also given in Table 3.

At the second stage the Gediz plain is classified with the training samples of Little Meander which are completed with the samples out of the Gediz region. The results of the maximum likelihood classification of the region from the 2048 by 512 pixels are given in Table 3.

At the third stage the Gediz plain is classified with the method of Euclidian distance. The results are presented in Table 3 in the fourth column. The feature spaces of the both classification are to be seen in the Figures 1 and 2.

Conclusions:

The main difference in the result of the classification is seen at the new classes such as soil with high content of water, and mud area. These new classes were misclassified as forest, maquies and water. It is very interesting too, that the Marmara sea in the plain of Gediz was classified as a mixed class consisting of sand and stone, cotton and forest. The reason of this could be the turbidity of water and the underwater topography of the sea.

As it is not possible to prepare valid photo-interpretation keys for each region in visual interpretation, it is also not possible to obtain general valid statistical values for digital classification.

References

1. Lillesand, T.M.-Kiefer, R.W., 1979, Remote Sensing and Image Interpretation, John Wiley, Chichester, 612 p.

2. Örmeci, C., 1982, Eignung der Euklidischen-Distanz-Klassifikation von LANDSAT-Daten für kleinflachige, gemischte Landnutzung--untersucht an Hand eines Beispiels aus dem Tal von K.Menderes/Türkei, Bildmessung und Luftbildwesen 50, Heft 4, p. 131-137.

3. Örmeci, C., 1984, Klassifizierung von LANDSAT-Daten für kleinflachige, gemischte landnutzung, untersucht an Hand der Beispiele aus der Türkei, ISPRS Int. Archives of photogrammetry and Remote Sensing XV th Congress, Rio de Janeiro, Vol. XXV, Part A8, p. 526-530.

4. Swain, P.H. -Davis, S.M., 1978, Remote Sensing: The Quantitative Approach, McGraw-Hill, New York, 396 p.

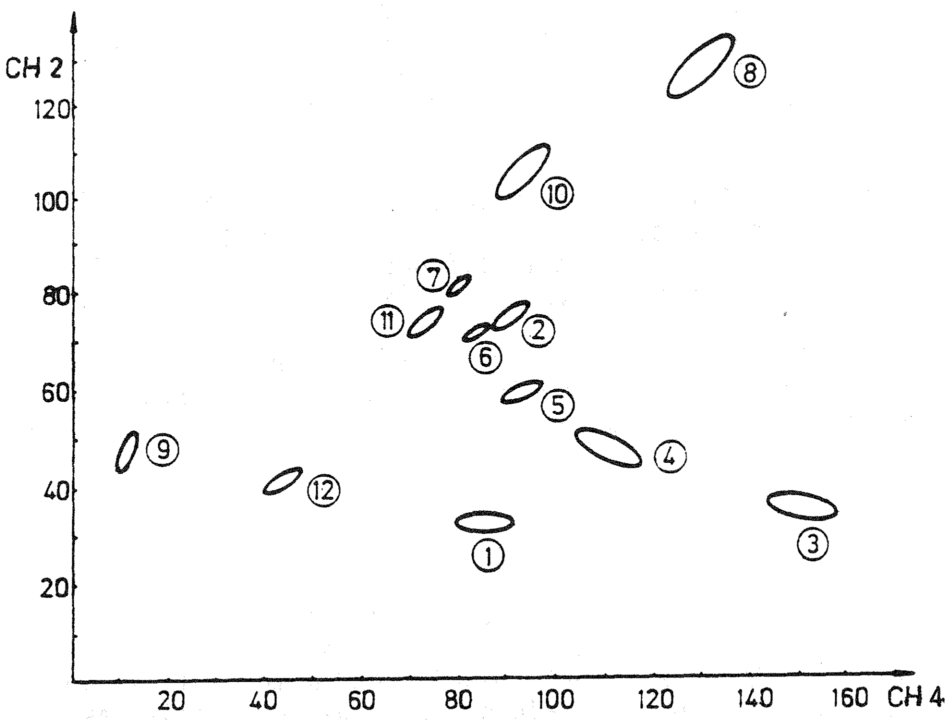

Figure 1. Equiprobability Elipses of the Classes

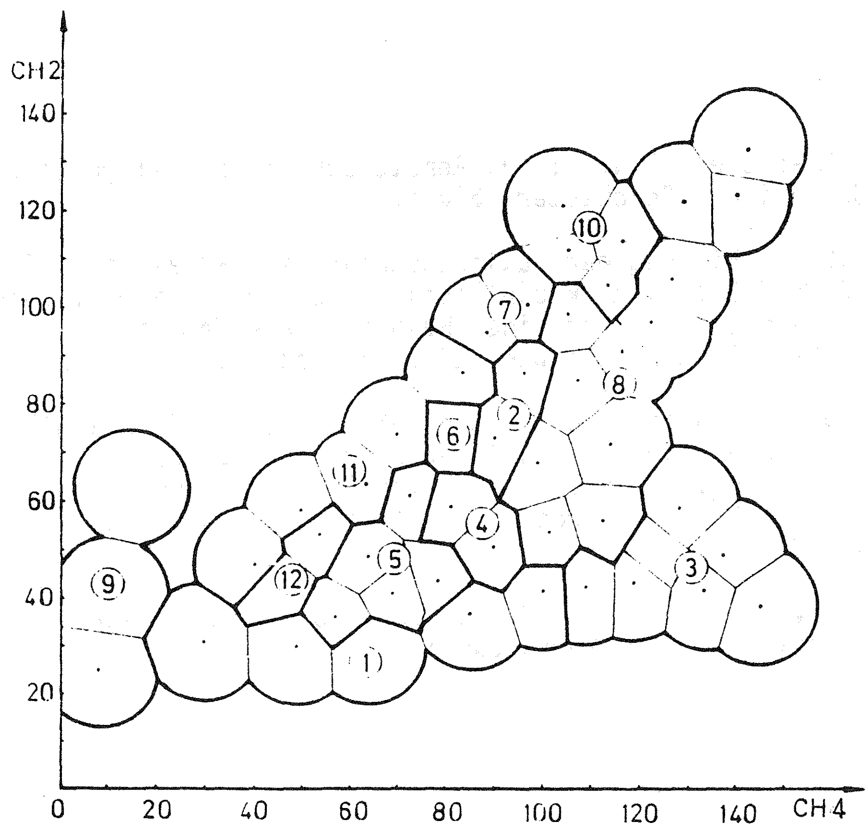

Figure 2. Clusters which define the Classes