Automatic Cloud Classification

B M MEHTRE, N N MURTHY
CMC Limited, 115 Sarojini Devi Road, Secunderabad 500 003 India

BOJAN LIPOVSCAK
Hydrometeorological Institute, CRIC 3, 41000 Zagreb, Yugoslavia

AND

B CHATTERJEE
E & ECE Deptt, Indian Institute of Technology, Kharagpur 721 302 India

In this paper, design and implementation of an automatic cloud classifier is discussed. It is assumed that the pattern classes obey the multivariate normal distribution. Maximum Likelihood with threshold and penalised misclassification algorithms have been implemented both on host computer VAX-11/750 and Array Processor FPS-100. This technique classifies not only different cloud types but land, sea and snow as well. The result of classification and the time required on host as well as on array processor for various complexities is presented. NOAA satellite pictures have been used for testing.

Indexing terms : Training phase, Classification phase, Maximum likelihood, threshold, Penalised misclassification Risk function

CLOUD classification is very useful in meteorology to identify the rain bearing clouds, tropical cyclones, estimating the extent of cloud coverage, snow coverage, and assessing the land erosion at sea coasts. One important by-product of the classification is its ability of data compression, which is very useful in data storage, transmission and processing.

The early classification attempts were manual and based on visible channel information [1]. The first attempt for computer based classification was based on multichannel measurements. Desbois *et al.* [2] used clustering method for classification of three channel Meteosat data to classify high level clouds.

PATTERN RECOGNITION MODEL

Following is the pattern recognition model for cloud classification (refer Fig 1).

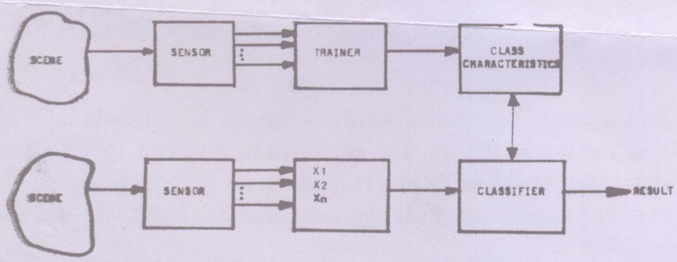

Fig 1 Pattern recognition model
Trainer is meant for training the classifier. This program estimates the class characteristics and stores them for use during classification. This is required to be done only once and the classifier continues to classify a given imagery correctly as long as there is no change in class characteristics

$X_1 \ldots\ldots X_n$ are the values of different channels of the multispectral scanner, as received from the satellite based sensor. Thus, each pixel is represented by an n-component measurement vector $\mathbf{X} = (X_1, X_2 \ldots\ldots X_n)\ T$. The classifier assigns the measurement vector to one of a prespecified class according to an appropriate classification rule.

CLASSIFICATION SCHEME

Classifying different cloud classes (including land, sea and snow) from a satellite imagery consists of the following two phases:

—To estimate the characteristics of each class

—To classify a given picture into different classes

The first phase is called the training phase (or learning phase) and the second phase is called the classification phase.

THE TRAINING PHASE

We have assumed that the patterns (cloud classes) obey normal distribution in each of the spectral band. The mean vector and covariance matrix completely characterise such types of patterns. Thus, the training phase consists of estimating the mean vector and covariance matrix for all classes of interest. Following is the list of cloud classes [3].

1. Cumulonimbus (Cb)
2. Cumulus congestus (Cucong)
3. Nimbostratus
4. Altocumulus (Ac)
5. Altostratus (As)
6. Cirostratus (Cs), Cirus (Ci), Cirocumulus (Cc)
7. Stratocumulus (Sc), Cumulushumuli (Cu)
8. Stratus (St), Fog

Manuscript received 1985 May 31; revised 1985 November 13.
(Accepted at International Geoscience and Remote Sensing Symposium (IGARSS), October 1985 at Amherst, USA)

9. Snow
10. Land
11. Sea

Classes 6 and 7 are families of cloud classes. These subclasses can not be identified by statistical method alone, as they vary in texture and shape. Thus, syntactic methods are suited to identify these subclasses. No attempt to identify these subclasses is made in this paper.

Estimation of the characteristics

Mean of a sample set having discrete values is given by [4].

$$M(N) = \frac{1}{n} \sum_{i=1}^{N} X(i)$$

where $X(i)$'s are the sample values

N is the number of samples

Similarly, the covariance can be estimated from the relation

$$C(N) = \frac{1}{N} \sum_{i=1}^{N} X(i)\, X(i)\, T - M(N)\, M(N)\, T$$

where $C(N)$ is the covariance of N elements

$X(i)\, T$ is the transpose of $X(i)$

and $M(N)T$ is transpose of $M(N)$

In the process of training, adequate number of sample elements may not be available at a time. In that case, it is required to update these parameters (characteristics) as and when more number of samples become available. This is necessary for better representation of the pattern class.

If we have two sets of Mean vectors $M(N)$ and $M(K)$ for N and K number of samples respectively, then the overall mean for $N+K$ samples is given by

$$M(N+K) = \frac{N^*M(N)+K^*M(K)}{N+K}$$

Similarly, if $C(N)$ and $C(K)$ are the covariance matrices for N and K number of samples respectively, then the overall covariance matrix is given by

$$C(N+K) = \frac{N}{N+K}(C(N)+M(N)^*(N)T)$$

$$+\frac{K}{N+K}(C(K)+M(K)^*M(K)T)$$

$$-\frac{(N^*M(N)+K^*M(K))\ ^*\ (N^*M(N)+K^*M(K))T}{(N+K)^2}$$

If n is the number of channels used, the Mean vector will have n elements, and covariance will be $n \times n$ symmetric matrix.

CLASSIFICATION PHASE

The job of a classifier is, given a pixel (n-dimensional measurement vector), assign the pixel to the class it belongs to. This is done by constructing some functions which can distinguish different classes. These functions are called discriminant functions. Following are the algorithms used to derive the discriminant functions.

MAXIMUM LIKELIHOOD PRINCIPLE (MLP)

Suppose there are M pixel classes. Let $p\,(X/w_i)$ be the probability density function associated with the measurement vector X, given that X is from class i. Let OP (w_i) be the *a priori* probability of class i. *A priori* probability is the probability of observing a pixel from class i, independent of any other information.

Then the MLP (5) states that assign X to class i, if and only if $p(X/w_i)\, P(w_i) > p(X/w_j)^*\, P(w_j)$ for all $j \neq i$

Thus, the discriminant function by MLP is given by $g_i(X) = p\,(w_i)^*\, P\,(X/w_i)$

$$= \frac{P(w_i)^* \exp\left(-0.5\,(X-M_i)T^*\, S_i^{-1*}\,(X-M_i)\right)}{(2^*\,PI)^{n/2}\ ^*\ (S_i)^{0.5}}$$

where M_i is the mean vector for class i,

S_i^{-1} is the inverse of covariance matrix and (S_i) is the absolute value of the determinant of covariance matrix.

An equivalent set of discriminant function can be derived by taking logarithm, which is simpler to implement.

$$G_i\,(X) = \log_e\, P\,(w_i) - 0.5\ \log_e\,(S)$$
$$-0.5\,((X-M_i)\,T^*S_i^{-1*}(X-M_i))$$

where $S = (2\,PI)^n\,(S_i)$

Once a problem has been specified and the statistical parameters estimated from training data, only the quadratic term (right most in the above equation) varies from pixel to pixel during the actual classification.

MLP WITH THRESHOLD

Notice that the pattern recognition system based on the MLP classifies every pattern presented into one of the classes it has been designed to recognise. But there are almost invarably a number of patterns in the area to be classified which in fact do not belong to any of the classes. Such patterns may belong to classes for which there are insufficient training patterns for estimating the parameters or the classes which have been completely overlooked. Although such patterns can not be correctly classified, since there is no discriminant function corresponding to correct class. We can at least design the classifier to detect them, provided they are spectrally very much different from the points in the 'valid' classes. As suggested by the one dimensional, two class example shown in Fig 2, the points to be detected correspond to those which have a very low probability of belonging to any of the trained classes. At the cost of 'rejecting' a very small percentage of the points actually belongingto the training classes, it is possible to reject a comparatively large number of points not belonging to any of the training classes. This is accomplished by

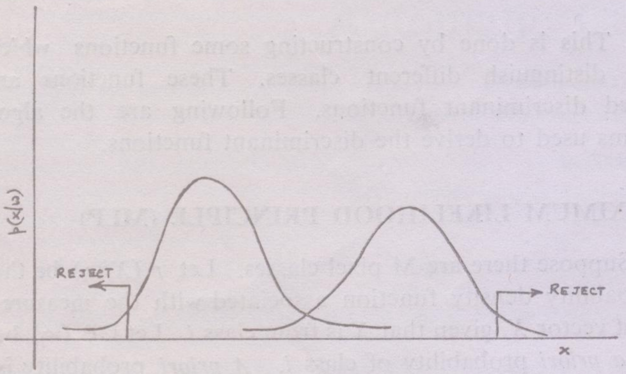

Fig 2 Thresholding principle

a technique known as 'Thresholding', in which the probability value $P(X/w_i)$ associated with the data vector and the class into which it would ordinarily be classified is compared with a user specified threshold. If the probability is below that threshold, the pixel is assigned to a 'Reject' class.

For the estimation of threshold value, consider the class conditional probability density function

$$P(X/w_i) = \frac{\exp\left(-0.5(X-M_i)T^*S_i^{-1*}\,(X-M_i)\right)}{(2\ PI)^{n/2*}\,(S_i)^{0.5}}$$

In this expression, the numerator varies from zero to one. This term is taken as the expression for cut-off probability.

Thus, the expression for estimating threshold value for a class becomes

Fig 3 NOAA satellite picture, VIS, Orbit 9514 dated 28-04-1983

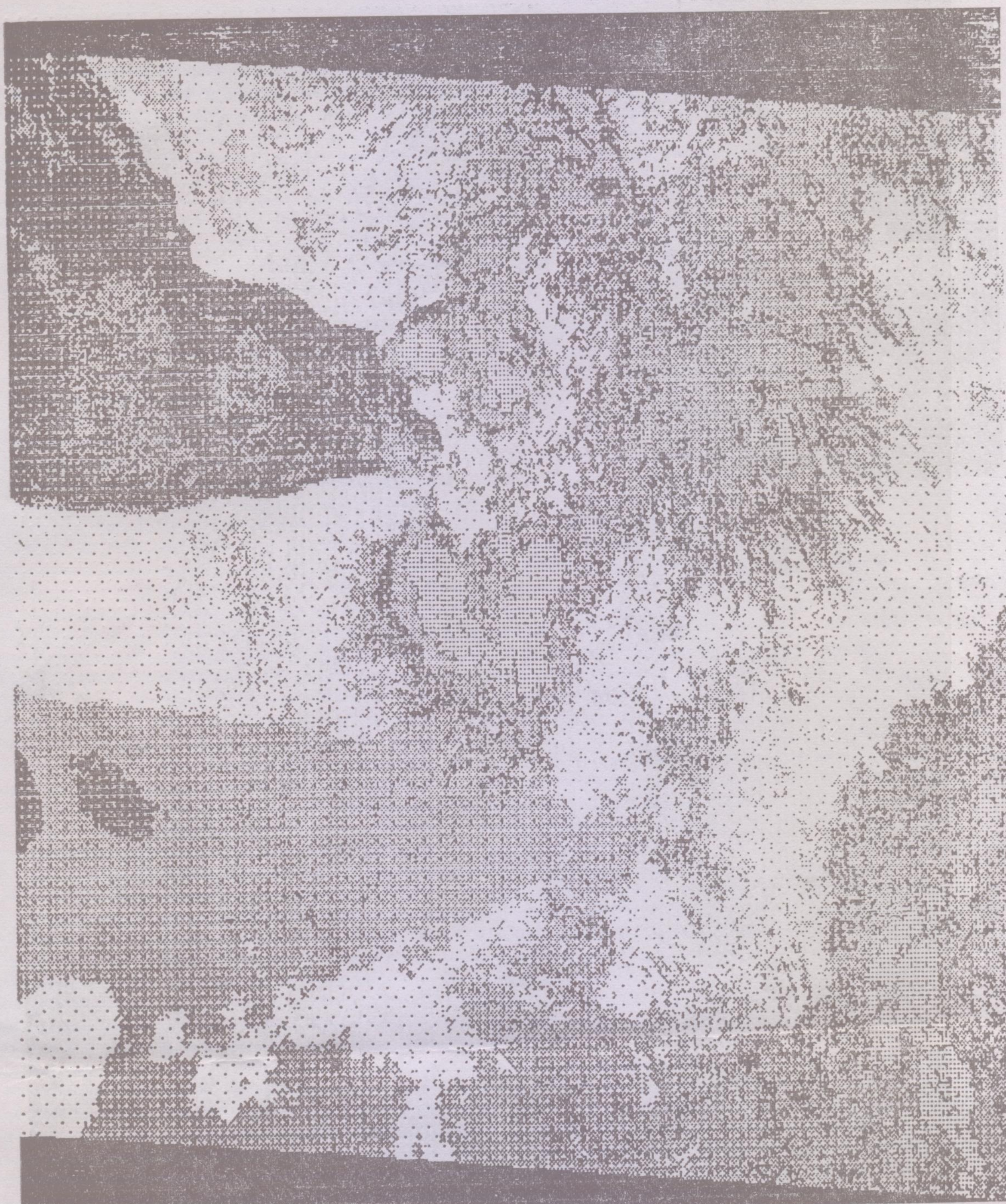

Fig 4 Classified picture with MLT algorithm

$$T(w_i) = \frac{\text{cprob}\ (i)}{(2^*\ PI)^{n/2*}\ (S_i)^{0.5}}$$

Where cprob (i) is the cut-off probability for class i, $0 < \text{cprob} < 1$

The cprob value is assigned by the user for each class. Then the threshold value is computed for each class. At the time of classification, the maximum probability class is compared against its threshold. If the maximum probability is greater than threshold, then it is assigned to that class, otherwise it is assigned to reject class.

PENALIZED MISCLASSIFICATION PRINCIPLE (PMP)

Using the same discriminat function, we can construct a slightly different discriminant function, which takes into account different losses involved in assigning a sample pattern to a particular class.

Suppose, a given measurement vector is assigned to a class i when its actual class is j, a loss $L\ (i, j)$ is incurred. If $P\ (w_j/X)$ is the probability that the actual class of the

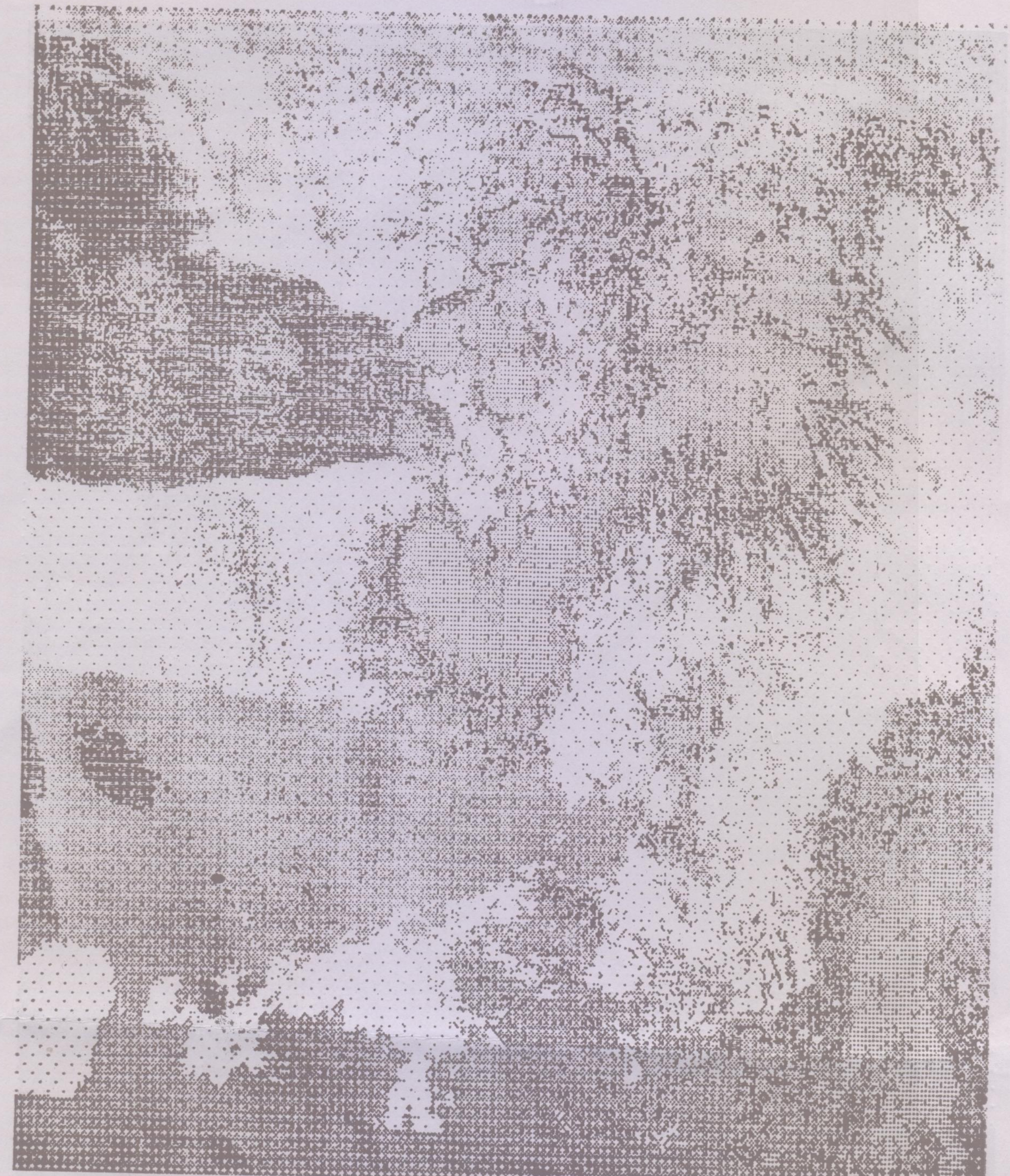

Fig 5 Classified picture with PMP algorithm

sample pattern is w_i (also called the posteriori probability), then the expected loss with the above decision rule is given by,

$$R(i) = \sum_{j=1}^{M} L(i, j)^* P(w_j/X)$$

In general $L(i, j) \leqslant 0$ for $i = j$

and $L(i, j) > 9$ for $i \neq j$ for

By Bayes Decision theory,

$$P(w_j/X) = \frac{P(X/w_j)^* P(w_j)}{P(X)}$$

where $P(X) = \sum_{i=1}^{M} P(X/w_j)^* P(w_i)$, this is constant and can be ignored.

Then the discriminant function is given by

$$R(i) = \sum_{i=1}^{M} L(i, j)^* P(X/w_j)^* P(w_j)$$

$R(i)$ is called the Risk function [6] risk involved in assigning a sample X belonging to class j to class i, $L(i, j)$ is also called the loss matrix or penalty matrix.

Table 1 Results after training for 8 classes

Cloud Classes	Number of Pixels	Class Mean	Class Covariance		
1	1052	202.34886	42.43945	59.26978	4.28564
		99.49905	59.26978	210.68347	—3.73413
		231.90114	4.28564	—3.73413	3.64355
2	0	0.00000	0.00000	0.00000	0.00000
		0.00000	0.00000	0.00000	0.00000
		0.00000	0.00000	0.00000	0.00000
3	0	0.00000	0.00000	0.00000	0.00000
		0.00000	0.00000	0.00000	0.00000
		0.00000	0.00000	0.00000	0.00000
4	0	0.00000	0.00000	0.00000	0.00000
		0.00000	0.00000	0.00000	0.00000
		0.00000	0.00000	0.00000	0.00000
5	866	108.08314	1.81287	—0.44836	—0.87598
		49.17090	—0.44836	34.90613	43.43176
		142.21709	—0.87598	43.43176	75.30835
6	549	135.97267	210.82495	62.06851	111.4886
		25.83789	62.06851	19.75333	35.84644
		176.07468	111.48682	35.84644	83.26416
7	297	152.44444	152.10547	131.09009	10.38989
		54.89562	131.09009	280.76688	—127.25354
		164.02356	10.38989	—127.25354	396.39990
8	892	95.96076	1.39868	—0.22531	2.74084
		23.17152	—0.22531	1.08157	—1.48853
		168.99214	2.74084	—1.48853	13.15381
9	1481	169.06549	125.87646	—42.91138	18.56348
		54.65631	—42.91138	149.96490	—29.64868
		231.01485	18.56348	—29.64868	47.67139
10	2369	89.42297	22.11084	0.30322	42.58789
		26.96581	0.30322	5.29175	—6.63379
		155.30815	42.58789	—6.63379	134.85938
11	4136	89.79763	12.27832	6.89954	—23.40234
		15.17626	6.89954	5.45854	—19.14600
		138.80995	—23.40234	—19.14600	75.90234

Table 2 The time factor

Details			Alg.	Host run time (Hrs : Mins : Seconds)		AP run time (Minute : Seconds)	
				CPU	Elapsed	CPU	Elapsed
1.	$N=1$	$C=2$	1	0:04:17.59	0:08:44.20	0:5.35	3:09.45
			2	0:05:43.74	0:07:05.53	0:5.24	3:57.44
2.	$N=1$	$C=4$	1	0:07:18.07	0:11:52.83	0:5.14	4:38.50
			2	0:14:49.01	0:37:26.60	0:5.39	4:35.96
3.	$N=3$	$C=8$	1	0:21:58.70	0:37:29.11	0:11.52	8:09.72
			2	0:47:33.57	1:07:35.19	0:11.49	8:40.19

(These timings refer to 512×512 size picture of 8 bit data)
N is the number of channels used for classification (and training).
C is the number of classes for which classifier is trained.
Alg. 1 : refers to maximum likelihood with threshold.
Alg. 2 : refers to penalized misclassification

A given sample pixel is assigned to that class for which the risk function is minimum.

RESULTS

NOAA satellite pictures have been used for testing the above classification scheme.

Figure 3 shows a NOAA-7 satellite picture in VIS channel. The TRAINER was run to acquire characteristics of different cloud classes. Table 1 shows the parameters estimated by the TRAINER program. In all, it was possible to train the classifier for eight classes from the available pictures. Channels 1,3 and 4 of the AVHRR channels (refer appendix) were used.

The Classifier was tested at various stages of training. To start with, it was trained for only one class using only one channel and tested for classification. Using Maximum likelihood with threshold, it gave two classes, one for which it was trained and the other the 'reject' class.

Figure 3 shows the input picture used for classification. Figures 4 and 5 are the classified pictures using the two algorithms.

Table 1 gives a glimpse at the time requirements of classification, with different input picture complexities, and also for different algorithms. It is clear from the table that the CPU time increases with the number of channels used and the number of classes involved in the classification. As expected, AP runs are much faster than the host. Also, the increase in CPU time for more number of channels on AP runs is much less compared to the host CPU time. It can also be observed from Table 2 that the PMP is taking almost double the time taken by MLT on the host. This difference is seen only on host and is missing conspicuously on AP runs.

CONCLUSION

The classifier is found working satisfactorily even with small set of training samples (as few as 250 pixels for one class). It is evident from the pictures shown that the classifier has identified the land, sea and different types of clouds clearly. It is felt that this technique is a good tool in the hands of researchers and meteorologists, in addition to being a general purpose classifier for satellite imagery. It will also be useful in building an expert system for satellite imagery.

ACKNOWLEDGEMENT

The authors are thankful to Mr. R.P. Jhunjhunuwala and Mr. S. Kapoor, Project INTERACT, CMC Limited, for providing all the facilities to carry out this work.

REFERENCES

1. B M Mehtre, Automatic cloud classification, M Tech thesis, Dec 1984, Indian Institute of Technology, Kharagpur.
2. Micheal Desbois *et al*, Automatic classification of clouds on METEOSAT Imagery : Application to high level clouds, *Journal of Applied Meteorology*, Mar 1982.
3. Bojan Lipovscak: Pattern recognition in meteorological satellite imagery; *CMC Technical Digest*, vol 5, pp 17–19, Aug 1983.
4. J Tou & R Gonzalez, *Pattern recognition techniques*, Addison Wesley, 1974.
5. P H Swain & S M Davis : *Remote sensing—The quantitative approach*; McGraw-Hill, 1978.
6. R O Duda and P E Hart, *Pattern classification and scene analysis*, New York, Wiley, 1973.
7. Floating Point Systems Inc., Array Processor Manuals, APMATH 38, vol 1–4.

APPENDIX : AVHRR Channels

In AVHRR, channels 1 and 2 (visible, as they are called) represent the reflectance in the spectral band, correspond to the thickness and composition of clouds. Channels 4 and 5 (called IR) represent the radiation in the spectral band, correspond to cloud temperature. And channel 3 (near IR) represents both reflectance and radiation, corresponds to the temperature and brightness (and in good for the land-sea boundary demarcation). Channels 3,4 and 5 are called the thermal channels. Following is spectrum of AVHRR channels:

CHANNEL	WAVELENGTH (in micrometre)
1	0.55— 0.68
2	0.72— 1.1
3	3.55— 3.93
4	10.3 —11.3
5	11.5 —12.5

Reprinted from IETE Technical Review of the Institution of Electronics and Telecommunication Engineers, Vol. 3, No. 6, 1986 pp, 272-278