Two-Dimensional Hydrometeor Image Classification by Statistical Pattern Recognition Algorithms

MIZANUR M. RAHMAN,[1] RAYMOND G. JACQUOT,[1] EDMUND A. QUINCY[1] AND RONALD E. STEWART[2,3]

University of Wyoming, Laramie, WY 82071

(Manuscript received 12 August 1980, in final form 17 January 1981)

ABSTRACT

The investigation reported here involves the automatic classification of binary (black and white) images of hydrometeors (ice particles and raindrops) taken from cloud samples. The goal is to classify such images (both complete and fractional) into the seven most common classes of hydrometeors by statistical pattern recognition techniques. Detailed investigation about the data acquisition system and preprocessing is made. Four moment invariants which yield good class separation were used as features for the classification process. A Bayes decision function which minimizes the probability of misclassification is used for classification.

Bayes theorem is employed to update mean vectors and covariance matrices involved in the decision function. A discrete Kalman filtering algorithm is developed for the on-line estimation of the probability of occurrence of each class. For such estimation a discrete adaptive Kalman filtering algorithm is also developed which adjusts the filter gain matrix such as to whiten the innovations sequence. These techniques were shown to work well but the adaptive algorithm was found to converge to the correct probability more rapidly.

The classification algorithm was modified to classify incomplete or fractional images and two metrics were successfully developed to detect the unclassifiable images. The adaptive Kalman filter with the Bayes decision function was employed to classify about 2000 images per minute of CPU time with about 10% error.

1. Introduction

Artificial precipitation or cloud seeding for enhanced precipitation is no longer a scientific fantasy. Several research and commercial cloud seeding projects have indicated that precipitation and resulting streamflow can be increased by seeding specific types of weather systems. For artificial cloud modification and control, it is very important for cloud physicists to understand the microphysical structure of clouds. An extensive investigation on hydrometeor class, feature extraction, feature selection and classifier has already been made (Rahman *et al.*, 1981).

The data acquisition system and preprocessing are discussed in Section 2. Detailed investigation is made on problems related to application of pattern recognition techniques to actual data in Section 3. Finally, the results of classification are given in Section 4.

2. Data acquisition system and preprocessing

The two-dimensional optical array spectrometer (2D probe) is used to record two-dimensional binary images of the hydrometeors. The 2D probe is packed in a cylindrical pod and requires a special data acquisition system for readout (Rahman, 1980). Two such probes, each with a different resolution are mounted on the wings of the University of Wyoming's King Air cloud physics aircraft which is used to collect data for this research as illustrated in Fig. 1. The data acquisition system is housed inside the aircraft. Examples of plotted image data are shown in Fig. 2.

The most important preprocessing needed for the interpretation of the actual images is to recognize artifacts. The artifacts are rejected from the raw data as proposed by Cooper (1978).[4] The four most common artifacts to be rejected are as follows:

1) Streakers (usually, water shedding across the upstream edge of the probe and across the aperture)

[1] Department of Electrical Engineering.
[2] Department of Atmospheric Science.
[3] Present affiliation: Atmospheric Environment Service, City of North York, Downsview, Ontario M3H5T4 Canada.

[4] Cooper, W. A., 1978: Cloud physics investigation by the University of Wyoming in Hiplex—1977. Dept. of Atmospheric Science, University of Wyoming, 30–31.

0021-8952/81/050536-11$06.75

are rejected if they are six times as long (longitudinal dimension) as they are wide (lateral dimension). Images are also rejected as streakers if they are three times as long as wide, if they are <150 μm wide and do not touch the edge of the scan region.

2) If the time interval between images corresponds to a spatial distance of 2.5 cm, the second image is not accepted. Series of images resulting from splashes from the edges of the aperture are eliminated in this way.

3) Cases in which the probe recording circuitry is triggered but no elements are recorded as shadowed are rejected. These often correspond to true objects near the minimum size detectable by the probe.

4) Images consisting of parts not connect-together in the longitudinal direction are rejected to avoid splash images.

3. Problems related to classification

a. Bayes estimation

The Bayes theorem can be used to update mean vector $\mathbf{m}_i$ and covariance matrix C_i of the decision function for time domain classifier (Rahman *et al.*, 1981) successively as the patterns are received (Fukunaga, 1972). To estimate $\mathbf{m}$ and C of a given class successively from normally distributed n-dimensional N samples $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$, we have to calculate the joint *a posteriori* density function $p(\mathbf{m}, \mathsf{C}|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$. Instead of calculating $p(\mathbf{m}, \mathsf{C}|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$, we can compute $p(\mathbf{m}, \mathsf{K}|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$, where $\mathsf{K} = \mathsf{C}^{-1}$. The reason for doing this is that the covariance matrix is always used in the inverse form for a normal distribution. After receiving N patterns the conditional probability function $p(\mathbf{m}, \mathsf{K}|\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N)$ becomes the Gaussian-Wishart distribution and the updated mean vector $\mathbf{m}_N$ and covariance matrix C_N are (Keenh, 1965)

$$\mathbf{m}_N = \frac{N^{-1}\sum_{i=1}^{N}\mathbf{X}_i + N^{-1}\mu_0\mathbf{m}_0}{1 + N^{-1}\mu_0}, \tag{1}$$

$$\mathsf{C}_N = \frac{1}{1+N^{-1}N_0}\left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\right) - (1+N^{-1}\mu_0)\mathbf{m}_N\mathbf{m}_N^{\mathrm{T}} + N^{-1}N_0\mathsf{C}_0 + (N^{-1}\mu_0)\mathbf{m}_0\mathbf{m}_0^{\mathrm{T}}\right], \tag{2}$$

where μ_0 and N_0 are the number of samples for initial sample mean vector $\mathbf{m}_0$ and sample covariance matrix C_0. The $\mathbf{m}_0$ and C_0 are given by

$$\mathbf{m}_0 = \mu_0^{-1}\sum_{l=1}^{\mu_0}\mathbf{X}_l, \tag{3}$$

$$\mathsf{C}_0 = \left[N_0^{-1}\sum_{l=1}^{N_0}\mathbf{X}_l\mathbf{X}_l^{\mathrm{T}}\right] - \mathbf{m}_0\mathbf{m}_0^{\mathrm{T}}. \tag{4}$$

FIG. 1. University of Wyoming's King Air cloud physics aircraft.

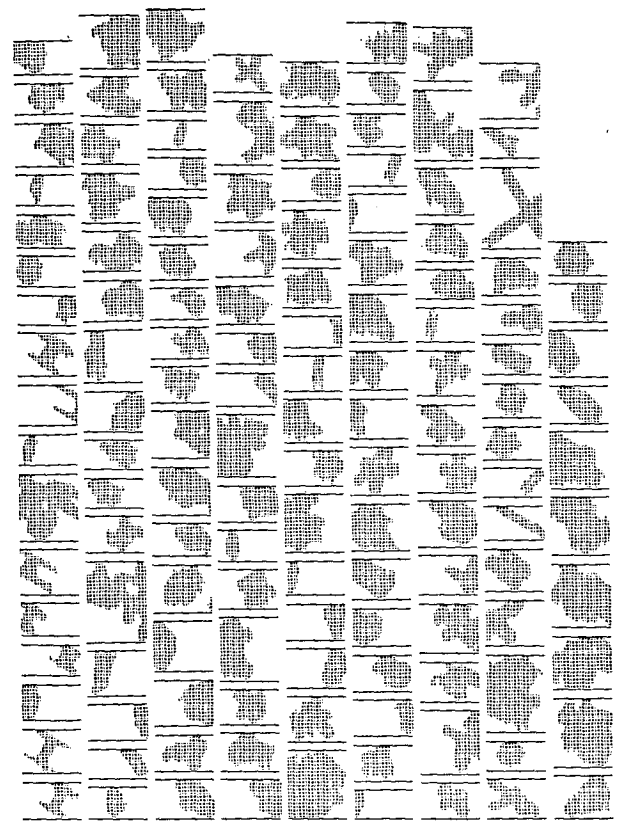

FIG. 2. A typical set of actual data.

b. Probability estimation by discrete Kalman filter

The discrete Kalman filter is a linear minimum-variance unbiased recursive estimator. We consider a discrete system whose state equation is given by

$$\mathbf{X}_k = \phi_{k-1}\mathbf{X}_{k-1} + \mathbf{W}_{k-1}, \quad \mathbf{W}_k \sim N(\mathbf{0}, \mathsf{Q}_k), \quad (5)$$

where $\mathbf{X}_k$ is the $(n \times 1)$ state vector at time t_k, ϕ_{k-1} is the $(n \times n)$ state transition matrix at time t_{k-1}, and $\mathbf{W}_k$ is a $(n \times 1)$ purely random Gaussian process noise vector with zero mean and covariance matrix Q_k. Measurements are taken as linear combinations of the system state variables, corrupted by uncorrelated additive noise, i.e.,

$$\mathbf{Z}_k = \mathsf{H}_k\mathbf{X}_k + \mathbf{V}_k, \quad \mathbf{V}_k \sim N(\mathbf{0}, \mathsf{R}_k), \quad (6)$$

where $\mathbf{Z}_k$ is an $(l \times 1)$ measurement vector, H_k is the $(l \times n)$ measurement matrix, and $\mathbf{V}_k$ is an $(l \times 1)$ Gaussian measurement purely random noise vector sequence with zero mean and covariance matrix R_k. Given a prior estimate of the system at time t_k, denoted $\hat{\mathbf{X}}_k(-)$, we seek an updated estimate, $\hat{\mathbf{X}}_k(+)$, based on use of the measurement $\mathbf{Z}_k$. The discrete Kalman filtering algorithm is most suitable for such estimation and is given by (Gelb, 1974)

$$\hat{\mathbf{X}}_k(+) = \hat{\mathbf{X}}_k(-) + \mathsf{K}_k[\mathbf{Z}_k - H_k\hat{\mathbf{X}}_k(-)], \quad (7)$$

$$\hat{\mathbf{X}}_k(-) = \phi_{k-1}\hat{\mathbf{X}}_{k-1}(+), \quad (8)$$

$$\mathsf{K}_k = \mathsf{P}_k(-)\mathsf{H}_k^{\mathrm{T}}[\mathsf{H}_k\mathsf{P}_k(-)\mathsf{H}_k^{\mathrm{T}} + \mathsf{R}_k]^{-1}, \quad (9)$$

$$\mathsf{P}_k(-) = \phi_{k-1}\mathsf{P}_{k-1}(+)\phi_{k-1}^{\mathrm{T}} + \mathsf{Q}_{k-1}, \quad (10)$$

$$\mathsf{P}_k(+) = [\mathsf{I} - \mathsf{K}_k\mathsf{H}_k]\mathsf{P}_k(-), \quad (11)$$

where K_k is referred to as the Kalman gain matrix and P_k is the error covariance matrix. From the definition (E denotes ensemble average)

$$\mathsf{P}_k = E[(\mathbf{X}_k - \hat{\mathbf{X}}_k)(\mathbf{X}_k - \hat{\mathbf{X}}_k)^{\mathrm{T}}]. \quad (12)$$

It is assumed that $E[\mathbf{W}_k\mathbf{V}_j^{\mathrm{T}}] = 0$ for all j and k.

A special version of the discrete Kalman filter is needed when the state vector $\mathbf{X}_k$ is an $(n \times 1)$ probability vector or vector of relative frequency of occurrence of the n classes. The vector being estimated at $(k-1)$th step is the relative frequency of occurrence of each class during the first $k-1$ samples, i.e.,

$$\mathbf{X}_{k-1} = \mathbf{Y}_{k-1}[1/(k-1)], \quad (13)$$

where $\mathbf{Y}_{k-1}$ is a vector of the number of occurrences of each class during the first $k-1$ samples. Then

$$\mathbf{X}_k = [\mathbf{Y}_{k-1} + \mathbf{U}][1/k], \quad (14)$$

where $\mathbf{U}$ is a vector of $(n-1)$ zeros and one unity, the order of which depends on the class to which the kth sample belongs. This vector will be a random vector because of the random occurrence of the patterns. Eq. (14) can be rewritten as (Rahman, 1980)

$$\mathbf{X}_k = \{[\mathbf{Y}_{k-1}][1/(k-1)] + [\mathbf{U}/(k-1)]\}\left[1 - \frac{1}{k-1} + \cdots\right]. \quad (15)$$

By substituting (13) into (15) and evaluating for large k

$$\mathbf{X}_k = \mathbf{X}_{k-1} + \mathbf{W}_{k-1}, \quad (16)$$

where $\mathbf{W}_{k-1} = [\mathbf{U}/(k-1)]$, which can be assumed as random because $\mathbf{U}$ is random.

We assume that the measure of every state is available and the $(n \times 1)$ vector $\mathbf{Z}_k$ is a measure of $\mathbf{X}_k$ in presence of additive random noise $\mathbf{V}_k$. Then

$$\mathbf{Z}_k = \mathbf{X}_k + \mathbf{V}_k, \quad (17)$$

since $\mathbf{X}_k$ is a probability vector the sum of its components must be unity or

$$\sum_{i=1}^{n} x_{ik} = 1 \quad \text{for all } k. \quad (18)$$

Since $\mathbf{Z}_k$ is also a probability vector, a similar restriction must hold, i.e.,

$$\sum_{i=1}^{n} z_{ik} = 1 \quad \text{for all } k. \quad (19)$$

Eqs. (16) and (17) may be assumed as state and measurement equations, respectively and the discrete Kalman filtering algorithm can be applied. By comparing (16) and (17) with (5) and (6), respectively, we have

$$\phi_{k-1} = \mathsf{I}, \quad (20)$$

$$\mathsf{H}_k = \mathsf{I}, \quad (21)$$

where I is the identity matrix. The components of the estimate $\hat{\mathbf{X}}_k(+)$ must satisfy (18), i.e.,

$$\sum_{i=1}^{n} \hat{x}_{ik}(+) = 1 \quad \text{for all } k. \quad (22)$$

Substituting (20) and (21) into (7)–(11) yields

$$\hat{\mathbf{X}}_k(+) = \hat{\mathbf{X}}_k(-) + \mathsf{K}_k[\mathbf{Z}_k - \hat{\mathbf{X}}_k(-)], \quad (23)$$

$$\hat{\mathbf{X}}_k(-) = \hat{\mathbf{X}}_{k-1}(+), \quad (24)$$

$$\mathsf{K}_k = \mathsf{P}_k(-)[\mathsf{P}_k(-) + \mathsf{R}_k]^{-1}, \quad (25)$$

$$\mathsf{P}_k(-) = \mathsf{P}_{k-1}(+) + \mathsf{Q}_{k-1}, \quad (26)$$

$$\mathsf{P}_k(+) = [\mathsf{I} - \mathsf{K}_k]\mathsf{P}_k(-), \quad (27)$$

Substituting (24) into (23) for $k = 1$ results in

$$\hat{\mathbf{X}}_1(+) = \hat{\mathbf{X}}_0(+) + \mathsf{K}_1[\mathbf{Z}_1 - \hat{\mathbf{X}}_0(+)]. \quad (28)$$

If K_1 is a diagonal matrix, that is, $\mathsf{K}_1 = g_1\mathsf{I}$, this equation will result n algebraic equations. By taking sum of these n equations gives

$$\sum_{i=1}^{n} \hat{x}_{i1}(+) = \sum_{i=1}^{n} \hat{x}_{i0}(+) + g_1\left[\sum_{i=1}^{n} z_{i1} - \sum_{i=1}^{n} \hat{x}_{i0}(+)\right]. \quad (29)$$

Now if

$$\sum_{i=1}^{n} \hat{x}_{i0}(+) = 1 \tag{30}$$

and with substitution of (19) into (29), the result is

$$\sum_{i=1}^{n} \hat{x}_{i1}(+) = 1. \tag{31}$$

From (25) K_1 will be a diagonal matrix if $\mathsf{P}_1(-)$ and R_1 are diagonal matrices. Also, from (26) $\mathsf{P}_1(-)$ will be a diagonal matrix if $\mathsf{P}_0(+)$ and Q_0 are diagonal matrices. With these assumptions and by repeating the above steps for $k = 2, 3, \ldots$ we get similar results. Therefore, (22) will be satisfied if (30) and the following conditions are satisfied:

$$\mathsf{Q}_k = q_k\mathsf{I}, \tag{32}$$

$$\mathsf{R}_k = r_k\mathsf{I}, \tag{33}$$

$$\mathsf{P}_0(+) = p_0(+)\mathsf{I}, \tag{34}$$

where q_k, r_k, and $p_0(+)$ are scalars and I is the identity matrix.

c. *Probability estimation by discrete adaptive Kalman filtering*

The innovation sequence contains information on the performance of the Kalman filter. The innovation sequence is defined as

$$\boldsymbol{\nu}_k = \mathbf{Z}_k - \mathsf{H}_k\hat{\mathbf{X}}_k(-). \tag{35}$$

The innovation property (Kailath, 1968) states that the innovation sequence, $\boldsymbol{\nu}_k$, is a "white" noise sequence. Heuristically, there is no information left in $\boldsymbol{\nu}_k$, if $\hat{\mathbf{X}}_k$ is an optimal estimate. That is, its autocovariance matrix is zero, i.e.,

$$E[\boldsymbol{\nu}_k\boldsymbol{\nu}_{k-j}^{\mathrm{T}}] = 0 \quad \text{for all } j \neq 0. \tag{36}$$

By substituting (6) into (35), we have

$$\boldsymbol{\nu}_k = \mathsf{H}_k[\mathbf{X}_k - \hat{\mathbf{X}}_k(-)] + \mathbf{V}_k. \tag{37}$$

Substituting (37) into (36) for $j = 0$, after some manipulation (Gelb, 1974), yields

$$E[\boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{T}}] = \mathsf{H}_k\mathsf{P}_k(-)\mathsf{H}_k^{\mathrm{T}} + \mathsf{R}_k. \tag{38}$$

If it is assumed that our model is a stationary and ergodic random sequence, Mehra (1972) has shown that the time-average approximation of the actual covariance given by

$$E[\boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{T}}] \approx N^{-1} \sum_{i=k-N+1}^{k} \boldsymbol{\nu}_i\boldsymbol{\nu}_i^{\mathrm{T}} \tag{39}$$

may be employed. The number of data points N is chosen empirically to give some statistical smoothing. From (38) and (39), we can estimate R_k as

$$\hat{\mathsf{R}}_k = N^{-1} \sum_{i=k-N+1}^{k} \boldsymbol{\nu}_i\boldsymbol{\nu}_i^{\mathrm{T}} - \mathsf{H}_k\mathsf{P}_k(-)\mathsf{H}_k^{\mathrm{T}}. \tag{40}$$

To estimate probability by the discrete adaptive Kalman filter, the usual task of probability estimation is carried out as described in Section 3b but instead of using (33), real time estimation of R_k is made using (40).

d. *A priori probability estimation*

The Bayes decision function $d_i(\mathbf{X})$ is given by

$$d_i(\mathbf{X}) = p(\omega_i|\mathbf{X}), \tag{41}$$

where ω_i is the ith class and $p(\omega_i|\mathbf{X})$ is the *a posteriori* probability which is given as

$$p(\omega_i|\mathbf{X}) = \frac{p(\mathbf{X}|\omega_i)p(\omega_i)}{\sum_{j=1}^{M} p(\mathbf{X}|\omega_j)p(\omega_j)}, \tag{42}$$

where M is the number of classes, $p(\omega_i)$ is the *a priori* probability and $p(\mathbf{X}|\omega_i)$ is the probability density function. In this research effort, sufficient prerecorded training patterns exist in each class and a good estimate of $p(\mathbf{X}|\omega_i)$ is available. It was assumed $p(\omega_i) = 1/M$ but after examining a few actual data tapes it is clear that this assumption is not always true. Depending on the cloud conditions the patterns may belong to only two or three classes. The discrete Kalman filtering algorithm developed can be used for on-line estimation of $p(\omega_i)$. If the initial estimates of $p(\omega_i)$ are not accurate, this algorithm will adjust these estimates so that they become more representative of the patterns being classified. The $(M \times 1)$ state vector is defined as

$$\mathbf{X}_k = [p(\omega_1), p(\omega_2), \ldots, p(\omega_M)]^{\mathrm{T}}, \tag{43}$$

and the $(M \times 1)$ measurement vector is assumed as

$$\mathbf{Z}_k = [p(\omega_1|\mathbf{X}_k), p(\omega_2|\mathbf{X}_k), \ldots, p(\omega_M|\mathbf{X}_k)]^{\mathrm{T}}. \tag{44}$$

Then (23)–(27) with the associated conditions can be used for on-line estimation of the *a priori* probabilities.

To start this algorithm, we need the values of M, $\hat{\mathbf{X}}_0, p_0, q$ and r. Since four selected features are used for classification of patterns into seven classes, $M = 7$ and $\hat{\mathbf{X}}_0 = [1/7, 1/7, 1/7, 1/7, 1/7, 1/7, 1/7]^{\mathrm{T}}$ are used. Since the magnitude of the largest error component in $\hat{\mathbf{X}}_0$ and $\mathbf{V}_k$ are both less than unity, p_0 and r should have magnitude less than unity. Also, p_0, q and r are the diagonal elements of the diagonal covariance matrix, so all of these must be positive. From the development of (16) it is clear that q is much smaller than p_0 and r because $\mathbf{W}_{k-1}$ decreases as $1/(k-1)$. The exact value of p_0, q and r is found by trial and error. Three sets of training data were selected first. Then the filtering algorithm was used to classify every set with different values of these parameters. Those values of parameters were selected which produce overall minimum error in classification.

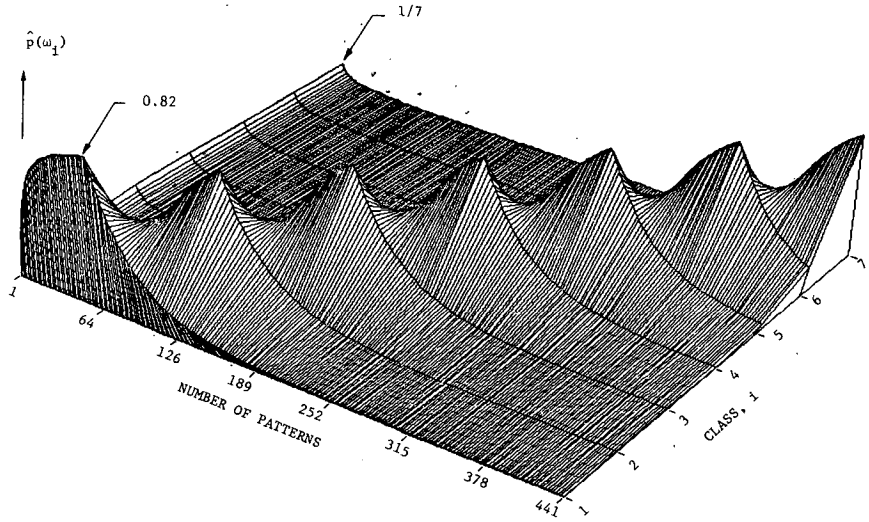

FIG. 3. Estimates of *a priori* probabilities by discrete Kalman filtering.

The values are $p_0 = 0.00172$, $q = 0.0000068$ and $r = 0.0096$.

To test the algorithm the worst cases were selected:

1) This algorithm, together with these parameters, is applied to a synthetic image set (Rahman *et al.*, 1981) with 63 images in each class. In other words, the first 63 images belong to class one and next 63 images belong to class two and so on. The estimated *a priori* probabilities are shown in Fig. 3. For better graphical display estimates of the alternate patterns, that is, first, third, fifth, etc., are shown in the figure.

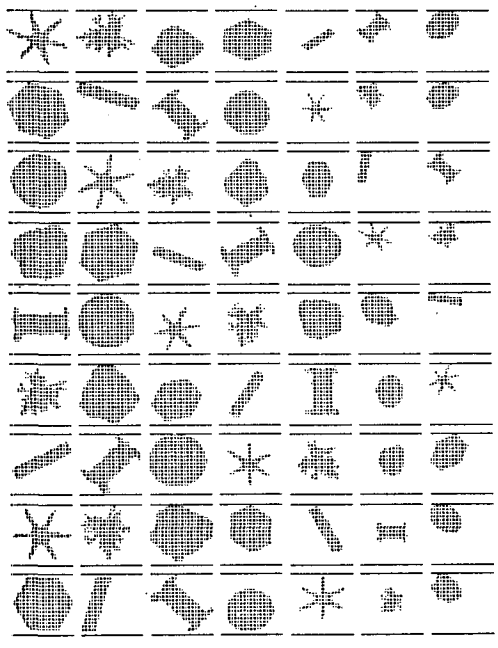

FIG. 4. A set of synthetic images.

2) The same algorithm is then applied to another set of synthetic images shown in Fig. 4 while the estimated probabilities are shown in Fig. 5. Fig. 3 shows that when the estimator receives patterns one after another from the same class, the *a priori* probability of that class increases while the others decrease while Fig. 5 shows that when the estimator receives patterns, one from a class and the next from another class, the estimates oscillate around 1/7.4. Both cases show a good indication of the performance of the discrete Kalman filter as applied to this type of classification.

Instead of using a constant value of r throughout the classification, it is possible to use an adaptive Kalman filtering procedure as developed above to update value of r during classification. Substituting Eqs. (21), (33) and (34) into (40) yields

$$\hat{r}_k| = N^{-1} \sum_{i=k-N+1}^{k} \boldsymbol{\nu}_i \boldsymbol{\nu}_i^{\mathrm{T}} - p_k(-)|. \qquad (45)$$

This equation with $N = 15$ is used to update r because it was felt that this value of N would provide sufficient averaging without requiring an unnecessary lag. Better results are obtained when this adaptive algorithm is applied to those two sets of data. Fig. 6 shows the comparison of $\hat{p}(\omega_4)$ obtained by these two algorithms when applied to first set of data. This figure is a clear indication that the adaptive procedure is much faster. A similar result is obtained for the second set of data but in this case the estimates oscillate around 1/7.

4. Classification

a. Incomplete patterns

All the images considered so far are completely recorded; in other words, the hydrometeors are

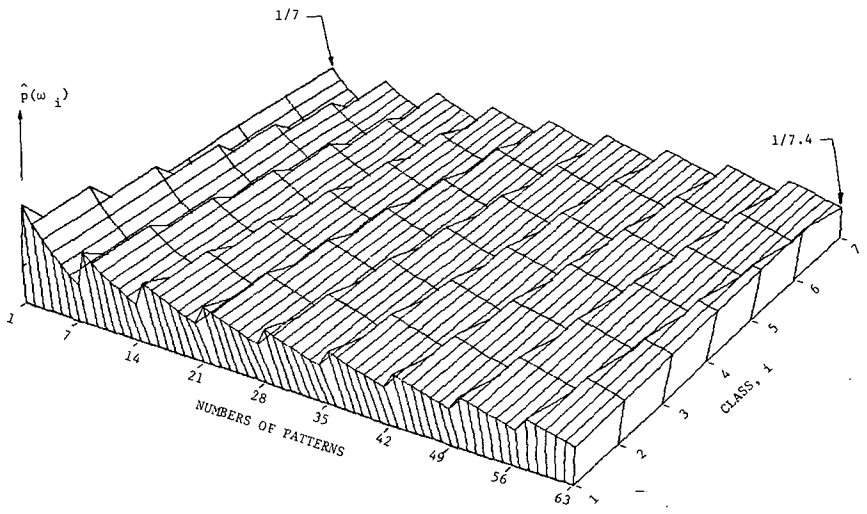

FIG. 5. Estimate of *a priori* probabilities by discrete Kalman filtering.

completely within the sampling area of the probe. The hydrometeors which are only partially within the sampling area of the probe will produce fractional images which are in this investigation defined as incomplete patterns. After examining thousands of actual patterns, it was concluded that incomplete patterns comprise a significant fraction of the total population sampled. Eliminating these incomplete patterns from the population significantly reduces the efficiency of the probe, but including those with the complete patterns results in serious deterioration in the ability to classify. Fig. 2 shows a typical set of actual data recorded by the probe. It is evident from these data that the number of incomplete patterns is significant compared to the total population. Two different methods are used to solve this problem. The first method is to separate the incomplete patterns from the population and to classify them separately while the second method is to classify the population of complete and incomplete patterns directly, that is, without separation.

To test these methods, two new synthetic training sets were generated. The first set was composed of only incomplete patterns and the second set was composed of both complete and incomplete patterns. Figs. 7 and 8 show the synthetic training data for only class two of first and second set, respectively. One thousand incomplete synthetic images of known classification are then classified as discussed in Section 4. Two such classifications are made, one using the class statistics of the first synthetic training set and the other one using the class statistics of the second synthetic training set and the results are given in Table 1. It is obvious from the table that the second synthetic training set gives better results.

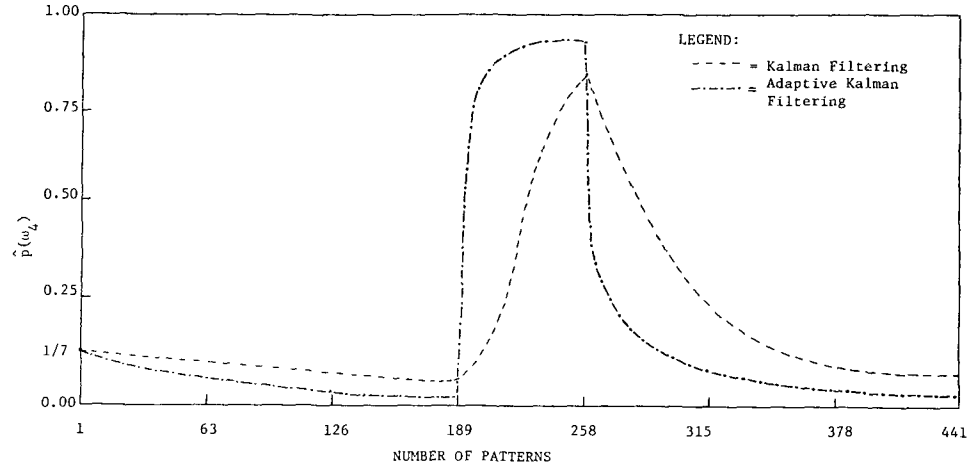

FIG. 6. Comparison between adaptive Kalman filtering and Kalman filtering.

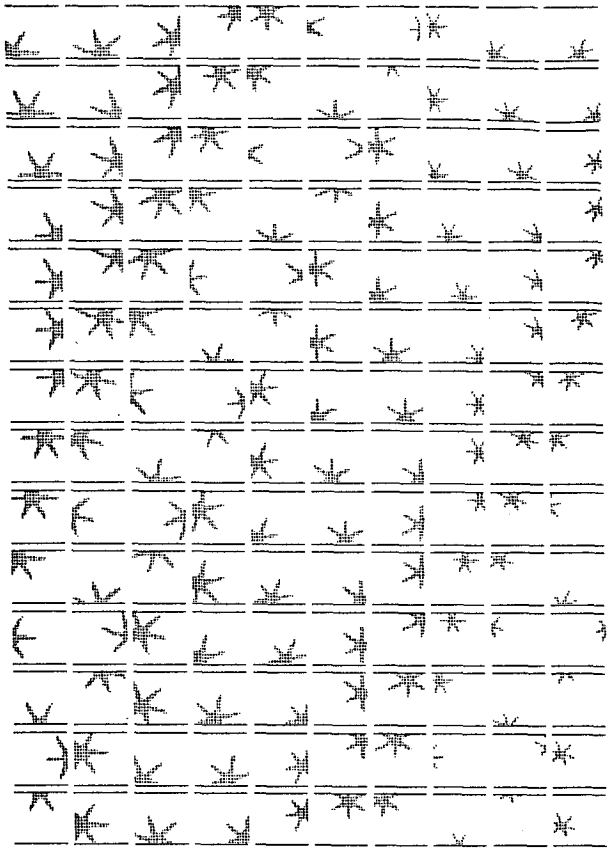

FIG. 7. Synthetic training set of incomplete patterns (class two).

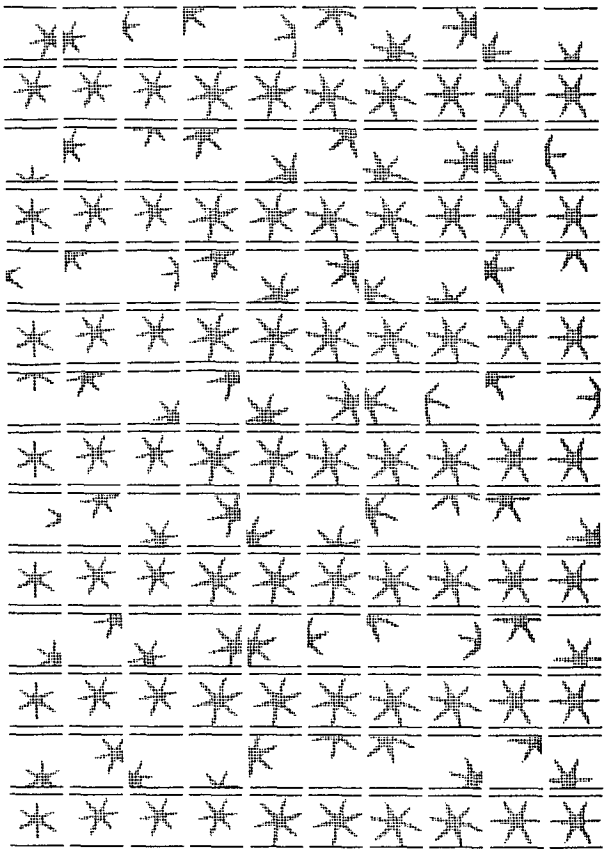

FIG. 8. Synthetic image set of both complete and incomplete patterns (class two).

Since the complete patterns can be classified with almost no error, it may be worth trying to reconstruct a complete pattern from the available incomplete pattern. Due to the wide variability in incomplete pattern, it takes several minutes of CPU time just to reconstruct one incomplete pattern. Since the computer time for this recognition scheme was prohibitive, reconstruction of incomplete patterns was not investigated in detail. The template matching method (Rahman, 1980; Rahman *et al.*, 1981) can be successfully applied to classify the incomplete patterns, but for the same reason this method also was rejected for classification of incomplete patterns.

b. Unclassifiable pattern detection

In this investigation, classification of patterns into the seven different most common classes is considered but the actual data belong to more than these seven classes. It is apparent that there are some patterns in Fig. 2 which do not belong to any of the seven classes already defined (Rahman *et al.*, 1981). In this research effort, these types of patterns are defined as unclassifiable patterns. The Bayes classifier assigns a pattern to the class with largest probability. So there will be no reject class in the classification scheme.

Two methods are used to detect the unclassifiable patterns. After the Bayes classifier assigns a pattern $\mathbf{X}_j$ to the ith class ω_i, either of the following two metrics are calculated and compared to a threshold value to detect the unclassifiable patterns. The metrics are

$$\mathrm{MHD}_{ij} = [\mathbf{X}_j - \mathbf{m}_i]^{\mathrm{T}} \mathbf{C}_i^{-1} [\mathbf{X}_j - \mathbf{m}_i] \qquad (46)$$

and

$$\mathrm{PRB}_{ij} = \frac{p(\mathbf{X}|\omega_i)|\mathbf{X} = \mathbf{X}_j}{p(\mathbf{X}|\omega_i)|\mathbf{X} = \mathbf{m}_i}, \qquad (47)$$

where $\mathbf{m}_i$ and $\mathbf{C}_i$ are the mean and covariance of the ith class, probability density function $p(\mathbf{X}|\omega_i)$. The metric MHD_{ij} also is known as Mahalanobis distance.

TABLE 1. Comparison of results of classification using two synthetic training sets.

Synthetic training set	Percent error of classification
Only incomplete patterns	28.71
Both complete and incomplete patterns	7.32

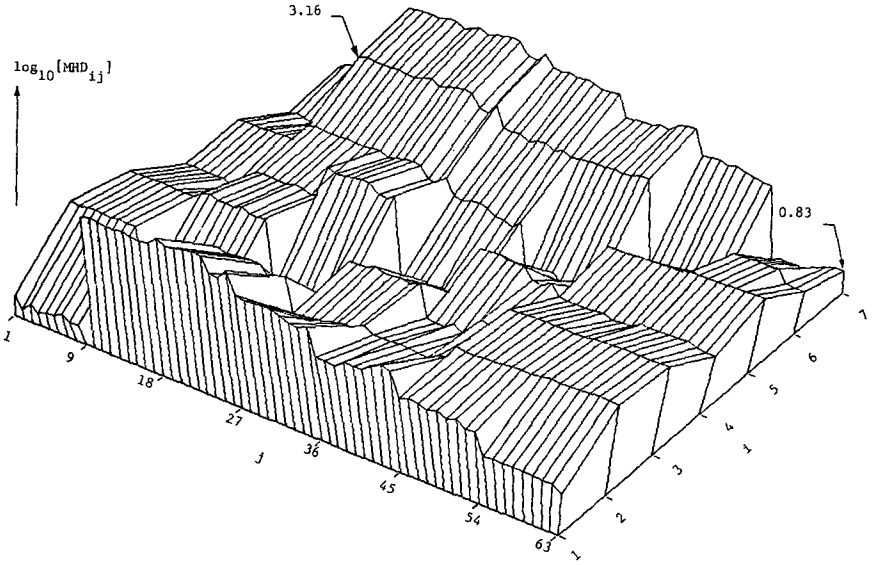

FIG. 9. Plot of the metric MHD_{ij}.

The threshold values depend on the type of patterns to be classified and are calculated from the training patterns. As an example, the threshold values of the synthetic image set (Rahman *et al.*, 1981) are calculated here. The metric MHD_{ij} for each of the 63 images for every class, i.e., MHD_{ij}, $i = 1, 2, \ldots, 7$ and $j = 1, 2, \ldots, 63$ is calculated and plotted in the Fig. 9. This figure clearly shows the metric MHD_{ij} has smaller values for the patterns which belongs to the ith class only. The threshold vector THVM calculated for seven classes from these data is

$$\text{THVM} = [thvm_1, thvm_2, thvm_3, thvm_4, thvm_5, thvm_6, thvm_7]^T = [10, 60, 20, 50, 70, 15, 100]^T. \quad (48)$$

A pattern X_j is detected as unclassifiable if

$$MHD_{ij} > thvm_i. \quad (49)$$

Similar calculations are made for the metric PRB_{ij} and plotted in Fig. 10. The minimum value of PRB_{ij} calculated is lower than 0.999×10^{-50} and for better visual display and $PRB_{ij} < 0.100 \times 10^{-3}$ is set to zero. It is apparent from the figure that this metric PRB_{ij} is higher for the patterns which belong to the ith class only. The threshold vector THVP for this metric is

$$\text{THVP} = [thvp_1, thvp_2, thvp_3, thvp_4, thvp_5, thvp_6, thvp_7]^T = [0.001, 0.001, 0.001, 0.001, 0.0001, 0.0001, 0.0001]^T. \quad (50)$$

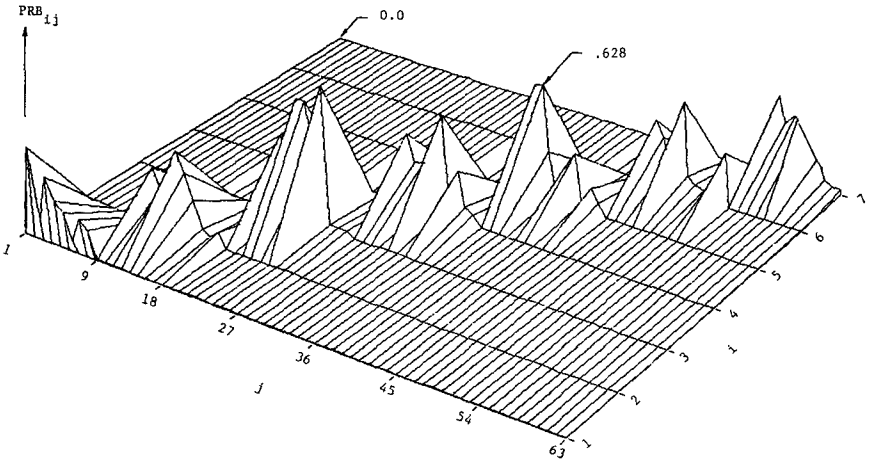

FIG. 10. Plot of the metric PRB_{ij}.

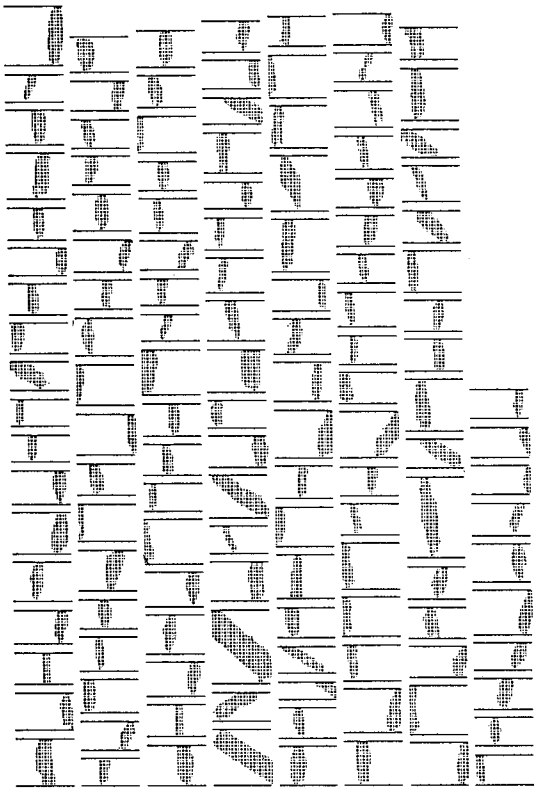

FIG. 11. Training patterns for class three.

A pattern $\mathbf{X}_j$ is detected as unclassifiable if

$$\mathrm{PRB}_{ij} < thvp_i. \tag{51}$$

c. Results

To classify actual data, training patterns from the actual data are needed for every class. The number of training patterns required, NUM, is given by (Tou and Gonzalez, 1974)

$$\mathrm{NUM} \geq 20(n + 1), \tag{52}$$

where n is the dimension of the pattern vector. Since four features were selected for classification from (52) the required number NUM $\geq$ 100. In this investigation, 500 training patterns for each class were selected. Some of the selected training patterns of class three and seven are shown in Figs. 11 and 12, respectively. The values of the four selected features for every class are calculated and the class statistics (sample mean vectors and sample covariance matrices) are determined. The class separation plots (Rahman, 1980; Rahman *et al.*, 1981) for the selected features MEU1, MEU2, MEU3 and *TR* are given in Figs. 13, 14, 15 and 16, respectively. It is apparent from these figures that the degree of class separation is less and this is mainly due to the presence of some incomplete patterns.

The Bayes classifier with the class statistics of the training patterns is then applied to classify the random patterns shown in Fig. 2. The results of classification together with the patterns are shown in Fig. 17. The numerals indicate the class to which the patterns are assigned. The letters *M* and *U* indicate misclassified patterns and unclassifiable patterns, respectively. The error of classification is 9.65%. The decision of true classification is made by the group discussion between authors and experienced observers from the Department of Atmospheric Science, University of Wyoming. Depending on cloud type the number of unclassifiable patterns will vary. The unclassifiable patterns are not included in classification error calculation. The classification error will change depending on cloud type and will increase significantly if the number of incomplete patterns is higher. For better error analysis a set of ten thousand random preprocessed actual images were selected. When the Bayes classifier with the class statistics of the training patterns is applied, images are classified at the rate 2120 min^{-1} of CPU time with 15.14% error of classification. The Bayes classifier with mean vectors and covariance matrices updated by (1) and (2), respectively, is then used to classify the image set. The sample mean vectors and sample covariance matrices of the training patterns are used as the initial estimate

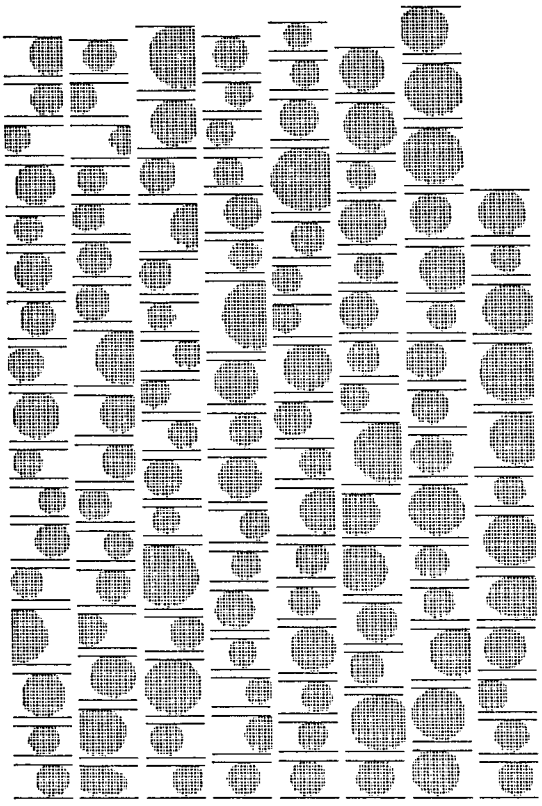

FIG. 12. Training patterns for class seven.

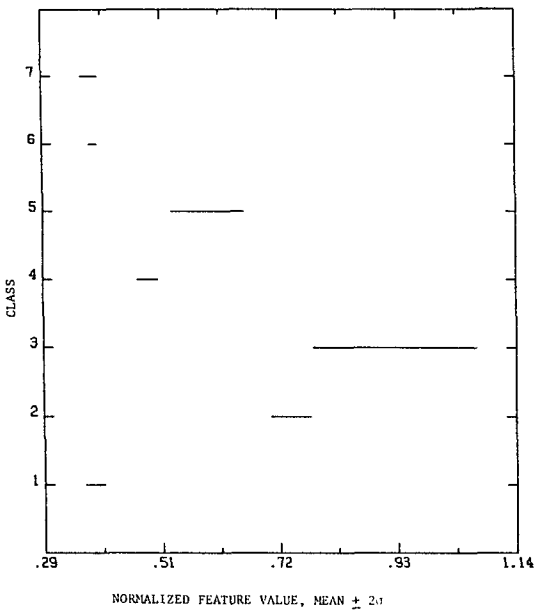

FIG. 13. Class separation plot of feature MEU1.

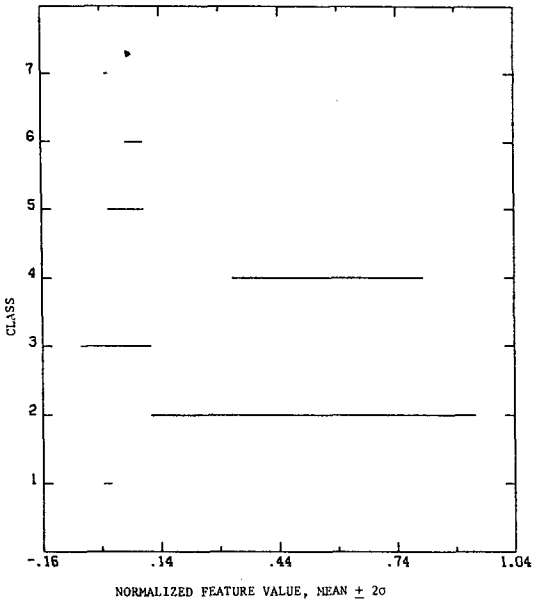

FIG. 15. Class separation plot of feature MEU3.

for these two equations. About 1511 images are classified per minute with 12.12% error. The reduction in error is insignificant compared to the increase in CPU time taken because the class statistics of the training patterns are good estimates and no significant improvement in the error results. The Bayes classifier with on-line estimates of *a priori* probabilities by the discrete Kalman filtering algorithm as discussed in Section 3 is then applied to classify the same image set. The images are classified at the rate 2009 min^{-1} with 11.03% of error which is a significant reduction without sacrificing CPU time. Then the discrete adaptive Kalman filtering algorithm developed in Section 3 is used for on-line estimates of *a priori* probabilities and the Bayes classifier classified images at the rate of 1901 per minute with 9.82% error. By this method, minimum percent error is obtained without excessive additional CPU time. The results of the above classifications are summarized in Table 2. The error is mainly

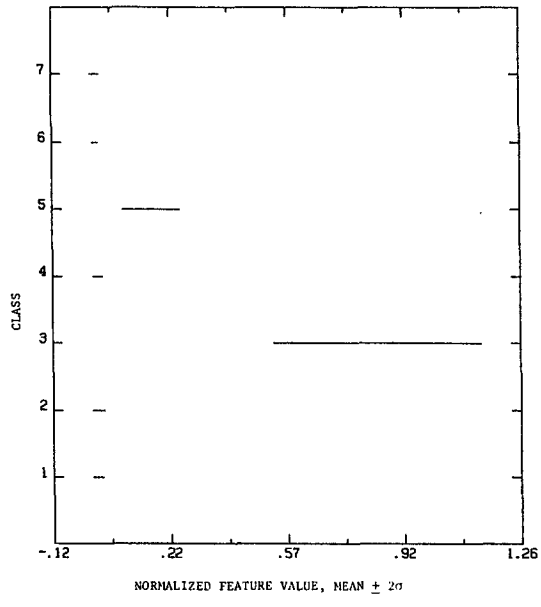

FIG. 14. Class separation plot of feature MEU2.

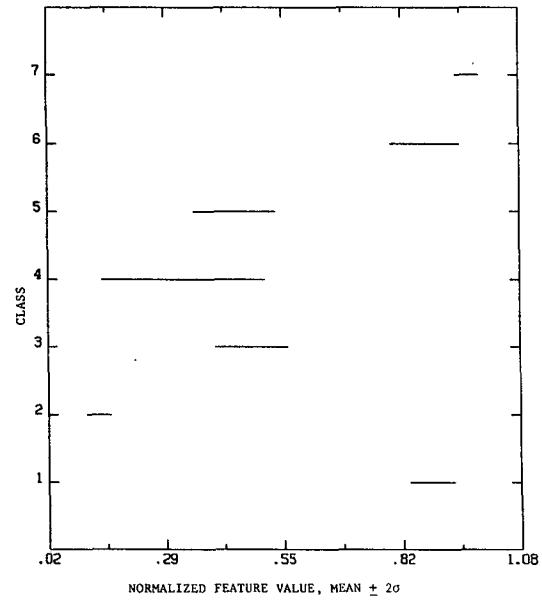

FIG. 16. Class separation plot of feature TR.

contributed by incomplete patterns. The complete and incomplete patterns which belong to class two can be classified with almost no error and, on the other hand, it is very difficult to classify the incomplete patterns belonging to classes one, six and seven.

5. Conclusions

In this research effort the main object was to classify actual images of hydrometeors with minimum error in classification using as little computing time as possible and this goal has been achieved successfully.

The description of the data acquisition system was given in brief. The feature selection made from synthetic data (Rahman, 1980; Rahman *et al.*, 1981) was also used for this investigation. Several problems related to the application were discussed. The Kalman filtering algorithm and adaptive Kalman filtering algorithm were modified for on-line estimation of *a priori* probabilities needed for the Bayes classifier. To classify incomplete patterns, the training patterns composed of both complete and incomplete images were found more suitable. To detect the unclassifiable patterns two matrices were successfully developed. Detailed error analysis of different methods of classification was made and the Bayes classifier with on-line estimation of *a priori* probabilities by the adaptive Kalman filter was found to be most efficient.

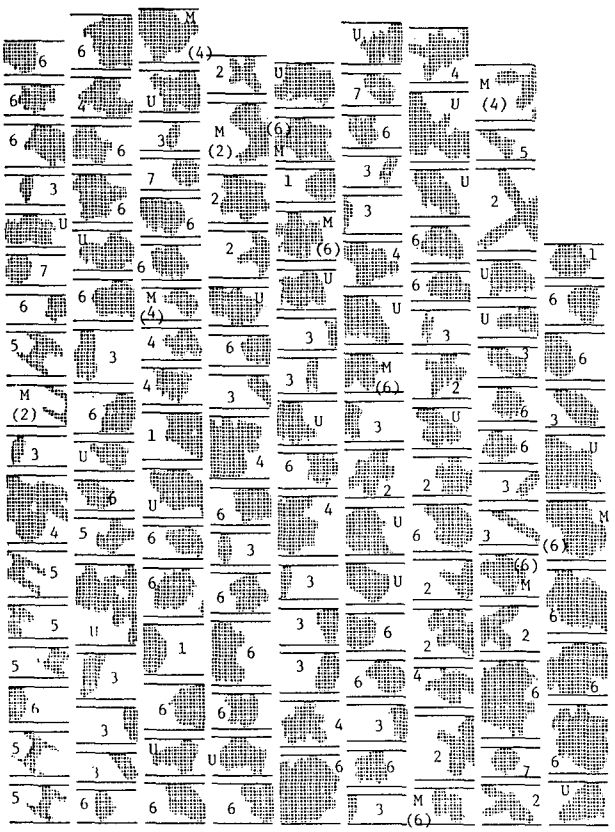

FIG. 17. Classification of random pattern.

TABLE 2. Results of classification of actual data.

Method of classification	Number of patterns classified per min of CPU time (CDC CYBER 173/730 computer)	Percent error of classification
Bayes classifier	2120	15.14
Bayes classifier with mean vector and covariance matrix update	1511	12.12
Bayes classifier with discrete Kalman filtering algorithm	2009	11.03
Bayes classifier with discrete adaptive Kalman filtering algorithm	1901	9.82

Acknowledgments. The authors are indebted to Dr. M. J. Magee with the Department of Computer Science and Dr. W. A. Cooper with the Department of Atmospheric Science, University of Wyoming, for their valuable suggestions, technical contributions and interest shown during this work.

REFERENCES

Fukunaga, K., 1972: *Introduction to Statistical Pattern Recognition.* Academic Press, 369 pp.

Gelb, A. Ed., 1974: *Applied Optimal Estimation,* The MIT Press, 374 pp. (see pp. 102–132).

Kailath, T., 1968: An innovations approach to least squares estimation—Part I: Linear filtering in additive white noise. *IEEE Trans. Auto. Control.*, **AC-6,** 646–655.

Keenh, D. G., 1965: A note on learning for Gaussian properties. *IEEE Trans. Inform. Theory*, **IT-1,** 126–132.

Mehra, R. K., 1972: Approaches to adaptive filtering. *IEEE Trans. Auto. Cont.*, **5,** 693–698.

Rahman, M. M., 1980: Statistical classification of ice particles and raindrops by pattern recognition techniques. Ph.D. dissertation, University of Wyoming, Laramie 192 pp. (see pp. 26–150).

——, E. A. Quincy, R. G. Jacquot and M. J. Magee, 1981: Feature extraction and selection for pattern recognition of two-dimensional hydrometeor images. *J. Appl. Meteor.*, **20,** 521–535.

Tou, J. T., and Gonzalez, R. C., 1974: *Pattern Recognition Principles.* Addison-Wesley, 377 pp. (see pp. 110–140).