Pergamon

Pattern Recognition, Vol. 29, No. 3, pp. 519–531, 1996
Elsevier Science Ltd

0031-3203/96 $15.00+.00

0031-3203(95)00103-3

A FAST TWO-CLASS CLASSIFIER FOR 2D DATA USING COMPLEX-MOMENT-PRESERVING PRINCIPLE

SOO-CHANG PEI and CHING-MIN CHENG
Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, Republic of China

(*Received* 9 *August* 1994; *in revised form* 20 *June* 1995; *received for publication* 7 *July* 1995)

Abstract—A new moment-preserving classifier for two-class clustering is suggested. Based on preserving the complex moments of two-dimensional (2D) input data, an analytic, non-iterative and unsupervised classifier is proposed. This new classifier is suitable for applications requiring fast automatic two-class clustering of 2D data or fast automatic hierarchical clustering. Furthermore, the computation time is of order of data size and hence much faster than the well known iterative k-means algorithm. Experimental results show that the proposed classifier can acquire acceptable clustering results.

Clustering　Patterns　Complex moments　Moment-preserving principle
Two-class classifier

1. INTRODUCTION

Clustering, the unsupervised process of finding homogeneous groups in data,[1] plays an essential role in many fields, including image processing, pattern recognition, artificial intelligence, geosciences, medical research, economics, etc. Specifically, two-class clustering problems are frequently encountered in real applications of the above fields, for example, block truncation coding for color image compression, cell tissue analysis and forecast of economic behaviours, etc. The optimum approach to this problem is the Bayes classifier. The job of the Bayes classifier is to find an optimum decision boundary which minimizes the average risk or cost. To implement this parametric classifier, *a priori* probability distributions of classes have to be known beforehand. However, the estimation of class statistical parameters is not an easy task. Non-parametric clustering methods have thus been designed to overcome this difficulty. Most of non-parametric clustering methods developed so far, such as the k-means method,[2] the ISODATA method,[3] the fuzzy c-means method,[4] and the agglomerative hierarchical method using a dissimilarity matrix,[5] etc., are iterative and thus unsuitable for performing fast automatic two-class clustering. It is therefore desirable to develop a fast automatic non-parametric clustering method which can be employed to partition an input set Q of N patterns into two classes.

One efficient method of achieving this goal is based on the moment-preserving principle proposed by Delp and Mitchell.[6,7] It avoided iterative computation by using mathematical formulae to express the decision boundary, which separates the two classes in terms of the input patterns directly. When the N input patterns are one-dimensional (1D), say, forming a set $Q = \{q(i)\}_{i=1}^{N}$, the partition of Q into two disjoint clusters Q_A and Q_B has been described in reference (7) by the moment-preserving principle. Through the derivation of reference (7), two disjoint clusters Q_A and Q_B are obtained such that the first three moments, $\mathbf{E}[q]$, $\mathbf{E}[q^2]$, $\mathbf{E}[q^3]$:

$$\mathbf{E}[q^k] = \frac{1}{N}\sum_{i=1}^{N} q(i)^k, \quad \text{for } k = 1, 2 \text{ and } 3, \qquad (1)$$

are preserved, where $\mathbf{E}[]$ represents the expectation. The good performance of this clustering method has been reported when it is applied to monochrome image segmentation.[8] However, when the input patterns are two-dimensional (2D), $Q = \{[q_0(i), q_1(i)]\}_{i=1}^{N}$, Lin and Tsai[9] have shown that preserving five joint-moments of $[q_0(i), q_1(i)]$:

$$\mathbf{E}[q_0^k \cdot q_1^l] = \frac{1}{N}\sum_{i=1}^{N} q_0(i)^k \cdot q_1(i)^l, \quad \text{for } k + l = 1 \text{ and } 2, \qquad (2)$$

could not generate two separated classes. Instead of preserving the joint-moments,[9] some other features $\{\mathbf{E}[q_0], \mathbf{E}[q_1], \mathbf{E}[r], \mathbf{E}[\theta], \phi]\}$ were preserved during the process of two-class clustering. Among these features, r and θ represent the polar coordinates of the point (q_0, q_1) and ϕ is the directional angle of the principle axis. When there are well-elongated shapes in the data, Lin and Tsai have reported that their algorithm may fail if two clusters are too close to each other.

The divisive approach is another approach to reduce the computational complexity of clustering problems and produces an acceptable solution in the meantime. It partitions the input pattern space sequentially into K disjoint sub-regions. Concerning 2D patterns, the partition line is assumed to be perpendicular to one of the coordinate axes of pattern space. One implementation of this approach is Heckbert's median-cut algorithm.[10] The partitioned axis of reference (10) is chosen to be the axis with the largest variance. Also the partition line passes through the median point of the projected pattern distribution along the partitioned axis. The centroids of the separated classes are then chosen to be the class representatives. One drawback of a med-cut algorithm is that it can only partition the pattern space in the direction perpendicular to that of the coordinate axis.

In this paper we propose a complex-moment-preserving (CMP) clustering algorithm to partition the input 2D patterns into two classes. Instead of using the joint-moments as designated by equation (2), we express the input 2D pattern space as a complex-valued space and define complex moments of input patterns. Through this definition, the moment-preserving principle for 1D data clustering can be extended to 2D data. An analytical solution is also obtained. The computation time is of the order N, the data size, and hence much faster than any interactive two-class clustering algorithms.

This paper is organized as follows: Section 2 first describes complex moments. Section 3 then presents the binary CMP clustering algorithm for two-class classifying problems. In Section 4 we compare the performance of the proposed classifier with some two-class classifiers. The extension of the binary CMP clustering algorithm to multi-class clustering problems is also discussed. Finally, a conclusion is made in Section 5.

2. COMPLEX MOMENTS

To generalize the clustering method of the moment-preserving principle from 1D data to 2D data, we adopt the complex moments[11] to explicitly express the statistical parameters of 2D data.

For the set of input patterns $Q = \{[q_0(i), q_1(i)]\}_{i=1}^{N}$, we first designate a 2D pattern $[q_0(i), q_1(i)]$ as the complex number $\tilde{q}(i)$:

$$\tilde{q}(i) = q_0(i) + jq_1(i), \tag{3}$$

with j being denoted as the imaginary unit and $j^2 = -1$. Based on the above expression of 2D patterns, we extend the first three 1D real moments as defined by equation (1) to the complex moments as follows:

$$\begin{aligned} \tilde{m}_1 &= \mathbf{E}[\tilde{q}] \\ \tilde{m}_2 &= \mathbf{E}[\tilde{q}\tilde{q}^*] \\ \tilde{m}_3 &= \mathbf{E}[\tilde{q}\tilde{q}^*\tilde{q}] \end{aligned} \tag{4}$$

where $\tilde{q}^*$ is the complex conjugate of $\tilde{q}$ and the definition of the third-order complex moment $\tilde{m}_3$ is adopted from high-order statistics.[12] Equation (4) can be reduced to:

$$\begin{aligned} \tilde{m}_1 &= \mathbf{E}[q_0] + j\mathbf{E}[q_1] \\ \tilde{m}_2 &= \mathbf{E}[q_0^2 + q_1^2] \\ \tilde{m}_3 &= \mathbf{E}[q_0^3 + q_1^2 q_0] + j\mathbf{E}[q_1^3 + q_0^2 q_1], \end{aligned} \tag{5}$$

that is, the complex moments can be seen as 2D vector moments. The complex-valued $\tilde{m}_1$ represents the centroid of the input pattern $[q_0(i), q_1(i)]$. The real-valued $\tilde{m}_2$ expresses expected value of the vector length of $[q_0(i), q_1(i)]$. Also, the complex-valued $\tilde{m}_3$ consists of the sum of joint third-order moments among $[q_0(i), q_1(i)]$. Since the principle of moment-preserving is to preserve the statistical distribution of input patterns, it is usually necessary to maintain at least the location of distribution and to hold its dispersion in some degrees. In applied statistics[13] the information about the dispersion of input patterns can be described by the central moments. That is, input patterns will be shifted to its centroid first and then calculated from the associated $\tilde{m}_k$ defined by equation (5), for $k = 1, 2, 3$. The complex moments in equation (5) become:

$$\begin{aligned} \tilde{m}_1 &= 0 + j0 \\ \tilde{m}_2 &= \mathbf{E}[(q_0 - \mathbf{E}[q_0])^2 + (q_1 - \mathbf{E}[q_1])^2] \\ \tilde{m}_3 &= \mathbf{E}[(q_0 - \mathbf{E}[q_0])^3 + (q_1 - \mathbf{E}[q_1])^2(q_0 - \mathbf{E}[q_0])] \\ &\quad + j\mathbf{E}[(q_1 - \mathbf{E}[q_1])^3 + (q_0 - \mathbf{E}[q_0])^2(q_1 - \mathbf{E}[q_1])]. \end{aligned} \tag{6}$$

We will implement the proposed clustering method in the next section by preserving the input central complex moments. The complex moments mentioned later indicate central complex moments. Concerning 1D input patterns, it can be proved that the preservation of central real moments is equivalent to the preservation of non-central real moments. Nevertheless, this relationship of equivalence will not hold when considering the definitions of complex moments [equations (5) and (6)].

3. BINARY COMPLEX-MOMENT-PRESERVING CLUSTERING

The problem of performing the binary CMP clustering on a 2D data set Q is to select a line as a threshold to separate Q into a two-level data set G, such that the first three complex moments of Q are preserved. The resultant two levels in G are defined as $\tilde{z}_0$ and $\tilde{z}_1$. After the binary CMP clustering, all below-threshold patterns in Q are replaced by the complex-valued $\tilde{z}_0$ and those above-threshold patterns in Q are set to be the complex-valued $\tilde{z}_1$.

Let p_0 and p_1 denote the probability of the below-threshold and the above-threshold patterns in Q after the binary CMP clustering. Therefore:

$$p_0 + p_1 = 1. \tag{7}$$

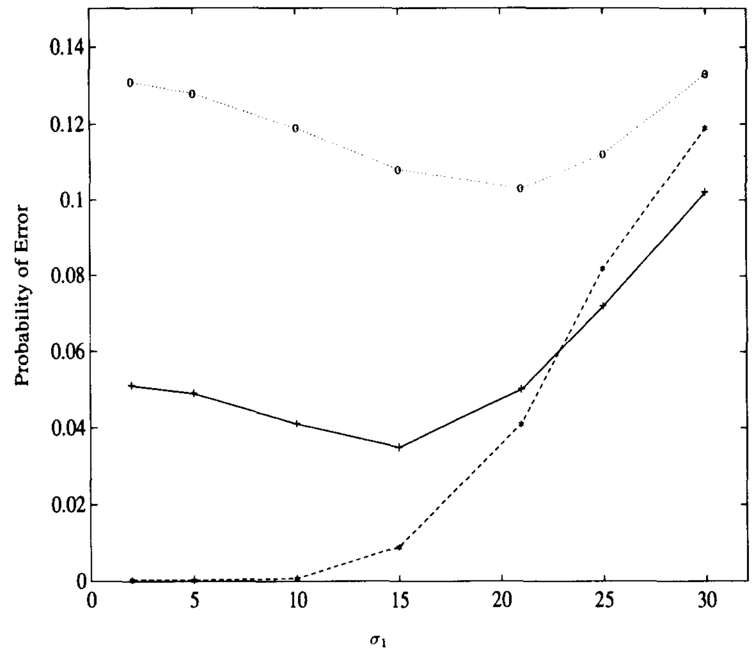

Fig. 1. The error probability distributions of the CMP classifier vs the variances, σ_0 and σ_1, when ρ, μ_0 and μ_1 are fixed to be 0.1, 100 and 150, respectively. The "*", "+" and "o" curves correspond to $\sigma_0 = 11$, $\sigma_0 = 20$ and $\sigma_0 = 31$, respectively.

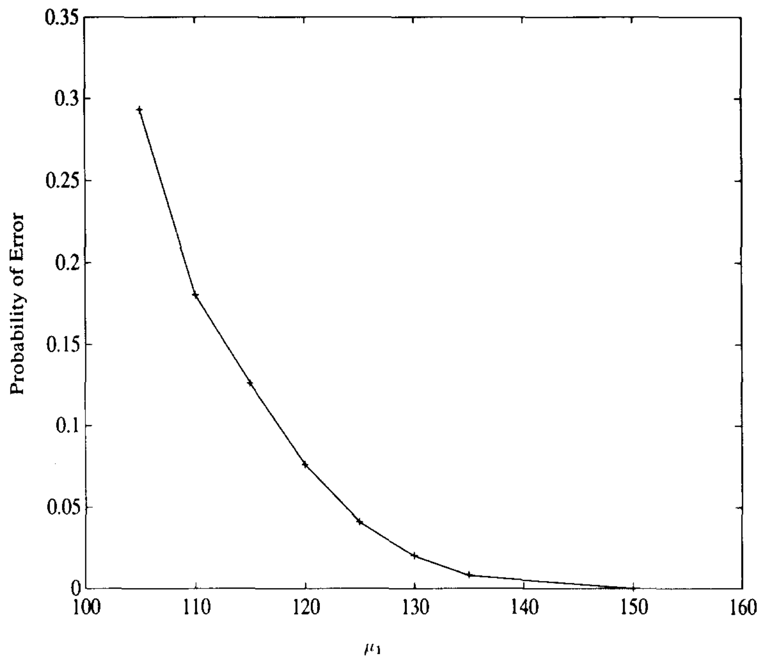

Fig. 2. The error probability distribution of the CMP classifier vs the mean values, μ_0 and μ_1, when ρ, σ_0 and σ_1 are fixed to be 0.1, 5 and 10, respectively. The "+" curve corresponds to $\mu_0 = 100$.

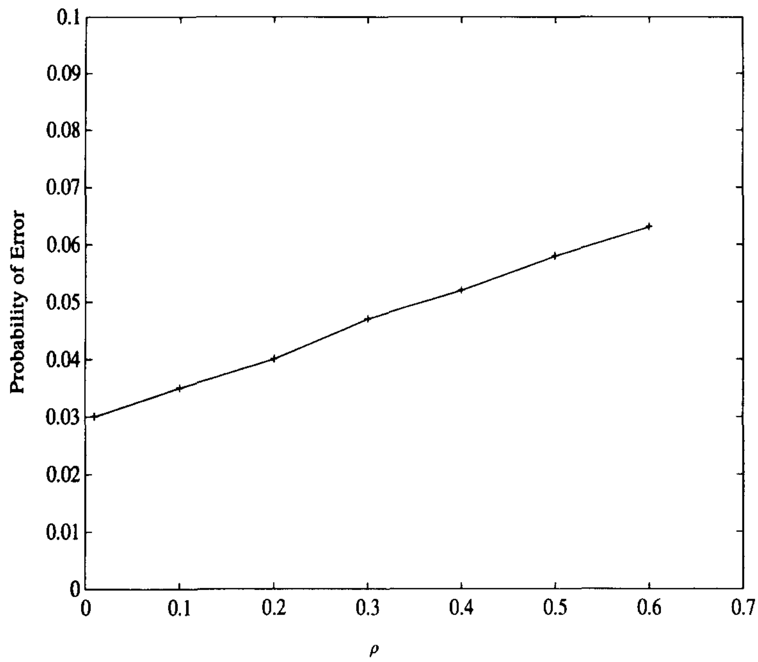

Fig. 3. The error probability distribution of the CMP classifier vs the correlation coefficient, ρ, when μ_0, μ_1, σ_0 and σ_1 are fixed to be 100, 150, 20 and 15, respectively.

Then the first three complex moments of the two-level data set G are given by:

$$\tilde{m}'_1 = \sum_{k=0}^{1} p_k \tilde{z}_k$$

$$\tilde{m}'_2 = \sum_{k=0}^{1} p_k \tilde{z}_k \tilde{z}_k^* \quad (8)$$

$$\tilde{m}'_3 = \sum_{k=0}^{1} p_k \tilde{z}_k \tilde{z}_k^* \tilde{z}_k .$$

If we let the first three complex moments of the two-level data set G equal those of Q, $\tilde{m}_k = \tilde{m}'_k$, $k = 1, 2, 3$, we obtain the following moment-preserving equations:

$$\begin{aligned} p_0 \tilde{z}_0 + p_1 \tilde{z}_1 &= \tilde{m}_1 \\ p_0 \tilde{z}_0 \tilde{z}_0^* + p_1 \tilde{z}_1 \tilde{z}_1^* &= \tilde{m}_2 \\ p_0 \tilde{z}_0 \tilde{z}_0^* \tilde{z}_0 + p_1 \tilde{z}_1 \tilde{z}_1^* \tilde{z}_1 &= \tilde{m}_3 . \end{aligned} \quad (9)$$

Thus, using the complex moments, moment-preserving principle can still be maintained for 2D data. The solution of reference (7) is a special case of the proposed clustering operator. The moment preserving equations (8) can be solved indirectly by adopting the following polynomial of complex number $\tilde{z}$:

$$\begin{aligned} C(\tilde{z}) &= \tilde{z}^* \tilde{z} + \tilde{z}^* \tilde{c}_1 + \tilde{c}_0 \\ &= (\tilde{z} - \tilde{z}_0)(\tilde{z} - \tilde{z}_1), \end{aligned} \quad (10)$$

where $\tilde{z}_k$ for $k = 0, 1$, the representatives of the thresholded classes by the binary CMP clustering operator are the roots of $C(\tilde{z})$, i.e. $C(\tilde{z}_k) = 0$. If we multiply $C(\tilde{z}_k)$ with equation (7) and the equation defining $\tilde{m}'_1$, respectively, we obtain:

$$\begin{cases} p_0 C(\tilde{z}_0) + p_1 C(\tilde{z}_1) = 0 \\ p_0 z_0 C(\tilde{z}_0) + p_1 z_1 C(\tilde{z}_1) = 0. \end{cases}$$

Then using equation (8) we have:

$$\begin{cases} \tilde{c}_0 + \tilde{m}_1^* \tilde{c}_1 + \tilde{m}_2 = 0 \\ \tilde{m}_1 \tilde{c}_0 + \tilde{m}_2 \tilde{c}_1 + \tilde{m}_3 = 0. \end{cases} \quad (11)$$

From equation (11), $\tilde{c}_0$ and $\tilde{c}_1$ can be expressed as follows:

$$\tilde{c}_1 = \frac{(\tilde{m}_3 - \tilde{m}_1 \tilde{m}_2)}{(\tilde{m}_1 \tilde{m}_1^* - \tilde{m}_2)} \quad (12)$$

$$\tilde{c}_0 = -(\tilde{m}_1^* \tilde{c}_1 + \tilde{m}_2). \quad (12)$$

Since $\tilde{m}_1 = 0$ in the proposed CMP clustering, $\tilde{c}_0$ and $\tilde{c}_1$ will be reduced to $-\tilde{m}_2$ and $-(\tilde{m}_3/\tilde{m}_2)$, respectively. Using $\tilde{c}_0$ and $\tilde{c}_1$, the roots of $C(\tilde{z}_k) = 0$ can be solved. For convenience, we define:

$$\begin{aligned} \tilde{z}_k &= z_{k0} + j z_{k1} \\ \tilde{c}_1 &= c_{10} + j c_{11} \\ \tilde{c}_0 &= c_{00} . \end{aligned}$$

By the above definitions, the real and imaginary parts of $C(\tilde{z}_k) = 0$ can be represented as the following equations:

$$z_{k0}^2 + z_{k1}^2 + z_{k0} c_{10} + z_{k1} c_{11} + c_{00} = 0 \quad (13)$$

$$z_{k0} c_{11} - z_{k1} c_{10} = 0 \quad (14)$$

Substituting equations (14) into (13), we have a second-order equation of number z_{k1}:

$$\left(1+\left(\frac{c_{10}}{c_{11}}\right)^2\right)z_{k1}^2+\left(\frac{c_{10}^2+c_{11}^2}{c_{11}}\right)z_{k1}+c_{00}=0. \quad (15)$$

From equation (15), the two solutions, z_{01} and z_{11}, can be obtained directly as:

$$z_{01}=\frac{-1}{2a}(ac_{11}+|c_{11}|\sqrt{a^2-4c_{00}})$$
$$z_{11}=\frac{-1}{2a}(ac_{11}-|c_{11}|\sqrt{a^2-4c_{00}}), \quad (16)$$

where $a=\sqrt{c_{10}^2+c_{11}^2}$. Since $c_{00}=-\tilde{m}_2$, it can be readily shown that the values of z_{01} and z_{11} are real. After z_{k1} is obtained, the corresponding z_{k0} can be acquired from equation (14).

3.1. *Algorithm for the binary CMP clustering*

Since we preserve the central complex moments in implementing the binary CMP clustering operator, $\tilde{z}_0$ and $\tilde{z}_1$ should be transformed back to the old coordinates by the inverse translation with $\Delta q_0=-\mathbf{E}[q_0]$ and $\Delta q_1=-\mathbf{E}[q_1]$ after $\tilde{z}_0$ and $\tilde{z}_1$ are obtained. Then we choose the line l^i perpendicular to and bisecting the line segment $\overline{\tilde{z}_0\tilde{z}_1}$ as the decision boundary to separate the two classes. This segmentation of two classes is equivalent to the nearest-neighbour clustering of two class representatives, $\tilde{z}_0$ and $\tilde{z}_1$. We notice that the computational complexity of the binary CMP cluster-

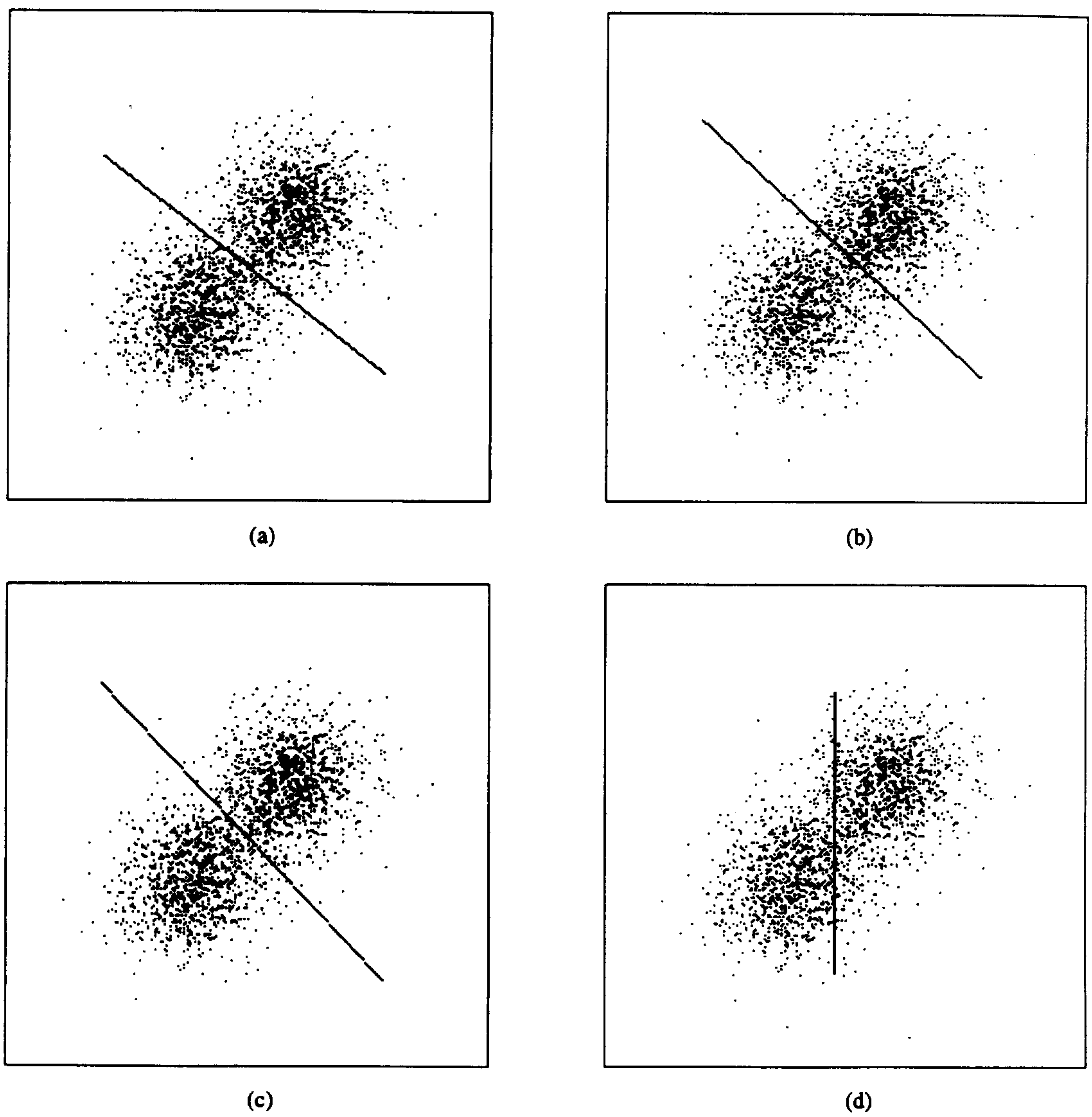

Fig. 4. Clustering results of the generated data listed in Table 1: (a) the proposed CMP classifier; (b) Lin's classifier; (c) the Bayes classifier; (d) the MC classifier.

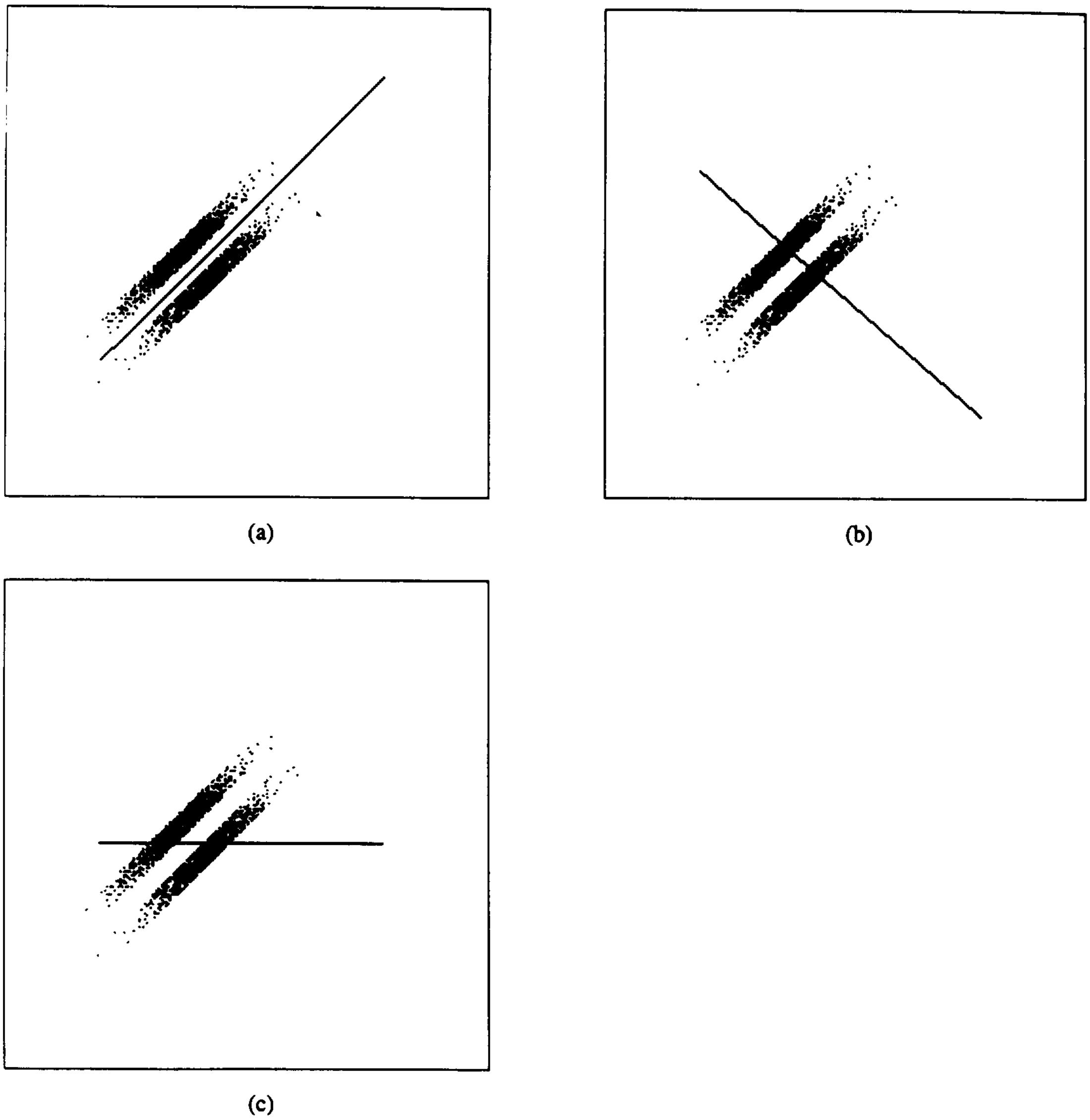

Fig. 5. Clustering results of the generated data listed in Table 2: (a) the proposed CMP classifier; (b) Lin's classifier; (c) the MC classifier.

ing operator is mainly dominated by the calculation of complex moments $\tilde{m}_k$, for $k = 1, 2, 3$, which is of the order N, the data size.

To classify a set of 2D data points using the CMP principle, we summarize the proposed clustering algorithm as follows:

(a) The values of 2D data points are represented by complex numbers. Then the moment preserving equations (8) are solved to obtain $\tilde{z}_0$ and $\tilde{z}_1$.

(b) The line l' perpendicular to and bisecting the line segment $\overline{\tilde{z}_0\tilde{z}_1}$ is chosen to separate the two classes of data points.

(c) A two-class bit-map, which identifies the membership of classified data points, is constructed such that each data point is coded as "one" or "zero", depending on whether that data point is on the right of the decision boundary or not.

As an example, suppose a 2D data set Q is provided and arranged in the following manner:

$$Q = \{(1, 11), (2, 12), (3, 13), (4, 14), (5, 15), (6, 16), (7, 17), (8, 18), (60, 30), (61, 31), (62, 32), (63, 33), (60, 34), (61, 35), (62, 36), (63, 37)\},$$

thus

$$\tilde{z}_0 = (7.4, 7.7)$$
$$\tilde{z}_1 = (58.3, 40.1),$$

and the corresponding two-class bit-map is:

$$\{0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1\}.$$

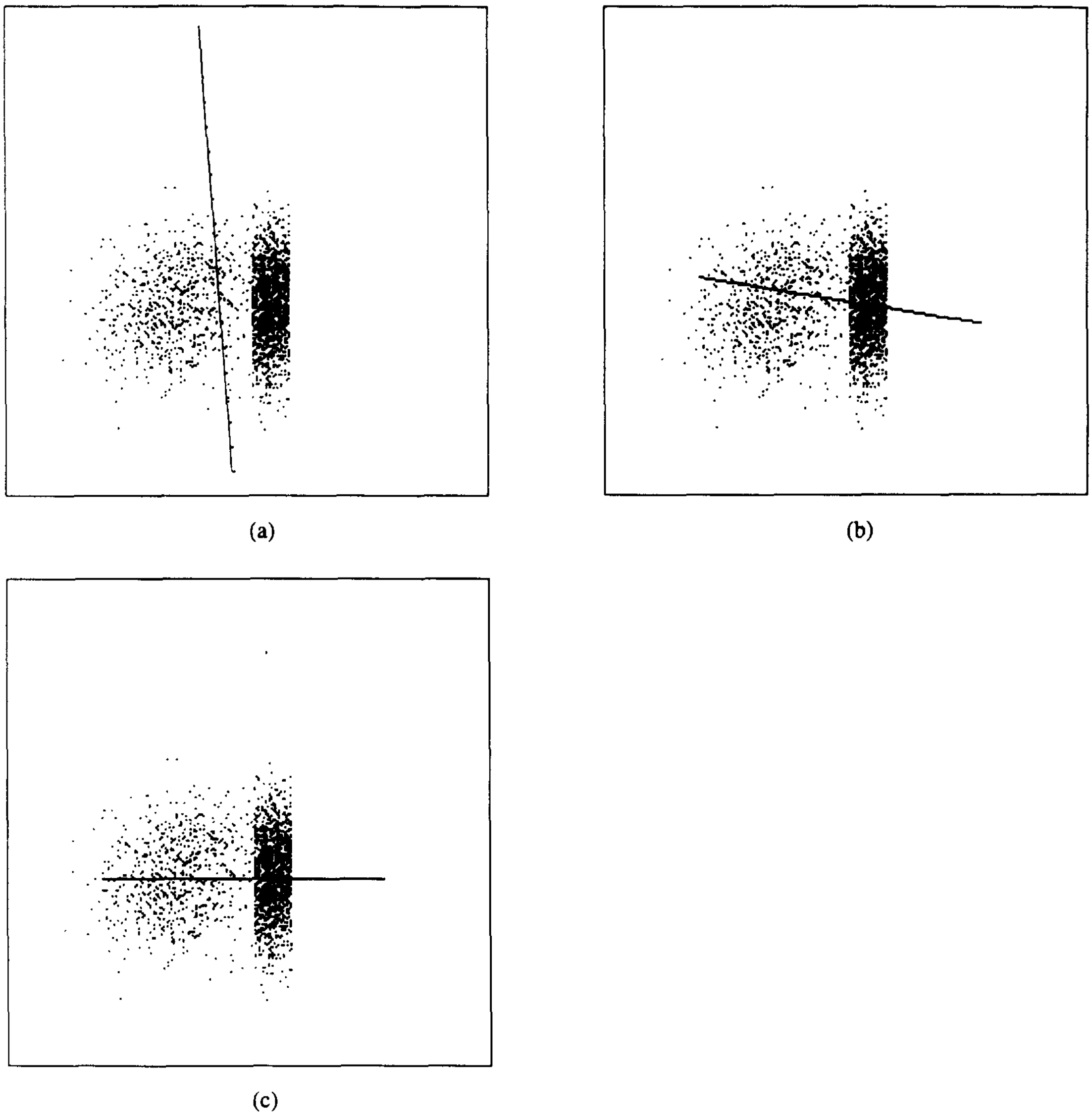

Fig. 6. Clustering results of the generated data listed in Table 3: (a) the proposed CMP classifier; (b) Lin's classifier; (c) The MC classifier.

3.2. *Statistical analysis of performance*

Since it is commonly assumed that a region of high local concentration of patterns, known as the core, is associated with each cluster, we use the normal distribution to approximate the probability distribution of each cluster. In view of the central-limit theorem,[13] this approximation seems to have a good degree of reliability. Thus, in this sub-section, we analyse the performance of the binary CMP clustering operator for the case that the input data set is composed of two equiprobable 2D normal distributions with equal correlation coefficients. For each distributions, the components of data have equal variances. The covariance matrix, $\mathbf{C}_i$, of each distribution is given by:

$$\mathbf{C}_i = \begin{pmatrix} \sigma_i^2 & \rho\sigma_i^2 \\ \rho\sigma_i^2 & \sigma_i^2 \end{pmatrix}, \quad \text{for } i = 0, 1,$$

where σ_i^2 and ρ represent the variance and correlation coefficient, respectively. The probability density functions of these normal distributions are then denoted as:

$$f_i(\mathbf{q}) = (2\pi|\mathbf{C}_i|^{1/2})^{-1} \exp\left(-\frac{1}{2}(\mathbf{q} - \mathbf{m}_i)^t \mathbf{C}_i^{-1}(\mathbf{q} - \mathbf{m}_i)\right), \quad \text{for } i = 0, 1,$$

where $\mathbf{q} = [q_0, q_1]$ is the vector representing the data point and $\mathbf{m}_i = [\mu_i, \mu_i]$, for $i = 0, 1$, are the mean vectors for each distribution. The value of $|\mathbf{C}_i|$ is determinant of the covariance matrix $\mathbf{C}_i$. According to the moment theorem for normal distribution,[11] complex moments $\tilde{m}_2$ and $\tilde{m}_3$ can be calculated as:

$$\tilde{m}_2 = \sigma_0^2 + \sigma_1^2 + \tfrac{1}{2}(\mu_0 - \mu_1)^2$$

$$\tilde{m}_3 = (\sigma_0^2 - \sigma_1^2)(\mu_0 - \mu_1)\left(1 + \frac{\rho}{2}\right) + j(\sigma_0^2 - \sigma_1^2)(\mu_0 - \mu_1)\left(1 + \frac{\rho}{2}\right). \quad (17)$$

Then, $\tilde{z}_0$ and $\tilde{z}_1$ can be obtained by using equation (16). The linear equation, which formulates the decision boundary, is thus described as:

$$q_0 = s - tq_1, \quad (18)$$

with

$$s = \frac{(z_{00}^2 + z_{01}^2 - z_{10}^2 - z_{11}^2)}{2(z_{00} - z_{10})}$$

$$t = \frac{(z_{01} - z_{11})}{(z_{00} - z_{10})}.$$

To determine the total probability of classification error, we have to obtain P_{e0}, the probability of classifying an input pattern as cluster 1 when it is not, and P_{e1}, the probability of classifying an input pattern as cluster 0 when it is not. Assuming μ_0 is smaller than μ_1, these error probabilities, P_{e0} and P_{e1}, can be expressed as:[14]

$$P_{e0} = \int_{-\infty}^{+\infty} \int_{s-tq_1}^{+\infty} f_0(q_0, q_1)\,\mathrm{d}q_0\,\mathrm{d}q_1 \quad (19)$$

$$P_{e1} = \int_{-\infty}^{+\infty} \int_{-\infty}^{s-tq_1} f_1(q_0, q_1)\,\mathrm{d}q_0\,\mathrm{d}q_1, \quad (20)$$

where $f_0(q_0, q_1)$ and $f_1(q_0, q_1)$ are the probability density functions for each cluster, respectively. Once P_{e0} and P_{e1} are obtained, the total error probability, which is estimated as the ratio of the number of misclassified data points to the total number of available data points, can be denoted as:

$$P_e = \tfrac{1}{2}(P_{e0} + P_{e1}). \quad (21)$$

In Figs 1, 2 and 3 we investigated the behaviours of the proposed clustering operator when the parameters of two input 2D normal distributions are varied. In Fig. 1, ρ, μ_0 and μ_1 are set as 0.1, 100 and 150, respectively. The variances of two distributions are changed. We observed that the minimum total error probability happened when the variances of two distributions were near. In Fig. 2, we fixed ρ, σ_0 and σ_1 to be equal to 0.1, 5 and 10, respectively. The mean vectors of two distributions then were changd to indicate the closeness of two distributions. We noticed that the farther away the two distributions, the smaller the total error probability. Finally, we displayed the relationship between ρ and total error probability in Fig. 3 by setting μ_0, μ_1, σ_0 and σ_1 to be 100, 150, 20 and 15, respectively. Figure 3 showed that when the

Table 1. The statistical parameters of the two 2D distributions for the example of Fig. 4 and performance comparison of different two-class clustering algorithms

	Generated data			Baysian classifier			Lin's classifier			MC classifier			CMP classifier		
	Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix	
Population 1 (1050)	150.23	386.4	51.4	149.99	369.8	30.3	151.30	349.7	−3.23	150.98	308.0	140.3	149.96	370.3	21.2
	150.17	51.4	404.7	150.14	30.3	378.0	151.99	−3.23	325.3	144.92	140.3	621.3	150.59	21.2	348.1
Population 2 (1050)	99.85	402.0	65.1	98.53	365.0	21.1	100.48	417.9	62.6	96.75	312.5	137.2	99.01	390.9	32.0
	99.55	65.1	417.9	98.07	21.1	366.9	99.52	62.6	400.8	102.75	137.2	625.7	98.02	32.0	372.9
Error rate					0.049			0.054			0.105			0.052	
Confusion matrices					1004	46		965	85		952	98		996	54
					57	993		29	1021		124	926		56	994

correlation between the components of the data point, q_0 and q_1, was increased, the total error probability increased.

4. EXPERIMENTAL RESULTS

To illustrate the effectiveness of the proposed two-class clustering operator, several two cluster examples have been tested. For these examples, we have used a random number generator based on normal distributions[15] to create a 2D set S_A and used the same generator to create another 2D set S_B. These two sets were then merged together to form a data set Q. The different two-class classifiers were finally applied to Q. For the examples shown in Figs 4 and 5, the number of points in each subset, S_A and S_B, is 1050 points. In the example of Fig. 6, however, one of the subsets contains 1400 points and the other subset contains 700 points. The two generated data subsets in Fig. 4 are two identical normal distributions specified in Table 1. Figures 5 and 6 consider the cases when a well-elongated-shape data set exists. The distributions of two generated data sub-sets in Figs 5 and 6 are specified in Tables 2 and 3, respectively.

For these examples, we utilized (1) the Median-Cut (MC) classifier,[10] (2) Lin's classifier[9] and (3) the proposed classifier to carry out the task of the clustering data set Q. For the optimum Bayes classifier, we compared its performance with the above three classifiers only using the example of Fig. 4, in which its optimum decision boundary can be decided. The respective computed decision boundaries for each classifier are shown as the solid straight lines in each figure. The classified results were evaluated by means of the total error probability or equivalently the classification error rate. The confusion matrix is also a well-known indicator that can be used to evaluate clustering results.

As we observe, the MC classifier cannot classify data in each examples well. Lin's classifier fails in the examples shown in Figs 5 and 6 with the elongated-shape data set and the distance between the two clusters is too close. This situation has been reported in reference (9). The proposed CMP classifier does not separate very well the data in the example of Fig. 6. Nevertheless, the proposed classifier shows better results in each example compared with the MC classifier and Lin's classifier. These clustering results are further enhanced by the error rate and confusion matrices in

Table 2. The statistical parameters of the two 2D distributions for the example of Fig. 5 and performance comparison of different two-class clustering algorithms

	Generated data			Lin's classifier			MC classifier			CMP classifier		
	Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix	
Population 1	91.46	211.9	200.3	90.81	155.3	37.1	100.13	237.3	80.8	91.46	211.9	200.3
(1050)	126.33	200.3	205.0	110.99	37.1	139.4	128.15	80.8	107.8	126.33	200.3	205.0
Population 2	105.52	201.2	192.0	113.17	120.4	4.7	88.02	225.1	66.9	105.52	201.2	192.0
(1050)	111.91	192.0	199.4	134.67	4.7	105.3	104.11	66.9	95.7	111.91	192.0	199.4
Error rate					0.52			0.29			0.00	
Confusion				385	665		735	315		1050		0
matrices				714	336		313	737		0		1050

Table 3. The statistical parameters of the two 2D distributions for the example of Fig. 6 and performance comparison of different two-class clustering algorithms

	Generated data			Lin's classifier			MC classifier			CMP classifier		
	Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix		Mean vector	Covariance matrix	
Population 1	140.02	31.5	7.4	128.18	501.7	−56.2	120.01	652.2	9.1	138.41	50.7	36.5
(1400)	99.84	7.4	401.0	117.09	−56.2	135.1	112.11	9.1	152.8	99.77	36.5	405.8
Population 2	90.63	402.7	61.2	118.93	816.2	−42.3	120.13	759.0	21.1	85.00	256.7	−1.8
(700)	99.92	61.2	422.2	85.11	−42.3	172.2	80.05	21.1	163.2	98.36	−1.8	412.9
Error rate				0.46			0.49			0.045		
Confusion				684	716		693	707		1400	0	
matrices				252	448		340	360		96	604	

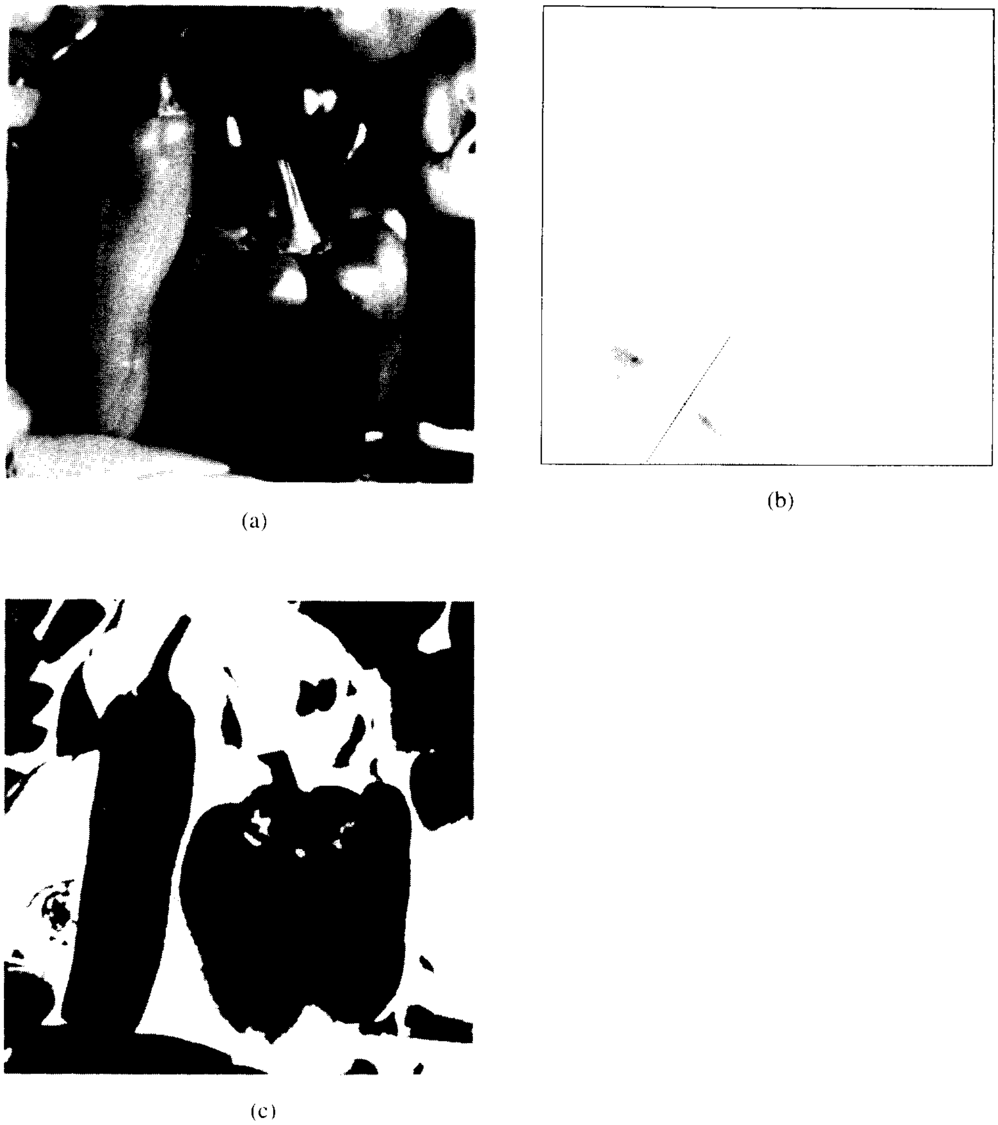

Fig. 7. Clustering results of the test image 'Peppers': (a) original colour image; (b) clustering result in (r, g) space by the CMP classifier; (c) colour segmentation result of the test image.

Tables 1, 2 and 3. These tables also compare the statistical properties in each classified sub-sets of the above classifiers for the test examples. As we can see, the proposed CMP classifier is better than the MC and Lin's classifiers on average, although one classified sub-set of Lin's classifier in the example of Fig. 4 resembles the original sub-set. In addition, we notice in Table 1 that the classifying result of the proposed classifier is close to the Bayes classifier. The error rate of the proposed classifier in this table is 0.052, whereas it is 0.049 with the Bayes classifier.

In the last two-class classifying example, we applied the proposed CMP clustering operator to the problem of colour-image segmentation. Since colour information can improve the performance of traditional monochrome image segmentation, colour image segmentation has become one subject of object recognition in recent years. The test colour image, named "Peppers", consists of three components that correspond to the "red", "green" and "blue" colour. The image is 480 × 480 size coded at 8 bits/pixel/component and shown in Fig. 7(a). The input 2D patterns chosen are from the (r, g) space:

$$r = \frac{R}{(R+G+B)}$$
$$g = \frac{G}{(R+G+B)}, \tag{22}$$

where R, G and B represent colour components red, green and blue of the test colour image. In this space, the r and g components represent the normalized chromatic information, respectively. Some previous works[16] have supported using these colour features

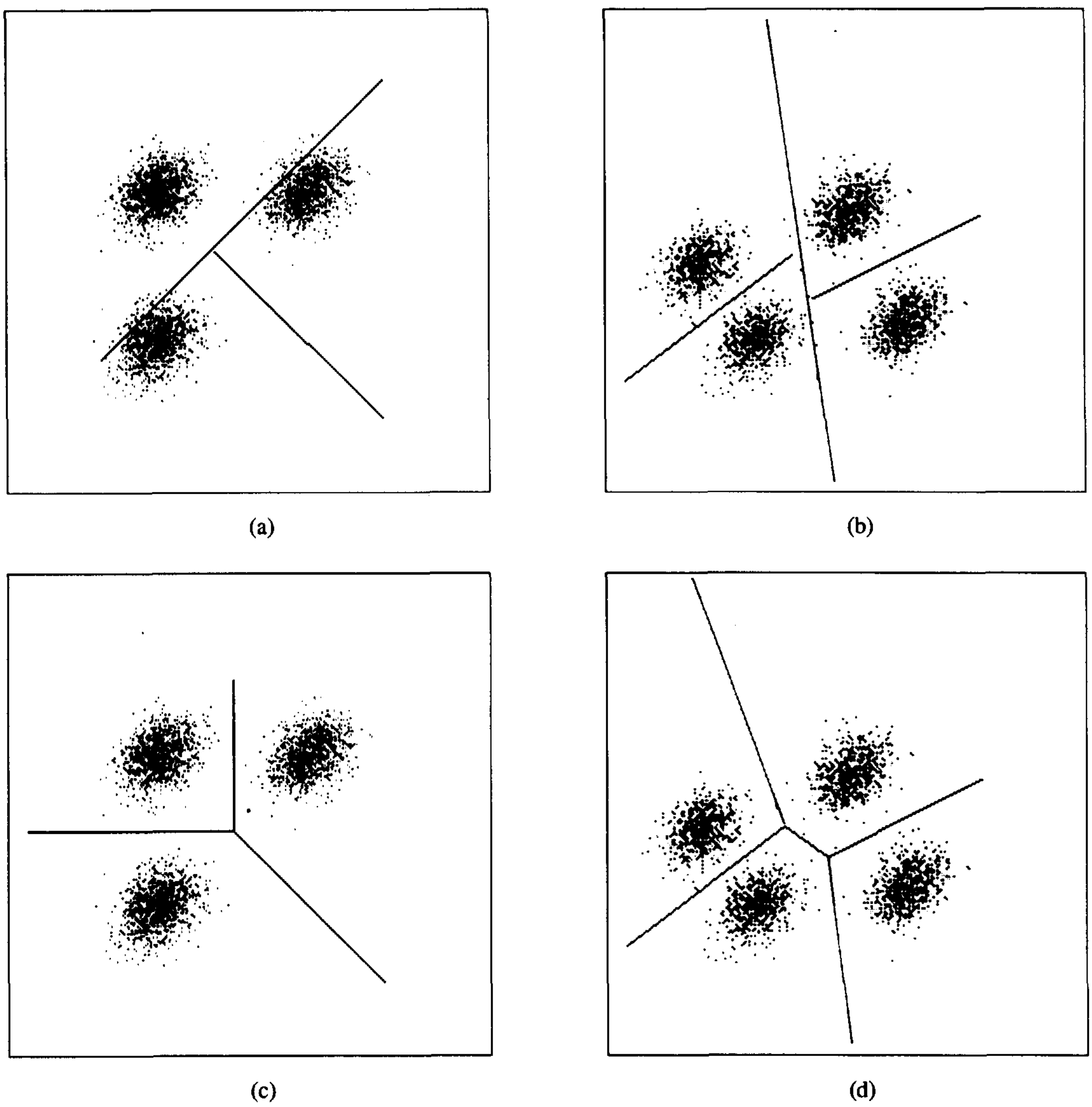

Fig. 8. Multi-class clustering results of the generated data listed in Table 4: (a) three-cluster result generated by the proposed CMP classifier; (b) four-cluster result generated by the proposed CMP classifier; (c) three-cluster result generated by the k-means classifier; (d) four-cluster result generated by the k-means classifier.

for colour-image segmentation. The distribution of (r, g) patterns for the test colour image is illustrated in Fig. 7(b), where the horizontal coordinate represents the r component and the vertical coordinate represents the g component. The colour segmentation result is shown in Fig. 7(c), in which two grey-levels were used to represent two classified classes. As we can see, the proposed classifier separates the colour pixel (picture elements) of the test image into red- and green-coloured objects. The highlights in the test image are classified as green-coloured objects.

Concerning the multi-class clustering problem, the proposed binary CMP clustering operator can be directly extended by using the binary decision tree approach. The proposed scheme will try to split the 2D data set Q until the pre-determined number of clusterings, M, is reached. The resultant partitioning of Q will exhibit the structure of a binary tree. Each node of the tree represents a subset of Q and the children of any node split the members of the parent node into two sets. The operator used for splitting nodes is the binary CMP clustering operator. Whether a node can be split further or not is indicated by the number of members in its child sets. If any of the child sets is empty, then this node will be declared as unsplitable. For these splitable nodes, the criterion to determine which node should be split in the next stage in the variance of the node. This scheme may now be described as:

(1) Input the 2D set Q.

(2) Do the following M-1 times or when no nodes are splitable:

Table 4. The statistical parameters of the multiple 2D distributions for the example of Fig. 8.

	Generated data for Figs 8(a) and (c)			Generated data for Figs 8(b) and (d)		
	Mean vector	Covariance matrix		Mean vector	Covariance matrix	
Population 1	80.13	97.0	30.5	80.15	97.1	30.6
(1050)	80.08	30.5	101.1	80.07	30.6	101.3
Population 2	159.89	101.9	34.4	159.89	101.9	34.5
(1050)	159.78	34.4	104.4	89.78	34.5	104.5
Population 3	80.23	95.8	28.3	50.23	95.6	28.4
(1050)	159.99	28.3	94.3	120.00	28.4	94.0
Population 4				129.7	104.8	35.8
(1050)				149.5	35.8	99.8

(a) Find a splitable node such that its variance is maximum;

(b) Use the binary CMP clustering operator to form two new nodes.

(3) Assign the membership of each cluster formed by step 2 to the data points in Q.

Figure 8 shows examples of applying the above multi-class clustering scheme when the generated data are in fact formed of more than two clusters. The generated data sub-sets are all identical normal distributions, whose statistical parameters are listed in Table 4. For these data, the classified results of the k-means classifier[2] are also obtained. From Fig. 8(a) and (b) we observe that the proposed scheme can obtain the desired clusters. Although some clusters might be improperly cut by the proposed scheme in Fig. 8(a), some merging operations after thresholding may be introduced to overcome this situation. The results achieved by the k-means classifier is presented in Fig. 8(c) and (d). We see that the k-means classifier has good decision boundaries to separate data into the desired clusters. Although the k-means classifier performs better than the proposed scheme, the complexity of the k-means classifier is greater than that of the proposed scheme. For instance, the k-means classifier should consider the number of iterations needed and the problem of being convergent or not. For the data size and the number of iterations being equal to N and I, respectively, the computational load of the k-means classifier is of the order IN. In addition, unless the initial guess is well chosen, the results of the k-means classifier are easily trapped inside the local minimum, which degrades the performance of the k-means classifier.

5. CONCLUSIONS

In this paper, we present a two-class classifying operator based on the complex-moment-preserving principle. The traditional 1D moment-preserving classifier[7] can be considered as a special case of the proposed classifier. Through preserving three complex moments, an analytical solution can be obtained. The computational complexity of the proposed classifier is proportional to the data size. Moreover, the performance of the clustering results is satisfactory from the experiments shown in Section 4. As compared with the other non-interactive algorithms mentioned in Section 4, the proposed algorithm becomes attractive for fast automatic two-class clustering or any other-fields requiring fast automatic hierarchical clustering.

REFERENCES

1. G. Nagy, State of the arts in pattern recognition, *Proc. IEEE* **56**, 836–862 (May 1968).
2. S. Z. Selim and M. A. Islam, K-Means type algorithms: a generalized convergence theorem and characterization of local optimality, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 81–87 (1984).
3. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York (1973).
4. R. J. Hathaway and J. C. Bezdek, Recent convergence results for the fuzzy C-means clustering algorithms, *J. Class.* **2**, 29–39 (1988).
5. P. A. Devijver and J. Kittler, *Pattern Recognition: a Statistical Approach.* Prentice-Hall, London (1982).
6. E. J. Delp and O. R. Mitchell, Image compression using block truncation coding, *IEEE Trans. Commu.* **COMM-27**, 1335–1341 (September 1979).
7. W. Tsai, Moment preserving thresholding: a new approach, *Comput. Vis. Graphics Image Process.* **29**, 377–393 (1985).
8. P. K. Sahoo, S. Selenate and A. K. C. Wong, A survey of thresholding techniques, *Comput. Vis. Graphics Image Process.* **41**, 233–260 (1988).
9. J. C. Lin and W. Tsai, Feature preserving clustering of 2D data for two-class problems using analytic formulas: an automatic and fast approach, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-5**, 554–560 (May 1994).
10. P. Heckbert, Colour image quantization for frame buffer display, *Comput. Graph.* **16**(3), 297–307 (July 1982).
11. A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw Hill, New York (1984).

12. C. K. Papadopoulos and C. L. Nikias, Parameter estimation of exponentially damped sinusoids using high order statistics, *IEEE Trans. Acoust., Speech Sign. Process.* **ASSP-38**(18), 1424–1435 (August 1990).
13. A. M. Mood, F. A. Graybill and D. C. Bose, *Introduction to the Theory of Statistics*. McGraw Hill, New York (1974).
14. H. L. Van trees, *Detection, Estimation, and Modulation Theory*. Wiley, New York (1968).
15. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*. McGraw Hill, New York (1991).
16. G. Healey, Segmentation images using normalized colour, *IEEE Trans. Syst., Man Cybernet.* **SMC-22**, 64–73 (1992).

About the Author—SOO-CHANG PEI was born in Soo-Auo, Taiwan, in 1949. He received the BSEE from National Taiwan University in 1970 and MSEE and Ph.D. from the University of California, Santa Barbara, in 1972 and 1975, respectively.

He was an engineering officer in the Chinese Navy Shipyard from 1970 to 1971. From 1971 to 1975 he was a research assistant at the University of California, Santa Barbara. He was Professor and Chairman in the EE Department of Tatung Institute of Technology from 1981 to 1983. Presently, he is the Professor and Chairman of the EE Department at National Taiwan University. His research interests include digital signal processing, image processing, optical information processing and laser holography. Dr Pei is a member of the IEEE, Eta Keppa Nu and the Optical Society of America.

About the Author—CHING-MIN CHENG was born in Taipei, Taiwan, in 1959. He received the BSEE from National College of Marine Science and Technology, Keelung, Taiwan, in 1982 and MSEE from University of California, San Diego, in 1986. He is currently working towards the Ph.D. degree in electrical engineering at the National Taiwan University. From 1983 to 1984 he was an engineering officer in the Chinese Airforce Anti-aircraft Corps. From 1986 to August 1989 he served as a patent examiner in the National Bureau of Standards. Since September 1989, he joined Telecommunication Labs, Ministry of Communications, Taiwan, as a research engineer. His research interests includes digital signal processing, video compression and multimedia communication.