The complete form of the transformation equations which map the scene coordinates onto the image coordinates is therefore

$$x' = (x + \Delta x) + b$$

and

$$y' = c \cdot (x + \Delta x) + d \cdot (y + \Delta y) + e.$$

In order to solve for the unknown parameters, we need to consider the greatest source of random error. Due to the process of quantization involved in digitizing a video image, the measured image coordinates are going to be in error. Therefore the unknown parameters can be computed by using a least squares techniques which aims to minimize the difference between the measured image coordinates and the computed image coordinates. A suitable technique would be the generalized form of the Newton-Raphson method. The equations to solve are

$$x'_m - (x + \Delta x) - b = 0$$

and

$$y'_m - c \cdot (x + \Delta x) - d \cdot (y + \Delta y) - e = 0$$

where (x'_m, y'_m) are the measured image coordinates.

To obtain the reverse transform from the image coordinates (x', y') to the corrected image coordinates (x, y), the image-to-image mapping equations are reversed to provide the following transformation equations

$$x = (x' + b') + \Delta x'$$

and

$$y = (c' \cdot x' + d' \cdot y' + e') + \Delta y'$$

where b', c', d', and e' are constants and $\Delta x'$ and $\Delta y'$ are systematic distortion corrections.

References

[1] Y. Shah, R. Chapman, R. Babaei-Mahani, and N. E. Lindsey, "Extraction of range information from stereo images," *Opt. Lasers Eng.*, vol. 6, no. 2, pp. 125-127, 1985.

[2] S. Nagata, "New versatile stereo (NS-type) display system and measurement of binocular depth perception," *Japan J. Med. Electron. Biol. Eng.*, vol. 20, no. 3, pp. 154-61, 1982.

[3] I. J. Dowman and A. Haggag, "Digital image correlation along epipolar lines," in *Proc. Int. Symp. Image Processing Interactions with Photogrammetry and Remote Sensing, Graz.*, 1977, pp. 47-49.

[4] R. R. Real and Y. Fujimoto, "Digital processing of dynamic imagery for photogrammetric applications," *IEEE Trans. Instrum. Measurement*, vol. IM-33, no. 1, pp. 45-51, 1987.

[5] M. D. Levine, D. A. O'Hanley, and G. M. Yagi, "Computer determination of depth maps," *Comput. Graphics Image Processing*, vol. 2, pp. 131-150, 1973.

[6] A. M. Chong, M. S. Beck, M. B. Brown, I. S. MacKenzie, and R. T. Ritchings, "Micro-computer based stereoscopic rangefinding," in *Proc. IRE Conf. Manchester*, 1982, pp. 95-102.

[7] V. P. Bennet and N. Balasubramanian, "Detection of image coincidence using correlation coefficient measurement," in *Proc. 39th Conf. Amer. Soc. Photogrammetry*, Washington, DC, 1973, pp. 1-19.

[8] J. G. Hardy and A. T. Zavodny, "Automation reconnaissance-based target-coordinate determination," in *Proc. Techniques and Applications of Image Understanding*, Soc. Photo-Opt. Instrum. Eng., 1973, pp. 95-104.

[9] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 4, pp. 333-340, 1980.

[10] R. D. Arnold and T. O. Binford, "Geometric constraints in stereo vision," in *Proc. Image Processing for Missile Guidance*, Soc. Photo-Opt. Instrum. Eng., vol. 238, 1980, pp. 281-290.

[11] L. Sidney, "Geometric constraints for interpreting images of common structural elements: Orthogonal trihedral vertices," in *Proc. Techniques and Applications of Image Understanding*, Soc. Photo-Opt. Instrum. Eng., vol. 281, 1981, pp. 332-338.

[12] M. Benard, "Extracting 3-D coordinates of an object from a digital stereopair: An automated method," in *Proc. Signal Processing II: Theories and Applications*, EUSIPCO, Erlangen, Germany, 1983, pp. 227-230.

[13] H. H. Baker and T. O. Binford, "Depth from edge and intensity based stereo," in *Proc. 7th Int. Joint Conf. Artificial Intelligence*, vol. 2, 1981, pp. 631-636.

[14] B. Julesz, *Foundations of Cyclopean Perception.* Chicago, IL: University of Chicago Press, 1971.

[15] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, pp. 283-287, 1976.

[16] —, "A theory of human stereo vision," MIT Artificial Intell. Lab., AI Memo 451, 1977.

[17] J. E. W. Mahew and J. P. Frisby, "The computation of binocular edges," *Perception*, vol. 9, pp. 69-86, 1980.

[18] S. A. Lloyd, "A dynamic programming algorithm for binocular stereo vision," *GEC J. R.*, vol. 3, no. 1, pp. 18-24, 1985.

[19] L. H. Quam, "Hierarchical warp stereo," in *Proc. Darpa Image Understanding Workshop*, New Orleans, LA, 1984, pp. 211-215.

[20] D. C. Brown, "Calibration of close range cameras," in *Proc. 12th Congress Int. Soc. Photogrammetry*, Ottawa, Canada, 1972.

[21] —, "Analytical calibration of close range cameras," in *Proc. Symp. Close-Range Photogrammetry*, Melbourne, FL, 1971.

[22] K. W. Wong, *Geometric Calibration of Television Systems for Photogrammetric Applications* (Civil Eng. Studies, Photogrammetry Series no. 16). Urbana, IL: University of Illinois Press, 1968.

Unsupervised Optimal Fuzzy Clustering

I. GATH AND A. B. GEVA

***Abstract*—Many algorithms for fuzzy clustering depend on initial guesses of cluster prototypes, and on assumptions made as to the number of subgroups present in the data. This study reports on a method for carrying out fuzzy classification without *a priori* assumptions on the number of clusters in the data set. Assessment of cluster validity is based on performance measures using hypervolume and density criteria. The new algorithm is derived from a combination of the fuzzy *K*-means algorithm and the fuzzy maximum likelihood estimation (FMLE). The UFP-ONC (unsupervised fuzzy partition-optimal number of classes) algorithm performs well in situations of large variability of cluster shapes, densities, and number of data points in each cluster. It has been tested on a number of simulated and real data sets.**

***Index Terms*—Clustering of sleep EEG, fuzzy clustering, hyperelliptoidal clusters, performance measures for cluster validity, unequally variable features, unsupervised tracking of cluster prototype.**

I. Introduction

Cluster analysis is based on partitioning a collection of data points into a number of subgroups, where the objects inside a cluster (a subgroup) show a certain degree of closeness or similarity. Hard clustering assigns each data point (feature vector) to one and only one of the clusters, with a degree of membership equal to one, assuming well defined boundaries between the clusters. This model often does not reflect the description of real data, where boundaries

Manuscript received June 26, 1987; revised May 5, 1988. Recommended for acceptance by A. K. Jain. This work was supported by the Kennedy-Leigh fund for Biomedical Engineering Research.

The authors are with the Department of Biomedical Engineering, Technion, Haifa 32000, Israel.

IEEE Log Number 8927500.

0162-8828/89/0700-0773$01.00

between subgroups might be fuzzy, and where a more nuanced description of the object's affinity to the specific cluster is required. Thus, numerous problems in the life sciences are better tackled by decision making in a fuzzy environment [1]–[4]. Bezdek [5] developed a family of clustering algorithms, based on fuzzy extension of the least-square error criterion, and proved the convergence of the algorithms to a local minimum [6]. Related algorithms, taking into account differences in cluster shapes have been proposed by Bezdek and Dunn [7], Bezdek *et al.* [8], [9], and Gustafson and Kessel [10].

There are three major difficulties encountered during fuzzy clustering of real data: 1) The number of clusters can not always be defined apriori, and one has to find a cluster validity criterion [11], in order to determine the optimal number of clusters present in the data. 2) The character and location of cluster centroids is not necessarily known *a priori*, and initial guesses have to be made. 3) The presence of large variability in cluster shapes, variations in cluster densities, and variability in the number of data points in each cluster. A good example which demonstrates the complexity of handling real data is classification of EEG recordings [12], [13]. Fuzzy clustering of sleep EEG in order to classify the signal into various sleep stages is reported in [1], [4], [14].

In the present study an algorithm for fuzzy classification into optimal number of clusters will be described. Optimality is restricted here to the notion of optimizing new performance measures, based on cluster hypervolume and density criteria. The algorithm accounts for variability in cluster shapes, cluster densities, and the number of data points in each of the subsets. Classification prototypes for initiation of the iterative process are generated through a process of unsupervised learning. The new algorithm will be tested on different classes of simulated data, and on a real data set derived from sleep EEG signal.

II. Unsupervised Fuzzy Partition

In order to obtain satisfactory solution to the problem of large variability in cluster shapes and densities, and to the problem of unsupervised tracking of classification prototypes, a two-layer clustering strategy has been developed. During the first step, a modification of the fuzzy K-means algorithm [5] is carried out. There are no initial conditions on the location of cluster centroids, and classification prototypes are identified during a process of unsupervised learning. Using these prototypes, the second step involves the utilization of a second clustering algorithm in order to achieve optimal fuzzy partition. This scheme is iterated for increasing number of clusters in the data set, computing performance measures in each run, until partition into optimal number of subgroups is obtained:

1) Cluster with fuzzy K-means (Section II-A), using unsupervised tracking of initial classification prototypes (Section II-B).

2) Cluster with the fuzzy modification of the maximum likelihood estimation (refinement of step 1, Section II-A).

3) Compute-performance measures (Section II-C).

4) Increase K (number of subgroups) and repeat steps 1–3 until optimum value of performance measure is obtained.

A. The Fuzzy K-Means Algorithm and Its Derivatives

The fuzzy K-means algorithm [5] is based on minimization of the following objective function, with respect to U, a fuzzy K-partition of the data set, and to V, a set of K prototypes:

$$J_q(U, V) = \sum_{j=1}^{N} \sum_{i=1}^{K} (u_{ij})^q d^2(X_j, V_i); \qquad K \le N \tag{1}$$

where q is any real number greater than 1, X_j is the jth m-dimensional feature vector, V_i is the centroid of the ith cluster, u_{ij} is the degree of membership of X_j in the ith cluster, $d^2(X_j, V_i)$ is any inner product metric (distance between X_j and V_i), N is the number of data points, K is number of clusters. The parameter q is the weighting exponent for u_{ij} and controls the "fuzziness" of the resulting clusters [11].

Fuzzy partition is carried out through an iterative optimization of (1) according to [5]:

1) Choose primary centroids V_i (prototypes).

2) Compute the degree of membership of all feature vectors in all the clusters:

$$u_{ij} = \frac{\left[\dfrac{1}{d^2(X_j, V_i)}\right]^{1/(q-1)}}{\sum_{k=1}^{K}\left[\dfrac{1}{d^2(X_j, V_k)}\right]^{1/(q-1)}}. \tag{2}$$

3) Compute new centroids $\hat{V}_i$:

$$\hat{V}_i = \frac{\sum_{j=1}^{N} (u_{ij})^q X_j}{\sum_{j=1}^{N} (u_{ij})^q} \tag{3}$$

and update the degree of memberships, u_{ij} to $\hat{u}_{ij}$, according to (2).

4)

$$\text{if } \max_{ij} \left[|u_{ij} - \hat{u}_{ij}| \right] < \epsilon \text{ stop, otherwise goto step 3} \tag{4}$$

where ϵ is a termination criterion between 0 and 1.

Computation of the degree of membership u_{ij} depends on the definition of the distance measure, $d^2(X_j, V_i)$, [11]:

$$d^2(X_j, V_i) = (X_j - V_i)^T A (X_j - V_i). \tag{5}$$

The inclusion of A (an $m \times m$ positive-definite matrix) in the distance measure results in weighting according to the statistical properties of the features [10]. In the following, two different distance measures will be defined, to be used in the two different layers of the clustering process:

1) For the case where A equals the identity matrix the distance is Euclidean. The resulting algorithm is the fuzzy K-means.

2) For hyperellipsoidal clusters, as well as in the presence of variable cluster densities and unequal numbers of data points in each cluster, an "exponential" distance measure, $d_e^2(X_j, V_i)$, based on maximum likelihood estimation [7], [11], [15] is defined. This distance will be used in calculation of $h(i|X_j)$, the posterior probability (the probability of selecting the ith cluster given the jth feature vector):

$$h(i|X_j) = \frac{1/d_e^2(X_j, V_i)}{\sum_{k=1}^{K} 1/d_e^2(X_j, V_k)} \tag{6}$$

$$d_e^2(X_j, V_i) = \frac{[\det(F_i)]^{1/2}}{P_i} \exp\left[(X_j - V_i)^T F_i^{-1} (X_j - V_i)/2\right] \tag{7}$$

where F_i is the fuzzy covariance matrix of the ith cluster, and P_i, the *a priori* probability of selecting the ith cluster.

Comparison of (6) and (2) shows that for $q = 2$ $h(i|X_j)$ is similar to u_{ij}. Thus, substituting (6) instead of (2) in step 2 of the fuzzy K-means algorithm results in the fuzzy modification of the maximum likelihood estimation (FMLE). Step 3 of the FMLE algorithm includes, in addition to computation of the new centroid, calculation of P_i, the *a priori* probability of selecting the ith cluster:

$$P_i = \frac{1}{N} \sum_{j=1}^{N} h(i|X_j) \tag{8}$$

and of F_i, the fuzzy covariance matrix of the ith cluster:

$$F_i = \frac{\sum_{j=1}^{N} h(i \mid X_j)(X_j - V_i)(X_j - V_i)^T}{\sum_{j=1}^{N} h(i \mid X_j)}. \tag{9}$$

Due to the "exponential" distance function incorporated in the FMLE algorithm it seeks an optimum in a narrow local region. It therefore does not perform well, and might be even unstable during unsupervised identification of classification prototypes described in Section II-C. Its major advantage is obtaining good partition results in cases of unequally variable features and densities, but only when starting from "good" classification prototypes. The first layer of the algorithm (unsupervised tracking of initial centroids) is therefore based on the fuzzy K-means algorithm, whereas in the next phase optimal fuzzy partition is being carried out with the FMLE algorithm.

B. Unsupervised Tracking of Cluster Prototypes

The algorithms described in the previous section start with initial guesses of classification prototypes, and the iterative process results in convergence of the cluster centroids to a local optimum. Different choices of classification prototypes may lead to convergence to different local optima, i.e., to different partitions. In many practical situations *a priori* knowledge of the approximate locations of the initial centroids does not exist, and in order to achieve optimal partition unsupervised tracking of classification prototypes is required.

Given a partition into k clusters, the basic idea is to place the $(k + 1)$st cluster center in a region where data points have low degree of membership in the existing k clusters. The following scheme describes the steps for the selection of initial cluster centers, incorporated in the fuzzy K-means algorithm:

1) Compute average and standard deviation of the whole data set.

2) Choose the first initial cluster prototype at the average location of all feature vectors.

3) Choose an additional classification prototype equally distant (with a given number of standard deviations) from all data points (a nonphysical location).

4) Calculate a new partition of the data set according to steps 1 and 2 of the scheme outlined in Section II.

5) If k, the number of clusters is less than a given maximum, goto 3, otherwise stop.

C. Performance Measures for Cluster Validity

During clustering of real data one usually has to make assumptions as to the number of underlying subgroups present in the data set. When no *a priori* information exists as to the internal structure of the data, or in case of conflicting evidence about the optimal number of subgroups, performance measures for comparison between the goodness of partitions with different numbers of clusters need to be formulated.

A goal-directed approach [16] to the cluster validity problem can be chosen, where the goal is classification, in the sense of minimization of the classification error rate. Hence, one may accept the basic heuristic that "good" clusters are actually not very fuzzy [11]. Therefore, the criteria for the definition of "optimal partition" of the data into subgroups were based on three requirements:

1) Clear separation between the resulting clusters.

2) Minimal volume of the clusters.

3) Maximal number of data points concentrated in the vicinity of the cluster centroid.

Thus, although the environment is fuzzy, the aim of the classification is generation of well-defined subgroups, and hence these requirements lead to a "harder" partitioning of the data set.

The performance measures were based on criteria for hypervolume and density. Fuzzy hypervolume, F_{HV}, is defined by:

$$F_{HV} = \sum_{i=1}^{K} \left[\det (F_i)\right]^{1/2} \tag{10}$$

where F_i is given by (9).

Average partition density D_{PA} is calculated from:

$$D_{PA} = \frac{1}{K} \sum_{i=1}^{K} \frac{S_i}{\left[\det (F_i)\right]^{1/2}} \tag{11}$$

where S_i, the "sum of central members," is given by:

$$S_i = \sum_{j=1}^{N} u_{ij} \quad \forall\, X_j \in \left\{X_j : (X_j - V_i) F_i^{-1} (X_j - V_i)\right] < 1\right\} \tag{12}$$

taking into account only those members within the hyperellipsoid, whose radii are the standard deviations of the cluster features.

The partition density P_D is calculated from

$$P_D = \frac{S}{F_{HV}} \tag{13}$$

where

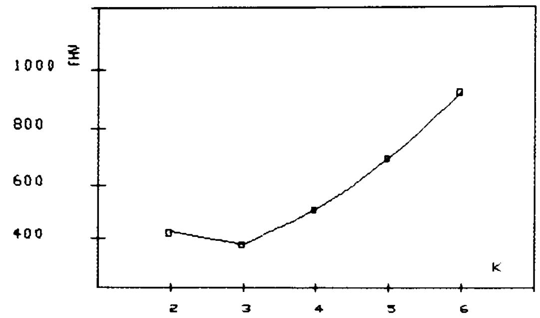
$$S = \sum_{i=1}^{K} \sum_{j=1}^{N} u_{ij} \quad \forall\, X_j \in \left\{X_j : (X_j - V_i) F_i^{-1} (X_j - V_i) < 1\right\}. \tag{14}$$

An example for estimating the optimal number of subsets in a data set, using the performance measures, is demonstrated in Fig. 1. The data set is the 150 patterns describing three iris subspecies [17], [18]. Plotting the performance measures F_{HV} and P_D as a function of the number of subgroups in the data set shows points of extremum at $k = 3$, in accordance with the botanically correct number of classes.

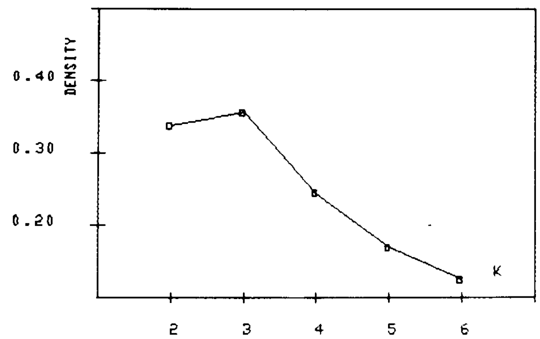
The F_{HV} criterion shows a clear extremum in most of the cases. However, the density criteria will be more sensitive as performance measures when there are substantial overlapping between the clusters and when large variability in compactness of the clusters exists. The D_{PA} criterion reflects the presence of single dense clusters (the fuzzy density is calculated for each cluster and then averaged over all clusters), and thus, partition resulting in both dense and loose clusters is considered a "good" partition because of the dense substructures. The P_D criterion expresses the general partition density according to the physical definition of density.

III. Sample Runs

In order to test the performance of the algorithm a simulation program was written, generating N artificial m-dimensional feature vectors from a multivariate normal distribution. The input to the program consisted of: 1) N, the number of data points. 2) m, dimension of feature space. 3) K, the number of required subsets in the data. 4) The required m-dimensional cluster prototypes. 5) The variance of each feature in each of the clusters. 6) The relative number of data points in each subset. By choosing the distances between cluster prototypes to be near each other, and controlling the variance of the features, overlapping between clusters could be obtained, resulting in a fuzzy environment. The features had unequal variance generating hyperellipsoidal clusters. The number of subgroups in the data, their density, and number of data points in each subgroup were subject to variation. Another artificial data set was taken from Gustafson and Kessel [10], and the algorithm was also tested on the iris data of Anderson [17] and Fisher [18], and feature vectors derived from sleep EEG. As to the algorithmic parameter q, a theoretical basis for an optimal choice of the weighting exponent is so far not available [11], [19], [20]. A value of $q = 2$ [7], [10], [11], [21] was chosen for the UFP-ONC algorithm.



(a)



(b)

Fig. 1. Performance criteria for the Iris data. (a) FHV—Fuzzy hypervolume as a function of the number of subgroups in the data. (b) Partition density as a function of number of subgroups. Extrema are seen for $k = 3$.

Example 1: This example demonstrates optimal partition of touching clusters with large variability in cluster densities and number of data points in each cluster. In Fig. 2(a), an artificial data set with two-dimensional feature vectors drawn from a bivariate normal distribution is demonstrated. One of the subgroups in the data is large and loose, while the other is small and shows a much higher density of the data points. There is no clear border between the subgroups.

Using the fuzzy K-means algorithm alone results in misclassification of boundary data points, Fig. 2(b). Peripheral data points generated by the loose cluster will be misclassified as belonging to the high density cluster. Application of the UFP-ONC algorithm classifies correctly all 200 data points, Fig. 2(c).

Example 2: This example demonstrates successful partition of linear substructures. Fig. 3(a) demonstrates two linear clusters, generated from a uniform distribution by Gustafson and Kessel [10]. The two subsets consist of two long and narrow formations, at right angle to each other. The cluster centers were generated to coincide exactly with each other. Running the UFP-ONC algorithm gives the two clusters, Fig. 3(b), with no misclassification of any of the points.

Example 3: This example shows optimal partition of a data set with multiple substructures. Twelve different clusters are generated from a multivariate normal distribution, Fig. 4(a). The feature space is five-dimensional. There is a significant variability of shapes, densities, and number of patterns in each cluster. The performance measures for estimating the number of subgroups in the data set are depicted in Figs. 4(b) and (c). A minimum for $k = 12$ is clearly seen for the F_{HV} criterion, as well as a maximum for $k = 12$ for the partition density criterion. The partition, running the UFP-ONC algorithm is shown in Fig. 4(d). All the patterns have been correctly classified.

Example 4: The iris data set of Anderson [17] and Fisher [18] has three subgroups, two of which are overlapping. Estimation of the optimal number of substructures in the data set (whether it is 2 or 3) is the crucial point here. The patterns are depicted in Fig. 5(a), and the F_{HV} and P_D curves in Fig. 1. The optimal number of subgroups in the data set is given by the minimum of the F_{HV} curve and the maximum of the P_D curve at $k = 3$ clusters. Partition into the three clusters is shown in Fig. 5(b). There are 4 misclassifications within the 150 patterns (an error of 2.7 percent). Three plants of Iris Versicolor have been classified as Iris Virginica, whereas only one plant of Iris Virginica has been attributed to Iris Versicolor. All 50 Iris Setosa plants have been correctly classified.

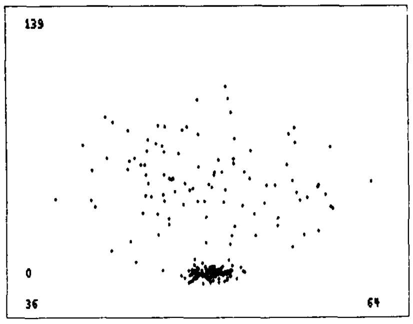
Example 5: Computerized scoring of sleep EEG into various stages [1], [4], [14], [22] represents a typical example of handling real data by fuzzy clustering. The patterns characterizing sleep EEG segments generate a fuzzy environment, with some traits complicating any process of classification:

1) Physiologically, there are continuous transitions between the sleep stages, i.e., the subgroups *are not well separated.*

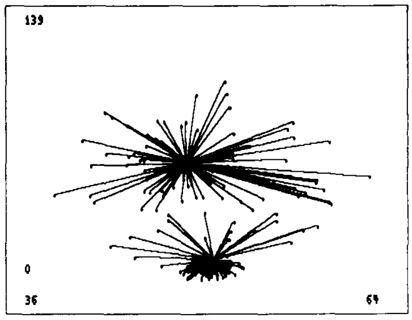
2) There is a great deal of intersubject variability of the spectral features of the various sleep stages, and *the features have unequal variance* (large variability in cluster shapes).

3) The number of stationary EEG segments and the variability of their features differ for the various sleep stages (variability in cluster densities and number of data points in each cluster).

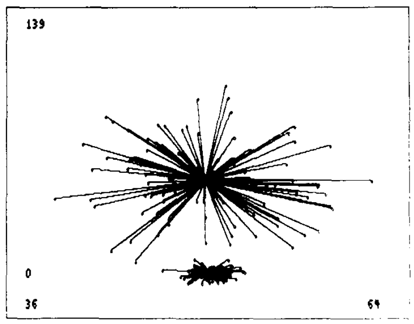
4) The number of sleep stages might vary between subjects (depends on age, pathological conditions, etc.), i.e., the number of subgroups in the data *is not known a priori.*



(a)



(b)



(c)

Fig. 2. Partition of simulated data with unequally variable features. (a) Two hundred data points generated from a bivariate Gaussian distribution. There are two subgroups in the data, one large and loose and the other small and dense. (b) Partition using the fuzzy K-means algorithm. Peripheral points generated by the loose cluster are misclassified as belonging to the smaller cluster. (c) The UFP-ONC algorithm classifies correctly all 200 data points.

Thus, even if one fixes the number of subgroups in the data set, fuzzy clustering of sleep EEG using either of the algorithms described in [5], [10], [19] does not guarantee optimal partition.

Patterns representing a whole night's sleep EEG segments from a 30 year old female are shown in Fig. 6(a). The five features, derived by adaptive segmentation and time-dependent clustering of the signal [4], are the relative power in the physiological frequency bands, delta, theta, alpha, sigma, and beta. One of the criteria for estimating the optimal number of classes in the data F_{HV} is plotted in Fig. 6(b). From the minimum in the curve it can be concluded that there are five subgroups. The partition and classification histogram (hypnogram) are depicted in Figs. 6(c) and (d), respectively. For comparison, a hypnogram scored manually by a physician is given in Fig. 6(e). There is a clear similarity between the two classification histograms. Due to the scanty number of EEG segments belonging to sleep stage I, sleep stage wake and sleep stage I have been classified by the UFP-ONC algorithm as being one class.

The CPU requirements of the UFP-ONC algorithm on the IBM AT personal computer, analyzing four-dimensional 150 patterns (iris data), with K, the maximal number of clusters equal to 6, was 14 min.

IV. Conclusions

Implementing the strategy of unsupervised tracking of initial cluster centroids, the most flexible algorithm has been found to be

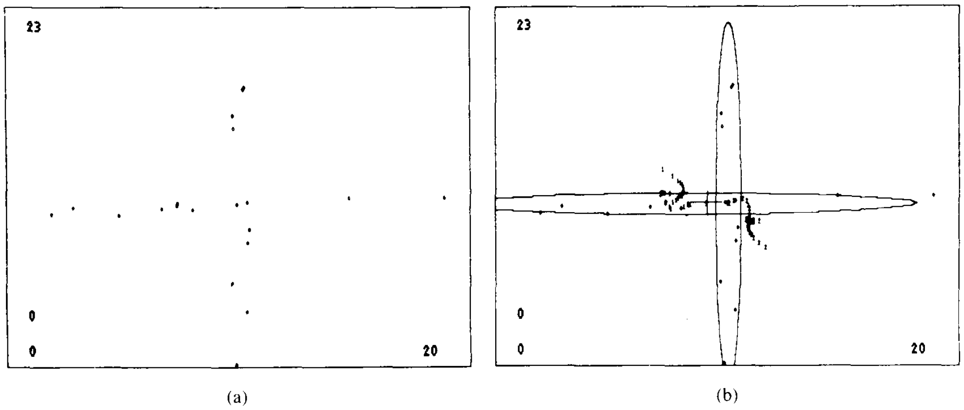


Fig. 3. Partition of linear clusters, data of Gustafson and Kessel [10]. (a) Twenty data points drawn from a uniform distribution. The two subsets consist of two long and narrow formations, at right angle to each other. (b) Convergence of the two centroids to their final locations running the UFP-ONC algorithm. The trajectories of the two centroids during the iterations can be followed by the points denoted by small numerals 1 and 2. Two standard deviations are drawn around the final centroids. All data points have been correctly classified.

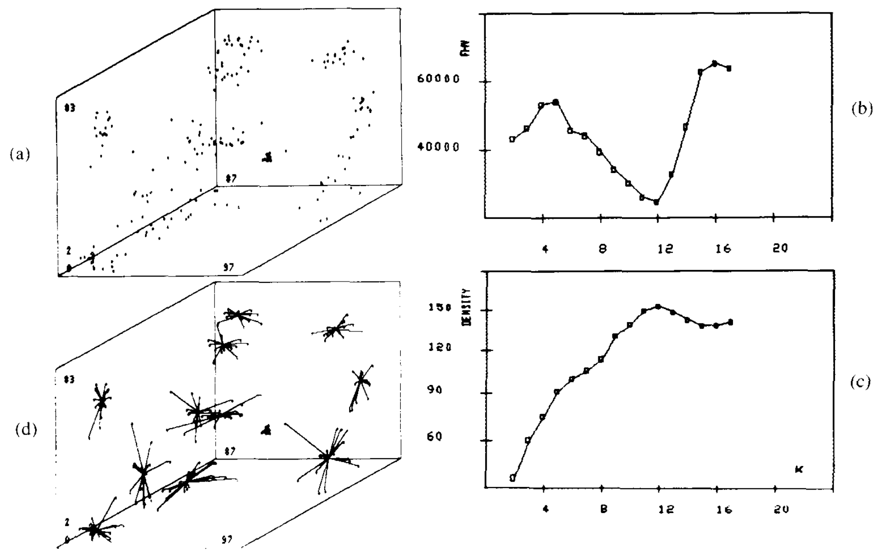
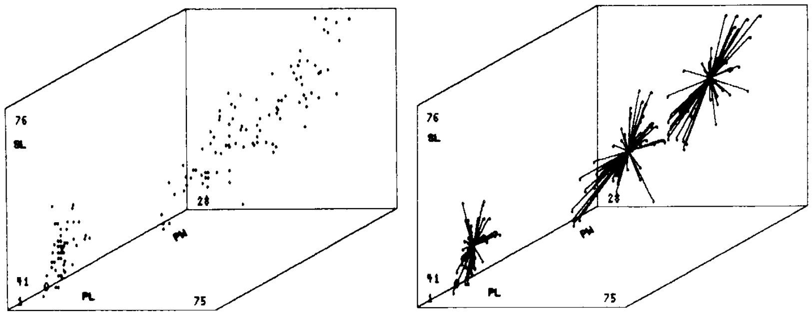


Fig. 4. Partition of 12 clusters generated from five-dimensional multivariate Gaussian distribution with unequally variable features, variable densities and variable number of data points in each cluster. (a) Data points before partition. Only three of the features are displayed. (b), (c) Fuzzy hypervolume (FHV) and partition density, as a function of the number of subgroups in the data. Extrema for $k = 12$ can be seen. (d) Partition of 12 subgroups using the UFP-ONC algorithm. All data points have been correctly classified.

the fuzzy K-means, although it does not give optimal partition in cases of variable cluster shapes and densities. On the other hand, using an "exponential" distance measure including the fuzzy covariance matrix (the FMLE algorithm [7], [11], [15]) results in optimal partition even when a great variability of cluster shapes and densities is present. An optimal performance of the FMLE algorithm requires starting from "good" seed points, because due to the "exponential" distance this algorithm converges to a local optimum in a rather narrow region. Taking this limitation into account, the FMLE algorithm is superior to Gustafson and Kessel's [10] fuzzy covariance algorithm, in that it does not require an extra volume constraint (the ρ_j of Gustafson and Kessel), limitation on the hypervolume being achieved through the exponent.

The new algorithm described in the present study combines the favorable features of both the fuzzy K-means algorithm and the FMLE, together with unsupervised tracking of classification prototypes. Optimal partition has been achieved with the UFP-ONC algorithm for several synthetic data sets, as well as for sleep EEG



(a) (b)

Fig. 5. Classification of the Iris data of Anderson and Fisher using the UFP-ONC algorithm. (a) 150 four-dimensional feature vectors. Only three of the features are displayed here: PL—Petal length. PW—Petal width. SL—Sepal length. (b) Fuzzy partition to three subgroups. There are a total of four errors, three items of Iris Versicolor have been misclassified as Iris Virginica whereas one Iris Virginica has been misclassified as Iris Versicolor.

classification, omitting the need for initial guesses on cluster prototypes. The iris data set [17], [18] is a well known example of overlapping substructures. The results of applying the UFP-ONC algorithm in this case were optimal, both from the point of view of estimating the number of underlying substructures, and that of classification error rate [23]–[25]. Extending the notion of hyperellipsoidal clusters to the extreme, by letting one feature vary much less than the others, gives rise to line-like clusters. Graph-theoretic methods have been proven to be successful in detecting linear substructures [26], but in general these methods fail on hyperellipsoidal clusters [8]. Due to the inclusion of the FMLE in the UFP-ONC algorithm, the new algorithm is also able to detect line-like clusters, as demonstrated on the data of Gustafson and Kessel [10].

Performance measures for assessing cluster validity have been proposed in the framework of ranking various partitions obtained from different clustering algorithms. Such a cluster validity strategy was implemented in [16], using performance measures based on fuzzy decomposition of the data. The search for a proper cluster validity criterion in the present study has been goal-oriented, with relation to the application domain [16], [27]. The aim was to estimate the optimal number of substructures in the data set for the purpose of classification (minimum classification error rate). It has been motivated by studies of automatic classification of sleep stages [4], [14], [22], where the number of subsets in the data is not necessarily known *a priori*, and where a large intersubject variability of the number of classes may be present.

In order to estimate the optimal number of subgroups present in the data the UFP-ONC algorithm incorporates performance measures based on hypervolume and density criteria. The hypervolume criterion is related to the within-cluster scatter, but due to its fuzzy characteristics the F_{HV}, unlike the square error criterion, is not a monotone function of k. These performance measures (and in particular the hypervolume criterion) plotted as a function of the number of clusters k show a clear extremum, from which conclusions as to the optimal number of substructures in the data can be drawn. This has been demonstrated for the iris data, where the botanically correct number of clusters was detected by the new algorithm.

Other performance measures, aimed at delineating the number of subgroups in the data set, are either monotone functions of k [11], [28], or show a very slight preference for a certain value of k, as is the case with Windham's proportional exponent and the UDF criterion [24], [21] applied to the iris data. The cluster separation measure of Davis and Bouldin [29] failed to uncover the botanically correct number of classes for the iris data, in addition to exhibiting two extra local minima, botanically meaningless. Jain and Moreau [30] developed a method for cluster validity based on the bootstrap technique, that could be used with any clustering algorithm. Using a criterion based on Davis and Bouldin's [29] cluster separation measure, and on cluster compactness measure (within-cluster scatter), both the K-means and Ward clustering algorithms succeeded in detecting the botanically correct number of classes for the iris data set.

Acknowledgment

The authors wish to thank Dr. E. Bar-On for valuable discussions, and to Prof. A. K. Jain for reading the manuscript and providing many useful suggestions. The research was supported by the Kennedy-Leigh fund for Biomedical Engineering Research.

References

[1] L. Larsen, E. Ruspini, J. McDew, D. Walter, and W. Adey, "A test of sleep staging system in the unrestrained chimpanzee," *Brain Res.*, vol. 40, pp. 319-343, 1972.

[2] J. C. Bezdek, "Feature selection for binary data: Medical diagnosis with fuzzy sets," in *Proc. 25th Nat. Computer Conf.*, 1976, pp. 1057-1068.

[3] J. C. Bezdek and W. A. Fordon, "Analysis of hypertensive patients by the use of fuzzy ISODATA algorithm," in *Proc. JACC*, vol. 3, 1978, pp. 349-356.

[4] I. Gath and E. Bar-On, "Computerized method for scoring of polygraphic sleep recordings," *Comput. Progr. Biomed.*, vol. 11, pp. 217-223, 1980.

[5] J. C. Bezdek, "Fuzzy mathematics in pattern classification," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1973.

[6] ——, "A convergence theorem for the fuzzy ISODATA clustering algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 1, pp. 1-8, 1980.

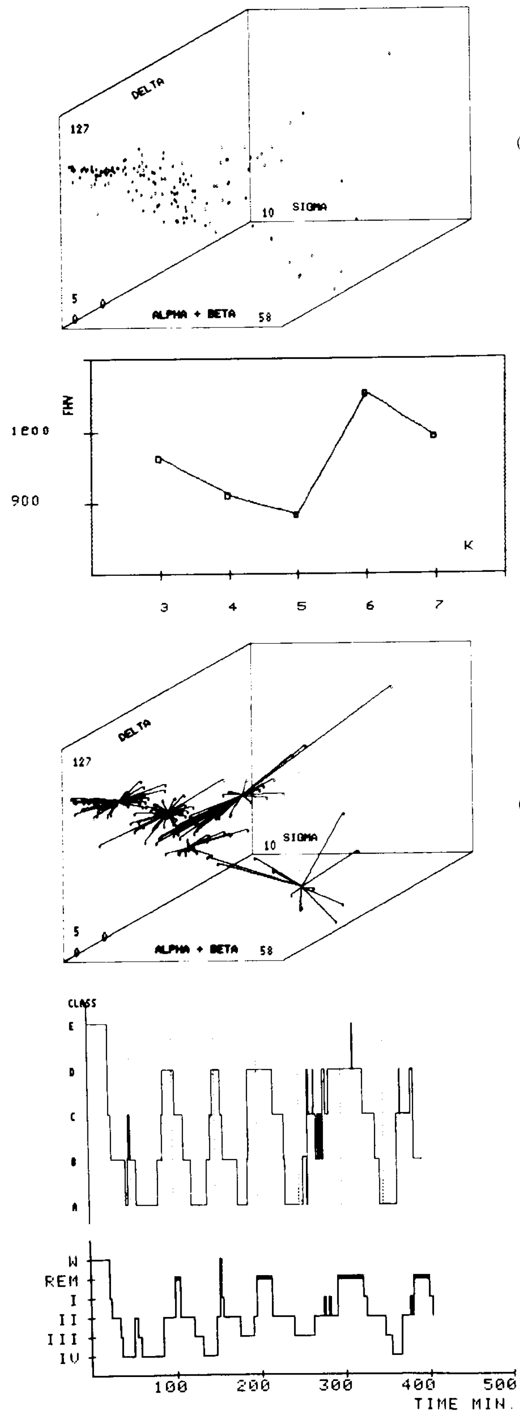
[7] J. C. Bezdek and J. C. Dunn, "Optimal fuzzy partition: A heuristic for estimating the parameters in a mixture of normal distributions," *IEEE Trans. Comput.*, vol. C-24, pp. 835-838, 1975.

[8] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure. I. Linear structure: Fuzzy c-lines," *SIAM J. Appl. Math.*, vol. 40, pp. 339-357, 1981.

[9] J. J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson, "Detection and characterization of cluster substructure. II. Fuzzy c-varieties and convex combinations thereof," *SIAM J. Appl. Math.*, vol. 40, pp. 358-372, 1981.

[10] E. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. IEEE CDC*, San Diego, CA, 1979, pp. 761-766.

[11] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.



(a)

(b)

(c)

(d)

(e)

Fig. 6. Fuzzy classification of sleep EEG segments derived from adaptive segmentation of a whole nights sleep EEG. The five-dimensional feature vectors include the relative power in the physiological frequency bands delta, theta, alpha, beta, and sigma. (a) Data points before partition. (b) Performance measures. Fuzzy hypervolume as a function of k, the number of subgroups in the data. A minimum for $k = 5$ can be seen. (c) Partition using the UFP-ONC algorithm. (d) Classification histogram. (a)–(e) are the various classes. (e) Manual scoring of the same EEG as in (a)–(d) by a physician into sleep stages. W—waking, REM—rapid eye movement sleep, I, II, III, IV—non-REM sleep stages.

[12] A. S. Gevins, "Pattern recognition of human brain electrical potentials," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 5, pp. 383-404, 1980.
[13] B. H. Jansen and W. K. Cheng, "Classification of sleep patterns by means of markov modeling and correspondence analysis," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 5, pp. 707-710, 1987.
[14] I. Gath and E. Bar-On, "Classical sleep stages and the spectral content of the EEG signal," *Int. J. Neurosci.*, vol. 22, pp. 147-155, 1983.
[15] N. E. Day, "Estimating the components of a mixture of normal distributions," *Biometrika*, vol. 56, pp. 463-474, 1969.
[16] E. Backer and A. K. Jain, "A clustering performance measure based on fuzzy set decomposition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, no. 1, pp. 66-74, 1981.
[17] E. Anderson, "The irises of the Gaspe peninsula," *Bull. Amer. Iris Soc.*, vol. 59, pp. 2-5, 1935.
[18] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179-188, 1936.
[19] K. Leszczynski, P. Penczek, and W. Grochulski, "Sugeno's fuzzy measure and fuzzy clustering," *Fuzzy Sets Syst.*, vol. 15, pp. 147-158, 1985.
[20] R. L. Cannon, J. V. Dave, and J. C. Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 2, pp. 248-255, 1986.
[21] M. P. Windham, "Cluster validity for the fuzzy c-means clustering algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-4, no. 4, pp. 357-363, 1982.
[22] I. Gath and E. Bar-On, "Sequential fuzzy clustering of sleep EEG recordings," in *Methods of Sleep Research*, S. Kubicki and W. Herrmann, Eds. Jena, Germany: Gustav Fischer, 1985, pp. 55-64.
[23] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," *J. Math. Biol.*, vol. 1-1, pp. 57-71, 1974.
[24] M. P. Windham, "Cluster validity for fuzzy clustering algorithms," *Fuzzy Sets Syst.*, vol. 3, pp. 1-9, 1980.
[25] T. Gou and B. Dubuisson, "A loose-pattern process approach to clustering fuzzy data sets," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 3, pp. 366-372, 1985.
[26] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *IEEE Trans. Comput.*, vol. C-20, no. 1, pp. 68-86, 1971.
[27] R. Dubes and A. K. Jain, "Validity studies in clustering methodologies," *Pattern Recognition*, vol. 11, pp. 235-254, 1979.
[28] P. H. A. Sneath and R. Sokal, *Numerical Taxonomy*. San Francisco, CA: Freeman, 1973.
[29] D. L. Davis and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, no. 2, pp. 224-227, 1979.
[30] A. K. Jain and J. V. Moreau, "Bootstrap technique in cluster analysis," *Pattern Recognition*, vol. 20, no. 5, pp. 547-568, 1987.