hold in the presence of a spatial illumination gradient," *Atti Ford. Ronchi*, vol. 10, pp. 371–379, 1955.
[5] A. Fiorentini, "Further measurements of the differential threshold in the presence of a spatial illumination gradient," *Atti Ford. Ronchi*, vol. 11, pp. 67–71, 1956.
[6] D. Teller, "The influence of borders on increment-thresholds," Ph.D. dissertation, Dep. Psyc., Univ. California, Berkeley, 1965.
[7] D. Kahneman, "Methods, findings and theory in studies of visual masking," in *Information-Processing Approaches to Visual Perception*, R. N. Haber, Ed. New York: Holt, Reinhart and Winston, 1969, pp. 90–112.
[8] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals based on spatial masking," in *Proc. IEEE*, vol. 65, pp. 536–548, April 1977.
[9] G. L. Anderson and A. N. Netravali, "Image restoration based on a subjective criterion," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-6, pp. 845–853, Dec. 1976.
[10] B. R. Frieden, "Image restoration by discrete deconvolution of minimal length," *J. Opt. Soc. Amer.*, vol. 64, pp. 682–686, 1974.
[11] B. E. A. Saleh, "Trade-off between resolution and noise in restoration by superposition of images," *Appl. Opt.*, vol. 13, pp. 1833–1838, 1974.
[12] M. J. McDonnell, "Nonrecursive image restoration using a finite filter array," *Optik*, vol. 43, pp. 159–174, 1975.
[13] T. E. Reimer and C. D. McGillem, "Constrained optimization of restoration filters," *Appl. Opt.*, vol. 12, pp. 2027–2029, 1973.
[14] T. E. Riemer and C. D. McGillem, "Optimum constrained image restoration filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-13, pp. 136–146, 1977.
[15] A. Papoulis, *Systems and Transforms with Applications in Optics.* New York: McGraw-Hill, 1968.
[16] J. H. McClennan, "The design of two-dimensional digital filters by transformations," in *Proc. 7th Annu. Princeton Conf. Inform. Sci. and Syst.*, Mar. 1973, pp. 247–251.
[17] J. O. Limb, "Vision oriented coding of visual signals," Ph.D. dissertation, Univ. West. Australia, 1966.
[18] Z. L. Budrikis, "Visual fidelity criterion and modeling," in *Proc. IEEE*, pp. 771–779, July 1972.
[19] J. O. Limb, "Picture coding: The use of a viewer model in source encoding," *Bell Syst. Tech. J.*, vol. 52, no. 8, pp. 1271–1302, Oct. 1973.
[20] T. G. Stockham, Jr., "Image processing in the context of a visual model," in *Proc. IEEE*, vol. 60, pp. 828–842, 1972.
[21] T. N. Cornsweet, *Visual Perception.* New York: Academic, 1970.
[22] C. F. Hall and E. L. Hall, "A nonlinear model for the spatial characteristics of the human visual system," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-7, pp. 161–170, Mar. 1977.
[23] W. K. Pratt, *Digital Image Processing.* New York: Wiley-Interscience, 1978.

# Disaggregative Clustering Using the Concept of Mutual Nearest Neighborhood

K. CHIDANANDA GOWDA AND G. KRISHNA

***Abstract*—A nonparametric, hierarchical, disaggregative clustering algorithm is developed using a novel similarity measure, called the mutual neighborhood value (MNV), which takes into account the conventional nearest neighbor ranks of two samples with respect to each other. The algorithm is simple, noniterative, requires low storage, and needs no specification of the expected number of clusters. The algorithm appears very versatile as it is capable of discerning spherical and nonspherical clusters, linearly nonseparable clusters, clusters with unequal populations, and clusters with low-density bridges. Changing of the neighborhood size enables discernment of strong or weak patterns.**

Manuscript received November 15, 1977; revised April 6, 1978 and July 24, 1978. This work was supported in part by a Q.I.P. Fellowship to K. Chidananda Gowda from the Government of India and the Government of Karnataka.

K. Chidananda Gowda is with the Department of Electrical Engineering, Indian Institute of Science, Bangalore, India 560012.

G. Krishna is with the School of Automation, Indian Institute of Science, Bangalore, India 560012.

## I. Introduction

The objective of cluster analysis is to group a set of elements (data units or samples) into clusters such that elements within a cluster have a high degree of similarity, while elements belonging to different clusters have a high degree of dissimilarity. The partitioning of a data set into subsets can be divided into hierarchical and nonhierarchical methods. The general rationale of a nonhierarchical method is to choose some initial partition of the data units and then alter cluster memberships so as to obtain a better partition according to some objective function. In hierarchical clustering methods, the sequence of forming groups proceeds such that whenever two samples belong (or do not belong) to the same cluster at some level, they remain together (or separated) at all higher levels. Hierarchical clustering procedures can be divided into agglomerative methods, which progressively merge the elements, and disaggregative methods, which progressively subdivide the data set.

There exists a variety of hierarchical disaggregative clustering techniques. Williams and Lambert [1] and Lance and Williams [2] consider the monothetic disaggregative methods under the heading "association analysis." In their method the population of individuals, specified by binary attributes are divided with respect to a single attribute, so that, in the two resulting subpopulations, this attribute is possessed by all members of one and lacked by all members of the other. Each of these groups may be further subdivided on any of the remaining attributes, and so on, until some satisfactory configuration is obtained. Crawford and Wishart [3], [4] present a variant on monothetic division based on an interaction measure. Macnaughton-Smith [5] uses dissimilarity analysis for hierarchical subdivision. In this method, a new set is grown by hierarchically transferring samples from the original set. The first sample used to grow a new set is selected such that it has the highest dissimilarity with respect to the rest of the samples of the original set. The second sample to be added to the new set is selected such that it has high similarity with respect to the elements of the new set and high dissimilarity with respect to the remaining elements of the original set. The process of growing the new set is stopped when no sample, which is more similar to the new set than to the rest of the original set, is available. Rose [6] outlines a statistical method to identify cut points and cut sets in a weakly connected graph of the data set. The removal of such points and lines divides a set into subsets. Edwards and Cavalli-Sforza [7] propose a method of examining $(2^{N-1} - 1)$ partitions of a data set containing $N$ elements and choose that division which makes between-cluster sum of squares maximum and within-cluster sum of squares a minimum. Scott and Symons [8] show that the number of partitions to be considered is less than $(2^{N-1} - 1)$. Casetti [9] and Hung and Dubes [10] describe disaggregative clustering approaches based on the use of discriminant analysis. The procedure in this analysis is to begin with an initial partition, compute a linear discriminant function, and then iteratively reassign points and recompute discriminant functions until the readjusted partition becomes the optimal. Mayer [11] describes a variant of the discriminant analysis approach which uses a single dominant variable to specify an initial partition and to impose an ordering on the samples.

Chidananda Gowda and Krishna [12], [13] have introduced the concept of a new similarity measure, called the mutual neighborhood value (MNV), and developed a versatile algorithm for

0018-9472/78/1200-0888$00.75 

agglomerative clustering. The disaggregative clustering method presented in this paper is also based on the concept of mutual neighborhood value. The method presented here is superficially similar in some aspects to the method based on the dissimilarity analysis of Macnaughton-Smith [5], although the underlying rationale is quite different.

Section II covers the philosophy of the approach and gives the definitions of new terms used in this correspondence. Section III presents the computational aspects of the algorithm. Section IV is devoted to the presentation of some typical examples of disaggregative clustering which justify the efficacy of the approach. Finally, in Section V, we give an overall summary.

## II. Concepts and Definitions

Cottam *et al.* [14], [15], Callaghan [16], Hamming and Gilbert [17], and Clark and Evans [18] make some allusions to the idea of mutual neighborhood. In fact, Cottam *et al.* use the concept of "paired neighbors" (which have each other as nearest neighbors) to measure the characteristics of plant communities. But our real motivation for considering the concept of mutual neighborhood comes from real-life observations.

Let us consider two persons, $A$ and $B$. If $A$ feels that $B$ is his closest friend, and $B$ also feels the same, then there exists a feeling of mutual closeness between them, and hence they will group together as close friends. On the other hand, if $A$ feels that $B$ is not such a close friend, then, even if $B$ feels that $A$ is his closest friend, the actual bond of friendship between them becomes weaker. As yet another possibility, if each feels that the other is not his friend at all, then they do not group together as friends. In other words, the strength of the bond of friendship between two persons is a function of mutual feelings rather than one-way feeling. By analogy it can be said that the possibility of clustering of two samples is a function of mutual nearness rather than conventional one-way nearness. This is the main philosophy behind the development of this clustering algorithm. The degree of mutual nearness of two samples is indicated by the mutual neighborhood value between them.

### *Mutual Neighborhood Value*

The mutual neighborhood value between any two samples is the sum of the conventional nearest neighbor ranks of these two samples, with respect to each other.

Let $X_1, X_2, \cdots, X_N$ be a set of $N$ $L$-dimensional vectors called samples, where the $X_i$'s take values in a metric space upon which is defined a metric $d$.

Let $X_k$ be the $m$th nearest neighbor of $X_j$ and $X_j$ be the $n$th nearest neighbor of $X_k$. Then the mutual neighborhood value between $X_j$ and $X_k$ is defined as $(m + n)$. That is, MNV $(X_j, X_k) = (m + n)$, where $m, n \in \{0, 1, 2, \cdots, N - 1\}$.

When $m = 0$ and $n = 0$, it is to be understood that each point is its own zeroth neighbor. Therefore,

$$\text{MNV}(X_j, X_k) \begin{cases} \in \{2, 3, 4, \cdots, 2N-2\}, & j \neq k \\ = 0, & j = k. \end{cases}$$

The MNV is a semimetric and satisfies the first two conditions of a metric:

1) MNV $(X_j, X_k) \geq 0$, and MNV $(X_j, X_k) = 0 \Leftrightarrow X_j = X_k$,
2) MNV $(X_j, X_k) =$ MNV $(X_k, X_j)$.

### *An Example*

We give a simple example to elucidate the concept of mutual neighborhood value. Consider a data set $\{A, B, C, D, E, F, G, H\}$ as shown in Fig. 1. Let the five nearest neighbors, according to Euclidean distance measure, of each sample point be as shown in Table I. In this neighborhood table, the first element of each row (written below sample) is the sample under consideration for which the five nearest neighbors are found. The remaining entries in that row are the conventional nearest neighbors (NN) of the first entry. For example, in row 1, the entries $B, C, D, F$, and $E$ are the first, second, third, fourth, and fifth nearest neighbors, respectively, of the first entry $A$. The number within the parenthesis under each sample indicates the mutual neighborhood value between that sample and the first sample in that row.

$B$ is the first NN of $A$ according to the first row, and $A$ is the third NN of $B$ according to the second row. Therefore, MNV between $A$ and $B$ is $(1 + 3)$ or 4. Hence, 4 is written below $B$ in row 1 and below $A$ in row 2. Similarly, $C$ is the second NN of $A$ according to row 1, and $A$ is the first NN of $C$ according to row 3. Therefore, the MNV between $A$ and $C$ is 3, and hence 3 is written below $C$ in row 1 and below $A$ in row 3. It is quite possible that, for the neighborhood width $k$ (here $k = 5$) considered, two samples may not be found in each other's neighborhood row. For example, according to row 7, $C$ is the first NN of $G$. But according to row 3, $G$ is not among the five NN's of $C$. Therefore, $G$ and $C$ are not mutual neighbors for the neighborhood width $k = 5$. So, some arbitrary large number (greater than $2k$), let us say 100, is written below $C$ in row 7.

TABLE I
Near Neighbors and Mutual Neighborhood Values

| Sample | First NN | Second NN | Third NN | Fourth NN | Fifth NN |
|---|---|---|---|---|---|
| A (0) | B (4) | C (3) | D (6) | F (6) | E (8) |
| B (0) | D (3) | E (4) | A (4) | C (7) | F (8) |
| C (0) | A (3) | F (3) | B (7) | D (8) | E (10) |
| D (0) | E (2) | B (3) | A (6) | C (8) | F (9) |
| E (0) | D (2) | B (4) | A (8) | F (9) | C (10) |
| F (0) | C (3) | A (6) | B (8) | D (9) | E (9) |
| G (0) | C (100) | H (3) | A (100) | F (100) | B (100) |
| H (0) | G (3) | C (100) | A (100) | B (100) | D (100) |

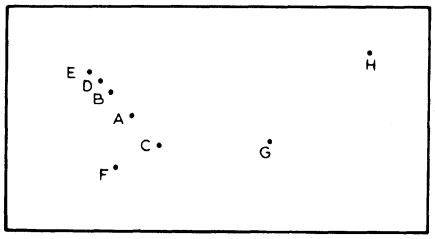


Fig. 1. Eight sample points.

## III. Computational Aspects

### *Algorithm*

The clustering algorithm proceeds as follows.

*Step 1:* For each sample in the data set $\{X_1, X_2, \cdots, X_N\}$, find the $k$ nearest neighbors using the Euclidean distance measure, and form an integer matrix $M_1$ with $N$ rows and $(k+1)$ columns. We have observed that a value of $k = 5$ is appropriate for discerning moderately strong clusters. In such a neighborhood matrix, the first entry in each row indicates the sample under consideration, the second entry indicates the first nearest neighbor, the third entry indicates the second nearest neighbor, and so on until the $(k+1)$th entry indicates the $k$th nearest neighbor.

*Step 2:* Set up an integer matrix $M_2$ with $N$ rows and $(k+1)$ columns, where an entry in the $i$th row and $j$th column is the MNV between samples at $(i, 1)$ and $(i, j)$ in matrix $M_1$. If two samples are not mutual neighbors for a given neighborhood width $k$, write their MNV as an arbitrary number greater than $2k$.

*Step 3:* Let $A$ initially represent the whole sample set and $B$ the null set. Let $T$ $(T \leq 2k)$ be the threshold of the MNV chosen for carrying out the first stage of the disaggregative process. Strong or weak clusters will be obtained depending on the value of $T$ chosen. Select $X \in A$ randomly, and make $B = B + X$ and $A = A - X$. Find all the mutual nearest neighbors (with MNV $\leq T$) of $X$ using the matrices $M_1$ and $M_2$. Add them to $B$, and delete them from $A$. Now consider the second element in $B$, and find all its mutual nearest neighbors (with MNV $\leq T$) using $M_1$ and $M_2$. Out of them assign all those that are not already present in $B$ to $B$, and delete them from $A$. This procedure is repeated for the third and subsequent elements in $B$. When no element in $B$ presents itself for consideration, the first bifurcation is complete, and the set $B$ now represents the first cluster that has been sifted from the whole sample set.

The set $A$, presently, comprises samples that will be considered for further division. So $B$ is made the null set $\varnothing$, a new sample $X$ belonging to the present set $A$ is randomly selected, and a new set $B$ is grown as before to obtain the second cluster. The first stage of the procedure terminates when $A$ becomes the null set.

*Step 4:* If we are interested in still stronger clusters, we can embark on the subdivision of each of the groups already sifted. This can be accomplished by choosing a lower value of $T$ and repeating Step 3 with the assumption that the set $A$ now consists of the group that is to be subdivided in this second stage. The procedure is not iterative as the matrices $M_1$ and $M_2$ are formed only once, and the clusters are sifted using these two matrices.

The entire algorithm can be written in a succinct form as follows.

1) Set up matrices $M_1$ and $M_2$. Choose an MNV threshold $T$ $(T \leq 2k)$. Let $A$ represent the whole sample set.
2) Make $B = \varnothing$, where $\varnothing$ is the null set.
3) Let $L = 1$.
4) Choose $X \in A$ randomly, COUNT = 1.
5) $A = A - X$, $B = B + X$.
6) Let $Y$ be the $L$th element of $B$.
7) Using matrices $M_1$ and $M_2$, find $J$ $(J \leq k)$ numbers of mutual neighbors (with MNV $\leq T$), $\{X_1, X_2, \cdots, X_J\}$, of $Y$ where $\{X_1, X_2, \cdots, X_J\} \notin B$.
8) $A = A - \{X_1, X_2, \cdots, X_J\}$, $B = B + \{X_1, X_2, \cdots, X_J\}$.
9) COUNT = COUNT + $J$.
10) $L = L + 1$.
11) If $L$ is greater than COUNT, then go to 12), else go to 6).
12) Store the samples in $B$ as the elements of one cluster.
13) If $A = \varnothing$, then go to 14), else go to 2).
14) If still stronger clusters are desired, then reduce $T$, take the group to be further subdivided as $A$, and go to 2); else END of algorithm.

### *Memory Requirement*

Matrices $M_1$ and $M_2$ are sufficient to carry out the divisive process, and no distances need be stored. $M_1$ and $M_2$ together require a memory of $(2k + 2)N$, and other bookkeeping arrays require a memory of about $6N$. So the major memory requirement for the implementation of this algorithm is of the order of $(2k + 8)N$. A neighborhood width of $k = 5$ with an MNV threshold of $T = 10$ is found to be quite adequate for discerning moderately strong clusters. For $k = 5$, for example, the major memory requirement is about $18N$.

### *Discerning Stronger Clusters*

If the investigator feels that the selection of an MNV threshold $T = 10$ has not yielded sufficiently strong clusters, then the existing clusters can be further subdivided to obtain stronger clusters by using a smaller value of $T$ for the divisive process.

### *Discerning Weaker Clusters*

If the matrices $M_1$ and $M_2$ are generated taking $k = 5$, then the maximum value of $T$ that can be selected for the divisive process is 10. Although $k = 5$ is sufficient for discerning moderately strong clusters, it is preferable to generate $M_1$ and $M_2$ taking a higher value of $k$. This facilitates the selection of an MNV threshold higher than 10 for discerning comparatively weaker clusters without the need to start all over again. But a price has to be paid in terms of storage requirements.

## IV. Examples

Nagy [19] and Zahn [20] have indicated typical cluster problems. According to Nagy, the major difficulties in cluster analysis are

1) unequal cluster populations,
2) bridges between clusters,
3) nonspherical clusters,
4) linearly nonseparable clusters.

Keeping this in mind, several data sets were generated manually. To bring out the efficacy of the proposed algorithm, a number of divisive clustering experiments were performed using this data. The results of the simulation done on the IBM 360/44 computer are interesting.

The expected number of clusters was assumed to be unknown for the analysis of the data sets. The application of our algorithm, for a neighborhood width of $k = 5$ and $T = 2k$, yielded the number of clusters and members belonging to each cluster.

The first example consists of a data set with tight as well as loose groups having substantially unequal cluster populations (2, 3, 5, 10, 15, 15, 25, and 50). The samples marked "1" in Fig. 2(a) represent the members of the first cluster resulting from the first division. The samples marked as "0" represent the remaining samples to be clustered in subsequent divisions. The members belonging to the second cluster, resulting from the second division, are represented by "2" in Fig. 2(b). Fig. 2(c) shows all eight clusters after the termination of the divisive process.

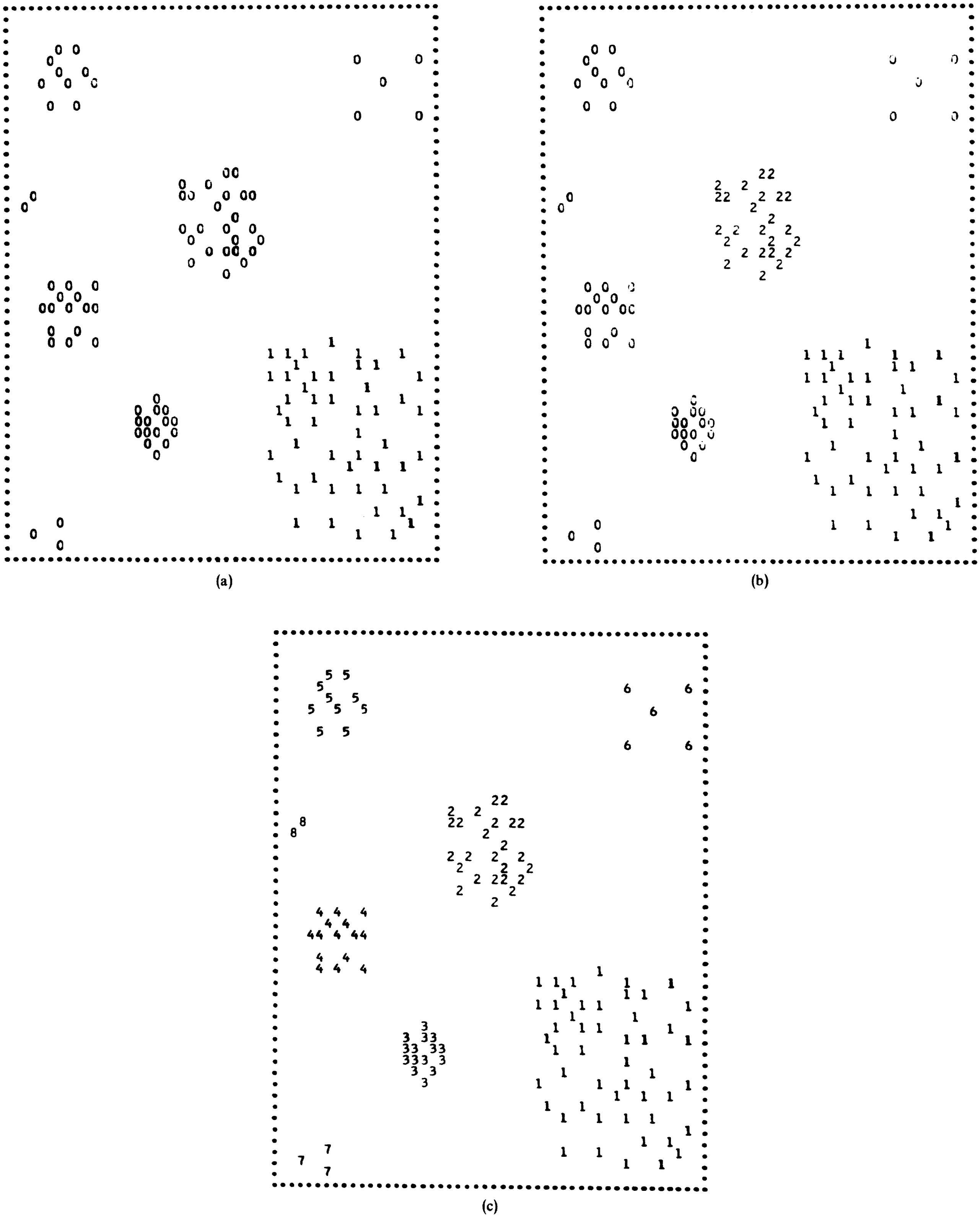


Fig. 2. Clusters with unequal population. (a) After first division. (b) After second division. (c) After final division.

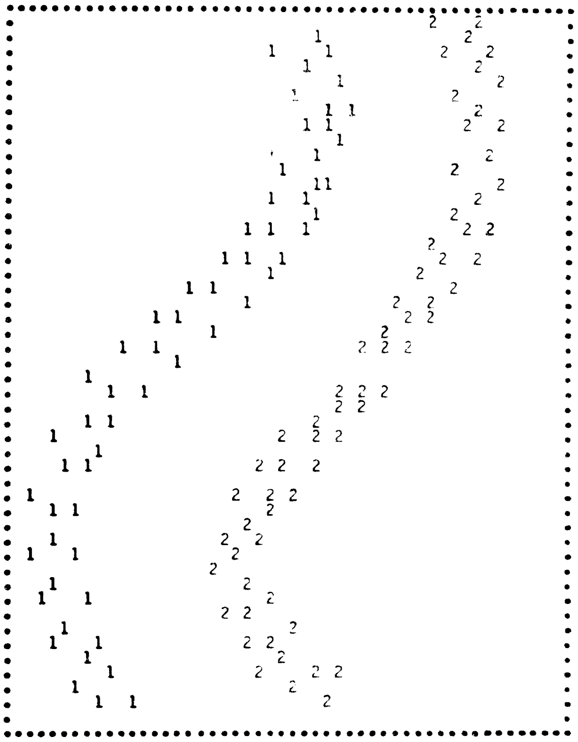

Fig. 3. Chain-like clusters.

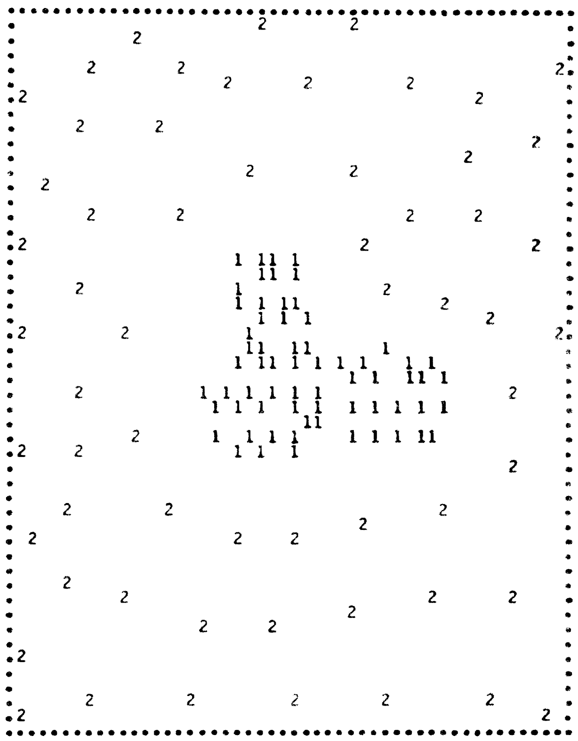

Fig. 5. Groups with unequal cluster densities.

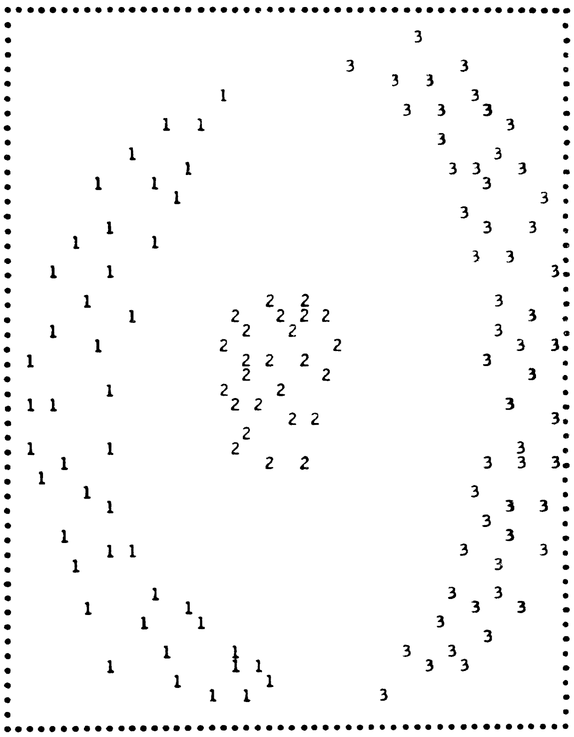

Fig. 4. One spherical and two nonspherical groups.

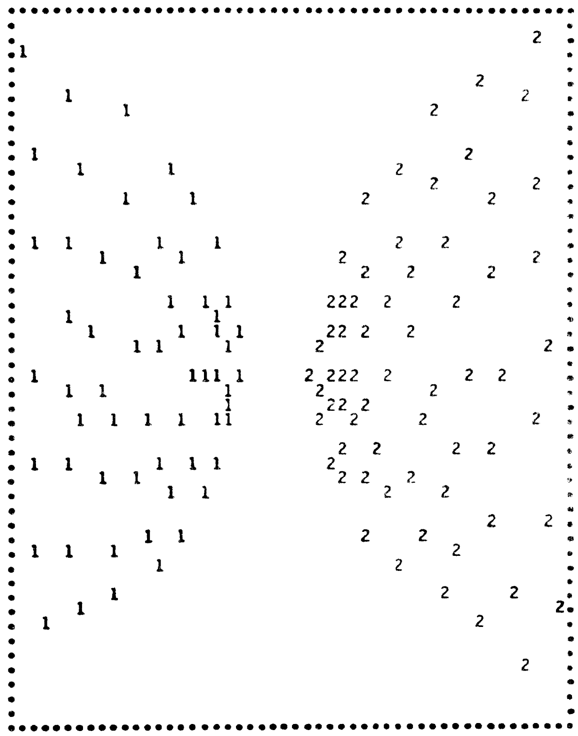

Fig. 6. Groups with nonhomogeneous cluster densities.

The clustering results for the second example are shown in Fig. 3, where the data set consists of two nonspherical chain-like groups.

The data set considered in the third example consists of one spherical group flanked by two nonspherical groups. The results are shown in Fig. 4.

The data set of the fourth example consists of an irregularly shaped high-density group surrounded by a population of low density. The results are shown in Fig. 5.

In the fifth example, the data set consists of groups with smoothly varying nonhomogeneous cluster densities with high densities near where the clusters approach each other. The results are shown in Fig. 6.

The data set of the sixth example consists of two distinct groups with a low-density bridge between them. The results are shown in Fig. 7, where the bridge appears as a separate cluster with two members.

The seventh example consists of a data set which is almost similar to that of example 6. But here the density of the bridge is high. The results are shown in Fig. 8. It is evident that the two distinct groups merge into a single cluster because of the high-density bridge.

The eighth example illustrates the discerning of strong or weak clusters by generating the matrices $M_1$ and $M_2$ taking a higher neighborhood width. Fig. 9(a) depicts the clustering results for an MNV threshold of $T = 10$. The data set now appears as consisting of three moderately strong clusters. If the disaggregative process is carried out using $T = 16$, then two clusters (one moderately strong and the other weak) are obtained as shown in Fig. 9(b). If $T$ is increased to 28, the data set appears as consisting of single weak cluster. This is shown in Fig. 9(c).

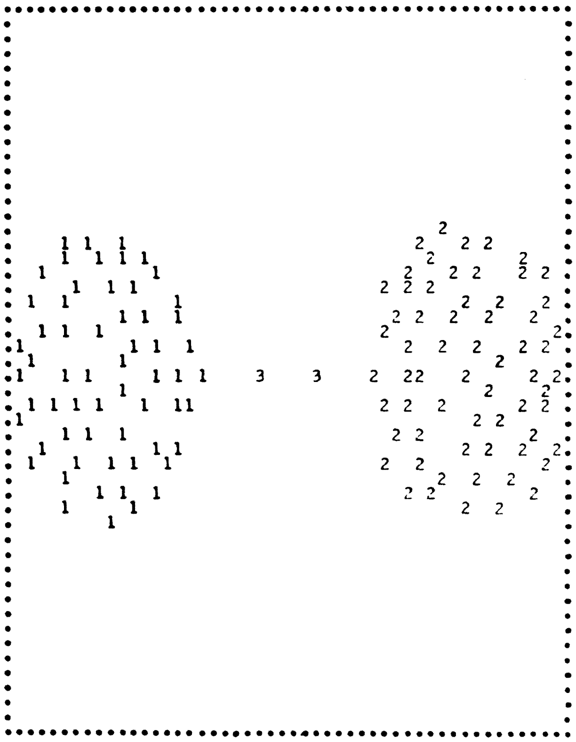

Fig. 7. Low-density bridge between two spherical clusters.

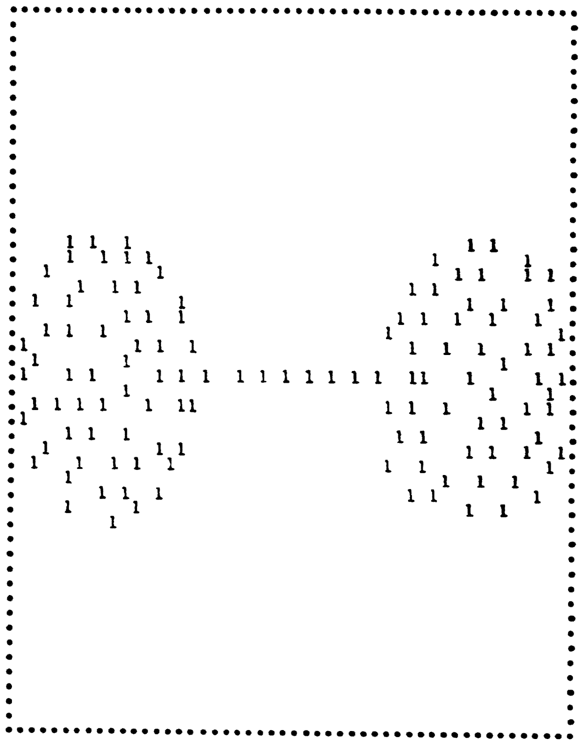

Fig. 8. High-density bridge between two spherical clusters.

## CONCLUSIONS

A new similarity measure, called the mutual neighborhood value, is defined. The MNV between any two samples is the sum of the conventional nearest neighbor ranks of these two samples, with respect to each other. A nonparametric, hierarchical, disaggregative clustering algorithm, based on the MNV, is developed. It is found that the algorithm is simple, noniterative, has low storage requirements, and is suitable for large sample size. The expected number of clusters need not be specified beforehand, and the clustering results are independent of the order in which the samples are processed. Strong or weak clusters can be discerned by changing the neighborhood width $k$. The application of the algorithm to a number of clustering problems has evinced its efficacy to discern difficult cases such as 1) nonspherical clusters, 2) linearly nonseparable clusters, 3) clusters with unequal populations, and 4) clusters with low-density bridges between them.

The MNV is better than the ordinary distance for the similarity measure because the former invokes two-way nearness and hence possesses the capacity to avoid the possibility of clustering a sparse group with a dense group. In a situation such as that shown in Fig. 5, where there are sparse and dense groups, the proposed method appears simpler than Zahn's method.

## ACKNOWLEDGMENT

The authors wish to thank Prof. K. S. Prabhu, Chairman of the Department of Electrical Engineering, Indian Institute of Science, Bangalore, for providing necessary facilities and giving encouragement from time to time.

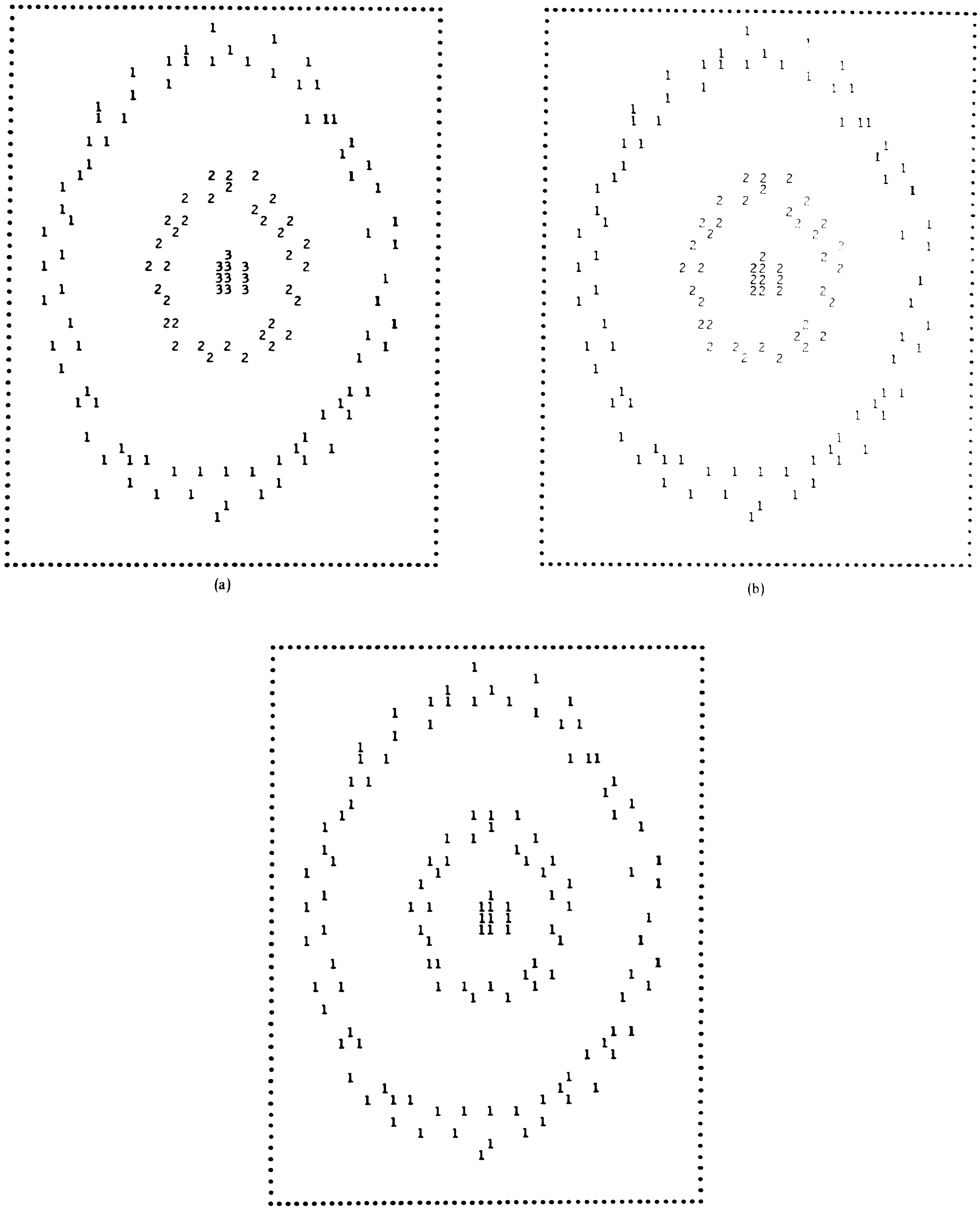

Fig. 9. Concentric clusters. (a) For $T = 10$. (b) For $T = 16$. (c) For $T = 28$.

## References

[1] W. T. Williams and J. M. Lambert, "Multivariate methods in plant ecology: I. Association analysis in plant communities," *J. Ecol.*, vol. 47, no. 1, pp. 83–101, 1959.

[2] G. N. Lance and W. T. Williams, "Computer program for monothetic classification (association analysis)," *Comput. J.*, vol. 8, no. 3, pp. 246–249, 1965.

[3] R. M. M. Crawford and D. Wishart, "A rapid multivariate method for the detection and classification of groups of ecologically related species," *J. Ecol.*, vol. 55, no. 2, pp. 505–524, 1967.

[4] R. M. M. Crawford and D. Wishart, "A rapid classification and ordination method and its application to vegetation mapping," *J. Ecol.*, vol. 56, no. 2, pp. 385–404, 1968.

[5] P. Macnaughton-Smith, "Dissimilarity analysis: A new technique of hierarchical subdivision," *Nature*, vol. 202, pp. 1034–1035, June 1964.

[6] M. J. Rose, "Classification of a set of elements," *Comput. J.*, vol. 7, pp. 208, 1964.

[7] A. W. F. Edwards and L. L. Cavalli-Sforza, "A method for cluster analysis," *Biometrics*, vol. 21, no. 2, pp. 362–375, 1965.

[8] A. J. Scott and M. J. Symons, "On the Edwards and Cavalli-Sforza method of cluster analysis," *Biometrics*, vol. 27, no. 1, pp. 217–219, 1971.

[9] E. Casetti, "Classificatory and regional analysis by discriminant iterations," TR-12, Contract-1228 (26), AD 608093, Northwestern Univ., Evanston, IL.

[10] A. Y. Hung and R. C. Dubes, "An introduction to multiclass pattern recognition in unstructural situations," Interim Sci. Rep. No. 12, AD 720812, Div. of Eng. Res., Michigan State Univ., East Lansing, MI.

[11] L. S. Mayer, "A method of cluster analysis when there exist multiple indicators of a theoretical concept," *Biometrics*, vol. 27, no. 1, pp. 143–155, 1971.

[12] K. Chidananda Gowda and G. Krishna, "Nonparametric clustering using the concept of mutual nearest neighborhood," Rep. No. EE/43, Dep. Electrical Engineering, Indian Institute of Science, Bangalore, India, Apr. 1977.

[13] ——, "Agglomerative clustering using the concept of mutual nearest neighborhood," to be published in *Pattern Recognition*, vol. 2, 1978.

[14] G. Cottam, J. J. Curtis, and B. W. Hall, "Some sampling characteristics of a population of randomly dispersed individuals," *Ecology*, vol. 34, pp. 741, 1953.

[15] G. Cottam and J. T. Curtis, "The use of distance measures in physiosociological sampling," *Ecology*, vol. 37, pp. 451, 1956.

[16] J. F. Callaghan, "An alternative definition for neighborhood of a point," *IEEE Trans. Comput.*, vol. C-24, pp. 1121, 1975.

[17] R. Hamming and E. Gilbert, "Probability of occurrance of a constant number of isolated pairs in a random population," *Univ. Wisconsin Comput. News*, no. 32, 1954.

[18] P. J. Clark and F. C. Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, pp. 445–453, 1954.

[19] G. Nagy, "State of the art of pattern recognition," in *Proc. IEEE*, vol. 56, pp. 836, 1968.

[20] C. T. Zahn, "Graph theoretical methods for detecting and describing Gestalt clusters," *IEEE Trans. on Comput.*, vol. C-20, no. 1, pp. 68, 1971.

[21] M. R. Anderberg, *Cluster Analysis for Applications*. New York and London: Academic, 1973.

[22] P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*. San Francisco: W. H. Freeman, 1973.


# A Note on the Use of Second-Order Gray-Level Statistics for Threshold Selection

NARENDRA AHUJA AND AZRIEL ROSENFELD, FELLOW, IEEE

***Abstract*—The gray-level histogram of an image is commonly used as an aid in selecting thresholds for segmenting the image. Various methods have been proposed for transforming the histogram so as to make threshold selection easier. This note describes a class of such methods that makes use of second-order gray-level statistics to define improved histograms.**

## I. Introduction

If an image contains dark objects on a light background or vice versa, its histogram (a plot of how often each gray level occurs in the image) may contain two peaks, representing the populations of object and background points, separated by a valley corresponding to the intermediate gray levels (which occur less frequently, primarily on object/background borders). If we threshold the image at a gray level near the bottom of this valley, the objects and background should be well separated. However, the valley bottom is sometimes difficult to locate.

Several methods have been proposed for transforming the histogram so that the valley is deepened, or is converted into a peak, with the result that good thresholds may become easier to select. These methods generally make use of edge value (e.g., gray-level gradient magnitude) in conjunction with the gray level itself. For example, if we histogram the gray levels only for those image points that have low edge value, we should obtain a deeper valley, since we are (hopefully) suppressing points that lie on object/background borders, and keeping primarily points that lie in the interiors. On the other hand, if we make a histogram of the gray levels of points that have high edge value, the valley should turn into a peak, since these points should primarily lie on borders. A number of methods of this type are reviewed in [1].

This note describes an alternative approach to histogram transformation, based on second-order gray-level statistics rather than on edge values. In tests on the same images used in [1], this approach gave somewhat better results.

It should be pointed out that the methods described here do not actually select a threshold, but only produce an improved histogram. In [2], a method of selecting thresholds to minimize the "busyness" of the thresholded image (which can be computed from the original image's second-order gray-level statistics) is described.

The gray-level cooccurrence matrices (or gray-tone spatial dependency matrices) used here and in [2] have been used by a number of investigators as a basis for computing textural properties of images; see, for example, [3].

## II. Approach

Given an image $I$ and a displacement vector $\delta \equiv (\Delta x, \Delta y)$, we define the *gray-level cooccurrence matrix* corresponding to $\delta$ as the matrix $M_\delta$ whose $(i, j)$ element is the number of times that a point having gray level $j$ occurs in $I$ in position $\delta$ relative to a point having gray level $i$. In particular, let

$$M = M_{(0,1)} + M_{(1,0)} + M_{(0,-1)} + M_{(-1,0)};$$

thus the $(i, j)$ element of $M$ is the number of times that gray level $j$ occurs as a neighbor of gray level $i$ in $I$. Note that $M$ is symmetric, since if gray levels $(i, j)$ occur in relative position $(1, 0)$, then levels $(j, i)$ occur in relative position $(-1, 0)$, and similarly for $(0, 1)$ and $(0, -1)$.

Distance from the main diagonal in $M$ corresponds to absolute gray-level difference; on the diagonal we have $i = j$, or $|i - j| = 0$, while in the upper right (and lower left) corner, $|i - j|$ is maximal. If the edge value at a point $P$ is low, $P$ cannot differ much in gray level from any of its neighbors, so that the (point, neighbor) pairs involving $P$ contribute to near-diagonal elements of $M$. Conversely, if $P$ has high edge value, at least some of its pairs contribute to elements of $M$ that lie far from the diagonal.

These remarks suggest that we can use $M$ to define histogram transformations analogous to those based on edge values that were described in Section I. In particular, if we can define a modified histogram based only on the points that contribute to near-diagonal elements of $M$, then this histogram should pri-

Manuscript received June 29, 1978; revised August 30, 1978. This work was supported by the Directorate of Mathematical and Information Sciences, U.S. Air Force Office of Scientific Research, under Grant AFOSR-77-3271.

The authors are with the Computer Science Center, University of Maryland, College Park, MD 20742.