Short Papers

Location- and Density-Based Hierarchical Clustering Using Similarity Analysis

Peter Bajcsy and Narendra Ahuja, *Fellow, IEEE*

Abstract—This paper presents a new approach to hierarchical clustering of point patterns. Two algorithms for hierarchical location- and density-based clustering are developed. Each method groups points such that maximum intracluster similarity and intercluster dissimilarity are achieved for point locations or point separations. Performance of the clustering methods is compared with four other methods. The approach is applied to a two-step texture analysis, where points represent centroid and average color of the regions in image segmentation.

Index Terms—Point patterns, clustering, hierarchy of clusters, spatially interleaved clusters, density-based clustering, location-based clustering.

1 INTRODUCTION

CLUSTERING explores the inherent tendency of a point pattern to form sets of points (clusters) in multidimensional space. Most of the previous clustering methods assume tacitly that points having similar locations or constant density create a single cluster (location- or density-based clustering). Location or density becomes a characteristic property of a cluster. Other properties of clusters are proposed based on human perception [1] or specific tasks (e.g, shape from texture [2]). The properties of clusters have to be specified before the clustering is performed and are usually a priori unknown.

This work presents a new approach to hierarchical clustering of point patterns. Two hierarchical, location- and density-based clustering algorithms are developed based on the new approach using similarity analysis. The similarity analysis relates intracluster dissimilarity with intercluster dissimilarity. The dissimilarity of point locations or point separations is considered for clustering and is denoted in general as a dissimilarity of elements $\mathbf{e_i}$. Each method can be described as follows.

1) Every element $\mathbf{e_i}$ gives rise to one cluster $C_{\mathbf{e_i}}$ having elements dissimilar to $\mathbf{e_i}$ by no more than a fixed number θ.
2) A sample mean $\overline{u}_{\mathbf{e_i}}$ of all elements in $C_{\mathbf{e_i}}$ is calculated.
3) Clusters would be formed by grouping pairs of elements if the sample means computed at the two elements are similar.
4) The degree of dissimilarity θ is used to form several (multiscale) partitions of a given point pattern. A hierarchical organization of clusters within multiscale partitions is built by agglomerating clusters for increasing degree of dissimilarity.
5) The clusters that do not change for a large interval of θ are selected into the final partition.

Experimental evaluation is conducted for synthetic point patterns, standard point patterns (80x-handwritten character recognition, IRIS-flower recognition) and point patterns obtained from image texture analysis. Performance of the clustering methods is compared with four other methods (partitional—FORGY, CLUSTER; hierarchical—single link, complete link [3]).

In the previous work on clustering [3], [4], all methods are divided into partitional and hierarchical methods. Partitional methods create a single partition of points while hierarchical methods give rise to several partitions of points that are nested. Partitional clustering methods can be subdivided roughly into

1) error-square clusterings [3], [5],
2) clustering by graph theory [6], [3], [7], and
3) density estimation clusterings [8], [3], [5].

Error-square clusterings minimize the square error for a fixed number of clusters. These methods require inputting the number of sought clusters as well as the seeds for initial cluster centroids. Clustering by graph theory uses geometric structures such as, minimum spanning tree (MST), relative neighborhood graph, mutual nearest neighborhood, Gabriel graph, and Delaunay triangulation. The methods using these geometric structures construct the graph first, followed by removal of inconsistent edges of the graph. Inconsistent edges and how to remove edges are specified for each method. The two most commonly used hierarchical clusterings are single-link and complete-link methods [3], [9]. The use of clustering methods can be found in many applications related to remote sensing [10], [11], image texture detection [12], [13], taxonomy [9], [14], geography [15], [16], and so on.

Proposed location-based clustering can be related to centroid clustering [3], and density-based clustering can be related to Zahn's method [1]. Centroid clustering achieves results identical to the proposed location-based clustering although the algorithms are different (see [3]). The only difference in performance is in the case of equidistant points, when the proposed method gives a unique solution, while the centroid-clustering method does not, due to sequential merging and updating of point coordinates. Zahn's method consists of the followings steps:

1) Construct the MST for a given point pattern.
2) Identify inconsistent links in the MST.
3) Remove inconsistent links to form connected components (clusters).

A link is called inconsistent if the link distance is significantly larger than the average of nearby link distances on both sides of the link. The proposed density-based clustering differs from Zahn's clustering in the following ways:

1) We use the average of the largest set of link distances rather than nearby link distances for defining inconsistent link, and this leads to more accurate estimates of inconsistent links.
2) We replace the threshold for removing inconsistent links ("significantly larger" in the definition of inconsistent links) with a simple statistical rule.
3) We work with all links from a complete graph[1] rather than a few links selected by MST (this is crucial for detecting spatially interleaved clusters).

Location- and density-based clusterings are suitable for texture analysis. A texture is modeled as a set of uniformly distributed identical primitives (see Fig. 1). A primitive is described by a set of photometric, geometrical, or topological features (e.g., color or shape). Spatial coordinates of a primitive are described by another set of features. Thus, the point pattern obtained from

• P. Bajcsy is with the Cognex Corp., Acumen Product Group, 7352 SW Durham Rd., Portland, OR 97062. E-mail: pbajcsy@cognex.com.
• N. Ahuja is with the University of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801.
E-mail: ahuja@vision.ai.uiuc.edu.

Manuscript received 24 Apr. 1997; revised 8 July 1998. Recommended for acceptance by G. Medioni.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107134.

1. Links between all pairs of points create a complete graph according to the notation in graph theory.

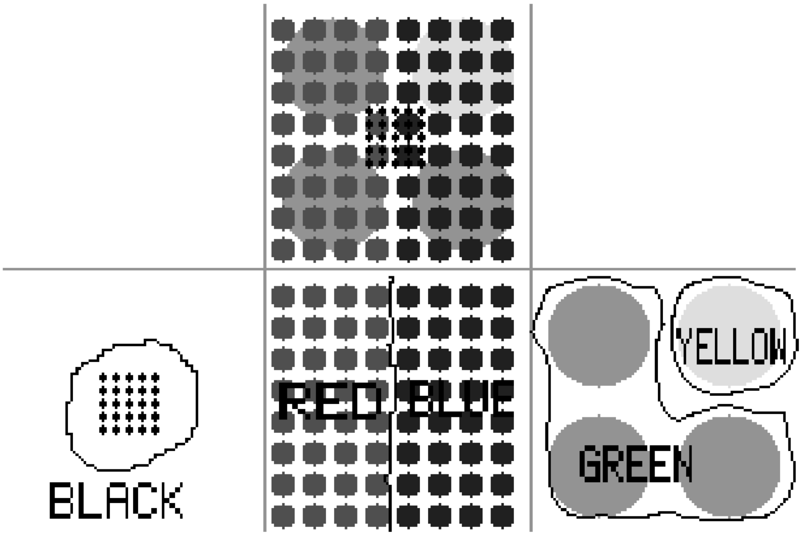


Fig. 1. Example of textures. Top: Original image. Bottom: The resulting five textures (delineated by black line) obtained by location- and density-based clusterings.

texture analysis consists of two sets of features (e.g., centroid location and average color of primitives) and has to be decomposed first. Location-based clustering is used to form clusters corresponding to identical primitives in one subspace (color of primitives). Density-based clustering creates clusters corresponding to uniformly distributed primitives in the other subspace (centroid locations of primitives). The resulting texture is identified by combining clustering results in the two subspaces. This decomposition approach is also demonstrated on point patterns obtained in other application domains. In general, it is unknown how to determine the choice of subspaces. Thus, an exhaustive search for the best division in terms of classification error is used in the experimental part for handwritten character recognition and taxonomy applications.

The objective of this work is to contribute to

1) the theoretical development of nonexisting clustering methods and
2) the use of clustering for texture detection.

The salient features of this work are the following:

1) A decomposition of the clustering problem into two lower-dimensional problems is addressed.
2) A new clustering approach is proposed for detecting clusters having any constant property of points (location or density).
3) A density-based clustering method using the proposed approach separates spatially interleaved clusters having various densities, thus is unique among all existing clustering methods.

The methods can be related to the graph-theoretical algorithms.

This paper is organized as follows. Theoretical development of the proposed clustering methods is presented in Section 2. Experimental performance evaluations of the clustering methods follow in Section 3. Section 4 presents concluding remarks.

2 Location- and Density-Based Clusterings

2.1 Mathematical Formulation

An n-dimensional point pattern is defined as a set of points $I = \{\mathbf{p_i}\}_{i=1}^{P}$ with coordinates $(p_1, p_2, \cdots, p_n)$. A general goal of unsupervised clustering is to partition a set of points I into nonoverlapping subsets of points $\{C_j\}_{j=1}^{N}$; $I = \cup_{j=1}^{N} C_j$, $C_k \cap C_j = 0$ and $C_j = \{\mathbf{p_i}\}_{i \in W_j}$, where W_j is an index set from all integer numbers in the interval $[1, P]$. The subsets of points are called clusters and are characterized in this work by the similarity (dissimilarity) of point locations or point separations. A notion of an element $\mathbf{e_i} \in \Re^n$ (a feature vector) is introduced to refer either to a point location $\mathbf{p_i}$ or a point separation $d\left(l_{\mathbf{p_1},\mathbf{p_2}}\right) = \left\|\mathbf{p_1} - \mathbf{p_2}\right\|$ (the Euclidean distance between two points also called length of a link $d\left(l_{\mathbf{p_1},\mathbf{p_2}}\right)$).

In general, every cluster of elements $C_j = \{\mathbf{e_k}\}$ can be characterized by its maximum intracluster dissimilarity $\theta = \max \{\| (\mathbf{e_i} \in C_j) - (\mathbf{e_k} \in C_j) \|\}$ (or minimum intracluster similarity θ) and minimum intercluster dissimilarity $\alpha = \min \{\|(\mathbf{e_i} \notin C_j) - (\mathbf{e_k} \in C_j)\|\}$ (maximum intercluster similarity α), where the dissimilarity (similarity) value of any two elements $\mathbf{e_i}$, $\mathbf{e_k}$ is defined as the Euclidean distance $\theta_{\mathbf{e_i},\mathbf{e_k}} = \left\|\mathbf{e_i} - \mathbf{e_k}\right\|$. Our goal is to partition a point pattern I into nonoverlapping clusters $\{C_j\}_{j=1}^{N}$ having a minimum intracluster and maximum intercluster dissimilarity of elements in order to decrease the probability of misclassification.

If clusters of elements are not clusters of points as in the case point separations, then a mapping from the clusters obtained to clusters of points is performed. The mapping from clusters of point separations (links) to clusters of points takes two steps:

1) Construct an MST from the average values of individual clusters of links.
2) Form clusters of points sequentially from clusters of links in the order given by the MST (from smaller to larger average values of clusters).

2.2 The Clustering Method

Given a set of elements I and the goal, the unknown parameters of the classification problem are the values θ and α for each final cluster, as well as the number of final clusters N. Two steps are taken to partition the input elements into clusters. First, a value of intracluster dissimilarity θ is fixed, and clusters characterized by θ are formed by grouping pairs of elements. The result of the first step is a set of clusters denoted as $\{CE_m^{\theta}\}_{m=1}^{M_\theta}$ since they are characterized only by θ. Second, values of θ and α are estimated for every final cluster $CE_j^{\theta,\alpha}$ by a maximization of intercluster dissimilarity α and a minimization of intracluster dissimilarity θ.

The final partition $\left\{CE_j^{\theta,\alpha}\right\}_{j=1}^{NE}$ is a subset of $\left\{CE_m^{\theta}\right\}_{m=1}^{M_\theta}$ and is aimed to be identical with a ground truth partition $\left\{C_j\right\}_{j=1}^{N}$, which is assumed to exist for the purpose of evaluating the classification accuracy (number of misclassified elements).

Step 1: Grouping pairs of elements into clusters $\left\{CE_m^{\theta}\right\}_{m=1}^{M_\theta}$

Let us assume that there exists C_j in the ground truth partition with intraclustering dissimilarity θ. The grouping of elements $\mathbf{e_i}$ in C_j is based on a certain estimate of the cluster C_j computed at each element $\mathbf{e_i}$. The best estimate of an unknown cluster C_j is obtained at a single element $\mathbf{e_i}$ if the element $\mathbf{e_i}$ gives rise to a cluster $C_{\mathbf{e_i}}$ identical with the unknown one C_j. It would be possible to create the cluster $C_{\mathbf{e_i}} = C_j$ if the unknown cluster C_j of elements is characterized by a value of intercluster dissimilarity α larger than a value of intracluster dissimilarity θ. Under the assumption $\alpha > \theta$, the cluster $C_{\mathbf{e_i}} = \left\{\mathbf{e_k}\right\}$ is obtained from any element $\mathbf{e_i} \in C_j$ by grouping together all other elements $\mathbf{e_k}$ satisfying the inequality $\|\mathbf{e_i} - \mathbf{e_k}\| \le \theta$. Thus, for any two elements $\mathbf{e_1}$ and $\mathbf{e_2}$ from the cluster $C_{\mathbf{e_i}}$, their pairwise dissimilarity is always less than 2θ; if $\mathbf{e_1} \in C_{\mathbf{e_i}}$ and $\mathbf{e_2} \in C_{\mathbf{e_i}}$, then $\|\mathbf{e_1} - \mathbf{e_2}\| \le 2\theta$. The last fact about 2θ intracluster dissimilarity leads to a notation $C_{\mathbf{e_i}} = C_{\mathbf{e_i}}^{2\theta}$. Furthermore, if $\mathbf{e_1}$ and $\mathbf{e_2}$ belong to an unknown cluster C_j, then $C_{\mathbf{e_1}}^{2\theta} = C_{\mathbf{e_2}}^{2\theta}$, otherwise $C_{\mathbf{e_1}}^{2\theta} \ne C_{\mathbf{e_2}}^{2\theta}$. Therefore, the clusters $\left\{CE_m^{\theta}\right\}$ characterized by θ can be obtained in the following way:

1) Create clusters $C_{\mathbf{e_i}}^{2\theta} = \left\{\mathbf{e_k}\right\}$, such that $\|\mathbf{e_i} - \mathbf{e_k}\| \le \theta$.
2) Compare all pairs of clusters $C_{\mathbf{e_i}}^{2\theta}$.
3) Assign elements into the final clusters of elements $\left\{CE_m^{\theta}\right\}$ based on the comparisons in (2).

For $\alpha \le \theta$, clusters $C_{\mathbf{e_i} \in C_j^{\theta}}^{2\theta}$ are not identical to an unknown cluster $C_j^{\theta,\alpha}$. A cluster $C_{\mathbf{e_i}}^{2\theta}$ is a superset of $C_j^{\theta,\alpha}$ because the cluster $C_{\mathbf{e_i}}^{2\theta}$ also contains some exterior elements of $C_j^{\theta,\alpha}$ due to $\alpha \le \theta$. Our analysis assumes that the case $\alpha \le \theta$ occurs due to random noise. This assumption about random noise leads to a statistical analysis of similarity of clusters $C_{\mathbf{e_i}}^{2\theta}$. Two issues are investigated next:

1) a statistical parameter of a cluster $C_j^{\theta,\alpha}$ that would be invariant in the presence of noise and
2) a maximum deviation of two statistically invariant parameters computed from clusters $C_j^{\theta,\alpha}$ and $C_{\mathbf{e_i}}^{2\theta}$.

First, let us assume that deterministic values of elements in $C_j^{\theta,\alpha}$ are corrupted by a zero mean random noise with a symmetric central distribution. Then a sample mean (average) $\overline{u}_j$ of elements in $C_j^{\theta,\alpha}$ would be a statistically invariant parameter because the mean of noise is zero. Second, a sample mean is computed from each cluster $C_{\mathbf{e_i}}^{2\theta}$ and is denoted as $\overline{u}_{\mathbf{e_i}}$. The deviation of $\overline{u}_{\mathbf{e_i}}$ from $\overline{u}_j$ is under investigation. If $C_{\mathbf{e_i}}^{2\theta}$ is a subset of $C_j^{\theta,\alpha}\left(C_{\mathbf{e_i}}^{2\theta} \subseteq C_j^{\theta,\alpha}\right)$, then the sample mean $\overline{u}_{\mathbf{e_i}}$ would not deviate by more than θ from $\overline{u}_j$; $\left\|\overline{\mu}_{\mathbf{e_i}} - \overline{\mu}_j\right\| \le \theta$. This statement is always true. If there are two arbitrary subsets $C_{\mathbf{e_i}}^{2\theta} \subseteq C_j^{\theta,\alpha}$ and $C_{\mathbf{e_1}}^{2\theta} \subseteq C_j^{\theta,\alpha}$, then their sample means would not be more than θ apart, as well; $\left\|\overline{\mu}_{\mathbf{e_i}} - \overline{\mu}_{\mathbf{e_1}}\right\| \le \theta$. If $C_{\mathbf{e_i}}^{2\theta}$ is a superset of $C_j^{\theta,\alpha}\left(C_{\mathbf{e_i}}^{2\theta} \supset C_j^{\theta}\right)$, then the same deviation of $\overline{u}_{\mathbf{e_i}}$ from either $\overline{u}_j$ or $\overline{u}_{\mathbf{e_1}}$ is assumed as before for $\mathbf{e_i}$, $\mathbf{e_1} \in C_j^{\theta,\alpha}$; $\left\|\overline{\mu}_{\mathbf{e_i}} - \overline{\mu}_j\right\| \le \theta$ and $\left\|\overline{\mu}_{\mathbf{e_i}} - \overline{\mu}_{\mathbf{e_1}}\right\| \le \theta$. The validity of the previous if statement depends on the ratio of elements from the true cluster $C_j^{\theta,\alpha}$ and other clusters exterior to $C_j^{\theta,\alpha}$. Thus for the second issue, the sample mean $\overline{u}_{\mathbf{e_i}}$ is not expected to deviate from $\overline{u}_j$ by more than θ; $\left\|\overline{\mu}_{\mathbf{e_i}} - \overline{\mu}_j\right\| \le \theta$, and any two elements $\mathbf{e_1}$ and $\mathbf{e_2}$ would be grouped together if their corresponding sample means $\overline{u}_{\mathbf{e_1}}$ and $\overline{u}_{\mathbf{e_2}}$ are not more than θ apart; if $\left\|\overline{\mu}_{\mathbf{e_1}} - \overline{\mu}_{\mathbf{e_2}}\right\| \le \theta$, then $\mathbf{e_1}$, $\mathbf{e_2} \in CE_j^{\theta}$. The inequality $\left\|\overline{\mu}_{\mathbf{e_1}} - \overline{\mu}_{\mathbf{e_2}}\right\| \le \theta$ used for $\alpha \le \theta$ can be applied to the case $\alpha > \theta$. There would be no classification error in the final partition $\left\{CE_m^{\theta}\right\}_{m=1}^{M}$ for the case $\alpha > \theta$ if the inequality was used. For $\alpha \le \theta$, the classification error is evaluated in a statistical framework as a probability $\Pr\left(\left\|\overline{\mu}_{\mathbf{e_i} \in C_j^{\theta,\alpha}} - \overline{\mu}_j\right\| > \theta\right)$.

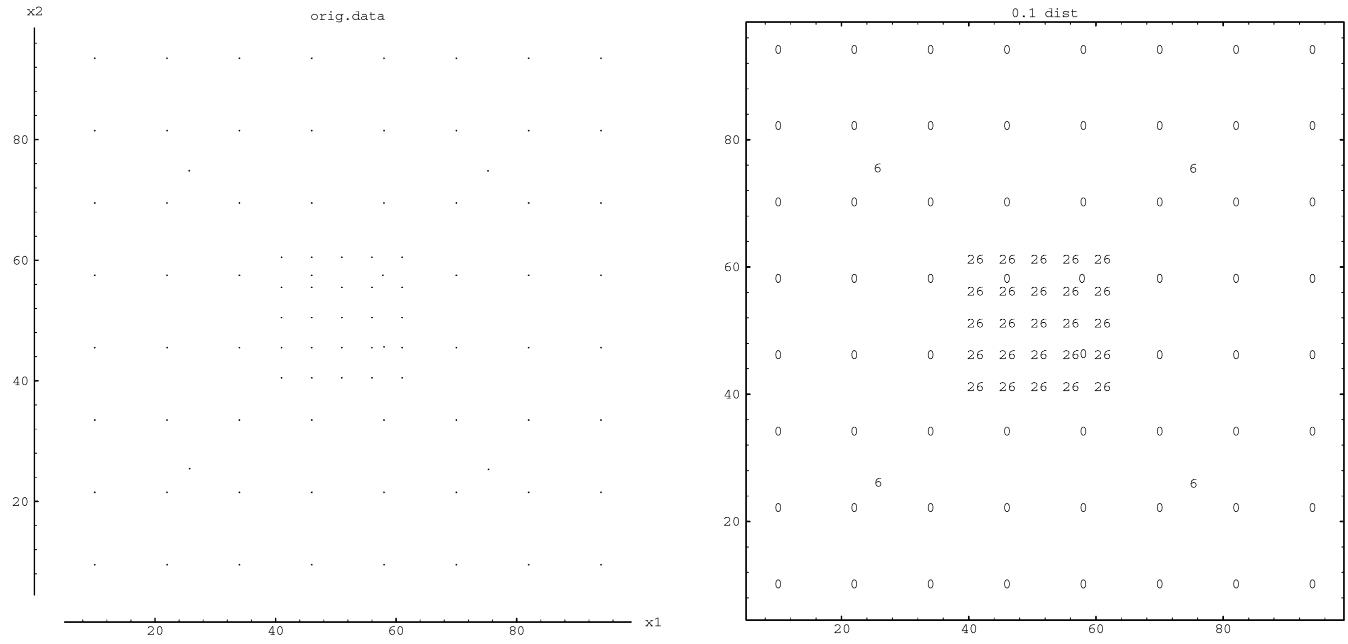
Step 2: Estimation of θ and α

θ and α are estimated by exploring the set of clusters $\left\{CE_m^{\theta}\right\}_{m=1}^{M_\theta}$. Clusters CE_m^{θ} that do not change their elements for a large interval of θ (maximize α and minimize θ) are selected into the final partition $\left\{CE_m\right\}_{m=1}^{NE}$. The set $\left\{CE_m\right\}_{m=1}^{NE}$ is an estimate of the ground truth partition $\left\{C_j\right\}_{j=1}^{N}$, and NE is the estimate of the number N.

The methods are called hierarchical because the set of clusters $\left\{CE_m^{\theta}\right\}_{m=1}^{M_\theta}$ is defined as a nested sequence of sets of clusters along the θ axis. The nested sequence is understood as follows: A cluster obtained at scale θ cannot split at scale $\theta + \Delta$ and cannot merge at scale $\theta - \Delta$ with other clusters. The hierarchy of multiscale classification results is guaranteed by modifying elements within the cluster CE_m^{θ} created at each scale θ to the sample mean of elements of the cluster (Proof: Two elements $\mathbf{e_1}$ and $\mathbf{e_2}$ which have identical values $\mathbf{e_1} = \mathbf{e_2}$ belong to the same cluster CE_m^{θ} for all scales $\theta \ge 0$.).

3 Performance Evaluation

We begin with an example of image texture analysis performed in two steps (see Figs. 1 and 2). First, texture primitives (color homogeneous image regions) are obtained using an image segmentation, and a point pattern is created by measuring an average color and centroid coordinates of each primitive. Second, features are decomposed into color and centroid coordinate corresponding feature sets, therefore two lower-dimensional patterns are created from these features. The location-based clustering is applied to the pattern shown in Fig. 2a consisting of the color feature, and the density-based clustering is applied to the pattern consisting of the centroid coordinate features. Clustering results are combined and shown in Fig. 2b. The method demonstrates its exceptional property of separating spatially interleaved clusters (marked by labels 0, 6, and 26 in Fig. 2) which is a unique property of the clustering methods described here.



(a) (b)

Fig. 2. Spatially interleaved clusters (original and labeled points).

Accuracy of both proposed clustering methods is tested using point patterns with known ground truth partitions. Quantitative evaluation is conducted by measuring the number of misclassified points with respect to the ground truth. Clustering accuracy is tested for

1) synthetic point patterns generated using location and density models of clusters (no interleaved clusters) and
2) standard test point patterns (80x, IRIS), which have been used by several other researchers to illustrate properties of clusters (80x is used in [3] and IRIS in [1], [3], [9]).

Experimental results are compared with four other clustering methods: two hierarchical methods—single link and complete link; and two partitional—FORGY and CLUSTER [3]. Decomposition of features followed by location- and density-based clusterings is explored for each point pattern (80X and IRIS). It is unknown how to determine the choice of features for the decomposition. The goal is to create lower-dimensional point patterns showing inherent tendency to form sets of points with similar locations or approximately constant density. An exhaustive search for the best division in terms of classification error was used. A summary of clustering results in terms of misclassified points is provided in Tables 1, 2, and 3 for synthetic and standard data.

The best method for a class of point patterns generated with a location model (Gaussian clusters) is the proposed location-based clustering (see Table 1). A class of point patterns generated with a density model (regular grid clusters perturbed by uniform noise of magnitude Δ or Gaussian noise with $\mu = 0$ and σ) was clustered the most accurately by the proposed density-based clustering (see Table 2). A combination of location- and density-based clusterings applied to 80X and IRIS data led to the best clustering results (see Table 3). The eight-dimensional point pattern 80X was decomposed experimentally into two lower-dimensional spaces; one n_s = four-dimensional subspace (features 1, 2, 7, 8) and one n_f = four-dimensional subspace (features 3, 4, 5, 6) in order to achieve the result stated in Table 3. By applying the location-based clustering to n_f = four-dimensional points followed by the density-based clustering applied to n_s = four-dimensional points, we could separate 0 from 8X and then 8 from X. The four-dimensional point pattern IRIS was decomposed experimentally as well, but the clustering results were not better than the results from location-based clustering applied alone. All six methods used for the comparison were applied to a class of point patterns with spatially interleaved clusters, e.g., Fig. 1. Proposed density-based clustering outperforms all other methods because it is the only method that is able to separate spatially interleaved clusters.

A time requirement for running each method is linearly proportional to the number of processed elements (N_{point} points, N_{link} links) and to the number of used elements for a sample mean calculation at each element ($N_{CF^{2\delta}_{\mathbf{p_i}}}$ and $N_{CL^{2\varepsilon}_{l_k}}$). The number of processed links N_{link} was reduced by sequential mapping of clusters of links to clusters of points, therefore the time requirement was lowered.

4 Conclusion

We have presented a new clustering approach using similarity analysis. Two hierarchical clustering algorithms, location- and density-based clusterings, were developed based on this approach. Location-based clustering achieves results identical to centroid clustering. Density-based clustering can create clusters with points being spatially interleaved and having dissimilar densities. The separation of spatially interleaved clusters is a unique feature of the density-based clustering among all existing methods. Performance of the clustering methods was compared with four other methods. Location- and density-based clusterings were used for image texture analysis. The contributions of this work can be summarized as

1) addressing a decomposition of the clustering problem into two lower-dimensional problems,
2) proposing a new clustering approach for detecting clusters having a constant property of interior points, such as location or density, and
3) developing a density-based clustering method that separates spatially interleaved clusters having various densities.

TABLE 1
NUMBER OF MISCLASSIFIED POINTS (LOCATION MODEL)

location model method / data	$\sigma = 3$ 30 pts	$\sigma = 3$ 60 pts	$\sigma = 5$ 30 pts	$\sigma = 5$ 60 pts	Σ	perform. order
locat. clus.	0	1	1	3	5	1
dens. clus.	1	11	8	11	31	6
single link	1	2	7	9	19	5
complete link	1	2	5	6	14	4
CLUSTER	2	1	2	3	8	3
FORGY	1	0	2	3	6	2

TABLE 2
NUMBER OF MISCLASSIFIED POINTS (DENSITY MODEL)

density model method / data	no noise	$\Delta = \pm 0.5$	$\sigma = 0.25$	Σ	order
locat. clus.	0	13	10	23	5
dens. clus.	0	3	1	4	1
single link	9	14	13	36	6
complete link	0	11	4	15	2
CLUSTER	6	6	6	18	4
FORGY	5	6	6	17	3

TABLE 3
NUMBER OF MISCLASSIFIED POINTS (80X AND IRIS)

80x and IRIS method / data	80x 45 pts	IRIS 150 pts	Σ	perform. order
locat. + dens. clus.	7	14	21	1
locat. clus.	24	14	38	4
dens. clus.	18	24	42	5
single link	24	25	49	7
complete link	12	34	46	6
CLUSTER	15	16	31	3
FORGY	7	16	23	2

ACKNOWLEDGMENTS

The authors gratefully acknowledge all people who provided data and clustering methods for experiments. Mihran Tuceryan, Texas Instruments; and Chitra Dorai and Anil Jain, Michigan State University. This research was supported in part by Advanced Research Projects Agency under grant N00014-93-1-1167 and U.S. National Science Foundation under grant IRI 93-19038.

REFERENCES

[1] C.T. Zahn, “Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters,” *IEEE Trans. Computers*, vol. 20, pp. 68-86, Jan. 1971.
[2] D. Blostein and N. Ahuja, “Shape From Texture: Integrating Texture-Element Extraction and Surface Estimation,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1,233-1,251, Dec. 1989.
[3] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall, 1988.
[4] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York, NY: Wiley, 1973.
[5] B.S. Everitt, *Cluster Analysis*. London: Edward Arnold, 1993.
[6] N. Ahuja and B.J. Schachter, *Pattern Models*. New York, NY: John Wiley, 1983.
[7] K.C. Gowda and G. Krishna, “Agglomerative Clustering Using the Concept of Mutual Nearest Neighborhood,” *Pattern Recognition*, vol. 10, pp. 105-112, 1978.
[8] M. Nadler and E. Smith, *Pattern Recognition Engineering*. Canada: John Wiley, 1993.
[9] P.H. Sneath and R.R. Sokal, *Numerical Taxonomy*. San Francisco, CA: W.H. Freeman, 1973.
[10] H. Hanaizumi, S. Chino, and S. Fujimura, “A Binary Division Algorithm for Clustering Remotely Sensed Multispectral Images,” *IEEE Trans. Instrumentation and Measurement*, vol. 44, pp. 759-763, June 1995.
[11] Y. Wong and E.C. Posner, “A New Clustering Algorithm Applicable to Multiscale and Polarimetric SAR Images,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 31, pp. 634-644, May 1993.
[12] M. Tuceryan and A. Jain, “Texture Segmentation Using Voronoi Polygons,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 211-216, Feb. 1990.
[13] P. Bajcsy and N. Ahuja, “Uniformity and Homogeneity Based Hierarchical Clustering,” *Proc. 13th Int’l Conf. Pattern Recognition*, vol. B, pp. 96-100, Vienna, Austria, 1996.
[14] M.G. Kendall, *The Advanced Theory of Statistics*, vol. 2. New York, NY: Hafner, 1951.
[15] A.S. Fotheringham and F.B. Zhan, “A Comparison of Three Exploratory Methods for Cluster Detection in Spatial Point Patterns,” *Geographical Analysis*, vol. 28, pp. 200-218, July 1996.
[16] A. Getis and B. Boots, *Models of Spatial Process: An Approach to the Study of Point, Line, and Area Patterns*. London: Cambridge Univ. Press, 1978.