

A BIOLOGICAL APPROACH TO CLUSTERING DATA SETS

MOHAMED BEN HADJ RHOUMA AND HICHEM FRIGUI

Received March 26, 2003

ABSTRACT. We introduce an efficient synchronization model that organizes a population of discrete Integrate and Fire oscillators into stable and structured groups. Each oscillator fires synchronously with all the others within its group, but the groups themselves fire with a constant phase difference. The structure of the synchronized groups depends on the choice of the coupling function. We show that by defining the interaction between oscillators according to the relative distance between them, our model can be used as a general clustering algorithm that is simple, efficient, robust, unbiased by the size of the clusters, and that can find an arbitrary number of clusters. In addition to helping the model self-organize into stable groups, the synergy between clustering and synchronization reduces the computational complexity significantly. The resulting clustering algorithm has several advantages over conventional clustering techniques. In particular, it can generate a nested sequence of partitions, and can determine the optimum number of clusters in an efficient manner. Moreover, since our approach does not involve optimizing an objective function, it is not sensitive to initialization, and can incorporate non-metric similarity measures. We illustrate the performance of our algorithms with several synthetic and real data sets.

1 Introduction Synchrony of coupled oscillators is a widespread phenomenon that manifests itself in mechanics [14], chemistry [9], and biology [11]. A characteristic feature of biological oscillators is that they interact with each other through sudden impulses. For example, the famous fireflies of south east Asia communicate through light flashes, while crickets exchange brief chirps. The behavior of a single biological oscillator is usually modeled with a singularly perturbed system of differential equations. The mathematical analysis of a large network of oscillators becomes then a complex task to accomplish especially when the connectivity of the graph is random, unknown, or non uniform. An alternative approach is to neglect the details of the shape of oscillators by modeling the network with a set of discrete one-dimensional Integrate and Fire (DIF) oscillators.

A DIF is characterized by a state variable which is assumed to be monotonically increasing toward a threshold. When this threshold is reached, the oscillator fires a pulse to its neighbors, and jumps back to its basal level and a new period begins. Unlike a *regular* integrate-and-fire oscillator, a DIF has no amplitude associated with it. This is obviously a costly simplification when it gets to a detailed analysis, however as we will show in the theoretical part of this paper, synchrony between oscillators, phase-locking, as well as the self-organizing of oscillators into groups are still present in the simplified network of DIF's.

In this paper, we show how to use the self-organizing property of oscillator networks and DIF networks in particular to cluster large data sets efficiently and without recurring to a costly optimization procedure. In particular, we show that by letting the coupling be a function of the similarity between oscillators, we can control the number and structure of the sub-groups, and thus use this model as a general clustering algorithm. Unlike existing

2000 *Mathematics Subject Classification.* 37N40, 62H30, 62H35, 68T10.

Key words and phrases. Oscillators, Synchronization, Phase-locking, Unsupervised learning, Clustering.

models, our model does incorporate techniques from relational and prototype-based clustering methods, and results in a clustering approach that is simple, efficient, robust, unbiased by the size of the clusters, and can find an arbitrary number of clusters. In addition to helping the model self-organize into stable structured groups, the synergy between clustering and synchronization reduces the computational complexity of our model significantly. This is because the number of competing oscillators shrinks progressively as synchronized oscillators get summarized by a single oscillator.

A brief outline of the rest of the paper is as follows. In Section 2, we introduce the discrete integrate and fire oscillators and outline a few theoretical results that are related to clustering. In Section 3, we present our model. First, we introduce the problem of clustering, and we present our approach to it. In Section 4, we demonstrate the performance of our clustering algorithm on several data sets and apply the developed algorithm to image segmentation.

2 Discrete Integrate and fire oscillators Let $\mathcal{O} = \{O_1, \dots, O_n\}$ be a set of n oscillators where each oscillator O_i is characterized by a phase ϕ_i and a state variable x_i given by

$$x_i = f_i(\phi_i)$$

where each function $f_i : [0, 1] \rightarrow [0, 1]$ is a smooth monotonically increasing function satisfying $f_i(0) = 0$ and $f_i(1) = 1$. When x_i reaches a threshold at $x_i = 1$, the i th oscillator fires and instantaneously resets to zero, after which a new cycle begins. In the presence of coupling, the phases of the adjacent oscillators change by an amount $\epsilon_{ij}(\phi_j)$ and we have

$$x_j(t^+) = x_j(t) + \epsilon_{ij}(\phi_j).$$

The coupling can be excitatory, inhibitory, or null depending on the firing oscillator. The magnitude of the coupling also depends on the affected oscillator and its phase at the time of firing. This is quite different from the work in [11] where the coupling is constant and where all oscillators are identical. However, for simplicity we will still maintain the assumption that all oscillators have the same frequency.

2.1 The case of two oscillators Let O_1 and O_2 denote the two oscillators, and f_1 and f_2 their corresponding characterizing functions. Let us also define the functions

$$\begin{aligned} (1) \quad u(\phi) &= 1 - f_2^{-1}(f_2(\phi) + \epsilon_{12}(\phi)), \text{ and} \\ (2) \quad h(\phi) &= 1 - f_1^{-1}(f_1(u(\phi)) + \epsilon_{21}(u(\phi))). \end{aligned}$$

The function $h(\phi)$ is the firing map of the system, in the sense that its iterates provide us with stroboscopic pictures of the system every time the first oscillator fires. Notice that if both couplings are null, then the function h will reduce to identity. In [15] we proved the following result.

Theorem 1. *Assume that there exists $\phi_* \in (0, 1)$ such that $h(\phi_*) = \phi_*$ and that for all $\phi \in (0, 1)$, we have*

$$(3) \quad \frac{f'_1(\phi) + \epsilon'_{12}(\phi)}{f'_1(f^{-1}_1(f_1(\phi) + \epsilon_{12}(\phi)))} \frac{f'_2(\phi) + \epsilon'_{21}(\phi)}{f'_2(f^{-1}_2(f_2(\phi) + \epsilon_{21}(\phi)))} > 1, \text{ for all } i = 1, \dots, n,$$

then, the two oscillators will perfectly synchronize for all initial conditions except for ϕ_ .*

The above theorem solves partially Peskin's conjecture [10] who postulated in 1975 that non identical integrate and fire oscillators will synchronize even when the coupling is not sufficiently small and when the coupling is asymmetric.

2.2 The case of n oscillators In the case of n oscillators, the firing map, h becomes a function of $n - 1$ variables. The dynamics of such a map can be quite complex especially when the oscillators are non identical. Tracking the behavior of the iterates of the map h would then require the following assumptions:

- (1) The difference function $\|f_i(\phi) - f_j(\phi)\| \ll 1$ for all $i, j = 1, \dots, n$ and $\phi \in [0, 1]$.
- (2) To obtain synchrony in the network, we also need the network to be globally coupled and that the firing of one oscillator affects all other oscillators in the same way: i.e.

$$\epsilon_{ij}(\phi) = \epsilon_{ik}(\phi) = \epsilon_i(\phi), \text{ for all } i, j, k \leq n.$$

Using the above assumptions we obtained [15]

Theorem 2. Assume that for all $\phi \in (0, 1)$, we have

$$(4) \quad \frac{f'_i(\phi) + \epsilon'_i(\phi)}{f'_i(f^{-1}_i(f_i(\phi) + \epsilon_i(\phi)))} > 1, \text{ for all } i = 1, \dots, n.$$

Then, the system of n oscillators will perfectly synchronize for all initial conditions except for a set of measure zero.

2.3 Clustering of discrete integrate and fire oscillators The clustering of globally coupled oscillators has been observed and studied by many authors [4], [5], [11], [13], and many others. For example, in a network of globally coupled Josephson junctions, some oscillators perfectly synchronize (their phases and amplitudes are identical at all times) while others phase-lock (their behavior is identical to the other oscillators but with a constant phase shift), thus forming clusters of oscillators [14]. To the best of our knowledge, there are no theoretical results on the necessary conditions for the occurrence of such phenomenon. The result presented in this section is a step in this direction.

For simplicity, we will assume that all the DIF oscillators are identical $f_i = f$, and that the set of oscillators \mathcal{O} is divided into k mutually disjoint groups G_1, \dots, G_k . We also assume that if O_i and O_j are two different oscillators, then

$$(5) \quad \begin{cases} \frac{f'(\phi_j) + \epsilon'_i(\phi_j)}{f'(f^{-1}(f(\phi_j) + \epsilon_i(\phi_j)))} > 1 & \text{if oscillators } O_i \text{ and } O_j \text{ belong to the same group,} \\ 0 < \frac{f'(\phi_j) + \epsilon'_i(\phi_j)}{f'(f^{-1}(f(\phi_j) + \epsilon_i(\phi_j)))} < 1 & \text{otherwise.} \end{cases}$$

Theorem 3. Under the above assumptions, for all initial conditions (except for a set of measure zero) any two oscillators will synchronize if and only if they belong to the same group G_k .

3 Data clustering with DIF oscillators Clustering is a process by which a data set is divided into different classes such that elements of the same cluster are as similar as possible and elements of different clusters are as dissimilar as possible. In a set of n elements, and assuming that one already knows that the desired number of clusters k , an exhaustive search of all possibilities requires a search time in the order of n^k . This, of course, makes this problem highly combinatorial prohibiting the use of exact optimization techniques for

any reasonable data set. Traditional clustering algorithms can be divided into two main categories [7]: hierarchical and partitional. Hierarchical clustering techniques yield a nested sequence of partitions and do not require the specification of the appropriate number of clusters. However, these algorithms are computationally expensive and cannot incorporate a priori knowledge about the global shape or size of clusters. Prototype-based clustering algorithms represent each cluster by a prototype and use the sum of the distances from the prototypes to the data points as an objective function. While this allows for incorporating knowledge about the size and shape of the clusters, finding the *appropriate* number of clusters remains a problem. Moreover, the major drawbacks of these algorithms are their sensitivity to noise, outliers, and initialization. Recently, a few algorithms that address these drawbacks have emerged [3] [2] [8]. However, these algorithms are computationally complex and are only suitable to simple distance measures. Moreover, since prototype-based algorithms essentially rely on an optimization procedure, they require parametric and differentiable distance measures, thus prohibiting subjective, and non metric similarity measures.

Terman and Wang [13], and Horn and Opher [6] realized the potential of coupled oscillators as a useful tool for data clustering. Both of their models associate one oscillator with one data point and define the coupling between the oscillators as a function of the similarity between the points. Local excitation was used to induce synchronization within clusters while global inhibition was used to create cluster competition and thus the phase locking between clusters. An inherent limitation of these models is that they tend to divide the data set into a small number of clusters of the same size. Terman and Wang [17] have devised an improved algorithm that overcomes these limitations through complex lateral interactions. This method however, cannot be used for general data sets and is computationally inefficient for very large data sets.

3.1 The proposed model Let $Y = \{y_1, \dots, y_n\}$ be the set of n objects to be partitioned. Object data can be represented by p features and the similarity measure d_{ij} between two points y_i and y_j is usually given by an explicit function in \mathbb{R}^p . Relational data on the other hand cannot be represented in \mathbb{R}^p and the objects y_i are not given explicitly. Instead the data is available only in a relational matrix $D = [d_{ij}]$ with zero diagonal terms. Usually, for both object and relational data, the similarity measure is normalized such that the $\text{Max}(d_{ij}) = 1$.

In our model, we represent each data point y_i by a DIF oscillator O_i . The problem of clustering reduces to partitioning the population of n oscillators into C subpopulations where C is unknown. Synchronized oscillators of the same group correspond to a data cluster, and should have a phase difference with oscillators of dissimilar groups. Obviously this dictates that the coupling strength $\epsilon_{ij}(\phi)$ should be a function of the similarity measure d_{ij} . To this purpose, we choose

$$\epsilon_{ij}(\phi) = \alpha(d_{ij})$$

where the function α is a strictly decreasing function satisfying $\alpha(0) > 0$ and $\alpha(1) < 0$. It is important to notice that the choice of the function α is not unique to cause clustering. Of course the choice of the function of α will affect the outcome in terms of the nature, size, and number of clusters. In most of our applications we used the function

$$(6) \quad \alpha_{ij} = \begin{cases} C_E \left[1 - \left(\frac{d_{ij}}{\delta_0} \right)^2 \right] & \text{if } d_{ij} \leq \delta_0 \\ C_I \left[\left(\frac{d_{ij}-1}{\delta_0-1} \right)^2 - 1 \right] & \text{if } d_{ij} > \delta_0 \end{cases} \quad \text{for all } j \neq i,$$

where C_E is the maximum excitatory coupling, C_I is the maximum inhibitory coupling, and δ_0 defines the neighborhood that encloses all oscillators O_j that are similar to oscillator O_i . If d_{ij} is close to δ_0 , then this is the *zone of doubt*, and the coupling should be very weak. On the other hand, if $d_{ij} \gg \delta_0$ or $d_{ij} \ll \delta_0$, then the relation between oscillators O_i and O_j is more certain, and the coupling should be strong. Typically, once an oscillator reaches the threshold, it excites few oscillators and it inhibits several others. In other words, during the evolution of the system, any oscillator receives inhibition more frequently than excitation. Therefore, it is preferable to let C_E be larger than C_I . Once a group of oscillators $G = \{y_{k_1}, \dots, y_{k_g}\}$ synchronize, they should be considered as a single point. Thus the distance d_{iG} from a single point i to the group G should be redefined appropriately. Some possible choices for d_{iG} are:

$$(7) \quad d_{iG} = \begin{cases} \min(d_{ik_1}, d_{ik_2}, \dots, d_{ik_g}) \\ \max(d_{ik_1}, d_{ik_2}, \dots, d_{ik_g}) \\ \frac{1}{g} \sum_{j=1}^g d_{ik_j} \end{cases} .$$

These three choices correspond to the single linkage, complete linkage and average linkage algorithms, respectively, in sequential agglomerative hierarchical nonoverlapping models [12]. When the clusters are compact and well separated all three choices yield the same results. Different results can be obtained otherwise. The average linkage is used in the applications shown in this paper. The resulting clustering algorithm, called Self-Organization of Oscillator Network SOON is summarized below:

Algorithm Self-Organization of Oscillator Network

Store (or construct) the similarity matrix $\{d_{ij}\}$;

Initialize phases ϕ_i randomly for $i = 1 \dots n$;

Repeat

Identify the firing oscillator i such that $\phi_i = \max(\phi_1, \phi_2, \dots, \phi_n)$;

$T = 1 - \phi_i$;

Reset $\phi_i = 0$;

For all oscillators $j \neq i$ DO

Compute ϵ_{ij} ;

Update the phases $\phi_j = \phi_j + T$;

$\phi_j = f^{-1}(f(\phi_j) + \epsilon_{ij})$;

End For

Identify synchronized groups and reset their phases;

Adjust the similarity measure matrix of newly formed or updated groups;

Until Synchronized groups stabilize.

The following figure illustrates the evolution of the phases for the objects of a data set consisting of two well separated clusters. The pairs of graphs display the phases of each oscillator and the corresponding objects partitioning at the corresponding stage. At first (a), the phases are random and the belonging of each object is still unknown. After 10 iterations, a few oscillators had already synchronized thus forming small clusters labeled by the same shape in the feature space(b). After 25 iterations, the clustering is complete and there are only two phases present in the phase space corresponding to the correct partitioning of the objects.

4 Applications

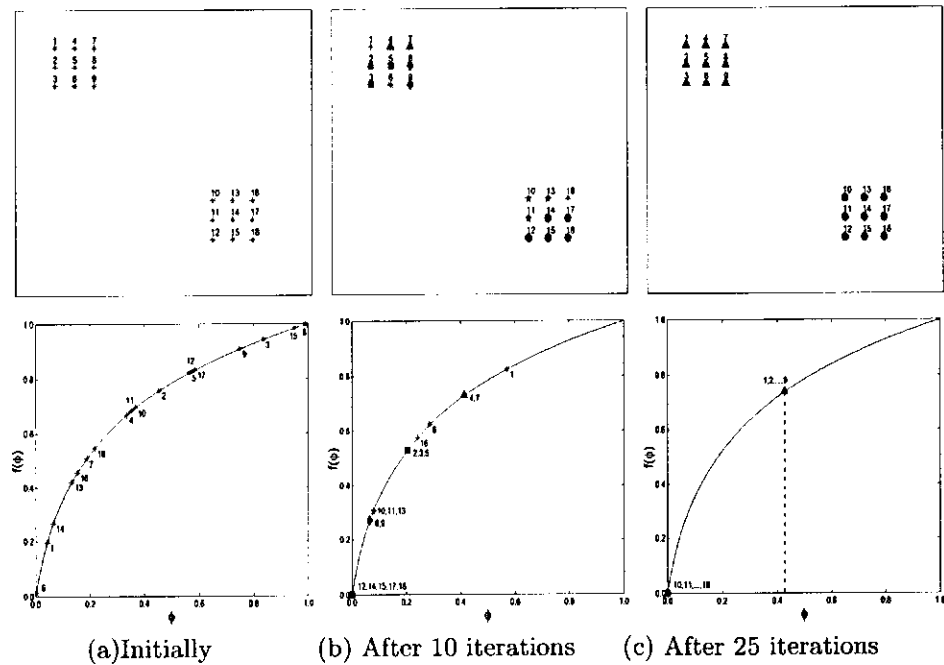


Figure 1: Evolution of the clustering algorithm: Phases and Clusters

4.1 Synthetic data We first start with a few examples from synthetic data to emphasize the robustness of SOON to noise and outliers and its ability to correctly identify the correct number of clusters in an unsupervised manner. In fact, for object data and when the clusters are expected to have ellipsoidal shapes, the Mahalanobis distance can be used as the dissimilarity measure. That is,

$$(8) \quad d^2(\mathbf{y}_j, G_k) = (\mathbf{y}_j - \mathbf{c}_k)^T \mathbf{C}_k^{-1} (\mathbf{y}_j - \mathbf{c}_k),$$

where \mathbf{c}_k and \mathbf{C}_k are the mean and covariance matrix of cluster k . The advantage of using this dissimilarity measure is that the distances of points belonging to the same cluster would have a χ^2 distribution with p degrees of freedom. This desirable feature will (i) automate the choice of the resolution parameter δ_0 , (ii) avoid the need to globally normalize the distances before using them to compute the coupling, and (iii) make the neighborhood of the excitatory region adaptive and cluster dependent. Figure 2 illustrates the evolution of SOON using a 2-D synthetic Gaussian mixture consisting of six clusters of various sizes and orientations. Fig. 2(a) displays the results after 5 iterations where 4 small groups have formed. The center of each group is indicated by the “+” sign and the ellipses enclose points with a Mahalanobis distance less than 9 (i.e. points within the excitatory neighborhood). The remaining points indicate the individual oscillators that did not synchronize yet. Fig. 2(b) displays the results after 500 iterations. As can be seen, most oscillators have synchronized, and there are several groups of various sizes. As the algorithm evolves further, similar groups get merged. Fig. 3(c) shows the result after 1200 iterations where the remaining groups are phase locked.

Figure. 3 illustrates the robustness of SOON to noise. Noise points will either form very small clusters that can be discarded based on cardinality or will not synchronize at all.

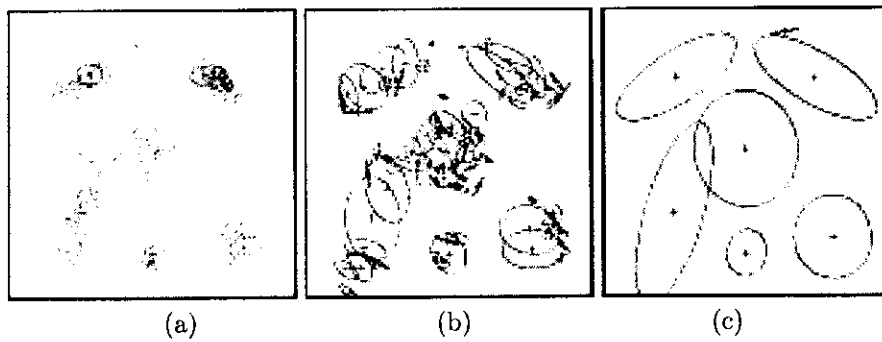


Figure 2: Detection of ellipsoidal clusters. Results at the end of (a) 5 iterations, (b) 500 iterations, and (c) 1200 iterations.

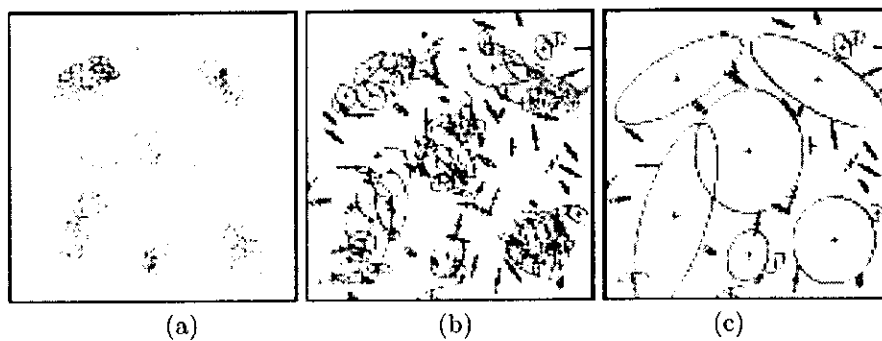


Figure 3: Detection of ellipsoidal clusters in noisy data. Results at the end of (a) 5 iterations, (b) 500 iterations, and (c) 1700 iterations.

4.2 Large databases and Scale-SOON Recently, the advent of the World Wide Web search engines, the problem of organizing massive multimedia databases, and the concept of "data mining" large databases has led to a renewal of interest in clustering and the development of algorithms that are suitable for large databases. The main reason for the renewal of interest is that running several passes over a database makes the run time prohibitive for many clustering algorithms. In [16] we modified the SOON algorithm presented here to Scale-SOON to efficiently cluster huge data sets. The algorithm Scale-SOON consists of successive runs of SOON over data sample sets that fill the memory buffer. After each run, the obtained clusters are summarized by a single oscillator with equivalent sufficient statistics and the synchronized oscillators are purged from the buffer. Our empirical evaluation has shown that Scale-SOON scales linearly with respect to the number of records and the number of attributes [16]. Our experiments have also indicated that even if large memory is available to hold the entire data, it is more efficient to process it incrementally. This is because the problem is much simpler when fewer oscillators are interacting, and also the cache memory is used more efficiently.

Figures 4 and 5 compare the performance of Scale-SOON with BIRCH [19], one of the best algorithms suitable for large data sets. BIRCH is a hierarchical algorithm that performs a preclustering phase that identifies and summarizes dense regions.

4.3 Color Image Segmentation Figure 6 illustrates the ability of SOON to segment real color images. We should note here that this problem is different from the image

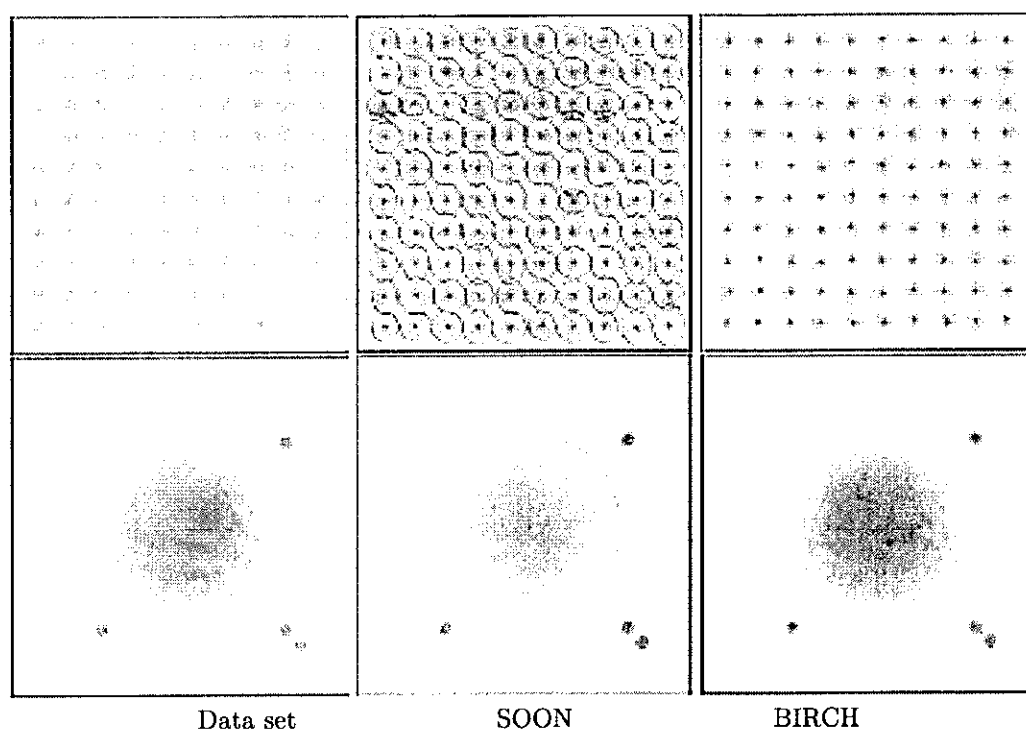


Figure 4: Comparison of SOON and BIRCH on two large data sets

segmentation achieved using pulse coupled neural networks where the spatial coordinates are used to define the neighborhood of the excitatory coupling and only the gray level is used to discriminate between pixels. In this application, each pixel in the image is mapped to an 8-D vector and the segmentation problem is transformed to clustering an 8-D dataset. The 8 features consist of three color, three texture, and two position features [1]. The images shown are from the Corel image database. Each image is to be segmented into several homogeneous regions by clustering the feature vectors mapped from the image pixels. In this application, there is no one correct number of clusters. In general, for each image, we can obtain several reasonable segmentations corresponding to different resolutions.

ACKNOWLEDGEMENTS The first author would like to acknowledge the support of the Mathematics Department of UMR. Partial support of this research was provided by the National Science Foundation under grant No. IIS-0133415.

REFERENCES

- [1] C. carson, S. Belongie, H. Greenspan, and J. Malik, "Color and Texture-Based Image Segmentation using EM and its application to Image Querying and Classification" *IEEE. Trans. Pattern Analysis Machine Intelligence*, pending publication.
- [2] T. Darrell and P. Pentland, "Cooperative Robust Estimation Using Layers of Support", *IEEE. Trans. Pattern Analysis Machine Intelligence*, vol. 17 no, 5, pp. 474-487, 1995.
- [3] H. Frigui and R. Krishnapuram, "A Robust Competitive Clustering Algorithm with Applications in Computer Vision", *IEEE. Trans. Pattern Analysis Machine Intelligence*, vol. 21 no, 5, pp. 450-465, 1999.

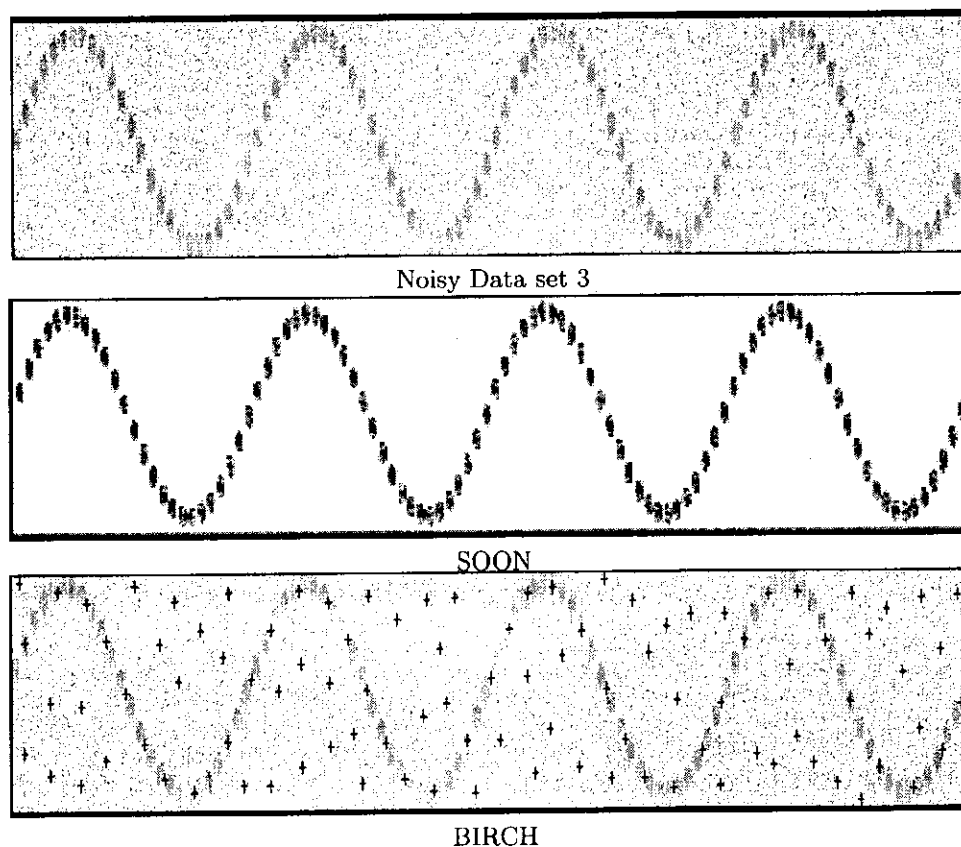


Figure 5: Comparison of SOON and BIRCH on a noisy data set

- [4] D. Golomb, D. Hansel, B. Shraiman, and H. Sompolinsky, "Clustering in Globally Coupled phase oscillators", *Physical Rev.A*, vol. 45, pp. 3516-3530, 1992.
- [5] E. Grannan, D Kleinfield, and H. Sompolinsky, "Stimulus Dependent Synchronization of Neural Assemblies", *Neural Computation*, vol. 5, pp. 550-569, 1993.
- [6] D. Horn, and I. Opher, "Collective Excitation Phenomena and their Applications", Pulsed Neural Networks, W. Maass and C.M. Bishop, eds., pp. 297-320, MIT Press, 1999.
- [7] A.K. Jain and R.C. Dubes, "Algorithms for Clustering Data." Prentice Hall, 1988.
- [8] J.M. Jolion, P. Meer and S. Bataouche, "Robust Clustering with Applications in Computer Vision", *IEEE. Trans. Pattern Analysis Machine Intelligence*, vol. 13 no, 8, pp. 791-802, 1991.
- [9] Y. Kuramoto, "Chemical Oscillations, Waves and Turbulence", Berlin: Springer, 1984.
- [10] C.S. Peskin, "Mathematical Aspects of Heart Physiology", Courant Inst. Math, 1975.
- [11] R.E. Mirollo and S.H. Strogatz, "Synchronization of Biological Oscillators", *SIAM J. Applied Math.*, vol. 50, pp 1645-1662.
- [12] P.H.A. Sneath and R.R. Sokal, "Numerical Taxonomy-The principles and Practices of Numerical Classification". San Francisco, W.H. Freeman, 1973.
- [13] D. Terman and D.L. Wang, "Global Competition and Local cooperation in a network of Neural Oscillators", *Physica D*, vol. 81, pp 148-176, 1995.
- [14] M.B.H. Rhouma, "On the Continuation of Periodic Orbits", Ph.D. Thesis, Univ. Missouri, Columbia 1999.

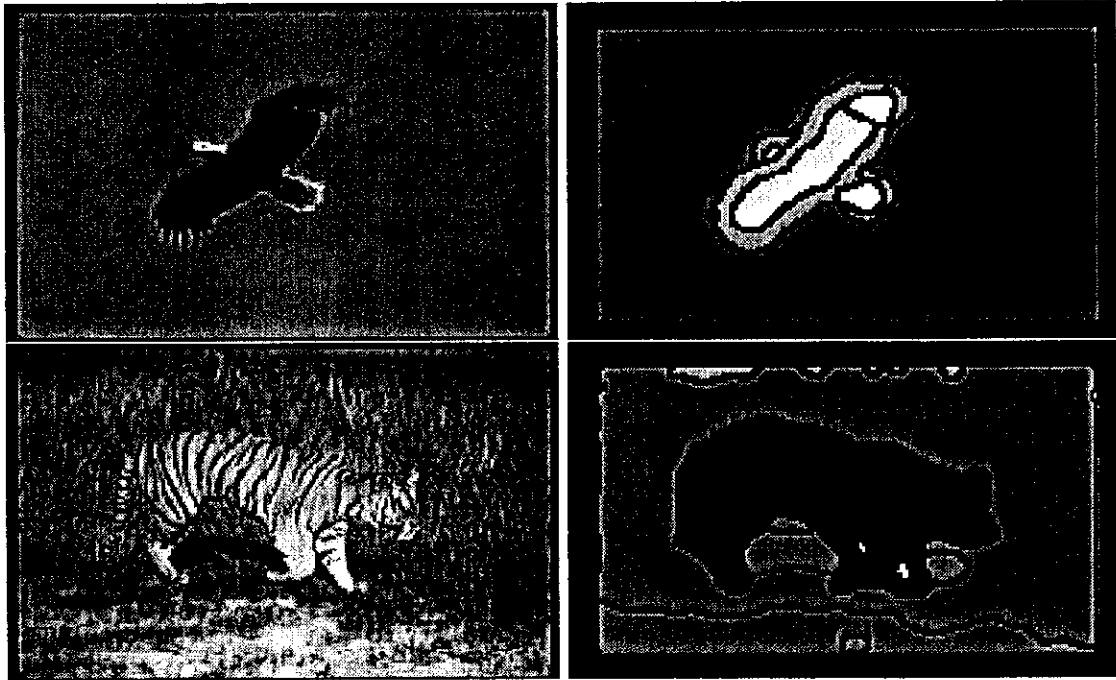


Figure 6: Image segmentation using SOON

- [15] M.B.H. Rhouma and H. Frigui, "Self-Organization of Biological Oscillators with Application to Clustering", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 23 no, 2, pp. 180-195, 2001.
- [16] H. Frigui and M.B.H Rhouma, "A Synchronization Based Algorithm for Discovering Ellipsoidal Clusters in Large Data Sets", *ICDM 2001*, pp.139-146, 2000
- [17] D.L. Wang and D. terman, "Image SegmentationBased on Oscillatory Correlation" *Neural Computation*, vol. 9, pp. 805-836, 1997.
- [18] A.T. Winfree, "Biological Rythms and the behavior of Populations of coupled Oscillators", *J. Theoretical Biology*, vol. 16, pp. 15-42, 1967.
- [19] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH, A New Data Clustering Algorithm and Its Applications", *Data Mining and Knowledge Discovery*, vol. 2, pp. 141-182, 1997.

Mohamed Ben Hadj Rhouma
 Dept. Mathematics & Statistics
 University of Missouri, Rolla
 mohamedb@unr.edu

Hichem Frigui
 EECE Dept.
 University of Memphis
 hfrigui@memphis.edu