Unsupervised Cluster Analysis Using Particle Swarms for Oculographic Signal Segmentation

Martin Macaš, Lenka Lhotská, *Member, IEEE,* and D. Novák

***Abstract*—The problem of real-world signal segmentation is very often difficult because of artifacts and noise. Furthermore, for each signal, a special method using supervised adaptation for every concrete type of segment must be mostly used.**

This paper proposes fully unsupervised approach using partitional clustering method with squared error criterion. The optimal partition is searched through the use of particle swarm optimization (PSO), which makes it possible to overcome local minima and find the near-optima solution with relatively good computational efficiency.

First, the PSO clustering is tested using an artificial benchmark data set and then, practical results of the method on electrooculographic (EOG) signal segmentation are described.

I. Introduction

THE medical domain is one of the areas in which Artificial Intelligence (AI) and machine learning methods are most frequently applied. This is quite natural because modern medicine generates huge amounts of data, but at the same time there is often a lack of explicit relations among this data and a lack of data understanding. In particular, data mining and knowledge discovery are tools than can help in dealing with this problem. In the medical domain, data exists in various forms - single numerical values, non-numerical expressions, measured signals (e.g. ECG, EEG, EMG, EOG). Therefore it is necessary to pre-process and transform the data into the most suitable form to serve as an input for a decision support system or a classification system.

Many data mining techniques applied to signals need the investigated signal to be fully or at least partially segmented to parts with similar characteristics and interpretation. This segmented signal is further processed using other multiple data processing methods leading to higher level interpretations, results and conclusions [1].

A common approach is to find the interesting parts of the signal using pattern classification methods that utilize supervised information depending strongly on signal character. Another approach is to detect some characteristic and significant parts of the signal. For example, time-frequency analysis can be applied using spectrograms or wavelet transform and the segmentation can be based on analysis of some properties of the time-frequency domain of the signal. Similar approach is based on splitting the time (or often frequency) domain of the signal into smaller windows, called segments, and describe each part using some extracted features [2]. However, the notion of segment is here different.

In this paper we describe a similar approach. Individual parts of the signal are characterized by real valued features and the resulting set of samples is partitioned into groups using a clustering method based on Particle Swarm Optimization (PSO) algorithm. The signal is segmented according to the resulting division into the corresponding groups, where each group represents certain type of segments, thus a certain cluster.

First, the PSO clustering technique is tested and verified on well-known artificial benchmark "IRIS" data-set. Further, the clustering method is applied to real data extracted from electrooculographic (EOG) signals. The PSO approach is compared with k-means algorithm.

II. Methods

A. Particle Swarm Optimization

The PSO method is one of optimization methods developed for searching global optima of a nonlinear function [3]. It is inspired by the social behavior of birds and fish. The method uses group of problem solutions. Each solution consists of set of parameters and represents a point in multidimensional space. The solution is called particle and the group of particles (population) is called swarm.

Each particle *i* is represented as a *D*-dimensional position vector $\vec{x}_i$ and has a corresponding instantaneous velocity vector $\vec{v}_i$. Furthermore, it remembers its individual best value of fitness function and position $\vec{p}_i$ which has resulted in that value. During each iteration *t*, the velocity update rule (1) is applied to each particle in the swarm. The $\vec{p}_g$ is the best position of the entire swarm and represents the social knowledge.

This research is supported by the research program No.MSM 6840770012 "Transdisciplinary Research in the field of Biomedical Engineering II"and by The University Development Foundation grant No. FRVS 2812/2006.

M. Macaš is with Czech Technical University, Prague, Technicka 2, Czech Republic (e-mail: mmacas@seznam.cz).

L. Lhotská is with Czech Technical University, Prague, Technicka 2, Czech Republic (e-mail lhotska@labe.felk.cvut.cz).

D. Novák is with Czech Technical University, Prague, Technicka 2, Czech Republic (e-mail xnovakd1@labe.felk.cvut.cz).

$$\vec{v}_i(t) = \alpha \vec{v}_i(t-1) + \varphi_1 R_1(\vec{p}_i - \vec{x}_i(t-1)) + \varphi_2 R_2(\vec{p}_g - \vec{x}_i(t-1)), \tag{1}$$

The parameter α is called inertia weight and during all iterations decreases linearly from α_{start} to α_{end}. The symbols R_1, R_2 represent the diagonal matrices with random diagonal elements drawn from a uniform distribution between 0 and 1. The parameters φ_1 and φ_2 are scalar constant that weight influence of particles' own experience and the social knowledge.

Next, the position update rule (2) is applied:

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t). \tag{2}$$

If any component of $\vec{v}_i(t)$ is less than $-V_{max}$ or greater than $+V_{max}$, the corresponding value is replaced by $-V_{max}$ or $+V_{max}$, respectively. The V_{max} is maximum velocity parameter.

The update formulas (1) and (2) are applied during each iteration and the $\vec{p}_i$ and $\vec{p}_g$ values are updated simultaneously. The algorithm stops if maximum number of iterations is achieved or any other stopping criterion is satisfied.

B. PSO Clustering

Data clustering is an important process in pattern recognition and machine learning that identifies natural groupings or clusters within multidimensional data based on a similarity measure. Clustering algorithms are used in many applications, such as image segmentation, vector and color image quantization, data mining, etc. Most clustering algorithms are based on two popular techniques known as hierarchical and partitional clustering. Formal partitional clustering procedures use a criterion function, such as the sum of the squared distances from the cluster centers, and seek the grouping that extremizes the criterion function [4]. Such optimization task could be solved using evolutionary optimization algorithms [5]. In this paper, the PSO approach is compared to k-means algorithm. That is why it was suitable to use the same criterion - the sum of squared errors.

Suppose that we have a set D of n samples $x_1,...,x_n$ that we want to partition into c disjoint clusters $D_1,...,D_c$. The sum of squared error criterion is defined:

$$J_e = \sum_{i=1}^{c} \sum_{x \in D_i} \left\| x_i - m_i \right\|^2, \tag{3}$$

where

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x \tag{4}$$

is the mean of samples from cluster i.

As it was mentioned above, the PSO clustering method was used here. The algorithm finds the centroids of a user specified number of clusters, where each cluster groups together similar patterns. In the context of data clustering, a single particle represents the c cluster centroids. That is, each particle $\vec{x}_i$ is constructed as c sequentially organized vectors of centroid positions.

Then, a swarm represents a number of candidate data clusterings (partitions). The quality of each particle is measured using function described in (3). A solution, which corresponds to the minima of the function is searched.

C. Signal Segmentation

Most of the real signals belong to stochastic signals. Stochastic signals can be divided into two basic groups, namely stationary and non-stationary signals. Stationary stochastic signals do not change their statistic characteristics in time. Non-stationary signals may have variable quantities in time, for example mean value, dispersion, or frequency spectrum. Such signals are considered stationary whose statistic parameters remain constant over sufficiently long time.

To avoid the problems with signal non-stationarity, the signal was divided into parts of constant length and each part was further described using feature extraction procedure. This method is often called segmentation, but in different meaning.

However, a disadvantage of this method is that the resulting parts are still not necessarily stationary. Our modification suggests to use overlapping of individual parts. Using small shift and suitable length of these parts, we can reach very good results for correct division directly into individual characteristic parts having natural interpretation.

As it was mentioned above, the cluster analysis searches some clusters of multidimensional data samples. Therefore, the data must be first extracted from the investigated signal. In this paper, constant segmentation was used.

Afterwards, a number of features was extracted from each part. The feature extraction procedure for the case of EOG signals is described in the next section. In this way, a data matrix was created, where columns correspond to features and rows represent particular parts of signal - samples. The unsupervised cluster analysis was further applied to the data matrix and each row (part of signal) was labeled according to its belonging to one of the clusters. Because of overlap of signal parts, each sample of signal could fall into more than one cluster.

The final classification was done in such a way that each sample was classified as member of the cluster into which it fell in most cases. The described process split the whole signal into c segment types with different patterns and hopefully with different interpretation - the unsupervised segmentation was done.

D. EOG Application

The human eye is never in entire calm. Its movement is a consequence of an ophthalmogyric muscle's work. There are two main kinds of eye movements - the big and the small. Saccades are rapid eye movements which allow binocular turning or version of the eyes from one fixation point to another. During these conjugated and volitional movements, the eye is browsing a visual field. The direction and the magnitude of the saccade can be influenced willingly.

The saccade alternates with the period of fixation made when the eyes are directed to a particular target. Fixation is a state in which the eye is not moving and the visual stimulus is perceived. However, even at fixation the eye is not completely motionless. It performs small movements (microsaccades, drifts, tremor).

Sequences of fixations and saccades (rapid eye movements between fixations) define scanpaths, providing a record of visual attention on a subject of interest.

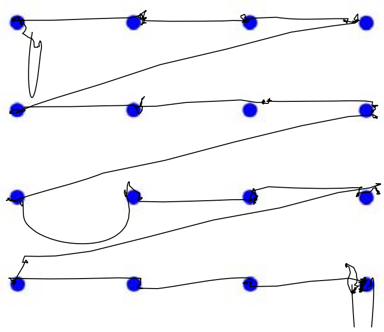

Fig. 1. The stimulus used for eye movement recording with XY plot of the EOG signal.

The recording of EOG signals during eye movements (eye tracking) finds many applications - for instance advertising, industry, biomedical engineering, or medical diagnostics. For these purposes, it is very important to find the particular components of signal. In fact, this process results in segmentation of EOG signal into parts with similar interpretation - fixation, saccades, and others.

The resulting segmentation could be used for statistical description of the signal or subsequent analysis. The EOG signals have big diagnostic potential. For example, the automatic segmentation and its results are useful for analysis of eye movements of dyslexic children [6]. Although the connection between dyslexia and eye movements has been examined in several studies, it is still not clear, if this connection exists. The functional significance of eye movements is far from being understood. Few studies carried out comparison between eye movements of dyslexics and matched control subjects. Many properties of dyslexic's eye movements were described, adverting to diagnostic potential of eye movements. Subjects with reading difficulties make a higher percentage of regressive eye movements during reading than normal readers, there could be also differences in length of fixation or directions of saccades. It could be possible, that the children with dyslexia have some specific patterns occurring in eye movement record. However, all these ideas are just hypotheses and in fact, the neurologists do not know what sort of patterns to look for. That is why the unsupervised approach to segmentation could be suitable and promising.

The eye movements of *76* female subjects were recorded using iView 3.0 videooculography system at the Department of Neurology, 2nd Medical Faculty, Charles University, Czech Republic. Only *52* of the *76* measured subjects were used for subsequent experiments because of difference in age or poor quality of signal records of the remaining subjects. There were three types of subjects - normal readers, retarded readers without dyslexia and dyslexics. The measurements were executed in dark. Subject with fixated head looked at a stimulus on the screen. The screen was placed *1* meter from the subject. Subjects were stimulated by a non-verbal stimulus consisting of bitmap picture - Fig. 1. The child was asked for browsing through dots on the screen and its horizontal and vertical eye movement signals were measured.

Let $S=s(1),s(2),..s(L)$ denotes the part of signal consisting of L samples. The feature extraction procedure computes a set of numbers describing the part of signal. The choice of suitable features has fundamental influence on results of clustering and final segmentation. During designing the feature set, the user can use some prior knowledge about the investigated signal and expected properties of the results. An example is the EOG signal segmentation. It is generally known, that the saccades and fixations differ in velocity. That is why the first feature F_1 was the average of absolute value of differences:

$$F_1 = \frac{1}{L}\sum_{i=2}^{L}\left|s(i)-s(i-1)\right|. \qquad (5)$$

The other features used in this paper were coefficients of 2nd order autoregressive model estimated by the use of covariance method [7].

The definition of the model is $s(t)=a_1s(t-1)+a_2s(t-2)+\varepsilon(t)$, where $\varepsilon(t)$ is assumed to be Gaussian white noise. The other features are thus $F_2=a_1$ and $F_3=a_2$.

As it was mentioned above, both the vertical and horizontal component of eye movement signals were recorded. The described features were extracted for both of them. Thus, the three features were extracted for each component, which gives in total *6* features describing one part of the two-dimensional signal.

III. Results

A. Testing The PSO Clustering

The suitability of the PSO approach to clustering was tested using the Iris data set. This well-known benchmark set has three subsets (i.e. Iris setosa, Iris versicolor, and Iris virginica). There are in total 150 data points in the data set. Each class has 50 patterns. The PCA plot of first two principal components is depicted in Fig. 2. The PSO clustering and k-means were compared using the Iris data set. The comparison can be done easily because both clustering approaches used the same criterion - the sum of squared errors. The second comparison criterion was the classification performance. It measures, how the samples (patterns) from one particular class are distributed in resulting clusters. We resolve the correspondence they have to the true cluster labels of the patterns. Ideally, each cluster should contain patterns that belong to only one class. The data were not normalized, because the clustering criterion was not scale invariant.

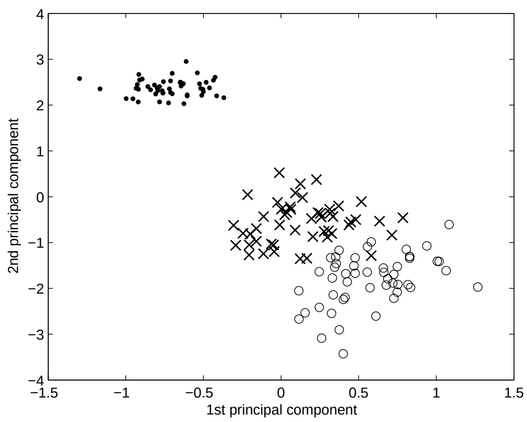

Fig. 2. The plot of first two principal components of the testing IRIS dataset.

During each run, both algorithms were initialized randomly. The PSO parameters were set experimentally as following. Inertia weight was decreasing linearly from value $\alpha=1$ down to the value $\alpha=0.3$ during first *80%* of iterations.

TABLE I
COMPARISON OF CLUSTERING RESULTS

	Method			
	k-means		PSO	
Patient 1	Mean	Std	Mean	Std
Fixations (%)	80.63	33.62	100.00	0.00
Saccades1(%)	85.45	32.09	100.00	0.00
Saccades2(%)	80.00	42.00	100.00	0.00
Criterion	134.81	7.47	130.52	0.83
Sufficient iterations	-	-	19.90	9.99
Patient 2	Mean	Std	Mean	Std
Fixations (%)	91.67	6.00	100.00	0.00
Saccades1(%)	100.00	0.00	100.00	0.00
Saccades2(%)	52.00	38.00	100.00	0.00
Criterion	107.40	6.54	102.87	0.00
Sufficient iterations	-		10.90	6.30

Next, it stagnated on constant value. Parameters φ_1 and φ_2 were set to 2 and the maximum velocity was $V_{max=}1$. The swarm with 20 particles and 40 iterations of PSO algorithm were used for all experiments. The same settings were used also for all experiments in the following section.

Fig. 3 shows the results of 30 runs of experiments with k-means and PSO. For each value of clustering criterion reached during the 30 runs, the relative occurrence is depicted here. The classification performance in percents is added to each column demonstrating a correspondence (or non-correspondence) of clustering criterion and classification.

It can be considered that the global minimum of the criterion is in the value *67.56%*. The value was reached by both the k-means and PSO clustering during more than fifty percents of the runs. The rest of the results show the difference between the two algorithms. The k-means reached the global optima relatively often and about third of the runs finished near the global extreme, however, 4 of 30 runs got stuck in very poor local optima where the criterion value was 85.18.

The PSO approach exhibited its ability to reach global optima. All runs of experiment converged to near vicinity of the criterion value $J_e= 67.56$ and all non-optimal results are distributed much more uniformly than in the previous case. Therefore, it can be concluded that the PSO approach had problems with accurate reaching the global optima which is given by stochastic component of the searching. In comparison with k-means, the PSO searching probably did not find a strong local minimum and only due to small number of iterations (40) it did not find the global optima exactly.

Fig. 3 tells also something about the classification results. It is clear, that the classification performance for this task should correspond to the clustering performance. The result for k-means shows, that if the poor local optima was reached, the classification was also quite poor. However, the classification did not correspond to clustering in very near vicinity of the global extreme. It could be justified by unsupervised character of clustering. The goal was not to reach the best classification, but to find the best clustering.

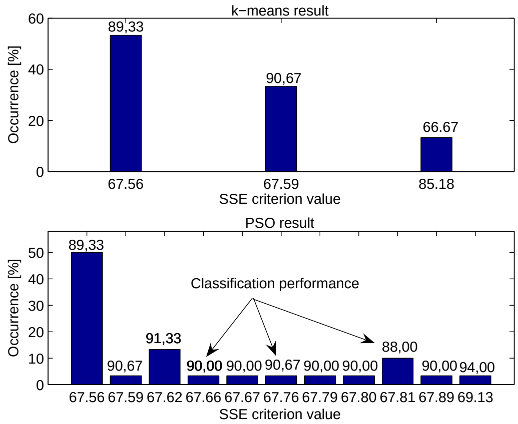

Fig. 3. The histogram of clustering criterion values obtained from 30 runs.

B. EOG Signal Segmentation

The results of EOG signal segmentation are described in this section. The PSO clustering technique is compared to k-means method using quality criteria. The main and most important criterion is the value of sum of squared error of clustering described by equation (3). The comparison can be done because both clustering approaches used this criterion.

The second comparative evaluation is the ability of the technique to reliably detect particular components of the signal and assign the signal segments to proper clusters. Nevertheless, the approach is unsupervised and we do not know, if optimal value of clustering criterion gives upon a segmentation corresponding to our idea (to separate fixations and saccades), hence this performance evaluation is the secondary one. The segmentation was performed using three clusters and splitting the signal into three parts with different interpretation. The first component are fixations - small eye movement appearing like a skein or smudge in X-Y graph. The remaining part of eye movements should be composed of saccades. Two sorts of saccades should be present in signals corresponding to the stimulus described above - the between-dot saccades (Saccades 1 in Table 1) appearing during moving the eyes from one dot to another and the between-row saccades (Saccades 2) appearing when the eye is moving between different rows.

For the method comparison, two horizontal and vertical signals measured on two different patients were used and 30 runs of the segmentation was repeated for each case. The Table 1 shows the percentage share of correctly segmented fixations and saccades of type 1 and 2.

As it was mentioned above, there are differences between basic parts of the EOG signal. Typical descriptive features of saccades are amplitude, velocity maximum, length, and latency. There is a relation between amplitude, length and velocity maximum that is characteristic and can be used for evaluation.

The signal segmentation is a crucial aim of the whole process. We understand under segmentation following operation. Segmentation is division of the set of signal samples X(i), i=1...N into certain number of disjunctive subsets, in our case into 3 disjunctive subsets $X_1,X_2,X_3,$ where $X_1 \bigcup X_2 \bigcup X_3 = X$. Individual subsets contain compact parts of the signal with the same or at least partially the same interpretation. Then we denote these compact parts as segments. Segment type or class is membership in one of the three subsets X_1, X_2, X_3,.

Correctly segmented fixation is such signal segment that a) belongs to one certain subset; b) corresponds to the main characteristic features describing fixation, and c) is separated on both sides with segments of other types (type saccade 1 or saccade 2). The definition of correctly segmented saccade is analogical.

During each run, both algorithms are initialized randomly. The PSO parameters were set experimentally as following. Inertia weight was decreasing linearly from value $\alpha=1$ down to the value $\alpha=0.3$ during first 80% of iterations. Next, it stagnated on constant value. Parameters φ_1 and φ_2 were set to 2 and the maximum velocity was $V_{max}=1$. The swarm with 20 particles and 40 iterations of PSO algorithm was used for all experiments.

Table 1 shows principal differences and advantages of PSO approach. The signals of two patients were used for comparison. First, it must be remarked, that the global minima of the clustering criterion for the two patients were $J_{emin1}=130.25$ and $J_{emin2}=102.87$ respectively. The average values of this criterion illustrate the main advantage of PSO arising from evolutionary character of this technique. It is evident, that the PSO is much less sensitive to random initialization than k-means. The global searching PSO method found the global minima J_{emin} in 90% of runs for Patient 1 and in all runs for Patient 2 and even if it did not approach the global minima accurately, the result could be supposed to be close to the global extreme. On the other hand, the k-means gave us quite different results. The big standard deviation indicates the disability of the algorithm to overcome local optima, which results from local character of the k-means method. Many runs finished in one of many shallow local minima and the resulting non-optimal segmentation did not matched our requirements consisting in similarly interpreted components separation. An example of such non-optimal segmentation using k-means clustering for the Patient 1 is depicted in Fig. 4. The segmentation corresponding to global optima $J_{emin1}=130.25$ obtained mostly when the PSO clustering was used is shown in Fig. 5. The three main components of eye movements are separated perfectly except the cluster 3 – between row saccades, into which some abnormal saccades are included.

IV. Discussion

From the experiments we can conclude that PSO clustering is suitable for such tasks in which the interpretation and meaning of clusters is not obvious and transparent.

PSO has also another advantage, namely it has better ability to reach global optimum and is not too sensitive on

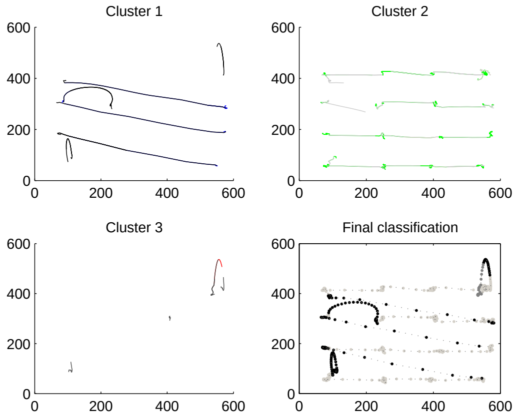

Fig. 4. An example of k-means result corresponding to local minima.

local optima. Another conclusion that can be obtained from

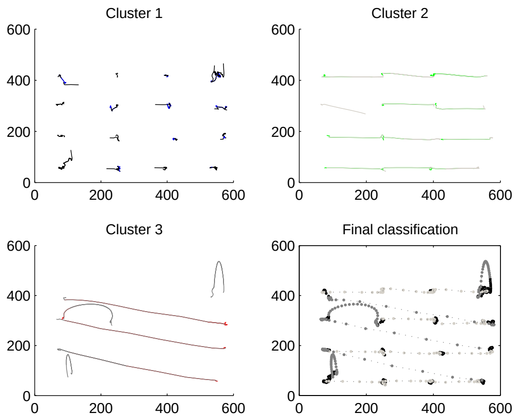

Fig. 5. An example of PSO result. The segmentation corresponds to classification of eye movements to three types.

the Table 1 are the convergence properties of PSO algorithm. The table shows the average number of iterations after which the global minimum was reached. Although, the algorithm is one of the evolutionary techniques and its main general drawback is the high computational cost, its convergence in this application is relatively fast. For the Patient 1, 26 iterations and for the Patient 2, 24 iterations were enough for reaching the global optima. For some particular runs, even 3 iterations were enough.

V. Conclusions

We investigated the possibility of applying the PSO to the field of biomedical signal processing. We presented a clustering approach using PSO. The proposed algorithm was applied to the problem of unsupervised classification of EOG signal. The performance of the algorithm was compared to the k-means algorithm. The PSO clustering outperformed the k-means algorithm.

The experiments presented here used the given number of clusters $c=3$. However, using additional clusters could enable to detect (and remove) artifacts or perform more detailed classification of fixation subtypes and saccade subtypes, possibly identifying new descriptive features that will characterize each subtype better.

However, there is still open space for further improvements. They may include, for example, adaptation of the basic PSO to the clustering task, extensions of basic PSO algorithm or the use of some alternative clustering criterion.

Finally, we must remark, that although the presented approach was applied on EOG signals, the same method could be used for other biomedical signals - to detect beats in ECG, to find spikes, sleep spindles and other patterns in EEG and, of course, to detect (and remove) some artifacts from signals. However, such applications would need different features to be extracted.

References

[1] J. Rieger, L. Lhotská, V. Krajča, and M.Matoušek, "Application of quantitative methods of signal processing to automatic classification of long-term EEG records," *Biological and Medical Data Analysis,* pp. 333-343, Springer, 2004.

[2] K. Kosař, L. Lhotská, and V. Krajča, "Classification of long term EEG recordings," *Biological and Medical Data Analysis*, pp. 322–332, Springer, 2004.

[3] J.Kennedy, and R.C. Eberhart "Particle swarm optimization," in *Proc. International Conference on Neural Networks IEEE*, vol. 4, pp. 1942–1948, 1995.

[4] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification (2nd Edition)*. Wiley - Interscience, 2000.

[5] L.O. Hall, B. Ozyurt, and J.C. Bezdek, "Clustering with a genetically supervised approach," *IEEE Trans. on Evolutionary Computation*, vol. 3, no. 2,103-112.

[6] G. Th. Pavlidis, "Eye movements in dyslexia – their diagnostic significance," *Journal of Learning Disabilities*, vol. 18, no. 1, pp. 42-49, 1985.

[7] S. L. Jr. Marple, "Digital spectral analysis with applications," *Prentice- Hall, Englewood Cliffs, 1987.*