UNSUPERVISED PATTERN RECOGNITION FOR DIGITAL WAVEFORM CLASSIFICATION FROM RADIATION DETECTORS

Jianwei Miao and Mark A. Clements

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

ABSTRACT

In this paper we have addressed the problem of analyzing the digital pulse waveforms of radiation detector outputs. With the availability of extremely high-speed A/D conversion with good resolution, it is now possible to look more deeply at the waveform shapes than is currently done. In our studies, a new technique of unsupervised pattern recognition has been applied which has demonstrated accurate classification (98.33% in probability) of digital pulse waveforms. To the best of our knowledge, application of such a technique is novel. The preliminary results of this system, which show clearly improved measurement conditions, are therefore very promising.

1. INTRODUCTION

Currently, prevailing methods of identifying particle capture events from a radiation detector simply detect the peaks of the pulse waveform for each event or perform an analog integration of the waveforms. Histograms are then plotted to examine the event population and assess the composition of the radiation sources [1, 2]. In effect, these histograms, or spectra, are usually 'clustered' in one dimension. These techniques are not able to classify two or more distinct peaks of particle events when the resolution of the radiation detector is less than the difference between two distinct peaks, or if to much noise is present.

In our study a new technique of unsupervised pattern recognition was investigated for classification of events acquired digitally from radiation detection instruments. In specific, waveforms were analyzed to separate different particle classes from an air source containing radon and plutonium. The rationale behind this study was to explore the possibility of performing direct digital analysis on these waveforms — acquired at sampling rate of 50 MHz — to classify events in a manner previously not possible by completely analog instruments. This system extracts key multidimensional features from each output pulse waveform and performs an analysis of those features. In the training phase, a cluster analysis was performed on vectors that consisted of features of the pulse shapes thought to be important for classification. In the recognition phase, pulse waveforms were analyzed to determine which population the events were drawn from. Results from multi-feature analysis suggest that methods using more detailed information than just pulse height or area offer considerable advantages over existing instruments.

This work was supported by U.S. Department of Energy under contract AA46420T.

2. THE CLASSIFICATION SYSTEM

This section presents the structure of the digital waveform classification system and the method to implement this system. As shown in Figure 1, the training classification system consists of five sub-systems, namely, a preprocessor, a selection of measurement features, a cluster analysis, a canonical discriminant transformation and an optimal decision function. The implementation of classification system is given in Figure 2. The functionalities of those sub-systems are described in the following sub-systems

2.1. PREPROCESSOR

The major task of the preprocessor is first to filter out 'bad' or double digital waveforms and then to extract measurement features from each digital waveform. Even this stage of analysis is something analog methods can not do. Each of the measurement features indicates a particular measurable characteristic of the digital waveform. In general, these measurement features do not share the same units due to the fact that they describe the physical objects from various aspects. In order to enable the application of sub-systems to make the en-

0-7803-2431-5/95 $4.00

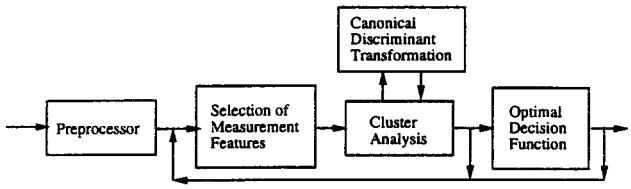

Figure 1: A training block of multi-feature based unsupervised pattern classification system.

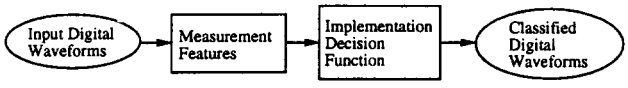

Figure 2: A block of implementation classifier.

ture feature space roughly 'circular' in spread, the values of the output features from the preprocessor are normalized.

2.2. SELECTION OF MEASUREMENT FEATURES

It has been showed [3] that there is a 'peaking' phenomenon in the finite-sample-size case. That is, misclassification error initially drops with addition of new features, then attains a minimum, and then begins to increase. The number of features at which the expected the probability of misclassification is minimal is called the optimal number of features [4]. This number depends on the design sample size, the type of classification rule, the class-conditional distributions of the pattern vector, and on the effectiveness of features and their ordering. Therefore, it is necessary to select the subset of discriminatory features from the initial features in such a way that they best reveal difference among the digital waveform classes and produce a minimum probability of misclassification error.

In our study, the selection subset of discriminatory features from the initial features of the digital waveforms is based on a statistical technique called stepwise discriminant analysis (SDA). The SDA procedure selects a subset of quantitative features using stepwise selection under the assumption that the within-class distributions are approximately multivariate Gaussian. The final measurement features chosen by SDA along with cross-validation is usually a good subset of possible features. We refer to this SDA as a pre-dimensional reduction technique.

2.3. CLUSTER ANALYSIS

A commonly used partitional clustering approach is the K-means algorithm. This algorithm evaluates the proximity between groups using the Euclidean distance between group centroids. The K-means algorithm is based on the minimization of a performance index which is defined as the cluster center. This iterative algorithm is guaranteed to converge to a locally optimal clustering and has been extensively addressed in the literature [5]. Therefore, further details of the K-means algorithm are omitted herein.

2.4. CANONICAL DISCRIMINANT TRANSFORMATION

The canonical discriminant transformation (CDT) is a dimensional reduction technique related to principal components transformation (PCT). This technique has certain maximal properties similar to the PCT [6]. However, whereas PCT considers interrelationship within a set of variables, the focus of CDT is on the relationship between two groups of variables.

When two or more groups of observations with measurements on several quantitative variables are provided, CDT can obtain a linear combination of the variables that summarizes between-class variation in much the same way that PCT summarizes total variation. The highest possible multiple correlation with the groups is called the first canonical correlation. The coefficients of the linear combination are the canonical coefficients. The first canonical component is defined by the variable of the linear combination.

The second linear combination uncorrelated with the first canonical component that has the highest possible multiple correlation with the groups is considered as the second canonical correlation. Until the number of canonical component equals the number of classes minus one, the process of extracting canonical component can be repeated. The detail mathematical treatments of the CDT can be found in reference [6].

An advantage of the CDT is that it not only has certain maximal properties similar to the PCT but it also can take class discrimination into consideration. Indeed, this procedure preserves the separation between the classes.

A disadvantage of the CDT is that it needs the pattern samples with label information before this approach can be employed. In other words, in the unsupervised pattern recognition, this approach can be applied only after a clustering algorithm has been employed on the data set.

2.5. OPTIMAL CLASSIFIER

When a large number of pattern vectors are available, it is reasonable to assume that the class density function of the feature vectors for each class is multivariate Gaussian density. Then the class density functions of the feature vectors for each class can be estimated from the pattern vectors and the Bayes' classifier that minimizes the probability of misclassification error can be derived.

Let $\hat{p}(\omega_i)$ be the estimated a priori probability, $\hat{\mathbf{m}}_i$ the estimated mean vector, and $\hat{\mathbf{C}}_i$ the estimated covariance matrix of the ith class, where i = 1, 2, ..., N. The quadratic classifier assigns a pattern vector $\mathbf{x}$ to the training digital waveform associated with the maximum likelihood. Thus the expressions,

$$\hat{f}_i(\mathbf{x}) = \frac{\hat{p}(\omega_i)}{\sqrt{|\hat{\mathbf{C}}_i|}} exp\left[-\frac{1}{2}[(\mathbf{x}-\hat{\mathbf{m}}_i)^T \hat{\mathbf{C}}_i^{-1}(\mathbf{x}-\hat{\mathbf{m}}_i)\right], \tag{1}$$

define a set of discriminant functions that optimally assign training digital waveform inputs to the correct distribution. Finally, the implementation of classification rule is

$$\text{choose } k\text{th class} \quad \text{if } \hat{f}_k(\mathbf{x}) = max_i[\hat{f}_i(\mathbf{x})]. \tag{2}$$

Taking the natural log of Equation (1) and multiplying by two yields an equivalent set of discriminant functions, $\hat{D}_i(\mathbf{x})$, given by

$$\hat{D}_i(\mathbf{x}) = 2\ln[\hat{p}(\omega_i)] - \ln|\hat{\mathbf{C}}_i| - (\mathbf{x}-\hat{\mathbf{m}}_i)^T \hat{\mathbf{C}}_i^{-1}(\mathbf{x}-\hat{\mathbf{m}}_i), \tag{3}$$

where $(\mathbf{x}-\hat{\mathbf{m}}_i)^T \hat{\mathbf{C}}_i^{-1}(\mathbf{x}-\hat{\mathbf{m}}_i)$ is the Mahalanobis distance from the pattern vector $\mathbf{x}$ to the mean vector associated with ith class. Note that the middle term on the right-hand side of Equation (3) represents the difference between statistical correlation and likelihood partitions. This term contains the information about the spread of each distribution in the multidimensional signal space.

3. DATA DESCRIPTION

The training data set was 149,988 samples consisting of 1,435 digital waveforms sampled at 50 MHz with 10-bit resolution. These were obtained from a radiation detector of an alpha-constant air monitor with radon and plutonium sources. Figure 3 shows a sample plot of waveforms acquired from a radiation detector. No a priori label information was provided in the training data set. Knowledge of the underlying physics dictates there should be three distinct populations of events.

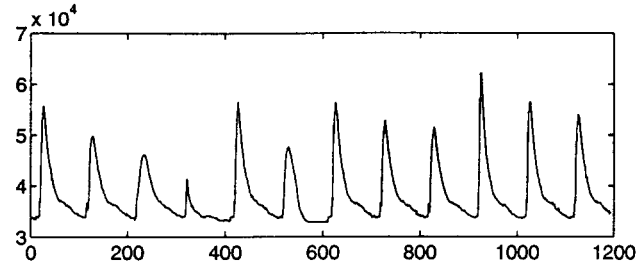

Figure 3: A plot of originally digital waveforms from a radiation detector of an alpha-constant air monitor with radon and plutonium sources.

4. EXPERIMENTAL RESULTS

An initial algorithm was developed to extract 15 measurement features from each digital waveform of incoming radon and plutonium events. The final selection of the subset of quantitative key features from an initial 15 measurement features was based on a stepwise discriminant analysis. Seven-dimensional measurement features that best revealed differences among digital waveform classes were finally generated.

The goal of the cluster analysis was to find a way of parametrically describing the digital pulse waveforms that would form distinct clusters in a multidimensional feature space. A modified K-means clustering algorithm was employed on the 50 MHz sampled data from the alpha-constant air monitor with the radon and plutonium source. The results of this analysis for 50 MHz sampled data are shown in the Table 1 (a) and (b). As can be seen, the within-cluster scatter and the between-cluster distances suggest that the clustering resulted in strong separability. In this case, the probability of misclassification error would have been reasonably low if Gaussian statistics are assumed.

The covariance of the multidimensional feature space has a complex structure, making visual interpretation of the digital waveforms more difficult. For our case, given a cluster variable and seven quantitative measurement features, a CDT was applied to derive a linear combination of the quantitative features (called canonical variables) that have 97.43% multiple correlation with the three clustering groups. The second canonical correlation (80.90%) is obtained by finding the linear combination uncorrelated with the first canonical variable. The first and second canonical variables together accounted for 100% of the total variation among the seven measuring features. This analysis can simplify the structure of the covariance and has been applied to the sampled data to make visual interpretation of

Table 1(a) Cluster Summary				
Cluster	No. of Vectors	RMS Std Deviation	Nearest Cluster	Centroid Distance
1	107	0.8786	2	6.0332
2	1256	0.6223	3	5.5381
3	72	0.8999	2	5.5381

Table 1(b) Distance Between Cluster Means			
Cluster	1	2	3
1	0.0000	6.0332	7.7055
2	6.0332	0.0000	5.5381
3	7.7055	5.5381	0.0000

cluster graphs more straightforward and to assess the results of clustering performance. Figure 4 shows a clustering graph for dimensionality reduction in the 50 MHz sampled data after employing the canonical discriminant technique. This again demonstrated that the three clusters are fairly separable.

The estimated probability of misclassification error of the Bayes' optimal classifier was evaluated based on the a posteriori probability of membership in each cluster. The projected accuracy of the Bayes' Gaussian model to classify three classes of pulse waveforms is 98.33% in overall.

An added benefit of the digital technique is that it can be used to detect corrupted pulse waveforms (such as Compton scattering or overlapping events) and eliminate them from the aggregate statistics. This property is particularly important in analyzing radiation sources with high count rates.

5. CONCLUSION

We have presented an novel unsupervised pattern recognition system to analyze and classify digital waveforms from a radiation detector of an alpha-constant air monitor with radon and plutonium source. The preliminary results of this new technique demonstrate clearly improved measurement conditions and are therefore very promising. This system not only improves current assessment methods of digital pulse waveforms but also provides a new useful tool in detection for digital radiation spectroscopy.

6. ACKNOWLEDGMENTS

Many thanks are due to Mr. D. Mackenzie C. Odell for helpful discussions and assistance in collecting all the digital waveform data.

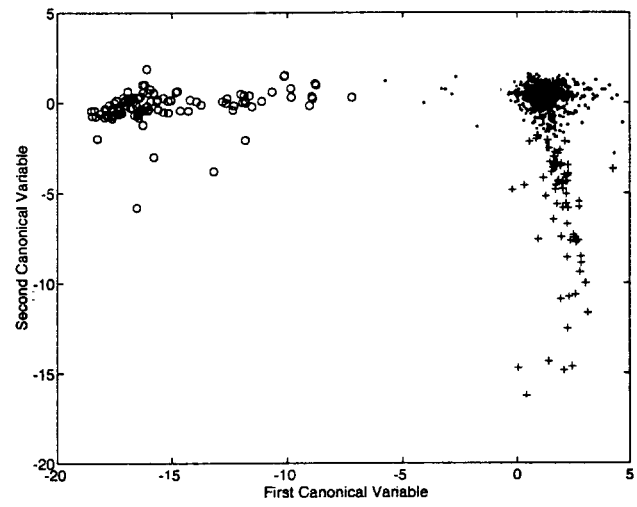

Figure 4: Clustering graph of dimensional reduction. o and + represent digital waveforms from radon. · represents digital waveforms from plutonium.

7. REFERENCES

[1] B. Zoghi, Y. Lee and D. C. Nelson, 'A Low-Cost Multichannel Analyzer with Data Reduction Assembly for Continuous Air Monitoring System,' IEEE Trans. Nucl. Sci., Vol. 39, No. 2, pp. 299-302, April, 1992.

[2] H. Takahashi, S. Kodama and J. Kawarabayashi, 'A New Pulse Height Analysis System Based on Fast ADC Digitizing Technique,' IEEE Trans. Nucl. Sci., Vol. 40, No. 4, pp. 626-629, 1993.

[3] K. Fukunaga and R. R. Hayes, 'Effects of Sample Size in Classifier Design,' IEEE Trans. Pattern Anal. Machine Intell., Vol. 11, No. 8, pp. 873-885, 1989.

[4] S. J. Raudys and A. K. Jain, ' Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners,' IEEE Trans. Pattern Anal. Machine Intell., Vol. 13, No. 3, pp. 252-264, 1991.

[5] J. T. Tou and R. C. Gonalez, *Pattern Recognition Principles*, Addison-Wesley publishing company, Fourth Printing, Massachusetts, 1981.

[6] K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis*, Academic Press, Inc., Third Printing, 1982.