UCRL-JC-126740
PREPRINT

Event Identification from Seismic/Magnetic Feature Vectors– A Comparative Study

J. K. Wolford, Jr.

This paper was prepared for submittal to the
11th Annual Conference on Aerospace/Defense
Sensing, Simulation and Controls
Orlando, FL
April 23, 1997

April 1997

Lawrence Livermore National Laboratory

MIMS Signature Analysis Task

Event Identification from Seismic/Magnetic Feature Vectors— A Comparative Study

James K. Wolford, Jr.
Proliferation Detection and Defense Division
Lawrence Livermore National Laboratory

Submitted to the 11th Annual Conference on Aerospace/Defense Sensing, Simulation and Controls
April 23, 1997

This report supports the Modular Integrated Monitoring System Project
Funded by the U.S. Department of Energy
Office of Nonproliferation and National Security

Lawrence Livermore National Laboratory
NAI Directorate/J Division
University of California • P.O. Box 808 • Livermore, CA • 94551-9900

Event Identification from Seismic/Magnetic Feature Vectors—A Comparative Study

James K. Wolford, Jr.[†]

Lawrence Livermore National Laboratory, MS L-183, Livermore CA 94551

ABSTRACT

The event identification problem plays a large role in the application of unattended ground sensors to the monitoring of borders and checkpoints. The choice of features and methods for classifying features affects how accurately these classifications are made. Finding features which reliably distinguish events of interest may require measurements based on separate physical phenomena. Classification methods include neural net versus fuzzy logic approaches, and within the neural category, different architectures and transfer functions for reaching decisions. This study examines ways of optimizing feature sets and surveys common techniques for classifying feature vectors corresponding to physical events. We apply each technique to samples of existing data, and compare discrimination attributes. Specifically, we calculate the confusion matrices for each technique applied to each sample dataset, and reduce them statistically to scalar scores. In addition, we gauge how the accuracy of each method is degraded by reducing the feature vector length by one element. Finally, we gather rough estimates of the relative cpu performance of the forward prediction algorithms.

Keywords: Signal processing, neural network, data fusion, event identification

1. APPLICATION DOMAIN

The MIMS (Modular Integrated Monitoring System) program at Livermore supports development of sensors for unattended detection applications. A challenge common to all of these applications is the need to discriminate local man-made events from environmental noise which may be partially coherent, and to do so within the confines of the power budget of an unattended sensor. Such applications may require the sensor to operate continuously for months or years before replacing the power supply. This affords very little energy per calculation and limits the complexity of the techniques one may bring to bear on the collected data in order to extract decisions. The low power requirement places a similar constraint on memory size, since the lowest power microprocessors have small word sizes and therefore small address spaces. One sensor system employs a combination of seismic and magnetic detectors, in an effort to discriminate among 4 events: a passing person, a passing group of people, a passing vehicle, and anything else, including background. This set of events is well suited for detecting and identifying traffic moving past important checkpoints or across borders. The small size of the event space makes it a convenient real world example for the testing of extraction and classification techniques. The sensor accepts signals on 4 channels from up to 4 detectors, allowing enough flexibility so that a broad range of scenarios may be accommodated. The in-sensor software was designed so that the portion of the data manipulation which is specific to the physical scenario is contained in a small, localized block of code. Since the identification process includes discriminating humans from metallic objects, one expects that a magnetometer would be useful as part of the detector set. Combining or fusing the extracted information from the detectors enhances the discriminating power of the overall system. To stay within the power budget, the algorithm used to extract features must necessarily be simple, as must the feature classifier.

2. DATA DESCRIPTION

The example sensor uses a single axis 10 Hz geophone with a useful range of 30 meters (dictated by the minimum signal-to-noise ratio) and a two axis magnetometer. Time series data are gathered from both types of detectors at up to 50 Hertz sampling rate. The seismic data are rectified prior to the extraction of features. Figure 1 shows time series of the sampled geophone measurement corresponding to each of the candidate events. The scale of the plotted quantities shows variation in the magnitude and duration of each of the events.

[†] Email: wolford@llnl.gov; Telephone: (510) 422-7236; Fax: (510) 422-9343

This implies that to be useful, a classification system must be insensitive to these gross properties, relying instead on smaller scale characteristics of the signals.

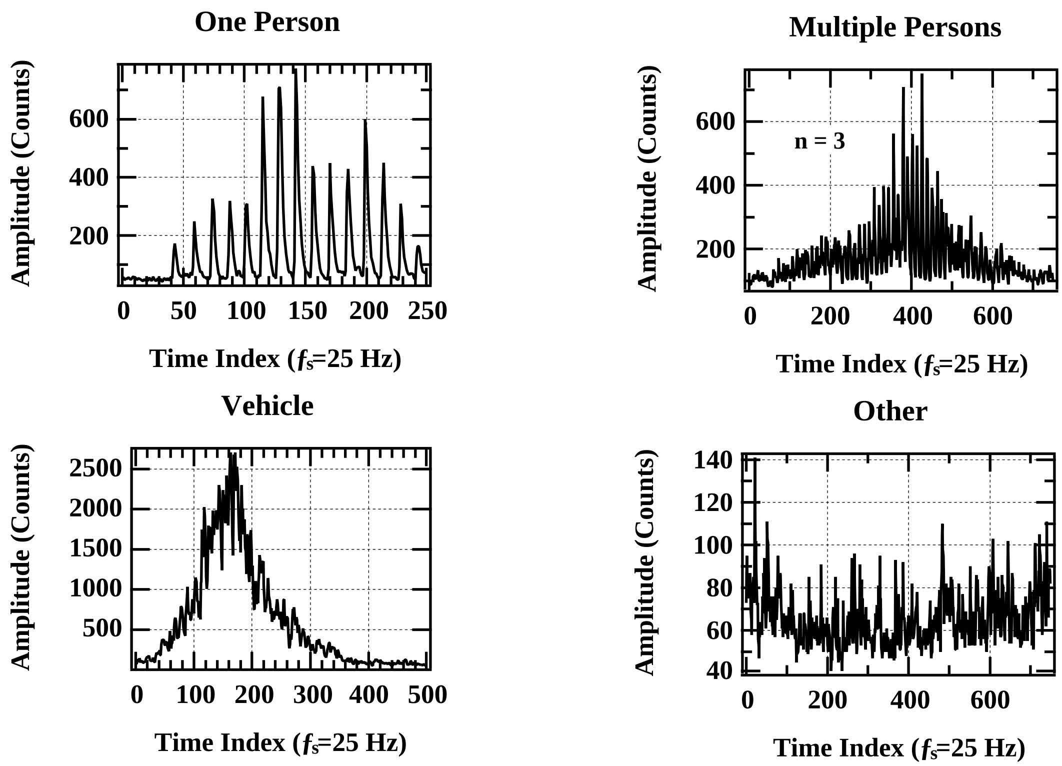

Figure 1. Gain corrected geophone measurements representing the full set of possible events. The shapes of the time series envelopes reflect the similarity between the vehicle and the multi-person cases, and challenge the designer to select features which distinguish them. Including a magnetic feature in the ensemble—one that is influenced by a passing metallic object—improves the classifier performance, as reflected in the calculated scores given in Table 1.

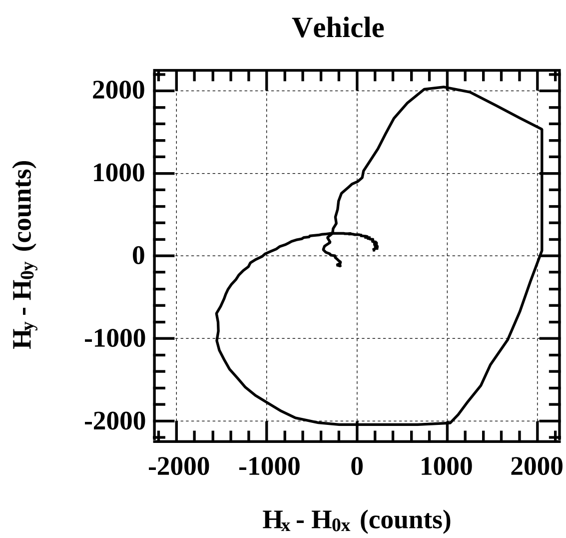

Figure 2. X and Y components of the magnetic field trace out a trajectory over the 6 second duration of a passing vehicle event. The "area" enclosed by the curve indicates the magnitude of the magnetic disturbance, which depends on the components of the dipole moment of the source along the magnetometer axes, and on the source's proximity. The algorithm uses equation (1). The calculation yields one value for the entire history of the triggered event, and its sign indicates the direction of motion of the source. Note that the ADC is driven into saturation on both axes. Saturation is common in scenarios where broad ranges of possible target distance require high detector sensitivity, but

saturation does not degrade the result, since the classification process depends only on the value of the area reaching or surpassing a threshold.

As a basis for comparison, we have also included an unrelated set of data which is published openly and used in classification research. The iris dataset (Fisher, 1988) came from the iris plants database, which is posted anonymously on several internet hosts and is widely used as sample data for studying neural and fuzzy techniques. Many publications in the open literature cite it as example data, and it has been exhaustively analyzed. Like the data collected in the domain of this project, it consists of four element feature vectors. The quantities themselves are length measurements of various parts of the flowers. The "events" are the three varieties of iris, "setosa", "versicolor", and "virginica". The additional dataset labeled “iris-3” below was derived by simply omitting one feature, the sepal length, from the ensemble.

3. FEATURE EXTRACTION

Several factors determine the best choice of features to extract from data. The two largest considerations are discrimination value and ease of computation. To assess the efficacy of a feature for discrimination, one typically evaluates the performance of the full feature set, them systematically deletes one feature at a time from the ensemble and repeats the analysis. Using this technique, careful analysis by Pearson at LLNL on data collected previously yielded a suite of candidate features for seismic recordings. They perform comparably, depending on the specific data set one extracts them from, and so no attempt was made to shorten the list. They share in common the virtues of fast computation and low memory requirement. They are as follows:

1. The duration of the triggered event (sample length divided by sampling rate).
2. The duration of a triggered event rising above a computed threshold.
3. The fraction of points within the highest 4 second window where the signal rises sharply through a 50 point moving average.
4. The mean number of samples between rising crossings, defined as instances where the present point rises above the moving average by a prescribed margin and the two previous points are below the moving average.
5. The standard deviation of the number of samples between rising crossings.
6. The mean slope at rising crossings.
7. The standard deviation of the slope at rising crossings.
8. The area under the “superior envelope” of the time series divided by the area under the 50 point moving average. The superior envelope is a derived function whose value is never *less* than the value of the time series, and whose value relaxes exponentially toward the instantaneous value of the time series with a time constant of 0.78 second (e-fold).
9. The area under the “inferior envelope” of the time series divided by the area under the 50 point moving average. The inferior envelope is a derived function whose value is never *greater* than the value of the time series, and whose value rises exponentially toward the instantaneous value of the time series with a time constant of 0.78 second (e-fold).
10. The area under the inferior envelope divided by the area under the superior envelope.
11. The fraction of measurements which exceed the 50 point moving average.
12. The peak (maximum) of the rectified signal
13. The area under the signal for the full duration of the event.

In the current analysis, datasets labeled “test1” and “test2” (see Table 1 below) were composed of features 1, 12, and 13, plus the magnetic area feature. The dataset labeled “test3” used features 4, 6, 8, and 9 and omitted the magnetic feature.

The output of a magnetometer transducer varies in a complicated way with the motion of a magnetic dipole in a background of fixed magnetic clutter. Efforts to model this behavior analytically proved fruitless. Instead, attention focused on gross characteristics of the two dimensional field vector which varied in a more consistent way, and in particular those which could be used to indicate whether motion of the dipole proceeded from left to right or from right to left. In the present work, we define the “magnetic area” as the sum of the cross products of successive field vectors:

$$A_H = \sum_{i=1}^{N-1} \left| \vec{H}_i \times \vec{H}_{i+1} \right| \tag{1}$$

Where i is the time index and N is the duration of the event in samples. This quantity depends in a reproducible way on the magnitude of and distance to the magnetic dipole moment of the moving target. Furthermore, its sign indicates the direction of motion, i.e. the trajectory in Figure 2 proceeds clockwise for one direction and counterclockwise for the opposite direction.

More sophisticated features may be accessible to sensor systems whose power budgets are more liberal than that of the system under study. For example, a wavelet based feature could take advantage of the fact that the events of interest are localized in both frequency and time. Discrete wavelet transforms of the time series exhibit high peak and average values when the wavelet family is well matched to the time scale of the recorded event. Wavelet families have been surveyed using the data treated in this study, and suitable matches found, however corresponding wavelet features have not yet been implemented in a classifier.

4. EVENT CLASSIFICATION

Once features are identified and extracted, classification may proceed. This section lists the techniques applied and a brief description of their attributes. Readers are referred to the references for more complete explanations.

4.1. Simple 1 and 2 Layer Perceptron Nets

The simplest technique uses a single layer of neurons with hard limit transfer functions, which are trained on examples of correct behavior. The single layer architecture was used to demonstrate—by its failure—the lack of linear separability for all of the datasets surveyed. The most obvious enhancement adds a second layer to pre-process the inputs and transform them into a linearly separable representation. The enhanced architecture usually fails for datasets of any complexity because it depends on a fortuitous initial choice of weights in the first layer.

4.2. Backpropagation with hidden layer(s)

Three (or more) layer networks which are trained using the error backpropagation algorithm allow the network to separate classes from complex overlapping patterns in feature space. The presence of a hidden layer removes all sensitivity to the lack of linear separability or an unfortunate choice for the initial weights. A log-sigmoid replaces the hard limit transfer function of the simple perceptrons and yields superior results. This architecture represents a standard in networks because of its performance, speed, and generality.

4.3. Simple Competitive Learning Network

In a competitive learning network[3], the neuron weights distribute themselves to recognize frequently presented feature vectors. The network consists of a single competitive layer containing $2M$ neurons, where M is the number of events measured. The synaptic weights are initialized randomly. The activation function maps to a unit vector whose values vanish in all dimensions but one, thus the output of a simulation event will be zero for all neurons except the "winner" whose value is 1. The training process does not require a training signal, and so the technique’s performance may be used as a criterion for discovering the most strongly discriminating features among those available to the designer.

4.4. Self Organizing Map

Like the competitive learning network, the self organizing map[4] (SOM) is an unsupervised technique which does not use a training signal to derive weights. Instead during classification, it computes a discriminant function for all neurons and identifies the highest function value. It then “activates” the selected neuron and its neighbors. (Possible output values are 0, 0.5, and 1.) It uses an adaptive process to strengthen the discriminant value of the activated neurons in response to the inputs. We estimated the number or neurons required for adequate performance to be M^n, where again M is the number of possible events, and n is the length of the feature vector. Testing on the relevant datasets indicates that using significantly fewer than this amount of neurons will fail to produce a successful classifier, and that a 4 dimensional map performs no better than the traditional 2D architecture. The obvious disadvantage of this estimate is a loss of computational efficiency, since only a portion of the neurons allocated for training actually remain live, and can thus participate in the classification operation. This efficiency penalty may be offset by the ability of the SOM to accommodate new event classes in situ. Also, as in the case of the simple competitive learning network below, the distinctiveness with which event classes emerge during the training process, can be used as a figure of merit for the content of the feature vector.

4.4. Learning Vector Quantization

A learning vector quantization (LVQ) network[5] consists of a competitive layer coupled to a linear output layer. The competitive layer (a version of which is described above) classifies the feature vectors into intermediate or "subclasses" which are mapped by the linear output layer to the user's target classes. LVQ networks work especially well for data which are not linearly separable. Since this version of the LVQ algorithm (known as LVQ1) does utilize a known training signal as part of the training process, it should be expected to perform well when presented with feature vectors used in training.

4.1 Fuzzy Cluster Means

Fuzzy c-means is a generalization of the c-means clustering technique for determining cluster centers and membership. In it, one minimizes a cost function, which depends on the membership value of each vector and its norm taken with respect to each cluster center. The membership function

$$J = \sum_{k=1}^{N} \sum_{i=1}^{M} u_{ik}^{m} \left\| x_k - v_i \right\|^2 \tag{2}$$

where the x_k are the feature vectors, the v_i are the cluster centers, N is the number of feature vectors, M is the number of cluster centers (= number of events), and the u_{ik} are the degrees of membership defined as

$$u_{ik} = \frac{1}{\sum_{j=1}^{M} \left(\frac{\left\| x_k - v_i \right\|}{\left\| x_k - v_j \right\|} \right)^{2/(m-1)}} \tag{3}$$

where in a rough sense, the m (>1) governs the fuzziness of the set membership, with a higher m denoting more fuzziness. The algorithm used in this study was written by Bezdek[2] (1981). It converges either to a local minimum or to a saddle point of the cost function. The quality of the FCM solution depends strongly on the initially chosen set of cluster centers, however in practice, the technique yielded highly reproducible results.

5. ANALYSIS

The figure of merit for this system is the degree of ability to classify events properly. A straightforward representation for this quantity is the confusion matrix. The confusion matrix is constructed by sorting the known events into columns whose rows consist of the predicted events. The set of predicted events is generated by submitting the input data to the trained analyzer. Figure 3 below shows an annotated example using the results from fuzzy c-means analysis of the Fisher iris data.

$$\begin{array}{cc} & \begin{array}{ccc} \text{Setosa} & \text{Versicolor} & \text{Virginica} \end{array} \\ C = & \begin{pmatrix} 50 & 0 & 0 \\ 0 & 48 & 4 \\ 0 & 2 & 46 \end{pmatrix} \begin{array}{l} \mathit{Setosa} \\ \mathit{Versicolor} \\ \mathit{Virginica} \end{array} \end{array}$$

Figure 3. A sample confusion matrix for the iris dataset as produced by the learning vector quantization (LVQ) method. The columns represent the known events (training input) and the rows contain the derived events obtained using the trained net on the training data. Note that the first event, "setosa" is linearly separable from the other two.

Following this prescription, a perfect classifier would yield a diagonal matrix. The confusion matrix of a real world classifier will in general include some off diagonal elements. If the sum of the off-diagonal elements in a given column exceeds the diagonal element, then the probability is higher that the algorithm will yield an incorrect classification than a correct one for that column's event, and the trial is considered a failure. A compact measure for the reliability of the classifier is the trace of the confusion matrix. Since different data sets in general contain differing numbers of data records, for the trace result to be meaningful, they must be normalized. Below we define scores, giving the expectation for the given method that an event will be classified properly. Score S_1 gives the mean trace divided by the maximum possible trace, which is equivalent to the number of feature vectors presented.

$$S_1 = \frac{\sum_{k=1}^{P} \frac{Tr(C)_k}{P}}{N} \tag{4}$$

where P is the number of successful training trials, N is the number of vectors in the training set, and the trace Tr is defined by

$$Tr(C) = \sum_i C_{ii} \tag{5}$$

If we may assume that future field modules will be pre-trained, then score S_2 corresponding to the maximum trace is the more significant quantity.

$$S_2 = \frac{\max[Tr(C)]}{N} \tag{6}$$

Score S_3 is the quotient of successful trainings over total trainings. This quantity gives an indication of the ease of training, and may indicate how close a method is to failing altogether.

$$S_3 = \frac{P}{R} \tag{7}$$

where R is the total number of training trials.

These quantities are tabulated in this order for the various methods in Table 1 below.

6. RESULTS

Inspection of the results in Table 1 shows that fuzzy c-means performs adequately only for the iris dataset and otherwise does not compete well with the other techniques included in this study. Attempts to apply this technique included a broad parameter study in the partitioning exponent, *m*. The fcm technique seems easily thwarted by a lack of linear separability in the data. As the remaining entries show, competitive learning, while fast, does not perform as well as the other techniques. The three layer perceptron and learning vector quantization show the highest consistent success rate, followed by the self organizing map. The second data column contains results for the iris dataset, but were obtained using only the first 3 feature vector elements. Note the slight degradation in the accuracy of discrimination over the corresponding four element computation. Similarly, the “test2 3-element” column shows lower scores as a result of omitting the magnetic feature.

7. CONCLUSIONS

This study treated events which are largely distinguishable using seismic feature classification alone. The single field deletion exercise showed that adding the magnetic area feature does improve discrimination between the passing vehicle event and the remaining events. One may expect further improvement by adding a feature specific to humans in motion, such as the area under a passive infrared signature.

A useful system for this application must be flexible. This clearly emerges as a concern for some of the methods treated in this study. Fuzzy c-means performs well where the clusters overlap only slightly, but will probably require a more favorable feature space distribution to succeed more generally. The self-organizing map appears to require many more neurons to produce a successful training than actually remain live to form the predictive network. Though it consumes a much larger share of computing resources than the other neural techniques, the self organizing map, like the competitive learning technique, may nevertheless be useful during the design phase for selecting features through preferred topological groupings. Furthermore, if the application requires in situ training, this may well be the best alternative. The most successful and most robust methods overall are the 3 layer perceptron and learning vector quantization, followed by the competitive learning layer, which may be thought of as precursor to LVQ. Of these two, the 3 layer perceptron is clearly the more robust.

Future work will further develop the multilayer perceptron, exploring the sensitivity of performance to the network architecture. It will also optimize the LVQ method for application in the project domain, and examine ways of making the feature vector more amenable to classification with this technique.

Performance Scores S_1, S_2, and S_3 for Several Classification Techniques

Method//Dataset	*iris (N=150)*	*iris - 3 elements*	*test1 (N=1530)*	*test2 (N=1255)*	*test2 - 3 elements*	*test3 (N=320)*	*relative cpu load*
single layer perceptron	fails	fails	fails	fails	fails	fails	n/a
2 layer perceptron	fails	fails	fails	fails	fails	fails	n/a
3 layer perceptron w/ backpropagation	0.975 **0.993** 1.00	0.941 **0.967** 1.00	0.719 **0.757** 1.00	0.730 **0.783** 1.00	0.673 **0.723** 1.00	0.806 **0.814** 1.00	2.16
Simple competitive learning network	0.946 **0.960** 1.00	0.879 **0.913** 0.969	0.505 **0.590** 0.252	0.556 **0.654** 0.300	0.537 **0.619** 0.172	0.621 **0.724** 0.156	1.00
self organizing map	0.907 **0.980** 0.900	0.888 **0.967** 0.938	0.592 **0.620** 0.448	0.647 **0.729** 0.415	0.612 **0.667** 0.234	0.805 **0.910** 0.990	30.2
learning vector quantization	0.959 **0.973** 1.00	0.936 **0.960** 1.00	0.666 **0.699** 0.710	0.711 **0.758** 0.930	0.657 **0.704** 0.516	0.766 **0.796** 0.969	2.04
fuzzy c-means	— **0.960** 1.00	— **0.880** 1.00	— **0.474** 1.00	— **0.516** 1.00	— **0.509** 1.00	— **0.458** 1.00	

Table 1. A compilation of the results of applying a subset of the available techniques to samples of data collected to date. Failure of a method indicates either lack of convergence of the algorithm (as in the case of simple perceptrons) or the method's inability to produce a confusion matrix which demonstrates successful discrimination. Some instances of mild failure (scores just under 50%) are also shown. The table entries consist of scores S_1, S_2 and S_3, indicating the degree of success of each method, as defined in section 4, equations 4-7. The maximal score, S_2, (in bold) is the most relevant measure for systems intended to be fully trained prior to deployment. The rightmost column gives the relative cpu requirement for classifying a single feature vector using the trained net. Note that these measurements took place on interpreted scripting code, which may have contained differing amounts of vectorization at low level. One may therefore expect these results to differ for compiled scalar code.

ACKNOWLEDGMENTS

This work was completed by Lawrence Livermore National Laboratory for the U.S. Department of Energy under Contract W-7405-ENG-48. The author also wishes to acknowledge and thank Dr. Peter K. Pearson for providing the archival data and source code for the back propagation neural network technique.

REFERENCES

[1] Chiu, S. (1994):"Fuzzy Model Identification based on Cluster Estimation" *Journal of Intelligent and Fuzzy Systems*, (2) pp 267-268

[2] Bezdek, J. C. (1981): Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum Press, p 69.

[3] Carpenter, G. A. (1991): "Neural Network Models for Pattern Recognition and Associative Memory", in Pattern Recognition by Self-Organizing Neural Networks, Carpenter and Grossberg, Ed., Cambridge MA: MIT Press, p22.

[4] Kohonen, T. (1987): Self-Organization and Associative Memory, 2nd Ed, Berlin: Springer-Verlag, p 69.

[5] Haykin, S. (1994): Neural Networks, A Comprehensive Foundation, New York, Macmillan College Publishing Company, p 429.

Technical Information Department • **Lawrence Livermore National Laboratory**
University of California • Livermore, California 94551