Underwater Target Classification in Changing Environments Using an Adaptive Feature Mapping

Mahmood R. Azimi-Sadjadi, *Senior Member, IEEE*, De Yao, Arta A. Jamshidi, and Gerry J. Dobeck

***Abstract*—A new adaptive underwater target classification system to cope with environmental changes in acoustic backscattered data from targets and nontargets is introduced in this paper. The core of the system is the adaptive feature mapping that minimizes the classification error rate of the classifier. The goal is to map the feature vector in such a way that the mapped version remains invariant to the environmental changes. A K-nearest neighbor (K-NN) system is used as a memory to provide the closest matches of an unknown pattern in the feature space. The classification decision is done by a backpropagation neural network (BPNN). Two different cost functions for adaptation are defined. These two cost functions are then combined together to improve the classification performance. The test results on a 40-kHz linear FM acoustic backscattered data set collected from six different objects are presented. These results demonstrate the effectiveness of the adaptive system versus nonadaptive system when the signal-to-reverberation ratio (SRR) in the environment is varying.**

***Index Terms*—Adaptive classification, feature mapping, in situ learning, neural networks, underwater target classification.**

I. INTRODUCTION

DETECTION and classification of underwater targets from the acoustic backscattered signals pose several technical problems (see [1] for a good literature review). These problems are mainly attributed to factors such as nonrepeatability and variations of the target signature, environmental changes, presence of new nontargets, competing clutter caused by reverberation and biologics, and lack of any *a priori* knowledge about the shape and geometry of the nontargets.

Owing to the variations in the target signature and the environmental conditions, the feature space will clearly undergo some variations. The variations in the feature space to some extent can be tolerated by the neural-network classifier. However, if these changes become substantial, the classifier may no longer be able to capture the temporal and spectral properties of the varying signals, thus causing misclassifications and false reports. In addition, new threats and nontargets that were not in the original training environments may enter the field of view of the sensor. As a result, a fixed classification system cannot handle such drastic changes in the feature space. These, coupled with the effects of competing clutter caused by surface and volume reverberation, create a very difficult and challenging signal processing problem.

There are generally two classes of approaches that can be devised depending on the nature and extent of the feature space changes. In the first class, a flexible neural-network structure such as adaptive resonance theory (ART) [2]–[4] can be used to accommodate the changes in the feature space. As far as the ART network is concerned, the variations in the feature space can be handled by a proper selection of the vigilance parameter which determines the degree of coarseness or fineness of the categories. This network can provide stability of the established categories while offering flexibility needed in these situations. However, if the changes in the feature space caused by the environmental, operational, and target conditions are substantial, this network fails to form proper categories and accommodate the changes in a more adaptive fashion. The second class of approaches involves using a bank of classifiers each trained to handle a specific environmental condition. The main problem, however, is the decision on how and when to switch between the classifiers depending on the changes. This problem makes this approach impractical for real-life applications.

All the drawbacks listed above gave birth to the adaptive feature mapping idea in order to accommodate the environmental variations in a more effective way. The principal idea behind this approach is to provide the neural-network classifier with a set of features that are invariant or insensitive to the changes in the original feature space hence maintaining the performance. This on-line adaptive feature mapping is done without the need to modify the neural-network classifier. This is an important benefit of this scheme since on-line modification of the neural-network classifier could, over time, affect the previous training and erase the established categories. The study presented here can be applicable to a multitude of other classification problems that involve variations in the feature space.

In this paper, a new biologically inspired adaptive underwater target classification system based on the idea of feature mapping is proposed and studied in order to provide robust and efficient *in situ* classification under the above-mentioned conditions. The feature extraction scheme in [1] that uses wavelet packet subband decomposition in conjunction with linear predictive coding (LPC) method is employed here. The LPC coefficients of each subband signal are then computed and used as features to the classifier. To select those features with high discriminatory ability and to reduce the dimensionality of the feature space, the Fisher criterion for feature selection and reduction [5] is applied. The heart of the adaptive system is the adaptive feature mapping subsystem that maps the original feature space

Manuscript received April 1, 2000; revised July 9, 2001. This work was supported by the Office of Naval Research, Bisonar Program under Contracts N00014-99-1-0166 and N00014-01-1-0307.

M. R. Azimi-Sadjadi, D. Yao, and A. A. Jamshidi are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: azimi@engr.colostate.edu).

G. J. Dobeck is with the Signal/Image Processing, NSWC, Dahlgren Division, Coastal Systems Station, Panama City, FL 32407-7001 USA.

Publisher Item Identifier S 1045-9227(02)05566-2.

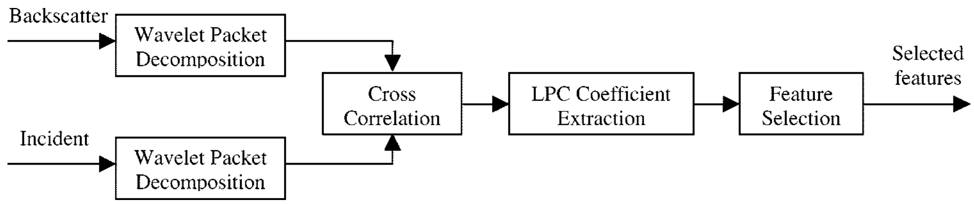


Fig. 1. Feature extraction and reduction processes.

such that the new feature vector remains invariant to the environmental conditions. The main goal is to minimize the classification error of the neural-network classifier on the changing data [6]. A backpropagation neural network (BPNN) is used as the nonadaptive (fixed) classifier. The feedback to the adaptation mechanism is provided by a K-NN, which is primarily used as a memory system to identify the K closest stored prototypes for an unknown pattern. In order to alleviate the problems caused by poorly scaled features and improve the overall performance, a scaled Euclidean distance K-NN [7] is used. Two different cost criteria are considered in the adaptive system. The first one is based upon the least-squares (LS) error criterion, whereas the other one is a two-dimensional (2-D) sigmoidal cost function. These two criteria are then combined together to improve the overall performance. The performance of the proposed system is examined on the 40 kHz acoustic backscattered data set collected from six different objects [1]. To test the robustness and ability of the system to adapt to new environmental conditions, the signal-to-reverberation ratio (SRR) was varied from 0 to 20 dB. The test results are then benchmarked against those of the nonadaptive fixed classification system. These results revealed that the performance of this adaptive classification system is much better than that of the nonadaptive classifier, especially at low SRR conditions.

The organization of this paper is as follows. Section II briefly describes the feature extraction and selection processes used in this study. In Section III, the description of the proposed adaptive target classification system and its components is provided. Section IV is dedicated to the derivations of the gradient descent-based updating equations for the two cost criteria. Section V will demonstrate and analyze the performance of the proposed adaptive classification system on the 40-kHz acoustic backscattered data set. Finally, Section VI gives some concluding remarks.

II. Feature Extraction and Reduction

The studies in [8]–[10] point to the interesting fact that the preprocessing and encoding of information in the cochlear of bats can be modeled relatively accurately by a filter bank consisting of several band-pass filters each tuned to pick up certain frequency information. This filter bank splits the frequency spectra for resolving different subtle spectral features while localizing the temporal information. In the spectrogram correlation and transformation (SCAT) model [8], the cochlear block operates in a spectrogram-like manner. Peak detection or threshold crossing afferents can then be applied in each filter channel prior to the spectrogram correlation process to determine the presence or absence of an event.

Inspired from these models, a wavelet packet-based scheme [11] is used to decompose the frequency spectra of the backscattered signals into several subbands that contain useful target information. The wavelet packet (WP) decomposition provides an optimal multiresolution decomposition of the signal spectrum in a manner very similar to the biological systems. The multiresolution property allows for capturing fine details or subtleties in the signals the same way as the zooming-in ability in the biological visual system in order to observe the small details in the detected objects. In WP decomposition, each subband extracts certain tonal features of the acoustic backscattered signals. Instead of the simple peak detection or threshold crossing in the SCAT model [8], the LPC method is employed in our model. LPC is widely used [12] for speech recognition applications as it provides an effective way of capturing the spectral peaks and behavior of these signals.

The feature extraction is then followed by a feature reduction process that reduces the dimensionality of the feature space according to the discriminatory ability of each selected feature. To select an appropriate set of features, a criterion function can be used to evaluate the discriminatory power of the individual features. In this study, the Fisher discriminant function [13] was used to evaluate the distance between the two classes for each feature, i.e.,

$$D_B(i,j) = \frac{|\mu_i - \mu_j|^2}{\sigma_i^2 + \sigma_j^2} \tag{1}$$

where μ_i and σ_i^2 represent the mean and variance of the features in class i, respectively. The features are sorted in a decreasing order of importance and the ones with higher discriminatory power were then selected to form the reduced feature vector. Note that the reduction of the subband features also removes noise to some extent. Fig. 1 depicts the overall block diagram of this feature extraction/selection system. For more detailed description, the reader is referred to [1].

III. Adaptive Target Classification System

Let us begin by posing two important questions. How do we perform information retrieval and association when we encounter an unknown or distorted pattern of an object? How does our memory get updated after this pattern is correctly recognized? When we encounter an unknown pattern our brain goes through cycles of information retrieval and association in order to recall and identify the corresponding pattern(s) already stored in the memory. It can be argued that the information retrieval process takes the form of an adaptive mapping of the unknown pattern to a pattern that is most familiar to the system hence recalling the closest match. The important point is that during this

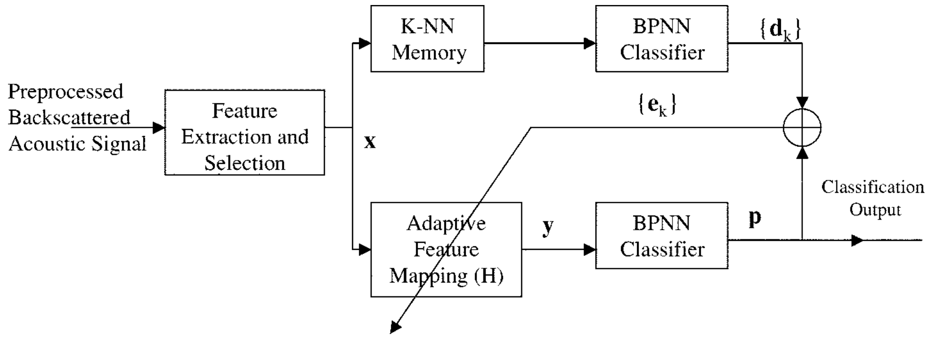


Fig. 2. Block diagram of proposed adaptive feature mapping system.

retrieval cycle the memory does not get modified or updated; rather it is the unknown pattern that gets manipulated in an iterative fashion until the "right" match is found. Updating the memory during this cycle would have had detrimental effects as it could cause loss of old information. Once the final decision is made and the pattern is associated and recognized, the memory can then be updated by including the "mapped" input pattern. This "soft updating" is performed without erasing or altering the previously established internal representation in the brain.

The adaptive classification system in Fig. 2 is designed to work in a somewhat similar fashion. Upon the application of an unknown pattern (e.g., target/nontarget pattern at different aspect/range or in different environmental condition), the system attempts to map the unknown pattern to a pattern already familiar to the neural-network decision making system. This mapping takes place iteratively and with the aid of the K-NN system, which plays the role of the memory to identify the potential matches from the memorized set of patterns. The matching operation in the K-NN is performed in the feature space and based upon some distance metric. The retrieval or recall process involves an iterative search or adaptation, in order to minimize the probability of error associated with the unknown pattern in presence of the memory by satisfying a certain performance criterion. This process in turn guarantees that the optimum match is reached.

The lower branch in this system performs feature mapping/manipulation using the mapping matrix H and decision making using a fixed BPNN classifier; whereas the function of the upper branch is to provide a set of possible matches and their corresponding decisions using a copy of the BPNN in the lower branch. Each pattern in the memory of the K-NN provides an item of "evidence" for possible class membership of the unknown pattern. However, since the distance in the feature space is not necessarily an effective measure for determining the class membership, the decision about the class membership of the unknown pattern is determined based upon the outputs of the BPNN for these K nearest neighbor patterns. More specifically, using the matrix mapping H, we choose to move the output of the BPNN in the lower branch to a location in the output space that maximizes the classifier's decision confidence according to the chosen cost criterion. For instance, one of the cost criteria chosen in Section IV moves the output of the BPNN to the centroid of all the outputs (i.e., average of all the labeled evidence) provided by the upper branch. In Section III-A, we shall show that this centroid indeed corresponds to the minimum of the classification error given the memory. Several possible scenarios are discussed in Section IV-D. In what follows, the functions of each branch of this system are described in more detail.

A. Upper Branch: Memory and Matching Processes

As mentioned before, the upper branch of the proposed system identifies the possible matches in the feature space depending on the distance metric utilized by the K-NN and then determines the associated decisions or labels using a copy of the BPNN.

For an applied unknown pattern $\mathbf{x}$, the K-NN is used as a memory to identify a set of K nearest neighbor patterns $\mathbf{x}_j$ for $j \in [1, K]$ as possible matches in the original training space. The training data set for this K-NN consists of a set of pairs $(\mathbf{x}_i, l_i)$ for $i \in [1, N]$ with $\mathbf{x}_i$ and l_i being the ith training pattern and the corresponding class label, respectively. The closeness is determined based upon a certain distance measure $d(\cdot,\cdot)$. In this particular application, K does not have to be an odd number and tie situations are allowed. The K-NN operates based on an assumption that most of the nearby (training) patterns of an unknown pattern are likely to be given the same classification label than the distant ones. Cases in which this assumption does not hold present difficulties for a K-NN [14]. In addition, when the number of training vectors is very large, i.e., the feature space is filled densely with samples, the empirical conditional class probabilities will be close to the true probabilities, and the performance of the K-NN asymptotically approaches the optimal Bayes classifier.

A useful feature of the K-NN that is very important for our application is that new patterns can be incorporated into the memory of the system and into the formed classes. This implies that an updated set of comparison pattern vectors is always available for the updating of the feature mapping matrix. Consequently, the changes in the environment, target signature and sensory system behavior can be incorporated into the K-NN as well as the adaptive feature mapping subsystem without the need to change the neural-network classifier responsible for final decision making. The incorporation of those new classified patterns that have high confidence will make the feature space denser hence improving the overall performance.

Several variants of K-NN with different distance measures have been suggested [7]. The standard K-NN uses the Euclidean distance measure $d(\mathbf{x}_i, \mathbf{x})$ between the training patterns $\{\mathbf{x}_i \,|\, \mathbf{x}_i \in R^n\}_{i=1}^N$ and the applied pattern $\mathbf{x}$ i.e., $d(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i - \mathbf{x})^T(\mathbf{x}_i - \mathbf{x})$, which defines a hyper-spherical region in the feature space. Clearly, this region does not take into account the statistical distribution of the data. Two generalization of this distance measure are the "scaled" and "Mahalanobis" distance measures [7] given, respectively, by

$$d(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i - \mathbf{x})^T(\mathrm{diag}(\Sigma))^{-1}(\mathbf{x}_i - \mathbf{x}) \tag{2}$$

and

$$d(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i - \mathbf{x})^T(\Sigma)^{-1}(\mathbf{x}_i - \mathbf{x}) \tag{3}$$

where Σ is the covariance matrix of all the patterns in the training data. The shape of the neighborhoods for both distance measures is hyper-ellipsoidal. The axes of the ellipses for the scaled Euclidean distance are parallel to the coordinate axes while for the generalized Mahalanobis Euclidean distance they are also rotated according to the distribution of the patterns. Our experimental studies indicated that for this application the scaled Euclidean distance is the better choice. Thus, in the sequel distance refers to the scaled Euclidean distance.

Once the possible candidate patterns have been selected, their corresponding decisions, $\mathbf{d}_j = F(\mathbf{x}_j)$, are determined by the BPNN in the upper branch (here $F(\cdot)$ represents the nonlinear function implemented by the BPNN). This BPNN classifier, which is a copy of the original BPNN classifier in the lower branch, is trained based on the same training data set used to train the K-NN. As mentioned before, the corresponding decision $\mathbf{d}_j$ for each $\mathbf{x}_j$ provides an item of evidence about the class membership of the unknown pattern $\mathbf{x}$ given by the K-NN. The main BPNN classifier in the lower branch also provides its decision, $\mathbf{p} = F(\mathbf{x})$, on this unknown pattern. If the confidence in the "collective evidence" provided by the upper branch is very strong, no updating is needed. Otherwise feature updating will take place.

A measure of confidence about each individual evidence $\mathbf{x}_j$ can be defined in terms of the probability that this evidence leads to the correct decision about the class membership of $\mathbf{x}$, i.e., $P(l = C_i, l_j = C_i \,|\, \mathbf{x}, \mathbf{x}_j)$, where l and l_j are the class labels of $\mathbf{x}$ and $\mathbf{x}_j$, respectively, and C_i, $i = 1, 2$, represents the two possible classes for targets and nontargets. Since $\mathbf{x}_j$ is one of the training patterns and its state is independent from the state of the environment when $\mathbf{x}$ was drawn during the classification process, we have $P(l = C_i, l_j = C_i \,|\, \mathbf{x}, \mathbf{x}_j) = P(l = C_i \,|\, \mathbf{x})P(l_j = C_i \,|\, \mathbf{x}_j)$ for $i = 1, 2$. However, the conditional probabilities $P(l_j = C_i \,|\, \mathbf{x}_j)$ and $P(l = C_i \,|\, \mathbf{x})$ are generated by the BPNNs in the upper and lower branches, respectively. Thus, the probability of error associated with wrong item of evidence is

$$P(e \,|\, \mathbf{x}, \mathbf{x}_j) = 1 - \sum_{i=1}^{2} P(l = C_i, l_j = C_i \,|\, \mathbf{x}, \mathbf{x}_j) = 1 - \mathbf{d}_j^T\mathbf{p} \tag{4}$$

where $[\mathbf{d}_j]_i = P(l_j = C_i \,|\, \mathbf{x}_j)$ and $[\mathbf{p}]_i = P(l = C_i \,|\, \mathbf{x})$ for $i = 1, 2$.

Thus far, we have defined a confidence measure for only one neighbor or one item of evidence. To extend this measure for multiple evidence defined in the K-NN memory M_x with $M_x = \{\mathbf{x}_j\}_{j=1}^K$, we assume that the influence of evidence $\mathbf{x}_k$ on $\mathbf{x}$ and that of $\mathbf{x}_l$ for $k \neq l$ on $\mathbf{x}$ are mutually exclusive. Using this assumption, we can write

$$P(e \,|\, \mathbf{x}, M_x) = \frac{\sum_{j=1}^K P(e \,|\, \mathbf{x}, \mathbf{x}_j)p(\mathbf{x}_j \,|\, \mathbf{x})}{\sum_{k=1}^K p(\mathbf{x}_k \,|\, \mathbf{x})}. \tag{5}$$

Note that the memory, M_x, is defined for every unknown pattern. Without loss of generality we can assume that $p(\mathbf{x}_j \,|\, \mathbf{x}) = 1$, since once an unknown pattern $\mathbf{x}$ is applied, the memory, M_x, and all its constituent patterns become known. This gives

$$P(e \,|\, \mathbf{x}, M_x) = \frac{1}{K}\sum_{j=1}^K \left(1 - \mathbf{d}_j^T\mathbf{p}\right) = 1 - \hat{\mathbf{p}}^T\mathbf{p} \tag{6}$$

where $\hat{\mathbf{p}} := 1/K \sum_{j=1}^K \mathbf{d}_j$ is the centroid of all the evidential decisions. Since $\mathbf{p}$ and $\hat{\mathbf{p}}$ are outputs of the BPNN with values in the range of $[0, 1]$, $P(e \,|\, \mathbf{x}, M_x)$ is minimized when $\mathbf{p} = \hat{\mathbf{p}}$. This means that the minimum of the probability of error occurs when the output of lower BPNN is moved to this centroid.

Remarks:

1) It must be pointed out that in a more general memory system, computing the density function $p(\mathbf{x}_j \,|\, \mathbf{x})$ requires generating the statistics of the error vector for every new unknown pattern $\mathbf{x}$. We may approximate this density function by a Gaussian, using the scaled distance measure in (2), i.e., $p(\mathbf{x}_j \,|\, \mathbf{x}) = e^{-(1/2)d(\mathbf{x}_j, \mathbf{x})}$. With this approximation, (5) becomes

$$P(e \,|\, \mathbf{x}, M_x) = \frac{\sum_{j=1}^K \left(1 - \mathbf{d}_j^T\mathbf{p}\right) e^{-\frac{1}{2}d(\mathbf{x}_j, \mathbf{x})}}{\sum_{k=1}^K e^{-\frac{1}{2}d(\mathbf{x}_k, \mathbf{x})}} = 1 - \hat{\mathbf{p}}^T\mathbf{p} \tag{7}$$

where in this case $\hat{\mathbf{p}} = \sum_{j=1}^K w_j\mathbf{d}_j$ is the weighted centroid of all the evidential decisions with weights that are $w_j := \left((e^{-(1/2)d(\mathbf{x}_j, \mathbf{x})})/(\sum_{k=1}^K e^{-(1/2)d(\mathbf{x}_k, \mathbf{x})})\right)$. In a similar fashion as before, one can deduce that $P(e \,|\, \mathbf{x}, M_x)$ takes its minimum when $\mathbf{p} = \hat{\mathbf{p}}$.

   It is interesting to note that this weighted measure is dependent on the distribution and distance of the evidence (neighboring pattern) from the unknown pattern both in the feature space through $p(\mathbf{x}_j \,|\, \mathbf{x})$ and in the output space via the term $(1 - \mathbf{d}_j^T\mathbf{p})$. Clearly, a wrong evidence (i.e., $\mathbf{d}_j^T\mathbf{p} \approx 0$) that is close to the unknown pattern in the feature space (i.e., $d(\mathbf{x}_j, \mathbf{x}) \approx 0$) produces a large contribution in this error measure.

2) Using either (6) or (7) the probability of error can be estimated and used as a measure of confidence of the collective evidence. If the confidence in the collective evidence is strong the probability of error $P(e \,|\, \mathbf{x}, M_x)$ will be very small. Thus a threshold, P_{Th}, should be chosen in order to decide when the feature updating is needed, i.e., updating will take place only if $P(e \,|\, \mathbf{x}, M_x) > P_{\mathrm{Th}}$.

B. Lower Branch: Feature Mapping and Classification

The adaptive feature mapping subsystem plays the central role in the whole system. Its main task is to ensure that the classification decision of the BPNN classifier remains invariant or insensitive to the changes that may have occurred in the original feature space. This is accomplished on the basis of the evidence on those candidate patterns that are determined by the upper branch. The main goal is to prevent the occurrence of misclassifications and false reports under changes in the environmental condition by minimizing the probability of error associated with input pattern in the presence of the memory. As shown in Fig. 2, the unknown pattern is mapped by the mapping matrix H before is passed on to the main classifier. This mapping matrix is initially identity (i.e., $H = I$) in order to determine whether or not feature updating is needed. If the collective evidence points to the conclusion that the feature updating can indeed lead to high confidence decision, this mapping matrix then transforms the input pattern in an attempt to minimize the error criterion or maximize the classifier's decision confidence. The direction and magnitude of this updating is determined in conjunction with the upper branch, which provides the evidential feedback signal to the feature adaptation process. The updating mechanism for the elements of the mapping matrix H is discussed in the next section.

The main classifier in the lower branch is responsible for making target versus nontarget discrimination. This is a two-layer BPNN classifier with two output cells corresponding to targets and nontargets. During training, the desired responses for these classes are $\mathbf{t_i} = [1\ 0]^t$, and $\mathbf{t_i} = [0\ 1]^t$, respectively. Once the network is trained, its function during the cycle of search and adaptation is to provide the decision, $\mathbf{p} = F(\mathbf{y})$ on the mapped feature vector, $\mathbf{y} = H\mathbf{x}$. As mentioned before, the error signals $\mathbf{e}_j = \mathbf{d}_j - \mathbf{p}, j = 1, \ldots, K$ between the outputs of the two BPNNs for the applied pattern $\mathbf{y}$ and the candidate patterns $\mathbf{x}_j$ selected by the K-NN, are then used to drive the adaptation process. As will be shown in the next section, this error is minimized in the mean squared (or the weighted mean square sense) by adjusting the mapping matrix elements.

IV. Adaptation Mechanisms

Two different adaptation mechanisms are derived in this section, using two different cost criteria. The first mechanism involves changing the elements of the mapping matrix H, in order to drive the output of the main BPNN to the centroid (or the weighted centroid) of the collective evidence to minimize the probability of error for the unknown pattern given the memory. This is accomplished by minimizing the sum squared errors $\{\mathbf{e}_j\}_{j=1}^{K}$ iteratively using a gradient descent algorithm. As described in the previous section, the updating is only performed if the probability of error is larger than the threshold P_{Th}.

There is yet another approach for updating in which no feedback signal from the upper branch is provided. In this mechanism, a 2-D sigmoid function (see Fig. 3) is used as the cost function. By minimizing this cost function the outputs of the BPNN for the applied pattern are pushed toward one of the minima points of the 2-D sigmoidal surface, which are located at $(0, 1)$ and $(1, 0)$. Consequently, minimization of this cost function via the adaptive mapping matrix will increase the gap between the outputs of the BPNN. Combining the two criteria can improve the overall confidence of the classification decision further.

Fig. 4 shows the block diagram of the lower branch of the adaptive feature mapping system in Fig. 2. The first layer

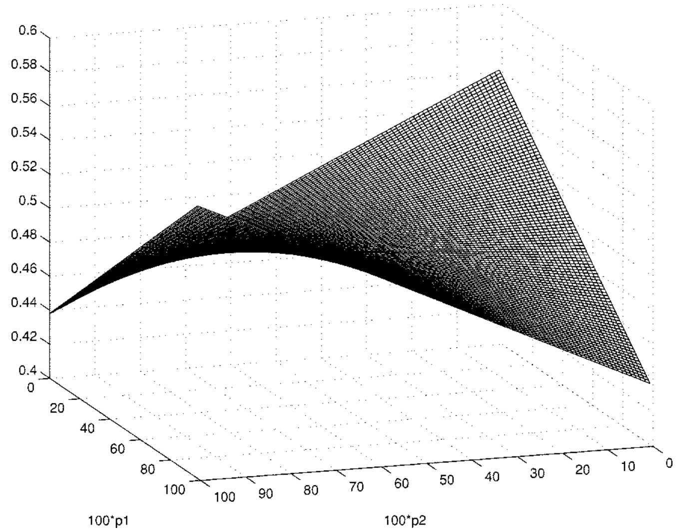


Fig. 3. 2-D sigmoid cost function.

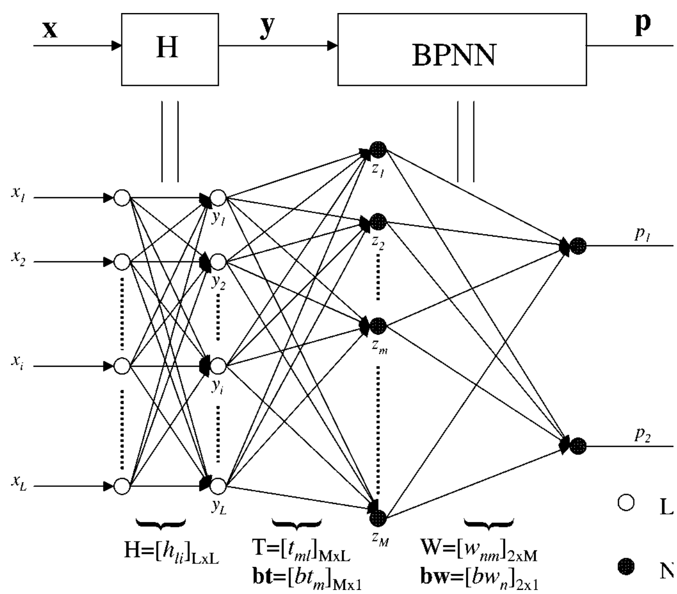


Fig. 4. Adaptive feature mapping system diagram.

with linear activation cells represents the adaptive mapping system with weight matrix H. The two subsequent layers represent the two layers of the main BPNN with nonlinear activation functions at cells. Let us define the input feature vector, the mapped feature vector, and the output vector by $\mathbf{x} = [x_1 x_2 \ldots x_L]^t, \mathbf{y} = [y_1 y_2 \ldots y_L]^t$, and $\mathbf{p} = [p_1 p_2]^t$, respectively. Since the BPNN is trained, the layer weight matrices T and W and the bias vectors $\mathbf{bt}$ and $\mathbf{bw}$ are already known. Thus, the goal is to find the optimal mapping matrix $H = [h_{li}], l, i \in [0, L-1]$ for the feature mapping system which would minimize the average squared error over all the candidate patterns in M_x. This is done iteratively using the gradient descent algorithm and the chain rule which relates the gradients of the error criterion to the weights h_{li}'s.

In the following sections, two different adaptation mechanisms are derived based upon two different error criteria and their combination.

A. Least Squares Error Criterion

As described in Section III-A, $P(e\,|\,\mathbf{x}, M_x)$ is minimized when the output of the BPNN in the lower branch is moved to the centroid $\hat{\mathbf{p}}$. It is easy to show that the appropriate cost function that achieves this upon its minimization is the average squared error over the evidential patterns $\{\mathbf{x}_j\}_{j-1}^{K}$. This average squared error is given by

$$\varepsilon = \frac{1}{K}\sum_{j=1}^{K}\|\mathbf{e}_j\|^2 = \frac{1}{K}\sum_{j=1}^{K}\|\mathbf{d}_j - \mathbf{p}\|^2 = \frac{1}{K}\sum_{j=1}^{K}\sum_{n=1}^{2}(d_{jn} - p_n)^2 \tag{8}$$

where $\mathbf{p}$ is the decision of the main BPNN for the mapped pattern $\mathbf{y}$, i.e., $\mathbf{p} = F(\mathbf{y})$ with $\mathbf{y} = H\mathbf{x}$. By minimizing this error with respect to the mapping matrix or its elements, the input pattern would be mapped toward a direction that minimizes this cost function over all the candidate patterns. As described before, this LS solution corresponds to the centroid of the outputs $\mathbf{d}_j$, i.e., the average of all the decisions. This is easy to show by taking the derivative of ε with respect to $\mathbf{p}$ and setting the result to zero, which gives $\hat{\mathbf{p}} = (1/K)\sum_{j=1}^{K}\mathbf{d}_j$. However, our goal is to find the corresponding feature vector $\mathbf{y}$ using an iterative search process and through the trained weights of the BPNN. This, effectively corresponds to inverse mapping from the output space to the feature space. It must be pointed out that although the forward mapping using the BPNN is unique, i.e., one input pattern leads to one output vector, the inverse mapping is not unique and one output could correspond to several points in the feature space. Nevertheless, if the confidence about the output $\mathbf{p}$ is high, it is insured that the inverse mapping selects a point in the feature space that corresponds to a pattern with the right class label. Clearly, the selection of the candidate patterns using the K-NN has a major impact on the result of this mapping. However, at present this is the simplest memory system that is available.

Remark: It must be pointed out that in the more general case, we may approximate the conditional density function using the scaled distance measure in (2) i.e., $p(\mathbf{x}_j\,|\,\mathbf{x}) = e^{-(1/2)d(\mathbf{x}_j,\mathbf{x})}$. The expression for ε will then be modified to

$$\varepsilon = \sum_{j=1}^{K} w_j\|\mathbf{d}_j - \mathbf{p}\|^2 \tag{9}$$

where $w_j = ((e^{-(1/2)d(\mathbf{x}_j,\mathbf{x})})/(\sum_{k=1}^{K} e^{-(1/2)d(\mathbf{x}_k,\mathbf{x})}))$ is the associated weight.

Having defined the error criterion, we can use the gradient descent adaptation rule to update the elements h_{li} for each candidate pattern iteratively.

The updating rule using the gradient descent scheme is given by

$$h_{li}(k+1) = h_{li}(k) - \frac{1}{2}\mu\nabla\varepsilon(k) \tag{10}$$

where k is the iteration index, μ is the step size and $\nabla\varepsilon(k)$ is the gradient of ε in (8) with respect to the weights, i.e., $(\partial\varepsilon(k))/(\partial h_{li}(k))$. It can easily be shown that

$$\nabla\varepsilon(k) = \frac{1}{K}\sum_{j=1}^{K}\frac{\partial\varepsilon_j(k)}{\partial h_{li}(k)} \tag{11}$$

$$\frac{\partial\varepsilon_j(k)}{\partial h_{li}(k)} = 2(d_{j1} - p_1(k))\left(-\frac{\partial p_1(k)}{\partial h_{li}(k)}\right) + 2(d_{j2} - p_2(k))\left(-\frac{\partial p_2(k)}{\partial h_{li}(k)}\right) \tag{12}$$

with

$$\frac{\partial p_n(k)}{\partial h_{li}(k)} = p_n(k)(1 - p_n(k)) \sum_{m=1}^{M} w_{nm}[z_m(1 - z_m)t_{ml}x_i] \quad n = 1, 2 \tag{13}$$

where t_{ml} and w_{nm} are the weights in the first and second layers of the trained BPNN and z_m represent the outputs of the hidden layer neurons for input vector $\mathbf{y}$. Thus, using (10), (11), and (13), the mapping matrix elements can be updated iteratively. To reach the LS solution that minimizes the index in (8), these updating equations are implemented in several iterations over the entire set of the evidential patterns.

B. 2-D Sigmoid Cost Function

This cost function is defined in terms of the decision, $\mathbf{p}$ of the main BPNN only, i.e.,

$$S = \frac{1}{1 + e^{-(p_1 - 0.5)(p_2 - 0.5)}}. \tag{14}$$

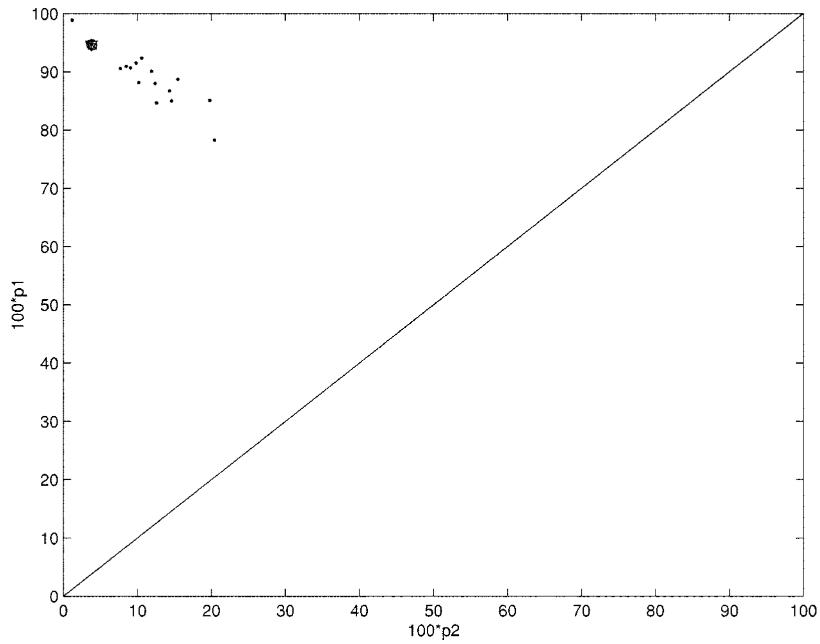
As a result, if this cost function is chosen there is no need to use the K-NN since there is no need for a feedback signal in the adaptation process. In this case, the goal of adaptation is to minimize this cost function by pushing the decision of the BPNN to the closest minimum point of this surface that is located at (0, 1) or (1, 0) (see Fig. 3). This, obviously, increases the gap between the outputs of the BPNN. The inverse mapping, using the adaptation mechanism, then finds the corresponding feature vector for this output.

Using a similar procedure as in the previous section, it can easily be shown that the updating equation for this cost function is given by

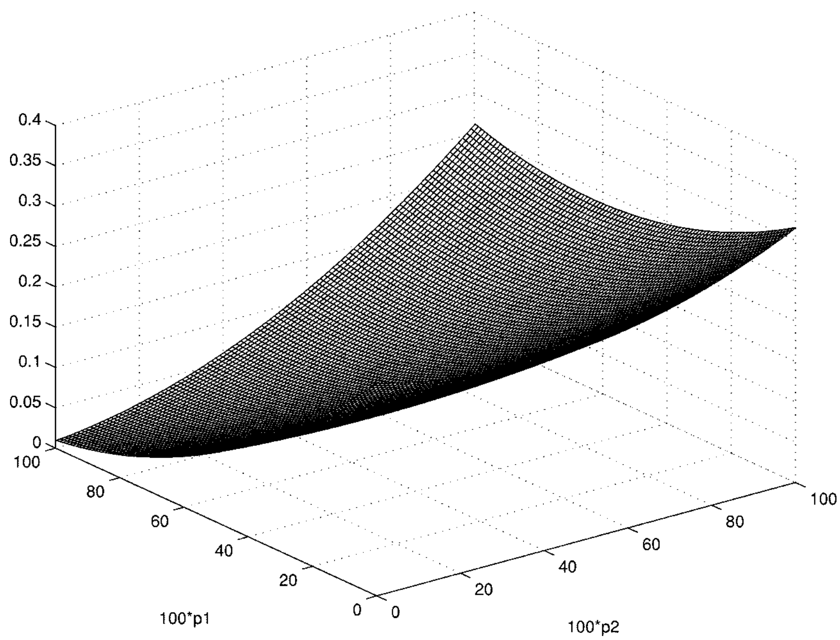
$$h_{li}(k+1) = h_{li}(k) - \frac{1}{2}\mu S(k)(1 - S(k)) \times \left[(p_2(k) - 0.5)\frac{\partial p_1(k)}{\partial h_{li}(k)} + (p_1(k) - 0.5)\frac{\partial p_2(k)}{\partial h_{li}(k)}\right]. \tag{15}$$

C. Combined Criterion

The two cost functions in the previous sections can be combined together to offer a tradeoff between minimizing the sum squared error in (8) and maximizing the BPNN output gap using



(a)



(b)

Fig. 5. Case 1: High confidence target scenario—all classifiers make right decisions (no updating). (a) Location of the outputs. (b) Error surface in the output space.

(14). This multicriteria [15] problem can be posed in terms of a single cost function given by

$$C = (1-\lambda)\varepsilon + \lambda(S - S_{\min}) \tag{16}$$

where λ is the weighting coefficient and $S_{\min}$ is the minimum value of S in (14) at $p_1 = 0, p_2 = 1$ or $p_1 = 1, p_2 = 0$. The gradient of this combined cost function is

$$\nabla C(k) = \frac{1-\lambda}{K}\sum_{j=1}^{K}\frac{\partial \varepsilon_j(k)}{\partial h_{li}(k)} + \lambda\frac{\partial S(k)}{\partial h_{li}(k)} \tag{17}$$

where $((\partial\varepsilon_j(k))/(\partial h_{li}(k)))$ and $((\partial S(k))/(\partial h_{li}(k)))$ can be found as above. Having computed this gradient, the gradient descent algorithm may then be applied.

In this multicriteria problem, changing the value of λ from zero to one provides classification "diversity." That is, when $\lambda = 1$ the first term in the cost function (16) associated with the collective evidence provided by K-NN is removed and the decision is solely determined by the fixed BPNN. On the other hand, when $\lambda = 0$ the decision of the main BPNN is ignored and the result is decided solely based upon the collective evidence of the K-NN. Thus, for every value of λ a different classifier is implemented. This is a very useful feature of the proposed

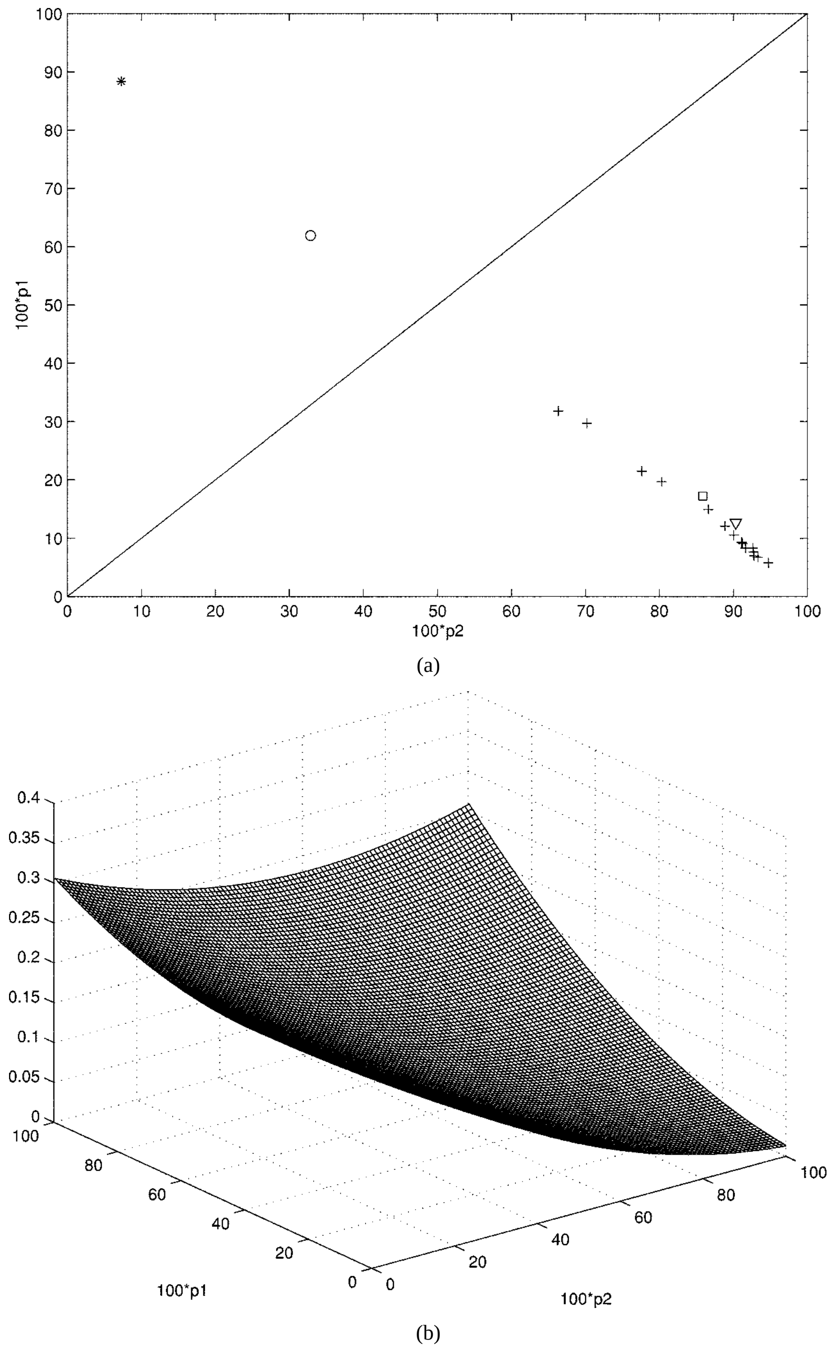


Fig. 6. Case 2: Low confidence nontarget scenario—false report of the BPNN is corrected by the adaptive system. (a) Location of the outputs. (b) Error surface in the output space.

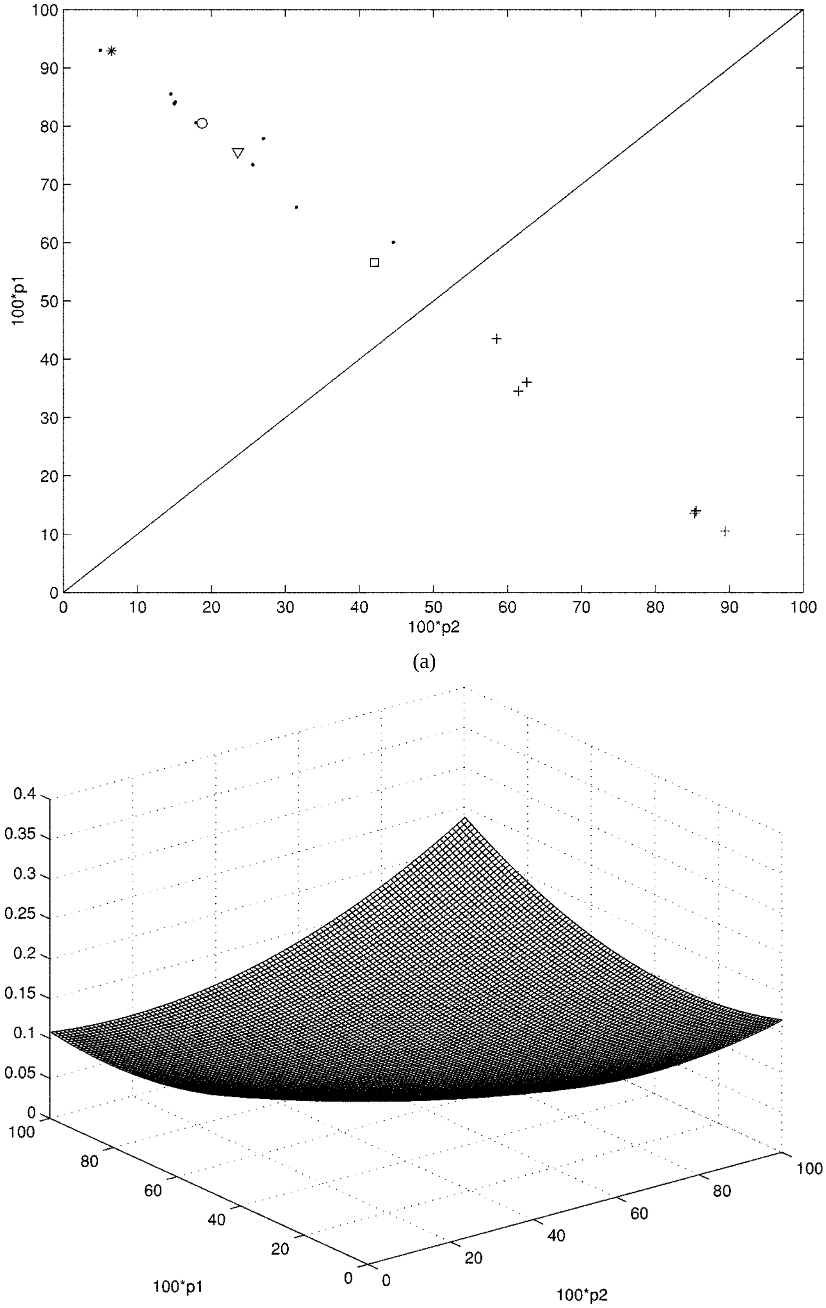
method. The case studies in the following section reveal some important properties of the adaptive feature mapping system.

D. Case Studies

In this section, different possible scenarios are considered to demonstrate the behavior of the adaptation mechanisms in the output space of the main BPNN. Moreover, for each case the shape of the cost function in (16) is provided in order to shed light on the goal of the adaptation process.

In the following figures dots and pluses represent the outputs for the target $(p_1 > p_2)$ and nontarget $(p_2 > p_1)$ patterns, respectively; while square, circle, star, and triangle are the results of LS alone $(\lambda = 0)$, initial BPNN, 2-D sigmoid function alone $(\lambda = 1)$, and the combined criterion $(\lambda = 0.8)$, respectively.

1) Case 1: This case corresponds to a strong target case where all the neighboring patterns identified by the K-NN memory are given the same class label by the BPNN as the unknown input pattern and the decisions are strong. Fig. 5(a) shows the locations of the outputs of the upper branch, $\mathbf{d}_j$'s, and the initial output of the main BPNN, i.e., $\mathbf{p} = F(\mathbf{x})$. Typically, this scenario corresponds to the case where the unknown pattern is far from the decision surface $(31 < d(\mathbf{x}_j, \mathbf{x}) < 42.5$ in the present example) and all the candidate neighboring patterns are of the same class. Clearly, in this case no adaptation is needed owing to the strong initial response of the BPNN and



(a)

(b)

Fig. 7. Case 3: Low confidence nontarget scenario—false report of the BPNN cannot be corrected by feature updating. (a) Location of the outputs. (b) Error surface in the output space.

the tightness of the evidence cluster $(P(e\,|\,\mathbf{x}, M_x) = 0.156)$. However, in order to show the shape of the surface $C(\mathbf{p})$ for $\lambda = 0.8$ in the output space (p_1, p_2), Fig. 5(b) is generated. As can be seen from both Fig. 5(a) and (b), in this case the final decision of the BPNN is moved to a point very close to (1, 0). The LS part of the cost function $\varepsilon(\mathbf{p})$ moves the output of the main BPNN toward the centroid of all the decisions for the candidate patterns selected by the K-NN, while the sigmoidal part tries to push that value toward (1, 0) output. Since the initial output (with initial $H = I$) of the BPNN is already very close to the final location, then obviously the mapping matrix will remain close to $H = I$, even if the updating is performed. Our test results indicated that a threshold of $P_{\mathrm{Th}} = 0.2$ is adequate for this target classification problem. This implies that since $P(e\,|\,\mathbf{x}, M_x) < 0.2$, no updating is needed as far as this particular case is concerned. In this case the tightness of the cluster in the output space explains the reasons for low probability of error.

2) Case 2: In this nontarget case, the majority of the neighboring patterns selected by the K-NN memory are assigned class labels by the BPNN in the upper branch that are opposite to those of the output of the main BPNN. The confidence measure in this case is low, even though the output cluster is somewhat tight! This is due to the fact that the BPNN initially classifies this nontarget as a target with a relatively strong response. In addition, the distances of the evidence from the unknown pat-

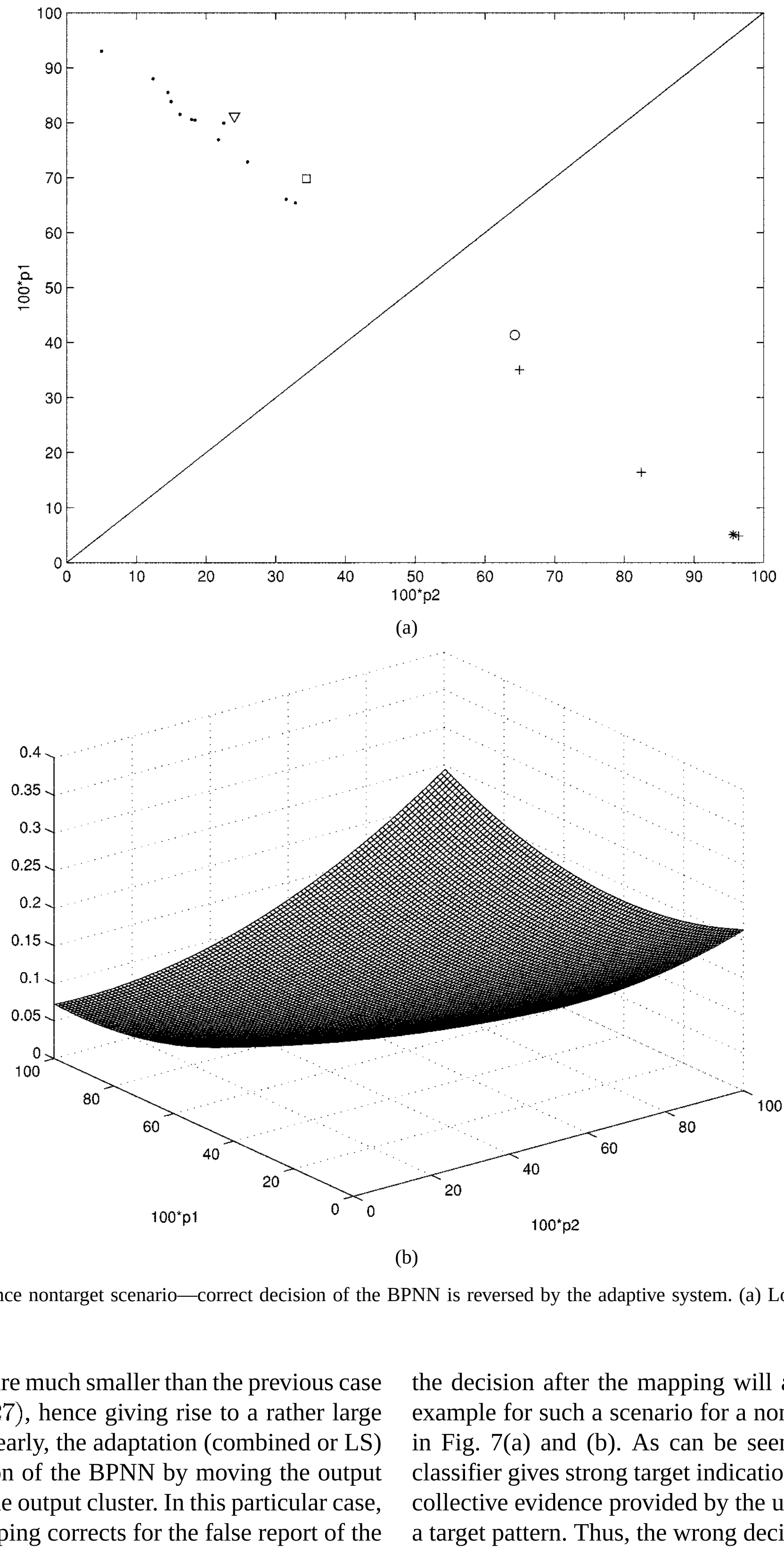


Fig. 8. Case 4: Low confidence nontarget scenario—correct decision of the BPNN is reversed by the adaptive system. (a) Location of the outputs. (b) Error surface in the output space.

tern in the feature space are much smaller than the previous case $(0.64 < d(\mathbf{x}_j, \mathbf{x}) < 1.27)$, hence giving rise to a rather large $P(e \mid \mathbf{x}, M_x) = 0.63$. Clearly, the adaptation (combined or LS) changes the final decision of the BPNN by moving the output close to the centroid of the output cluster. In this particular case, the adaptive feature mapping corrects for the false report of the original BPNN. Fig. 6(a) and (b) show the corresponding plots of the output locations and the error surface, respectively. Since $P(e \mid \mathbf{x}, M_x) > 0.2$, updating is performed.

3) Case 3: There are some cases that the decisions provided by both branches are wrong, i.e., not only the BPNN makes a wrong decision but also the majority of the KNN candidate patterns give wrong decisions in the output space. As a result, the decision after the mapping will also remain incorrect. An example for such a scenario for a nontarget pattern is provided in Fig. 7(a) and (b). As can be seen, in this case the BPNN classifier gives strong target indication and at the same time the collective evidence provided by the upper branch also points to a target pattern. Thus, the wrong decision of the main classifier will not be corrected in such scenarios. In this particular case, the probability of error is $P(e \mid \mathbf{x}, M_x) = 0.46$, and the distance varies in the range of $1.97 < d(\mathbf{x}_j, \mathbf{x}) < 2.88$.

4) Case 4: Finally, there can be some cases where the correct decision of the main BPNN will be impacted by the wrong collective evidence provided by the upper branch. An example for such a case is given in Fig. 8(a) and (b) for a nontarget pattern. Although, the initial decision of the BPNN was correct and strong, since the collective evidence strongly points to the opposite decision (target), the adaptive mapping reverses this correct decision. In this case, $P(e\,|\,\mathbf{x}, M_x) = 0.52$ and the distances are not very large $(6.1 < d(\mathbf{x}_j, \mathbf{x}) < 9.2)$. Obviously, two factors control the occurrence of this decision reversal, namely the choices of the parameter λ and the threshold, P_{Th}, for the confidence measure. However, such cases do occur, even though they may be rare.

In the next section, we will present and analyze the test results of the adaptive feature mapping system for underwater target classification in a changing environment.

V. Test Results

To test the effectiveness of the adaptive classification system developed in this paper, the 40 kHz data set [1] provided by Coastal Systems Station (CSS) in Panama City, FL, was used in this study. This data set contains backscattered signals corresponding to six different objects—two mine-like, namely a bullet-shaped metallic object and a truncated-cone-shape plastic object; and four nonmine-like, namely a water-filled drum, an irregular shape limestone rock, a smooth granite rock, and a water-saturated wooden log. The transmit signal was a linear FM up-sweep with frequency range from 20 to 60 kHz. Each object was insonified at aspect angles from 0 to 355° with 5° separation. This resulted in 72 aspect angles out of which the even-angles were used in the training data set while the odd-angle samples were used as the testing samples. As a result, for each object there are 36 patterns (at different aspect angles) in the training or testing data sets. The training data set contained the feature vectors of backscattered data with synthesized reverberation effects with SRR $= 12$ dB that corresponds to nominal operating conditions. The procedure for generating synthesized reverberation involves convolving the transmit signal with a random sequence and scaling the resultant signal according to the specified SRR [1]. This reverberation signal is then added to the backscattered signal to generate one "noisy realization." The process is repeated for every aspect angle multiple times in order to generate a statistically rich data set for determining the generalization ability of the classifier. In order to study the effectiveness of the adaptive feature mapping system in environments that differ from the training environment, a rich testing data set containing several sets of ten noisy realizations with varying SRR conditions (0 to 20 dB) was generated. Note that it is assumed that the environmental changes primarily cause variations in the SRR.

A five-level WP decomposition was applied to the noisy backscattered signals and six subbands that reside in the bandwidth of the transmit signal were selected. To avoid phase distortion and at the same time ensure the orthogonality of the representation, Symlet 4 wavelet [11] was used. The transmit signal was decomposed in the same way. Then, the cross-correlation of the subband signals of the backscattered and the corresponding subband signals of the transmit was performed. A fourth-order autoregressive (AR) model [12] was then used to represent the resultant signal and the model coefficients were used as features for classification. This leads to a total of 30 features out of which 22 features with high discriminatory power were selected according to the criterion in (1). These were then used as feature vector for classification of the data.

A two-layer 22-42-2 BPNN was trained to discriminate targets from nontargets. The BPNN was trained using an adaptive learning rate and momentum factor of 0.9 [16]. The sum squared error goal was ten and the maximum number of epochs was chosen to be 4000. Several training trials with different initial weights were tried and the network that had the best performance on the validation data set was chosen. The validation data set contained one noisy realization of odd-aspect data set with SRR $= 12$ dB. The K-NN classifier was also trained with the same training samples and for $K = 15$. In the adaptive feature mapping system, the step size was chosen to be $\mu = 0.005$ and the initial value of the mapping matrix was $H(0) = I$, i.e., an identity matrix. The stopping criterion for the iterative learning was $|\varepsilon(k) - \varepsilon(k-1)| < 0.0002$, where k is the iteration index. Several values of λ in the range of [0, 1] were tried and the best value for this parameter was determined empirically. In addition, the confidence threshold for updating was also determined experimentally to be $P_{\mathrm{Th}} = 0.2$.

In order to demonstrate the effectiveness of the adaptive system in a changing environment, the testing data sets with different SRR were applied to the system that was trained on the 12 dB SRR data. Fig. 9 shows the overall classification error rate plots for the adaptive classification system for different values of $(0 \leq \lambda \leq 1)$. Note that the performance of the adaptive system using the 2-D sigmoidal cost function alone $(\lambda = 1)$ corresponds to that of the fixed (nonadaptive) BPNN classifier with hard-limiter threshold decision (see the dotted plot in Fig. 9). On the other hand, using the LS error criterion alone $(\lambda = 0)$ the performance is mainly dependent on that of the K-NN (see the fine solid plot in Fig. 9). Consequently, using the combined criterion and changing λ offers diversity in configuring different classifiers. Moreover, the length of the locus of classification error as λ changes in the range of [0, 1] presents an indication for ambiguity in the decision making of the extreme classifiers. As can be seen in these plots, the fixed BPNN classifier gives its best performance at the 12 dB SRR, which is the training condition. The performance of BPNN degrades severely when the SRR decreases, i.e., with increasing reverberation. The performance degradation for higher SRR is not that severe. This is attributed to the fact that the training data set included only SRR $= 12$ dB cases. The best overall performance for the combined criterion was obtained for $\lambda = 0.8$, which retains the good performance of the system using the LS error criterion at lower SRR values. For this choice of λ, the performance remains somewhat invariant to the changes in the reverberation for 8 dB $\leq$ SRR ≤ 12 dB. Additionally, as can be observed at low SRR values (SRR ≤ 10 dB), the performance of all the combined criterion classifiers are substantially better than the fixed BPNN. It is also interesting to note that at SRR $= 12$ dB, the combined criterion system with $\lambda = 0.8$ achieves an error rate of 8.1% compared to 7.7% for the fixed BPNN $(\lambda = 1)$ and 9.6% for the system using the LS error

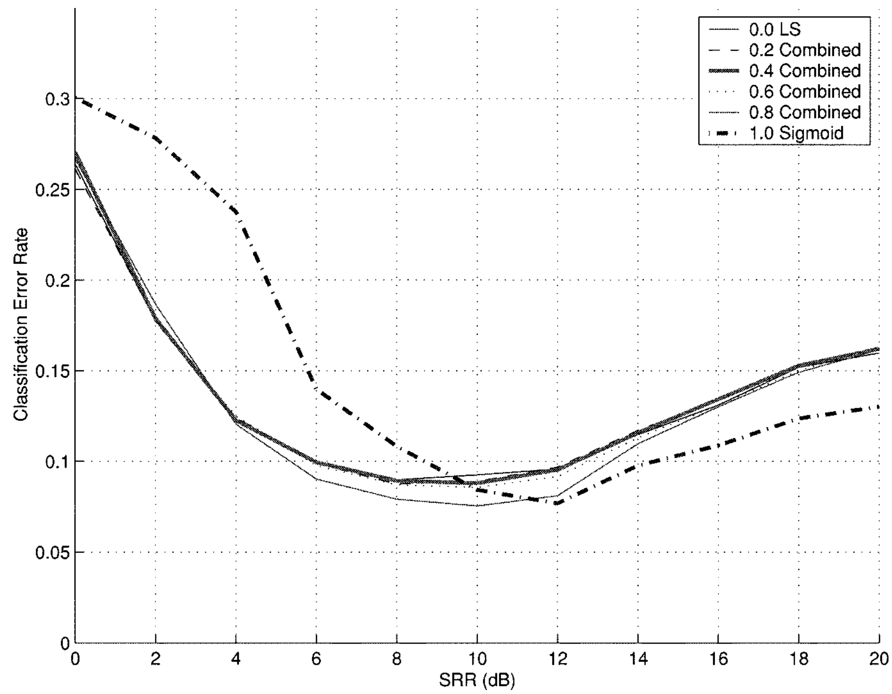


Fig. 9. Error rate versus SRR for adaptive classification system and different λ's: Ten noisy realizations.

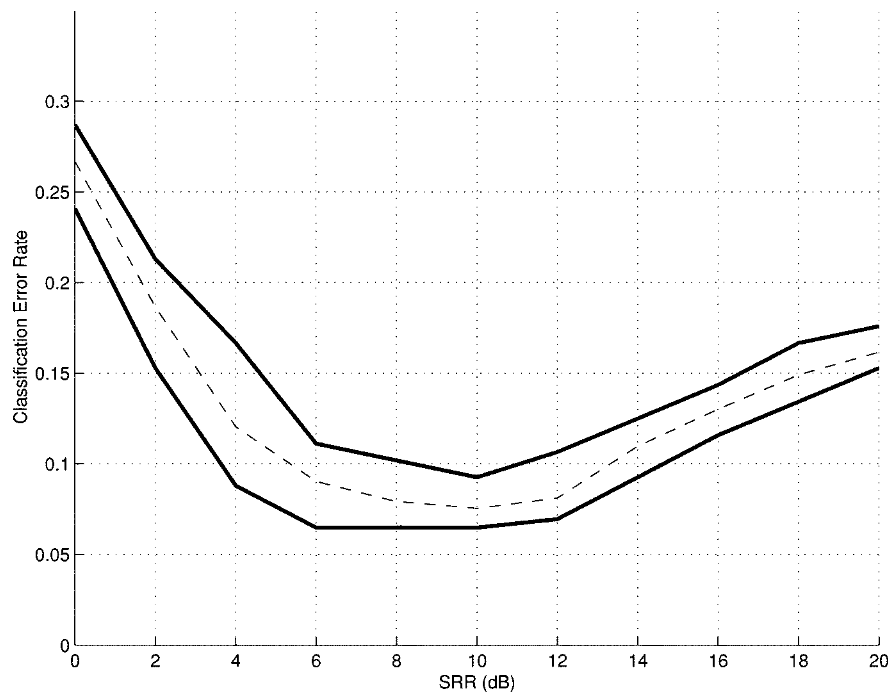


Fig. 10. Bounds on error rates for different SRR $(\lambda = 0.8)$: Ten noisy realizations.

criterion. Fig. 10 shows the bounds on the variations of the classification error rates for different SRR values and ten noisy realizations $(\lambda = 0.8)$. The dotted plot corresponds to the average over all the ten noisy realizations. Overall, the proposed system is very promising as it shows that the adaptive mapping is indeed effective in maintaining the feature space invariant to the environmental changes. Note that the improvement as well as the degradation are mostly attributed to the K-NN memory which provides the evidence patterns to the system.

VI. Conclusion

In this paper, a new adaptive classification system is presented which consists of two main branches, namely the upper and lower branches. The upper branch works as a memory system to identify the closest matches of an unknown pattern in the feature space and provide their corresponding decisions (labels). The lower branch performs feature mapping and classification. Based upon the collective evidence provided by the upper branch, the decision about the direction and magnitude of the feature mapping is made. A confidence measure based upon the probability of error is introduced that provides a means to determine when the feature mapping is needed and how should be directed. Two adaptation mechanisms were developed that use different error criteria or cost functions. Minimizing the first criterion leads to the LS solution among all the decisions provided by the upper branch and hence minimizing the probability of error given the memory; while minimizing the second criterion (2-D sigmoid function) maximizes the gap between the outputs of the BPNN. The test results on the acoustic backscattered data collected from six different targets/nontargets showed that the adaptive system using the combined error criterion offered very good performance in situations where the SRR is changing. The improved performance over the fixed classifier is particularly more prominent at very low SRR conditions where reverberation is dominant. Consequently, this scheme offers a very promising tool for underwater target classification in shallow water and under changing environmental conditions. It must be emphasized that the K-NN memory clearly has a significant impact on the performance of the entire adaptive system. As a result, further improvements in the overall performance are expected when a better matching system is designed. This could be achieved by either using more sophisticated nearest neighbor algorithms [17], [18] or mapping of the feature vector to a higher order space, prior to adaptive feature mapping, where the features can be separated better.

REFERENCES

[1] M. R. Azimi-Sadjadi, D. Yao, Q. Huang, and G. J. Dobeck, "Underwater target classification using wavelet packets and neural networks," *IEEE Trans. Neural Networks*, vol. 11, pp. 784–794, May 2000.

[2] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Trans. Neural Networks*, vol. 3, pp. 698–712, Sept. 1992.

[3] G. A. Carpenter, S. Grossberg, and D. B. Rosen, "ART2-A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks*, vol. 4, no. 4, pp. 493–504, 1991.

[4] T. Frank, K. Kraiss, and T. Kuhlen, "Comparative analysis of fuzzy ART and ART-2A network clustering performance," *IEEE Trans. Neural Networks*, vol. 9, pp. 544–559, May 1998.

[5] R. O. Duda, P. E. Hart, and D. Stork, *Pattern Classification*. New York: Wiley-Interscience, 2001.

[6] D. Yao, M. R. Azimi-Sadjadi, and G. J. Dobeck, "Adaptive feature mapping for underwater target classification," *Proc. IEEE Int. Conf. Neural Networks (ICNN'99)*, July 1999.

[7] L. G. L. Devroye and G. Lugosi, *A Probablistic Theory of Pattern Recognition*. New York: Springer-Verlag, 1996.

[8] P. A. Saillant, J. A. Simmons, S. P. Dear, and T. A. McMullen, "A Computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram and transformation receiver," *J. Acoust. Soc. Amer.*, vol. 94, pp. 2691–2712, Nov. 1993.

[9] R. C. Roverud, "Neural computations for sound pattern recognition: Evidence for summation of an array of frequency filters in an echolocating bats," *J. Neurosci.*, vol. 13, pp. 2306–2312, June 1993.

[10] J. M. Wotton and R. L. Jenison, "A back-propagation network model of the monaural localization information available in the bat echolocation system," *J. Acoust. Soc. Amer.*, vol. 101, pp. 2964–2972, May 1997.

[11] M. Vetterli and J. Kovacevic, *Wavelets and Sub-Band Coding*. Engelwood Cliffs, NJ: Prentice-Hall, 1995.

[12] L. Rabiner, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[13] S. Theodoridis and K. Koutroumbas, *Pattern Recognition.*. San Diego, CA: Academic, 1999.

[14] J. A. Anderson, *An Introduction to Neural Networks*. Cambridge, MA: MIT Press, 1995.

[15] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. New York: Wiley, 1993.

[16] S. Haykin, *Neural Networks A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[17] L. M. Zouhal and T. Denoeux, "An evidence-theoretic k-NN rule with parameter optimization," *IEEE Trans. Syst., Man, Cybern.*, vol. 28, pp. 263–271, May 1998.

[18] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 607–615, June 1996.

Mahmood R. Azimi-Sadjadi (S'81–M'81–SM'89) received the M.S. and Ph.D. degrees from the Imperial College of Science andTechnology, University of London, London, U.K., in 1978 and 1982, respectively, both in electrical engineering with specialization in digital signal/image processing.

He is currently a Full Professor at the Electrical and Computer Engineering Department at Colorado State University (CSU), Fort Collins. He is also serving as the Director of the Digital Signal/Image Laboratory at CSU. His main areas of interest include digital signal and image processing, target detection, classification, and tracking, adaptive filtering, and system identification, and neural networks. His research efforts in these areas resulted in more than 150 journal and refereed conference publications. He is the co-author of the book *Digital Filtering in One and Two Dimensions* (New York: Plenum, 1989).

Dr. Azimi-Sadjadi is the recipient of the 1999 the ABELL Teaching Award, the 1993 ASEE-Navy Senior Faculty Fellowship Award, the 1991 CSU Dean's Council Award, and the 1984 DOW Chemical Outstanding Young Faculty Award. He served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS.

De Yao received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in July 1992. He received the M.S. degree in electrical engineering from Colorado State University, Fort Collins, in December1999.

From August 1992 to April 1996, he was an Electronics Engineer with ShanDong Telecommunications Inc., ShanDong, China. From April 1996 to January 1998, he was with Goldtron Telecommunications Pte. Ltd., Singapore, as a Hardware Engineer. He is currently with AirNet Communications Corporation, Melbourne, FL. His research interests include digital signal processing, classification, and applications in communication systems.

Arta A. Jamshidi received the B.S. degree (summa cum laude) in electrical engineering from the University of Tehran, Tehran, Iran, in July 1998. Currently, he is pursuing the M.S. degree at the Department of Electrical and Computer Engineering at Colorado State University (CSU), Fort Collins.

He is a Research Associate at CSU. His research interests include signal/image processing, communications, and pattern recognition.

Gerry J. Dobeck received the B.S. degree in physics from the University of Massachusetts, Amherst, in 1970 and the M.S. and Ph.D. degrees in electrical engineering from the University of South Florida, Tampa, in 1973 and 1976, respectively.

Since 1976, he has been employed at the NAVSEA Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL. His current research interests include automatic detection and classification of targets in cluttered environments from synthetic/real aperture sonar imagery, the echo structure of acoustic returns, underwater electro-optic imagery, and gradiometer/magnetometer signals. Over the past eight years, he has been Project Leader for the development and application of processing technologies for automated mine detection and classification and data/sensor/algorithm fusion, sponsored by the Office of Naval Research 6.2/6.3 Mine Countermeasures Program. He has authored or coauthored more than 70 technical reports and papers.

Dr. Dobeck received the 1981 and 1996 Commanding Officer/Executive Director award for Science and Technology. In 2000, he received the National Defense Industrial Association Bronze Metal Award. He is a reviewer for IEEE, ASME, SPIE and the *Journal of Underwater Acoustics* and has been session chair at past IEEE and SPIE conferences.