

Comparative Analysis of Different Approaches to Target Classification and Localization with Sonar

Birsel Ayrulu, Billur Barshan

Abstract— This study compares the performances of different classification and fusion techniques for the differentiation and localization of commonly encountered features in indoor environments. Differentiation of such features is of interest for intelligent systems in a variety of applications.

Keywords— Target classification, Localization, Sensor data fusion, Sonar, Evidential reasoning, Voting, Artificial neural networks, Statistical pattern recognition.

I. INTRODUCTION

Differentiation of commonly encountered features in indoor environments is an important problem for intelligent systems for applications such as map building, navigation, obstacle avoidance, and target tracking. Since sonar sensors are light, robust, and inexpensive devices, they are a suitable choice for these applications.

Sensory information from a *single* sonar has poor angular resolution and is usually not sufficient to differentiate more than a small number of target primitives [1]. Improved target classification can be achieved by using multi-transducer pulse/echo systems and by employing both amplitude and time-of-flight (TOF) information. In this paper, the performances of different classification and fusion schemes in target differentiation and localization of commonly encountered features in indoor environments are compared. These include a target differentiation algorithm (TDA), statistical pattern recognition techniques [*k*-nearest neighbor (*k*-NN) classifiers, kernel estimator (KE), and parameterized density estimator (PDE)], linear discriminant analysis (LDA), fuzzy *c*-means clustering algorithm (FCC), and artificial neural networks (ANN). The fusion techniques used in this study are Dempster-Shafer evidential reasoning (DS), simple majority voting (SMV), and voting schemes with preference ordering and reliability measures.

II. BACKGROUND ON SONAR SENSING

In commonly used TOF systems, an echo is produced when the transmitted pulse encounters an object and a range value $r = ct_o/2$ is produced when the echo amplitude first exceeds a preset threshold level τ at time t_o back at the receiver. Here, t_o is the TOF and c is the speed of sound in air.

It is observed that the echo amplitude decreases with increasing target range (r) and absolute value of the target azimuth ($|\theta|$). The echo amplitude falls below τ when $|\theta| > \theta_o$, which is related to the aperture radius a and the resonance frequency f_o of the transducer by $\theta_o = \sin^{-1} \left(\frac{0.61c}{af_o} \right)$.

B. Ayrulu and B. Barshan are with the Department of Electrical and Electronics Engineering, Bilkent University, 06533 Bilkent, Ankara, Turkey. E-mail: {ayrulu, billur}@ee.bilkent.edu.tr

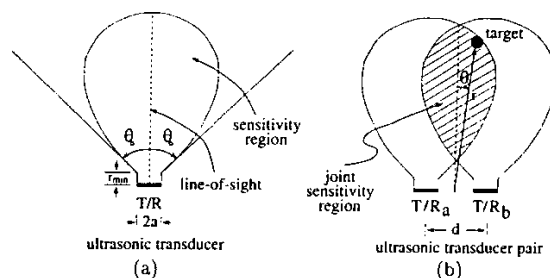


Fig. 1. (a) Sensitivity region of an ultrasonic transducer. (b) Joint sensitivity region of a pair of ultrasonic transducers.

In our system, two identical ultrasonic transducers a and b with center-to-center separation d are employed to improve the angular resolution of a single transducer. The transducers can operate both as transmitter and receiver to detect echo signals reflected from targets within their *joint sensitivity region* (Fig. 1).

The targets used in this study are *plane* (P), *corner* (C), *acute corner* (AC), *edge* (E) and *cylinder* (CY) (Fig. 2). Since the wavelength of operation ($\lambda \cong 8.6$ mm at $f_o = 40$ kHz) is much larger than the typical roughness of surfaces encountered in laboratory environments, targets in these environments reflect acoustic beams specularly, like a mirror. Detailed physical reflection models of these targets are provided in [2]. In the following, A_{aa} , A_{bb} , A_{ab} , and A_{ba} denote the maximum values of the sonar echo signals, and t_{aa} , t_{bb} , t_{ab} , and t_{ba} denote the TOF readings extracted from these signals. The first index in the subscript indicates the transmitting transducer, the second index denotes the receiver.

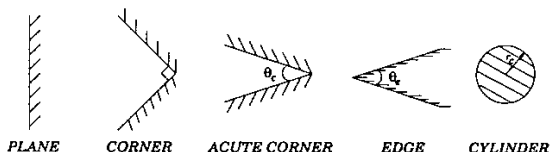


Fig. 2. Horizontal cross sections of the targets differentiated.

III. TARGET DIFFERENTIATION ALGORITHM

The TDA has its roots in the P/C differentiation algorithm developed in [1] based on the idea of exploiting amplitude differentials in resolving target type. In [2], it is extended to include other target primitives using both amplitude and TOF differentials:

if $[t_{aa}(\alpha) - t_{ab}(\alpha)] > k_t \sigma_t$ and $[t_{bb}(\alpha) - t_{ba}(\alpha)] > k_t \sigma_t$ then **AC**
 else if $[A_{aa}(\alpha) - A_{ab}(\alpha)] > k_A \sigma_A$ and $[A_{bb}(\alpha) - A_{ba}(\alpha)] > k_A \sigma_A$ then **P**
 else if $[\max\{A_{aa}(\alpha)\} - \max\{A_{bb}(\alpha)\}] < k_A \sigma_A$ and
 $[\max\{A_{bb}(\alpha)\} - \max\{A_{ab}(\alpha)\}] < k_A \sigma_A$ then **C**
 else **E, CY** or **U**

The $k_A(k_t)$ is the number of amplitude (TOF) noise standard deviations $\sigma_A(\sigma_t)$ and is employed as a safety margin to achieve robust differentiation. Differentiation is achievable only in those cases where the difference in amplitudes (TOFs) exceeds $k_A\sigma_A(k_t\sigma_t)$. If this is not the case, a decision cannot be made and the target type remains unknown.

IV. DEMPSTER-SHAFER EVIDENTIAL REASONING

In DS, each sensor's opinion is tied to a belief measure or basic probability mass assignment (BPMA) using belief functions [3]. These are set functions which assign numerical degrees of support on the basis of evidence, but also allow for the expression of ignorance. A frame of discernment, Ω , represents a finite universe of propositions and a BPMA, $m(\cdot)$, maps the power set of Ω to the interval $[0,1]$. The BPMA satisfies the conditions:

$$m(\emptyset) = 0, \quad \sum_{A \subseteq \Omega} m(A) = 1 \quad (1)$$

A set with a non-zero BPMA is a *focal element*. The belief or total support that is assigned to a set or proposition A is obtained by summing the BPMAs over all subsets of A :

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

The uncertainty in the measurements of each sonar pair is represented by a belief function having target type or *feature* as a focal element with BPMA $m(\cdot)$: $BF = \{feature; m(feature)\}$.

The BPMA is the underlying function for decision making using DS. It is defined based on the TDA and is thus dependent on amplitude and TOF differentials such that the larger the differential, the larger the degree of belief. The BPMA is made to plane, corner, and acute corner as $m(p)$, $m(c)$, and $m(ac)$, respectively. The remaining belief represents ignorance, or undistributed probability mass, and is assigned to an 'unknown' target type as $m(u) = 1 - [m(p) + m(c) + m(ac)]$.

The sequential combination of multiple bodies of evidence can be obtained for n sensor nodes as:

$$BF = (((BF_1 \oplus BF_2) \oplus BF_3) \dots \oplus BF_n) \quad (3)$$

Using Dempster's fusion rule:

$$m(h_k) = \frac{\sum \sum_{h_k=f_i \cap g_j} m_1(f_i)m_2(g_j)}{1 - \sum \sum_{h_k=f_i \cap g_j = \emptyset} m_1(f_i)m_2(g_j)} \quad (4)$$

where $\sum \sum_{h_k=f_i \cap g_j = \emptyset} m_1(f_i)m_2(g_j)$ represents conflict.

V. CONFLICT RESOLUTION THROUGH VOTING

Multi-sensor systems exploit sensor diversity to acquire a wider view of a scene under observation. This diversity can give rise to conflicts, which must be resolved when the system information is combined to reach a group decision or to form a group value or estimate. Voting has the advantages of being computationally inexpensive and, to a certain degree, fault-tolerant [4]. Its major drawback is the consistency problem of Arrow which states that there is no voting scheme for selecting from more than two alternatives that is locally consistent under all possible conditions [5].

A. Different Voting Schemes

In SMV, the votes are given equal weight and the group decision is taken as the outcome with the largest number of votes. Although SMV provides fast and robust fusion in some problems, there exist some drawbacks that limit its usage. For example, when all outcomes receive equal votes, a group decision cannot be reached. Moreover, it does not take into account the distribution of the decisions of dissenting classifiers. To overcome these drawbacks, more sophisticated voting schemes assigning preference ordering and reliability measures over the possible target types can be employed. Total preference is taken as the sum of the products of preference order assigned to each target type and the reliability measure assigned to the corresponding sensor node [6].

A.1 Reliability Measure Assignment

Assignment of belief to range and azimuth estimates is based on the observation that the closer the target is to the surface of the transducer, the more accurate is the range reading, and the closer the target is to the line-of-sight of the transducer ($\theta = 0^\circ$), the more accurate is the azimuth estimate. Therefore, belief assignments to range and azimuth estimates can be made as follows:

$$m(r) = \frac{r_{max} - r}{r_{max} - r_{min}}, \quad m(\theta) = \frac{\theta_o - |\theta|}{\theta_o} \quad (5)$$

Five different reliability measures are assigned to sensor node i which are different functional combinations of $m(r)$ and $m(\theta)$ [6]. Their performances are compared.

VI. STATISTICAL PATTERN RECOGNITION TECHNIQUES

An object coming from one of N classes is classified into class w_i if its vector representation \mathbf{x} is in region Ω_i . A rule which partitions a space into regions Ω_i , $i = 1, \dots, N$ is called a *decision rule*. Boundaries between regions are *decision surfaces*. The $p(w_i)$ are the *a priori* probabilities of an object belonging to class w_i , $i = 1, \dots, N$. To classify an object with vector representation \mathbf{x} , a *posteriori probabilities* $p(w_i|\mathbf{x})$, $i = 1, \dots, N$ can be compared and the object is classified to class w_k according to *Bayes minimum error rule*:

$$p(w_k|\mathbf{x}) > p(w_i|\mathbf{x}) \text{ for all } i \neq k \implies \mathbf{x} \in \Omega_k \quad (6)$$

Since these *a posteriori* probabilities are rarely known, they need to be estimated. Using Bayes' theorem $\left[p(w_i|\mathbf{x}) = \frac{p(\mathbf{x}|w_i)p(w_i)}{p(\mathbf{x})} \right]$:

$$p(\mathbf{x}|w_k)p(w_k) > p(\mathbf{x}|w_i)p(w_i) \text{ for all } i \neq k \implies \mathbf{x} \in \Omega_k \quad (7)$$

where $p(\mathbf{x}|w_i)$ are the class-conditional PDFs. The set of vector representations used to estimate these class-conditional PDFs is called *design or training set*. The performance of any decision rule can be assessed in a different set of vector representations which is called the *test set*.

More generally, classification rules can be written as

$$q_k(\mathbf{x}) > q_i(\mathbf{x}) \text{ for all } i \neq k \implies \mathbf{x} \in \Omega_k \quad (8)$$

where q_i is a *discriminant function*. Therefore, discriminant functions can be replaced with PDFs for computational simplicity. Linear discriminant functions result in additional computational advantages.

A. Kernel Estimator (KE)

KE is a class of PDF estimator first proposed by Parzen in 1962 [7]. In KE, the class-conditional PDF estimates are of the form:

$$\hat{p}(\mathbf{x}|w_i) = \frac{1}{n_i h^d} \sum_{j=1}^{n_i} K\left(\frac{\mathbf{x} - \mathbf{x}_j}{h}\right) \quad (9)$$

where \mathbf{x} is the d -dimensional vector at which the estimate is being made and \mathbf{x}_j 's, $j = 1, \dots, n_i$ are samples in the design set. In this equation, n_i is the total number of sample points in class w_i , h is called the *spread* or *smoothing* parameter or *bandwidth* of a KE, and $K(\mathbf{z})$ is a kernel function which satisfies the conditions $K(\mathbf{z}) \geq 0$ and $\int K(\mathbf{z})d\mathbf{z} = 1$. Usually, h is chosen as a function of n_i such that $\lim_{n_i \rightarrow \infty} h(n_i) = 0$. There are various approaches to select h if a constant h is to be used [8].

B. k -Nearest-Neighbor (k -NN) Method

Let k be the number of patterns from a combined set which are nearest neighbors of a pattern \mathbf{x} , and k_i of them come from class w_i . Then a k -NN estimator for class w_i can be defined as:

$$\hat{p}(w_i|\mathbf{x}) = \frac{\hat{p}(\mathbf{x}|w_i)\hat{p}(w_i)}{\hat{p}(\mathbf{x})} = \frac{k_i}{k} \quad (10)$$

The pattern \mathbf{x} is classified as belonging to class w_m if $k_m = \max_i(k_i)$.

Another interpretation of the k -NN estimator which relates it to the KE can be found in [9]. Let $r_k(\mathbf{x})$ be the Euclidean distance from \mathbf{x} to the k th NN of \mathbf{x} in the training set and $p(\mathbf{x}|w_i)$ estimates are taken as

$$\hat{p}(\mathbf{x}|w_i) = \frac{1}{n_i [r_k(\mathbf{x})]^d} \sum_{j=1}^{n_i} K\left(\frac{\mathbf{x} - \mathbf{x}_j}{r_k(\mathbf{x})}\right). \quad (11)$$

This estimator is referred as *generalized k -NN* estimator.

C. PDE with Normal Models

In this method, each class-conditional PDF is assumed to be a d -variate normal such that

$$p(\mathbf{x}|\mathbf{x}_1, \dots, \mathbf{x}_{n_i}) = \frac{1}{(2\pi)^{(d/2)} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)\right]$$

where $i = 1, \dots, N$, μ_i 's and Σ_i 's denote the class mean and covariance matrices, respectively. They must be estimated by using techniques such as the maximum likelihood estimator based on the design set [10].

Normal models (NM) used with PDE are divided into two as heteroscedastic and homoscedastic models. In the homoscedastic NM, all class-covariance matrices are equal ($\Sigma_i = \hat{\Sigma}$ for all $i = 1, \dots, N$). Usually $\hat{\Sigma}$ is taken as the weighted average of each class-covariance matrix estimate [11]. In the heteroscedastic NM, different class-covariance matrices are used for each class.

D. Linear Discriminant Analysis (LDA)

In LDA, discriminant functions $q(\mathbf{x})$'s are linear functions of x_i such that $q(\mathbf{x}) = a_0 + \sum_{i=1}^d a_i x_i = \mathbf{a}^T \mathbf{z}$ where $\mathbf{z} = (1, \mathbf{x}^T)^T$ is the augmented observation vector. The aim of LDA is to find the *weight* vector \mathbf{a} , based on the design set which consists of two class samples such that

- i) $\mathbf{a}^T \mathbf{z}_i > 0$, whenever \mathbf{x}_i is a sample from class w_1 , and
- ii) $\mathbf{a}^T \mathbf{z}_i < 0$, whenever \mathbf{x}_i is a sample from class w_2 .

Although the generalization of LDA to N classes can be done in different ways, here we have chosen to use $N - 1$ two-class decision rules, each one separating Ω_i , $i = 1, \dots, N - 1$ from *all* Ω_j , $j = 1, \dots, N$; $j \neq i$.

VII. FUZZY c -MEANS CLUSTERING (FCC)

FCC has been developed by Dunn [12] and extended by Bezdek. It minimizes the following objective function with respect to fuzzy membership μ_{ij} and cluster centers \mathbf{V}_i :

$$J_m = \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|\mathbf{X}_j - \mathbf{V}_i\|_{\mathbf{A}}^2 \quad (12)$$

where $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$. \mathbf{A} is a $d \times d$ positive definite matrix which specifies the shape of the clusters, d is the dimension of the input patterns \mathbf{X}_j ($j = 1, \dots, N$), N is the number of patterns, and $m > 1$ is the weighting exponent for μ_{ij} and controls the fuzziness of the resulting clusters. Usually, \mathbf{A} is chosen as the $d \times d$ identity matrix which leads the definition of Euclidean distance resulting in spherical clusters.

VIII. ARTIFICIAL NEURAL NETWORKS (ANN)

ANNs have been widely used in a variety of applications [13]. They consist of an input layer, one or more hidden layers, and a single output layer, each comprised of a number of neurons. ANNs have three distinctive characteristics: The model of each neuron includes a smooth nonlinearity, the hidden layers extract progressively more meaningful features, and the ANN exhibits a high degree of connectivity. Due to the presence of distributed form of nonlinearity and high connectivity, theoretical analysis of ANNs is difficult. ANNs are trained to compute the boundaries of decision regions in the form of connection weights and biases by using training algorithms such as back-propagation (BP) and generating-shrinking (GS).

A. Preprocessing of the Input Signals

A.1 Fourier Transform (FT)

The discrete Fourier transform (DFT) of a signal $f(n)$ is defined as:

$$F(k) = \mathcal{F}\{f(n)\} \triangleq \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) e^{-j \frac{2\pi}{N} nk} \quad (13)$$

where N is the length of the signal $f(n)$. The DFT can be represented in matrix notation as $\mathbf{f}_1 = \mathbf{F} \mathbf{f}$ where \mathbf{f} is an $N \times 1$ column vector, \mathbf{F} is the $N \times N$ DFT matrix, and \mathbf{f}_1 is the DFT of \mathbf{f} .

A.2 Fractional Fourier Transform (FRT)

The a th-order FRT $f_a(u)$ of the function $f(u)$ is defined for $0 < |a| < 2$ as [14]

$$f_a(u) \triangleq \int_{-\infty}^{\infty} K_a(u, u') f(u') du' \quad (14)$$

$$K_a(u, u') \triangleq A_\alpha \exp [j\pi(u^2 \cot \alpha - 2uu' \csc \alpha + u'^2 \cot \alpha)] \quad (15)$$

where

$$A_\alpha = \frac{\exp[-j(\pi \operatorname{sgn}(\alpha)/4 - \alpha/2)]}{|\sin \alpha|^{1/2}} \quad \text{and} \quad \alpha = \frac{a\pi}{2} \quad (16)$$

The kernel $K_a(u, u')$ approaches $\delta(u - u')$ and $\delta(u + u')$ as a approaches 0 and ± 2 , respectively, and are defined as such at these values. The FRT reduces to the ordinary FT when $a = 1$ and is linear and index additive; that is, the a_1 th-order FRT of the a_2 th-order FRT of a function equals the $(a_1 + a_2)$ th-order FRT. With a similar notation as in the case of DFT, the a th-order discrete fractional Fourier transform (DFRT) of \mathbf{f} , denoted \mathbf{f}_a , can be expressed as $\mathbf{f}_a = \mathbf{F}^a \mathbf{f}$ where \mathbf{F}^a is the $N \times N$ DFRT matrix which corresponds to the a th power of the ordinary DFT matrix \mathbf{F} .

A.3 Hartley Transform (HT)

HT [15] is another widely-used technique in signal processing applications [15]. The discrete Hartley transform (DHT) of a signal $f(n)$ is defined as:

$$H(k) = \mathcal{H}\{f(n)\} \triangleq \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) \operatorname{cas} \left(\frac{2\pi}{N} nk \right) \quad (17)$$

where $\operatorname{cas}(x) \triangleq \cos(x) + \sin(x)$ and N is the length of the signal $f(n)$. There is a close relationship between DFT and DHT such that if the DFT of a signal $f(n)$ is expressed as $F(k) = F_R(k) - jF_I(k)$, then its DHT is related to the real and imaginary parts of the DFT by $H(k) = F_R(k) + F_I(k)$. The DHT can also be represented in matrix notation as $\mathbf{h}_1 = \mathbf{H}\mathbf{f}$ where \mathbf{f} is an $N \times 1$ column vector, \mathbf{H} is the $N \times N$ DHT matrix, and \mathbf{h}_1 is the DHT of \mathbf{f} .

A.4 Wavelet Transform (WT)

The discrete wavelet transform (DWT) of a function $f(t) \in L^2$ is defined as [16]:

$$f(t) = \sum_{k=-\infty}^{\infty} c(k) \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j, k) \psi_{j, k}(t) \quad (18)$$

where $c(k) = \langle f(t), \varphi_k(t) \rangle = \int f(t) \varphi_k(t) dt$ and $d(j, k) = \langle f(t), \psi_{j, k}(t) \rangle = \int f(t) \psi_{j, k}(t) dt$. The coefficients $\{c(k)\}_{k=-\infty}^{\infty}$ and $\{d(j, k)\}_{j=0, k=-\infty}^{\infty}$ correspond to the DWT of signal $f(t)$. These coefficients completely describe the original signal and can be used in a way similar to Fourier series coefficients. The procedure of finding the DWT coefficients can be summarized as:

$$c_j(k) = \sum_{m=0}^{M-1} h(m - 2k) c_{j+1}(m) \quad (19)$$

$$d_j(k) = \sum_{m=0}^{M-1} g(m - 2k) c_{j+1}(m) \quad (20)$$

Here, $k = 0, 1, \dots, 2^j N - 1$ where N is the number of samples of the original signal that should be a power of 2. In these equations, $h(n)$, $n = 0, \dots, M - 1$ is a low-pass filter called the *scaling filter* and $g(n)$ is an high-pass filter called the *wavelet filter* related to the scaling filter by $g(n) = (-1)^n h(M - 1 - n)$, $n = 0, \dots, M - 1$. Usually, $c_o(k)$'s are taken as the samples of the original signal.

A.5 Self Organizing Feature Map

Self organizing ANNs are obtained by unsupervised learning algorithms that have the ability to form internal representation of the network that model the underlying structure of the input data without supervision. These networks are commonly used to solve the scaling problem encountered in supervised learning procedures. It is possible to achieve best results with these networks as feature extractors prior to a linear classifier or a supervised learning process [13]. Most commonly used algorithm for generating these networks is Kohonen's self organizing feature-mapping algorithm (KSOFM) [17] in which the weights are adjusted from the input layer to the output layer. The output neurons are interconnected with local connections and are geometrically organized in one, two, three, or even higher dimensions.

IX. EXPERIMENTAL STUDIES

Panasonic transducers with aperture radius $a = 0.65$ cm, resonance frequency $f_o = 40$ kHz, and $\theta_o \cong 54^\circ$ [18] are used (Fig. 1) with a center-to-center separation of $d = 25$ cm. The sensing unit is mounted on a stepper motor with step size 1.8° whose motion is controlled through the parallel port of a PC. Data acquisition is through a 12-bit 1 MHz PC A/D card. Starting at the transmit time, 10,000 samples of each echo signal are collected and thresholded. The amplitude information is extracted by finding the maximum value of the signal after the threshold is exceeded.

The targets employed in this study are: cylinders with radii 2.5, 5.0 and 7.5 cm, a planar target, a corner, an edge of $\theta_e = 90^\circ$, and an acute corner of $\theta_c = 60^\circ$. Amplitude and TOF patterns of these targets are collected with the sensing unit described above at 25 different training locations (r, θ) for each target (Fig. 3). The target primitive located at range r and azimuth θ is scanned by the sensing unit for scan angle $-52^\circ \leq \alpha \leq 52^\circ$ with 1.8° increments. For given scan range and motor step size, $58 (= \frac{2 \times 52^\circ}{1.8^\circ})$ angular samples of each of amplitude and TOF patterns $[A_{aa}(\alpha), A_{bb}(\alpha), A_{ab}(\alpha), A_{ba}(\alpha); t_{aa}(\alpha), t_{bb}(\alpha), t_{ab}(\alpha), t_{ba}(\alpha)]$ are acquired at each target location. Four similar sets of scans are collected for each target at each location, resulting in 700 ($= 4$ data sets $\times 25$ locations $\times 7$ target types) sets of signals to construct the training data. This training data is used in statistical pattern recognition techniques, LDA, FCC, and ANNs.

Two different sets of test data are constructed to assess and compare the performances of the different classification and fusion techniques. For test data I, each target is placed in turn in each of the 25 training positions in Fig. 3.

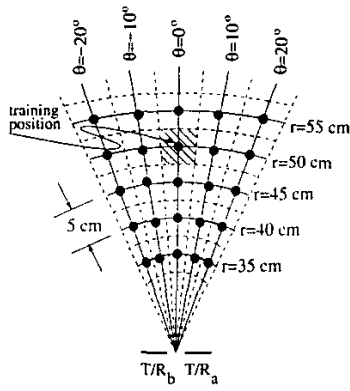


Fig. 3. Discrete network training locations.

Four sets of patterns are collected for each combination of target type and location, again resulting in 700 sets of patterns. While collecting test data II, targets are situated at randomly generated locations within the area shown in Fig. 3, not necessarily corresponding to one of the 25 grid locations.

Initially, the TDA is employed at each angular step to classify and localize the target. Then, DS and various voting schemes are used to fuse decisions made at each of the 58 angular steps to reach a single decision for a pattern set. Weighted averages of the 58 range and azimuth estimates in a pattern set are calculated to find the range and the azimuth of the target. In DS, these weights are the belief values assigned to range and azimuth estimates at each angular step. In SMV, these weights are taken as one for both range and azimuth at each angular step. In voting schemes including preference ordering and reliability measures, reliabilities assigned at each angular step are used as weights. In this case, preference orders are taken as the belief values assigned to each target at each angular step.

In statistical pattern recognition techniques, LDA, and FCC, three different training (design) sets are constructed using training data consisting of different vector representations. In each design set, seven classes each with 100 vector representations exist. These are:

$$\begin{aligned} \mathbf{x}_1 &: [A_{aa}(\alpha), A_{bb}(\alpha), \frac{A_{ab}(\alpha) + A_{ba}(\alpha)}{2}, t_{aa}(\alpha), t_{bb}(\alpha), \frac{t_{ab}(\alpha) + t_{ba}(\alpha)}{2}]^T \\ \mathbf{x}_2 &: [A_{aa}(\alpha) - A_{ab}(\alpha), A_{bb}(\alpha) - A_{ba}(\alpha), t_{aa}(\alpha) - t_{ab}(\alpha), t_{bb}(\alpha) - t_{ba}(\alpha)]^T \\ \mathbf{x}_3 &: [(A_{aa}(\alpha) - A_{ab}(\alpha))(A_{bb}(\alpha) - A_{ba}(\alpha)), (A_{aa}(\alpha) - A_{ab}(\alpha)) + (A_{bb}(\alpha) - A_{ba}(\alpha)), \\ & (t_{aa}(\alpha) - t_{ab}(\alpha))(t_{bb}(\alpha) - t_{ba}(\alpha)), (t_{aa}(\alpha) - t_{ab}(\alpha)) + (t_{bb}(\alpha) - t_{ba}(\alpha))]^T \end{aligned}$$

The first vector representation \mathbf{x}_1 is taken as the raw patterns without any processing, except for averaging the cross terms which should ideally be equal. The second vector representation \mathbf{x}_2 has been motivated by the TDA. The third vector representation \mathbf{x}_3 is motivated by the differential terms which are used to assign belief values to the targets classified by the TDA [2].

Thirdly, ANNs are employed. Performance of ANNs is affected by the choice of parameters related to the network structure, training algorithm, and input signals, as well as parameter initialization. We considered samples of the following different signal representations as alternative inputs to the ANNs:

- $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$
- DWT of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ at different resolutions j
- DFT of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$
- DFRT of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ at different orders a
- DHT of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$
- Features of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ extracted by KSOFM

The first three input signal representations are the same vector representations used above for statistical pattern recognition techniques, LDA, and FCC. In this case, they have been used both in their raw form and after taking their DFT, DFRT, DHT, and DWT, as well as after feature extraction by KSOFM. The a th-order DFRT of these three input signal representations for a values varying from 0.05 to 0.95 with 0.05 increments are used. DWT of each signal representation at different resolutions j are used. Finally, the features extracted by using KSOFM are used as input signals both prior to ANNs and linear classifiers.

Initially, ANNs are trained by using BP to classify and localize these target types. Next, modular ANNs for each type of input signal have been implemented in which three separate networks for target type, range, and azimuth, each trained with BP, are employed. ANNs using the same input signal representations are also trained with the GS algorithm. This algorithm can be applied to target classification but not localization [6].

To make a comparison of all differentiation and fusion techniques employed in this study, highest percentages of correct classification obtained with each method are given in Table X. With the TDA and the fusion techniques based on this algorithm (DS, and various voting schemes) which do not use training data, only three of the target types employed in this study (P, C, and AC) can be differentiated. However, all targets can be differentiated with the other methods. The fact that the other methods are able to distinguish all target types indicates that they must be making more effective use of the available data than the TDA. Statistical pattern recognition techniques, LDA, and FCC cannot be used for target localization, unlike all of the other methods. The highest percentages of correct classification is 100% and is obtained with non-modular ANN trained with BP employing the DFRT. However, best localization is achieved with the modular ANN trained with BP employing the DWT. The lowest percentage of correct classification is 61% and is obtained with the TDA. For most cases, vector representation \mathbf{x}_1 gives the best results, followed by \mathbf{x}_2 and \mathbf{x}_3 . Note that different vector representations are not applicable to TDA and fusion techniques based on this algorithm since they determine the target type by using differential signal \mathbf{x}_2 obtained by using original signal \mathbf{x}_1 and they assign belief values to their decisions using \mathbf{x}_3 . For most cases, the results obtained with test data I are the best for all methods except the TDA and the fusion techniques based on this algorithm. However, the gap between the results of test data I and II is higher for statistical pattern recognition techniques and LDA than that for all other methods.

X. CONCLUSIONS

In this study, different classification and fusion techniques have been compared for target classification and localization with sonar. These classification schemes include TDA, statistical pattern recognition techniques, LDA, FCC, and ANNs. ANNs are trained with different input signal representations. KSOFM is used for feature extraction both prior to ANNs and linear classifiers. Non-modular and modular ANNs and two different training algorithms are considered. The fusion techniques used in this study are DS, SMV, and various voting schemes with preference ordering and reliability measures.

Although all methods considered in this study can be used for target differentiation, statistical pattern classification techniques, LDA, and FCC cannot be used for localization. The performances of the methods are compared on two different test data sets that include patterns acquired from targets situated at training locations as well as arbitrary locations. TDA and fusion techniques employed in this study can only be applied to P, C, and AC differentiation. However, all seven target types can be differentiated using the other methods. Target localization performance of ANNs is better than the TDA and fusion techniques. Moreover, generalization capability of ANNs, FCC, and TDA and fusion techniques are better than that of statistical pattern recognition techniques and LDA. Among ANNs, best generalization ability is obtained with the ANNs trained with GS, followed by modular and non-modular ANNs trained with the BP in the given order. Among different input signal representations used with ANNs, DWT and DFRTed signals usually result in better performance than raw signals, DFT and HTs of these signals, and the features extracted by KSOFM. For statistical pattern recognition techniques, the target classification performance of non-parametric approaches (k -NN classifiers, and KE) is better than that of PDE using heteroscedastic and homoscedastic NM. Among non-parametric methods, the best results are obtained by generalized k -NN, followed by KE and k -NN.

REFERENCES

- [1] B. Barshan and R. Kuc, "Differentiating sonar reflections from corners and planes by employing an intelligent sensor," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.12, no.6, pp.560-569, June 1990.
- [2] B. Ayrulu and B. Barshan, "Identification of target primitives with multiple decision-making sonars using evidential reasoning," *Int. J. Robot. Res.*, vol.17, no.6, pp.598-623, June 1998.
- [3] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
- [4] B. Parhami, "Voting algorithms," *IEEE Trans. Reliab.*, vol.43, no.4, pp.617-629, December 1994.
- [5] K. J. Arrow, *Social Choice and Individual Values*, New York: Wiley, 1951.
- [6] B. Ayrulu, *A Comparison of Different Approaches to Target Differentiation with Sonar*, Ph.D. thesis, Bilkent University, Department of Electrical Engineering, Ankara, Turkey, 2001.
- [7] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol.33, pp.1065-1076, 1962.
- [8] D. J. Hand, *Discrimination and Classification*, John Wiley & Sons, New York, NY, 1986.
- [9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, NY, 1986.
- [10] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley & Sons, New York, NY, 1976.
- [11] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York, NY, 1992.
- [12] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol.3, pp.32-57, 1974.
- [13] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, New Jersey, NJ, 1994.
- [14] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier Transform with Applications in Optics and Signal Processing*, John Wiley & Sons, New York, NY, 2000.
- [15] R. N. Bracewell, *The Hartley Transform*, Oxford University Press, New York, NY, 1986.
- [16] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol.11, no.7, pp.674-693, July 1989.
- [17] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybern.*, vol.43, no.1, pp.59-69, 1982.
- [18] Panasonic Corp., Ultrasonic ceramic microphones, 12 Blanchard Rd., Burlington, MA. 1989.
- [19] S. W. Utete, B. Barshan, and B. Ayrulu, "Voting as validation in robot programming," *Int. J. Robot. Res.*, vol.18, no.4, pp.401-413, April 1999.
- [20] B. Barshan, B. Ayrulu, and S. W. Utete, "Neural network based target differentiation using sonar for robotics applications," *IEEE Trans. Robotics Automat.*, vol.16, no.4, pp.435-442, August 2000.

method	targets discriminated	differentiation accuracy (%)	r/θ est (%)	training data	learning	parametric
TDA [1], [2]	P,C,AC	61	yes 16/19	not used	no	no
DS [19]	P,C,AC,{E,CY,U}	89	yes 20/31	not used	no	no
voting	P,C,AC,{E,CY,U}		yes	not used	no	no
SMV [19]		82	16/19			
with pref. ordering and rel. measures [6]		94	29/32			
statistical pattern recognition [6]:	P,C,AC,E,CY		no	used	no	
k -NN		98		stored		no
generalized k -NN		99		stored		no
KE		99		stored		no
PDE with:						
heteroscedastic NM		81		not stored		yes
homoscedastic NM		76		not stored		yes
LDA [6]	P,C,AC,E,CY	71	no	used, not stored	no	no
FCC [6]	P,C,AC,E,CY	98	no	used, not stored	yes	no
ANNs [6], [20]:	P,C,AC,E,CY		yes(yes)[no]	used	yes	no
raw signal		95 (95) [95]	79/89(73/95)	not stored		
DWT		98 (99) [97]	71/90(80/92)			
DFT		97 (98) [97]	64/86(72/94)			
DFRT		100 (98) [97]	75/93(68/86)			
DHT		99 (97) [97]	67/84(62/84)			
KSOFM		78 (76) [13]	24/69(21/66)			
KSOFM with linear classifier [6]	P,C,AC,E,CY	85	yes 42/80	used, not stored	yes	no

TABLE I:

OVERVIEW OF THE METHODS COMPARED. THE TARGET TYPES ENCLOSED IN BRACES CAN BE RESOLVED ONLY AS A GROUP. THE NUMBERS IN PARENTHESES ARE FOR MODULAR NETWORKS, WHEREAS THE NUMBERS IN BRACKETS ARE FOR NETWORKS TRAINED WITH GS.