Acoustic Targets Identification Using Hybrid Classifiers

Hatem Khater, M.S.Zaghloul

Arab Academy for Science, Technology, and Maritime Transport, Electronic and Communication Department, P.O.1029, Alex, Egypt, hatem.a.khater@gmail.com, dr_mszaghloul@yahoo.com

Abstract-The goal of this paper is to develop classification techniques for acoustic targets. An extensive experimental study on real data of various classification and combination rules has been performed. It is applied to the problem of matching score fusion using data sets consisting of features extracted from underwater acoustics signals to produce a more effective system to identify and recognize naval targets (ships or submarine).This paper provides a comparative study between three well-known classifiers. They are, continuous hidden Markov model (CHMM), K-nearest neighbor (K-NN) and artificial neural network (ANN) and their combination (COMB) to identify and recognize the naval target. Mel frequency spectral coefficients (MFCCs) are chosen as the studied features. The general Gaussian density distribution HMM was developed for the CHMM system. We studied the effect of speed, distance and direction of the target on the identification process. The Results show that COMB has always best IR result in all experiments. In addition to the results show that CHMM gives the best identification rate (IR) at 91.67% while changing range, 100% while changing direction and 58.3% while changing the speed which is better than 75%, 83.33% and 41.67% of ANN for the same set of experiments using simulated targets data COMB achieves 100% IR which is higher than CHMM, K-NN and ANN.

Key words: *Combining Classifiers, Matching, Sonar Signal*

I. INTRODUCTION

Sonar system transmits and receives acoustic pulses of energy. Most of the energy arriving onto the seabed is scattered forward in the specular direction, with a small portion absorbed by the seabed itself. However a small portion of energy is scattered back to the sonar, amplified and recorded, the strength of the returning echo will be governed by several factors. Identification of an underwater acoustic signal is one of the important fields of pattern recognition and has been extensively studied in the last decade. Much work has been performed recently in this area to identify naval targets automatically using sonar signal (either passive or active) to reduce both operator load when confronted with many beams of data concurrently and time needed for the identification process manually [1, 2, 3, 4, 5] . Several techniques of feature extraction have been developed. Features from different classes are remarkably overlapped due to the complicated mechanism of vessel radiated noise [1]. Various kinds of classifiers such as probabilistic neural networks *(PNN)* [1], Kohonen Neural Networks [2] and back propagation neural network [3, 4] have been used as classifiers of targets radiated noises. Other areas of application where the same techniques could be used include, both military and for civilian purposes [1]. In all of these applications, underwater sound must be recorded, pre-processed, and classified into one of a series of possible classes. In this paper, an *ANN, K-NN, CHMM* and their *COMB* based identification techniques have been used to estimate the performance of naval targets identification using six simulated targets and five real targets. We evaluated the effect of the direction, the distance and the speed of the target on the performance of identification.

II. SIGNAL PREPROCESSING

In the proposed system a sound recorder is used to record and store the received analog data (sound signal) in a digital format at a sampling frequency of 11.25 KHz. Then, the signal is Pre-emphasized using a FIR high-pass filter to flatten the signal spectrum [6]. The pre-emphasized signal is then framed and windowed. The window function that is applied to the signal is preferably not rectangular, as this can lead to distortion due to vertical frame boundaries. A common choice for the non-rectangular window is the Hamming window. The frame length can vary, but based on empirical results, is often chosen from 20 to 30 ms. Using this length and an overlap of 30% to 50 % of the frame length is implemented to produce a smooth transition between frames [7]. *As Mel frequency spectral coefficients MFCCs* are well known features used to describe speech signal, they are the features we used to describe the sonar signal also.

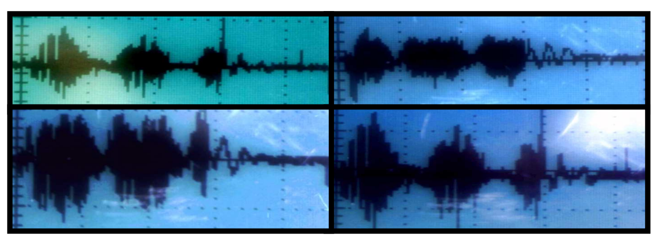

Fig.1. sound Echo Returned from different targets, sweep time 2 msec. per division, frequency 10 KHz, amplitude 10v/cm

II.1. PATTERN MATCHING

II.1.1 ARTIFICIAL NEURAL NETWORK (ANN) APPROACH

An artificial neural network is a mathematical model or computational model based on biological neural networks [8]. ANN architecture is classified into two principal types: recurrent and non-recurrent networks. Both feed forward and feedback paths between the layers characterize recurrent neural networks. The feedback paths enable activation at any layer to either be used as an input to a previous layer or be returned to that layer after one or more time steps. In this paper, we used the Elman network, which is a special kind of a recurrent network [9]. The Elman network, originally developed for speech recognition, is a two-layer network in which the hidden layer is recurrent. The inputs to the hidden layer are the present inputs and the outputs of the hidden layer, which are saved from the previous time-step in buffers, called context units. Hence, the outputs of the Elman network are functions of the present state, the previous state (as supplied by the context units) and the present inputs. This means that when the network is shown a set of inputs, it can learn to give the appropriate outputs in the context of the previous states of the network. The advantage of Elman networks over fully recurrent networks is that back propagation is used to train the network while this is not possible with other recurrent networks where the training algorithms are more complex and therefore slower [8].

II.1.2 K-NEAREST NEIGHBOR ALGORITHM (K-NN)

K-nearest neighbor *(K-NN)* is a supervised learning algorithm for classifying objects based on closest training examples in the feature space. Given a query point, one finds K (training points) closest to the query point [10]. The classification uses a majority vote among the classification of the K objects. The K nearest neighbor algorithm uses neighborhood classification as the prediction value of the new query instance. The best choice of K depends upon the data; generally, larger values of K reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good K value can be selected by various heuristic techniques, for example, cross-validation. The special case where the class is predicted to be the class of the closest training sample (i.e. when K = 1) is called the nearest neighbor algorithm [11] [12]. Distance (d) is defined as the Euclidean distance between two points P = (p_1, p_2, ...p_n) and Q = (q_1,q_2,....q_n) as follow:

$$\sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}=\sqrt{\sum_{i=1}^{n}(p_i-q_i)^2} \qquad (1)$$

From these distances, a distance matrix is constructed between all possible pairings of points (P, Q). Each data point within the data set has a class label in the set, C= {c_1,...,c_n}. The data points, K-closest neighbors (K being the number of neighbors) are then found by analyzing the distance matrix. The K-closest data points are then analyzed to determine which class label is the most common among the set. The most common class label is then assigned to the data point being analyzed.

II.1.3. HIDDEN MAARKOV MODEL (HMM) APPROACH

The *HMMs* used in this system are continuous *HMMs (CHMMs),* in which we measure the likelihood via conditional probability of the observations given for each model. But all of these output probabilities utilize random variables. The *CHMM* uses Gaussian density functions, which are the most widely efficient density functions without loss of generality. In fact, *CHMM* is more accurate than *DHMM* because it uses continuous observations to construct the model directly without a quantization. However, the computational complexity when then using the *CHMMs* is more than the computational complexity when using *DHMMs*. It normally takes more time in the training phase [6]. In this approach, we used the Baum-Welch algorithm during the training phase. The Viterbi algorithm was applied to compute the likelihood function of the testing data in the identification phase. The *HMM* with the largest likelihood is the identified unit.

III. PLATFORM

The recorded signals were stored using mono channel with 11.25 kHz sampling frequency and 8 bits quantization level. The recording time was 15 second for each signal. Each signal passed through a pre-emphasis high-pass filter $H(z) = 1 - \alpha z^{-1}$ where $\alpha = 0.97$ and segmented into 30 milliseconds length frames with 50% overlap. The feature vectors were extracted (each vector consists of 12 *MFCCs*) for each segment. A four-state left to right continuous density HMM identification engine was developed. We used a standard *k*-means initialization procedure to initialize the HMM parameters. The HMM-based identification engine was trained using the Baum-Welch algorithm. In the training phase, the feature vectors of the samples were extracted after the preprocessing procedures. At the next stage, each sound signal was divided into four equal length segments and the probability distribution density function of each segment was estimated using the EM algorithm with two mixtures. During the identification phase, the feature vectors sequence was segmented and the likelihood function was calculated using the Viterbi algorithm by assuming vectors independence. By comparing the likelihood functions that are derived from distinct models, the identified unit was determined [13]. On the other hand an Elman network with 12 neurons in the input layer , 40 neurons in the hidden layer and n neuron in the output layer (-depending on the number of classes) was developed. We used the hyperbolic tangent sigmoid transfer function for the hidden layer, and linear transfer function for the output layer. We used a training function updates weight and bias values according to Levenberg-Marquardt optimization, to train the network. The neuron with the highest output was selected to determine the recognized class (unit).Fig.2 indicates overall identification system

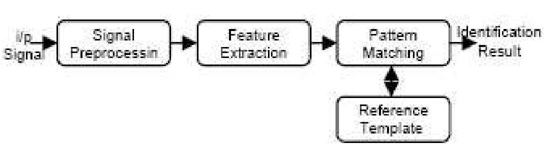

Fig.2. overall identification system

VI. SCORE NORMALIZATION AND COBINATION PROCESS

A Min-Max method is used for score normalization to map the raw scores to the [0,1] range as shown in Equation (2)[14]. We indicate a raw matching score obtained from features extraction as s and a normalized score as $\hat{s}$. The functions $max(s)$ and $min(s)$ return the end points of the score range. Fusing the scores of several classifiers has proved to be a promising approach to improve the overall accuracy of pattern recognition systems. We apply a weighted sum rule to combine $\hat{s}_{ANN}$, $\hat{s}_{KNN}$ and NN and CHMM classifiers. The output of - $\hat{s}_{CHMM}$, the ANN, K the combined classifiers can be obtained by multiplying the normalized matching score $\hat{s}$ by suitable weights, w_i, as shown in Equations (3) and (4). A threshold on the combined score is then applied to pick the most reliable matches.

$$\hat{s} = (s - \min(s)) / (\max(s) - \min(s)) \quad (2)$$

$$sum = \sum_i w_i \hat{s}_i \text{ , where } \sum w_i = 1 \quad (3)$$

$$sum = w_1 \hat{s}_{ANN} + w_2 \hat{s}_{KNN} + w_3 \hat{s}_{CHMM} \text{ ,}$$

Where:- $w_1 + w_2 + w_3 = 1$ (4)

V. EXPERIMENT & RESULTS

V.1. DATASETS

In our experiments in this research, two databases had been used. The first database consists of simulated targets data generated using the sonar simulator STU-3 (sonar training unit) [15, 16], which was used to train the sonar operators. This database contains six simulated targets; each having three groups of records. The first group recorded the target noise at different directions 0, 90 and 270 degrees while maintaining both distance and speed constant. The second group recorded the target noise at different distances 2, 4 and 8 miles while maintaining both direction and speed constant. The last group recorded the target noise at different speeds 2, 4 and 6 knots for submarines and 8, 12 and 16 knots for the surface targets while maintaining both distance and direction constant for both of them. Each target was named either 'S' for surface targets or 'U' for underwater targets, followed by the target number, letter indicating the variable factor ('d' for direction, 'r' for range and 's' for speed) and finally the value of the factor. For example, 'S2d270' is the file for surface target number 2 at direction 270. We constructed 54 file database (6 targets * 3 variant factors * 3 values for each factor). The second dataset is consisting of real targets data that had been recorded using different types of passive sonar equipments to record sounds generated. This dataset containing recorded noise coming out of five real targets at different distance and directions (see Table 1).

V.2. EXPERIMENTAL SETUP

In the first set of experiments, we examined the effect of the target direction on the IR. We trained the system with signal generated by simulated targets in direction 0 degree relative to the sonar platform and then testing the system using two other signals (for each target) received from directions 90 and 270 degree. In the second set of experiments, we examined the effect of the target range on the IR. We trained the system with signal generated at 4 Nm from the sonar platform and then testing the system using two other signals for (each target) received at range 2 and 8 Nm. In the third set of experiments, we examined the effect of the target speed on the IR. We trained the system with signal generated with speed 4 Knot for underwater targets and 12 Knot for surface targets and then testing the system using two other signals for each target, 2 and 6 Knot for underwater targets and 8 and 16 Knot for surface targets. The fourth set of experiments was done using the real targets. We trained the system with one signal from each class and tested with the rest of the signals.

V.3. RESULTS

Signals	class	Distance(Nm)	Direction(Deg.)
Target 1	U1	3	101
Target 2	U1	2.6	117
Target 3	U2	1.4	125
Target 4	U2	1.8	166
Target 5	U2	1.6	151
Target 6	U2	2.4	175
Target 7	U2	1.8	191
Target 8	U2	2	126
Target 9	U3	--	300
Target 10	U3	--	339
Target 11	U4	--	290
Target 12	U4	--	305
Target 13	U4	--	316
Target 14	U4	--	325
Target 15	U4	--	340
Target 16	U5	1.4	220
Target 17	U5	2	270
Target 18	U5	2.2	290
Target 19	U5	3	295

Table 1: Real Targets Data

In this paper we present the performance results for a number of classifiers and a combiner used to fuse the information from number of classifiers. The classifiers were the neural network classifier, the K-nearest neighbor classifier, the continuous hidden Markov model (CHMM) and their combination (COMB). This paper presents some experiments that underlines the reasons for expecting such improvements and quantifies the gains achieved. Results obtained from the first set of experiments showed that the IR was about 94% for *(CHMM) and (K-NN)* and 64 % *(ANN)* and 95% (*COMB*) (not shown in paper). Obtained results showed that IR for K-NN, ANN, CHMM and COMB classifiers hadn't been affecting by changing the direction of the target.

Results obtained from the second set of experiments showed that the IR was 93% for *(CHMM) and 94% (K-NN)* and 50 % for *(ANN)* and 95% for (*COMB)*. Obtained results showed that IR for K-NN, ANN, CHMM and COMB classifiers were affected by changing the distance of the target while *CHMM* and K-NN had better IR than *ANN*. Results obtained from the third set of experiments (not shown in paper) demonstrated that IR for K-NN, ANN, CHMM and COMB classifiers were affected by changing the speed of the target. We also noticed that target speed was the most effective factor on the identification process; we think the reduction caused in IR was due to decreasing of the SNR obtained in high speeds. Results obtained from the fourth set of experiments showed that the IR was 93 % (CHMM) and 63.68% (ANN) (see Table 2). Obtained results showed that IR in CHMM and K-NN system was better than NN system. According to the parameters of the targets, we had no chance to figure out the effect of the speed on the identification rate of real targets.

Targets	Classifier	Ul	U 2	U 3	U 4	U5
Target1	*CHMM*	**92.65**	84.83	87.44	35.23	55.13
	ANN	**47.5**	27.25	23.75	46.75	30.5
	KNN	**92.34**	84.92	87.55	35.43	55.44
	COMB	**92.78**	84.99	87.55	35.33	55.54
Target 2	*CHMM*	**87.49**	82.04	86.83	39	57.67
	ANN	41	41.5	**43.75**	23.25	38.25
	KNN	**87.96**	84.3	86.67	39.5	57.56
	COMB	**88.99**	82.65	86.89	39.8	57.88
Target 3	*CHMM*	87.73	**92.154**	89.78	44.36	60.91
	ANN	34.5	33.25	19	30.75	**36.25**
	KNN	87.83	**92.34**	89.94	44.84	60.93
	COMB	87.73	**92.88**	89.77	44.84	60.94
Target 4	*CHMM*	84.42	**89.67**	85.78	42.19	59.77
	ANN	38.75	**46.75**	14.75	40.75	33.75
	KNN	84.87	**89.88**	85.89	42.33	59.64
	COMB	84.98	**89.92**	85.83	42.34	59.92
Target 5	*CHMM*	85.12	**93.25**	87.89	44.17	62.37
	ANN	26.25	**47**	23.5	28.5	27.75
	KNN	85.43	**93.56**	87.76	44.54	62.89
	COMB	85.94	**93.54**	87.99	44.44	62.22
Target 6	*CHMM*	83.85	**91.05**	89.73	47.4	64.37
	ANN	27.75	**37.25**	30	37	25.5
	KNN	83.96	**91.5**	89.65	47.49	64.54
	COMB	83.95	**91.65**	89.88	47.65	64.54
Target 7	*CHMM*	82.77	**87.59**	82.92	35.43	60.93
	ANN	31.25	**60.25**	36	37.25	27
	KNN	82.89	**87.63**	82.48	35.76	60.43
	COMB	82.99	**87.94**	82.65	35.88	60.33
Target 8	*CHMM*	80.43	**90.21**	82.78	41.62	67.67
	ANN	23.75	**52.25**	23.75	39	10.25
	KNN	80.76	**90.44**	82.87	41.65	67.87
	COMB	80.87	**90.33**	82.88	41.77	67.88
Target 9	*CHMM*	80.86	75.66	**92.35**	44.68	58.44
	ANN	56	23	**65.75**	55.5	7
	KNN	80.97	75.77	**92.76**	44.54	58.65
	COMB	80.88	75.67	**92.87**	44.56	58.66
Target10	*CHMM*	82.21	79.9	**89.22**	47.62	53.72
	ANN	18.25	55.5	**61**	49.75	25.25
	KNN	82.55	79.96	**89.54**	47.76	53.74
	COMB	82.45	79.56	**89.55**	47.66	53.89
Target 11	*CHMM*	58.9	51.9	10.7	**88.71**	56.99
	ANN	25.5	23.25	50	**50.5**	42.75
	KNN	58.67	51.77	10.67	**88.56**	57.32
	COMB	58.66	51.87	10.77	**88.77**	52.5
Target 12	*CHMM*	14.5	20.7	20.71	**74.56**	57.81
	ANN	53.75	5.25	27.25	**59.25**	42.25
	KNN	14.56	20.8	20.89	**74.89**	57.87
	COMB	14.76	20.9	20.79	**74.99**	57.89
Target 13	*CHMM*	25.1	32.9	19.42	**64.71**	61.58
	ANN	60.75	5.5	**68.25**	65	40.5
	KNN	25.4	32.6	19.65	**64.68**	61.87
	COMB	25.3	32.4	19.76	**64.87**	61.98
Target 14	*CHMM*	43.9	44.4	5.39	**82.17**	62.65
	ANN	25	24.75	50	**52**	37.75
	KNN	43.87	44.7	5.9	**82.66**	62.77
	COMB	43.95	44.9	5.8	**82.58**	62.87
Target 15	*CHMM*	66.3	65	20.29	**88.98**	58.35
	ANN	5.75	27.5	36.75	**47.75**	18.75
	KNN	66.3	65	20.29	**88.98**	58.35
	COMB	66.87	65	20.29	**88.98**	58.35
Target 16	*CHMM*	33.35	76.89	8.11	33.86	**90.94**
	ANN	53.75	14.25	7.5	**56.75**	54.5
	KNN	33.76	76.87	8.98	33.97	**90.95**
	COMB	33.87	76.99	8.97	33.99	**92.65**
Target 17	*CHMM*	37.87	74.86	10.56	38.96	**90.87**
	ANN	38	23	43.25	36.25	**57.25**
	KNN	37.89	74.98	10.65	38.98	**90.90**
	COMB	39.99	76.97	19.50	38.99	**91.88**
Target 18	*CHMM*	40.18	77.65	7.01	34.45	**91.05**
	ANN	21.5	31	26	**54.75**	32.5
	KNN	40.18	77.65	7.01	34.45	**91.05**
	COMB	40.18	77.65	7.01	34.45	**91.05**
Target 19	*CHMM*	39.77	76.56	11.23	35.77	**90.96**
	ANN	25.5	37	52.75	62.5	65.5
	KNN	39.89	76.98	11.89	35.90	**91.54**
	COMB	39.66	76.85	11.85	35.89	**92.59**

Table 2: Identification (%)of Real Targets

VI. CONCLUSION

This paper investigated efficient fusion techniques for high-performance automatic naval targets identification using underwater acoustics. The ANN, CHMM and K-NN are well-known pattern matching techniques that are widely used in speech and speaker recognition. In this paper, we applied all of these classifiers and their combiner to identify naval targets. We used simulated targets to study the effect of the distance, bearing and speed of the target on the IR. Results show that COMB has always best IR result in all experiments. Based on these results, CHMM and K-NN had very good choice for maximizing the identification rate of the naval target identification. ANN did not provide as good results because of small number of datasets. We found that both direction and distance did not have as much effect as speed on the IR. We planned to do more studies to figure out how to minimize the effect of the speed on the identification rate. Also, we intended to study the effect of the ambient noise on the identification rate. (Due to the limited size for this paper some of the result is not shown but all are available upon request).

VII. REFERENCES

[I] J. Chen, S. Tang and H. Li, "Association Rules Enhanced Classification of Underwater Acoustic Signal," IEEE International Conference on Data Mining, 2001.

[2] V. José de Almeida, "Ship Noise Classification," Univ. of Lisbon, www.isegi.unl.pt, 2002.

[3] P. Chestnut and H. Landsman, "A sonar target recognition experiment," J. Acoustic Soc. Am. Vol. 66.

[4] H. Hashem, "Automatic classification of underwater sonar signals," Alexandria high institute of technology Egypt, MTS/IEEE techno–ocean'04 November 2004.

[5] M. R. Azimi-Sadjadi, D. Yao, Q. Huang, and G. J. Dobeck, "Underwater Target Classification Using Wavelet Packets and Neural Networks," IEEE Transactions on Neural Networks, Vol. 11, NO. 3, May 2000.

[6] M. Weddin, "Speaker Identification for Hearing Instruments", IMM, Denmarks Technical University 2005.

[7] J. Pool, "Investigation of the Impact of High Frequency Transmitted Speech on Speaker Recognition," M.Sc. thesis, University of Stellenbosch , Marsh 2002

[8] S. Kasuriya, C. Wutiwiwatchai and C. Tanprasert,"Comparative Study of Continuous Hidden Markov Models (CHMM) and Artificial Neural Network (ANN) on Speaker Identification System," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2001.

[9] F. Cecconi, M. Campenni, "Recurrent and Concurrenrt Neural Networks for Objects Recognition", international conference on Artificial intelligence and applications, 2006.

[10] Belur V. Dasarathy, editor Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, ISBN 0-8186-8930-7, 1991.

[11] Shakhnarovish, Darrell, and Indyk, Nearest-Neighbor Methods in Learning and Vision, edited by The MIT Press, ISBN 0-262-19547-X, 2005.

[12] Helena Mäkelä and Anssi Pekkarinen, Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. Forest Ecology and Management, Volume 196, Issues 2-3, 26 July 2004, Pages 245-255.

[13]F. Razzazi and A. Sayadiyan, "Evaluation Of Soft Segmen Modeling on a context Independent Phoneme Classification System", The Arabian Journal for Science and Engineering, Volume 32, Number 1B April, 2007.

[14] R. Mohr, Schmid, C. Bauckhage. "Evaluation of interest point detectors," International Journal of Computer Vision, 37 (2), 151-172, (2000).

[15] http://www.atlas-elektronik.de.

[16] Daily Shipping Newsletter, Number 068, Tuesday 18-05-2004