Sea Clutter Neural Network Classifier: Feature Selection and MLP Design

Jose Luis Bárcena-Humanes, David Mata-Moya, María Pilar Jarabo-Amores, Nerea del-Rey-Maestre, and Jaime Martín-de-Nicolás

Signal Theory and Communications Department,
Superior Polytechnic School, University of Alcala,
28805 Alcala de Henares, Madrid, Spain
{jose.barcena,david.mata,mpilar.jarabo,
nerea.rey,jaime.martinn}@uah.com

Abstract. The design of radar detectors in sea clutter environments is really a complex task. A neural network based automatic sea clutter classifier has been designed, as part of an adaptive detector capable of exploiting all the capabilities of detectors designed for specific clutter environments. The most extended sea clutter models have been considered (Gaussian, Weibull and K-distributed). Results show that an MLP with 3 inputs (the variance, the entropy of the modulus of the samples and the correlation coefficient), 6 hidden neurons and 4 outputs, is able to provide a performance similar to the $K-NN$ algorithm with $K = 10$ with a significant reduction in computational cost, a very important feature in real time applications.

Keywords: Sea clutter, radar detection, neural network, classifier, feature extraction, K-Nearest Neighbor.

1 Introduction

Marine radars are extensively used in maritime traffic monitoring and coastal surveillance tasks. One of the more important problems is related to the detection of small boats. Any object in the coverage area can intercept the transmitted signal and reradiate part of it towards the radar. This is the radar echo, that will be acquired by the receiver and applied to the detection stage. The objects to be detected are denoted as *targets*, while all the radar echoes from other non-desired objects are called *clutter*. In marine environments, the radar echo generated by the sea surface is of great interest. It can be significantly stronger than target radar echoes, and it is characterized by a high variability.

The objective is to decide between two hypotheses, target present (H_1) or target absent (H_0), fulfilling the Probability of Detection, P_D, and Probability of False Alarm, P_{FA}, requirements in the area of coverage. The P_D is the probability of detecting a target when it is present. The P_{FA} is the probability of deciding in favor of a target when there is no target. Under both hypotheses, the received signal has a noise component (receiver chain noise factor), and a

I. Rojas, G. Joya, and J. Cabestany (Eds.): IWANN 2013, Part I, LNCS 7902, pp. 575–583, 2013.

clutter component (radar echoes from the sea surface). The Neyman-Pearson, NP, detector is the most extended. It maximizes the P_D, maintaining the P_{FA} lower than or equal to a given value [1]. If $\tilde{\mathbf{z}}$ is the observation vector provided by the radar receiver, a possible implementation of the NP detector is the one based on the comparison of the likelihood ratio, $\Lambda(\tilde{\mathbf{z}})$, to a detection threshold selected according to the desired P_{FA} (1). This detector requires a complete statistical characterization of the observation vector under both hypotheses.

$$\Lambda(\tilde{\mathbf{z}}) = \frac{f(\tilde{\mathbf{z}}|H_1)}{f(\tilde{\mathbf{z}}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{P_{FA}} \tag{1}$$

In radar scenarios, target and clutter models are variable. Since the estimation of target parameters from observation vectors is difficult, a target model is usually assumed, and detection losses are suffered when actual target statistics differ from those assumed in the detector design [2,3]. Robust detectors with respect to target parameters have been analyzed and proposed, assuming constant and known clutter models [4,5,6]. Clutter parameters can be estimated from the environment.

In this paper, a sea clutter classifier is designed for allowing a dynamic selection of the best detection chain, among those designed assuming different clutter models. This classifier is composed of a feature extractor and a Multi-Layer Perceptron (MLP) designed for estimating the posterior probabilities of the classes, using the features provided by the feature extractor. The learning capabilities of the MLP are exploited for achieving performances similar to the ones provided by the K-Nearest Neighbor (KNN), with an important reduction in computational cost, a critical feature for real time implementations [7].

2 Characterization of the Observation Space

The general architecture of a coherent radar is presented in Fig. 1. After each antenna scan, at the output of the digital demodulator, a complex matrix is generated (Fig. 2). The observation vectors will be extracted from this matrix and applied to the detection stage. The anti-clutter processes are usually based on filtering approaches (MTI, Moving Target Indicator, or MTD, Moving Target Detector) and a detection threshold controlled by Constant False Alarm (CFAR) techniques, all of which assume a clutter model [8].

Sea clutter depends on multiple factors: the height of the waves, the wind speed, the transmitted signal frequency and polarization, the radar antenna pointing direction with respect to the direction of waves and the wind, the grazing angle (the complementary of the incident angle) and the size of the observation area (the radar resolution cell projected on the sea surface). In Fig. 3 the general geometry of a marine radar is shown.

The sea state is a term extensively used by mariners as a measure of the wave height, and it is commonly used to describe the roughness of the sea, although it doesn't provide a complete characterization of sea clutter [8].

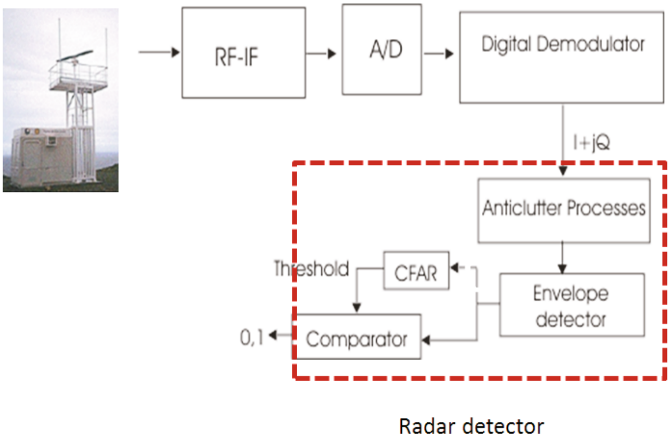

Fig. 1. General architecture of a coherent radar

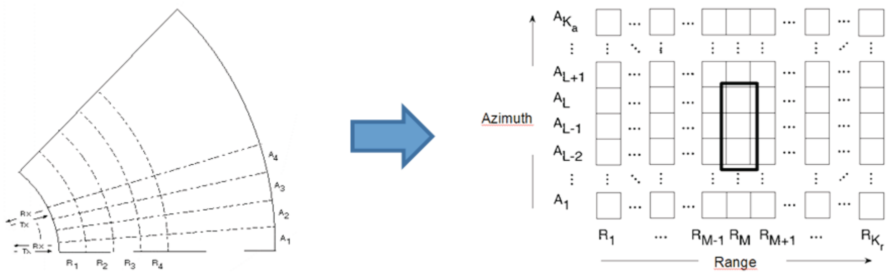

Fig. 2. Output of the digital demodulator for each radar scan

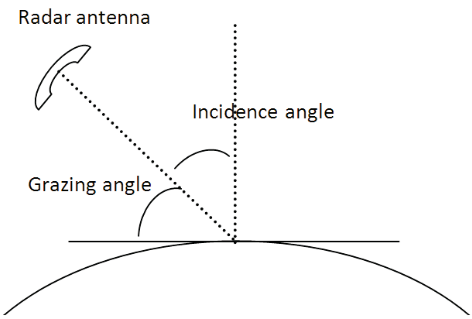

Fig. 3. Geometry of a marine radar

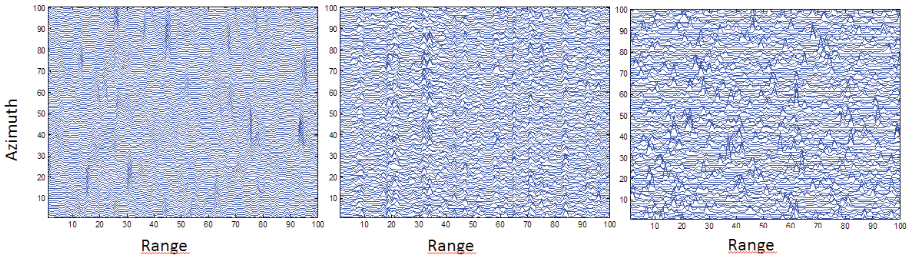

Fig. 4. Amplitude of clutter matrixes at the output of the digital demodulator: Correlated Gaussian (left), spiky K-distributed (center) and Weibull (right)

In low resolution radars (high resolution cells), sea clutter can be modeled as a Gaussian process. This model is also valid in systems with high grazing angles, and high resolution radars with low grazing angles and sea states 0 or 1 (wave height ranging from 0 to 0.1 meters). For higher sea states (wave height higher than 0.1 meters), the Weibull and the K-distributions have been proposed [9,10,11]. The probability density functions, pdfs, for K-distributed and Weibull processes are shown in (2) and (3), respectively. Their associated parameters are shape,ν, and scale, μ, for K, and shape,c, and scale, b, for Weibull. In Fig. 4, amplitude matrixes of clutter samples obtained at the output of the digital demodulator are presented.

$$f(r) = \frac{\sqrt{\frac{4\nu}{\mu}}}{2^{\nu-1}\Gamma(\nu)}\left(\sqrt{\frac{4\nu}{\mu}}r\right)^{\nu} K_{\nu-1}\left(\sqrt{\frac{4\nu}{\mu}}r\right)u(r) \tag{2}$$

$$f(r) = \frac{c}{b}\left(\frac{r}{b}\right)^{c-1} \exp\left(-\left(\frac{r}{b}\right)^{c}\right)u(r) \tag{3}$$

3 Proposed Scheme

Multiple detection strategies have been proposed for different clutter models. In order to exploit the potential of these detectors, a clutter automatic pre-classification stage is proposed, to select the more suitable detection strategy among those implemented (Fig. 5). Two operation modes are distinguished:

– *Off-line classifier design (dotted arrows).* A data base of pre-classified clutter matrixes is created. The most suitable sub-set of features is selected according to the estimated performance of classifiers trained in a supervised manner using different sub-sets of features.
– *On-line classifier operation (continuous arrows).* The selected sub-set of features is extracted from sub-matrixes of the actual scan to generate classifier inputs and select the best detector from the available bank.

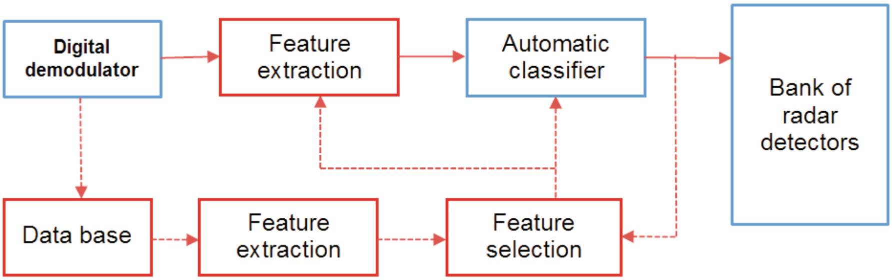

Fig. 5. Proposed detection scheme based on a bank of detectors and a clutter classifier capable of selecting the more suitable detector among those available

3.1 Training and Test Sets

A data base was generated synthetically for training and testing purposes, using the statistical models proposed in the literature and summarized in section 2. These models have been assessed using real data, because of that they can be used for generating synthetic data sets for design and testing purposes. As coherent models are considered, the generated samples are synthetic samples of the in phase and in quadrature components of the received radar echoes, which are provided by a synchronous detector in a real system. In Fig. 1, the digital demodulator encompasses the matched filter and the synchronous detector.

The data base was composed of clutter matrixes obtained at the output of the digital demodulator (Fig. 1). Assuming a radar system similar to SCANTER 2001 [12], the size of the matrixes was selected to guarantee a good estimation of the desired features, controlling the associated computational cost. Finally, matrixes of M = 625x1000 complex elements were synthetically generated for covering an azimuth sector of 60 degrees and a range interval of $7.5km$. The data base was composed of four clutter classes:

- Correlated Gaussian: 3000 patterns with correlation coefficient uniformly distributed in $[0.9; 0.95]$.
- Uncorrelated Gaussian: 3000 patterns.
- K-distributed: 3000 patterns with a shape parameter, ν, uniformly distributed in $[0.4; 4.1]$.
- Weibull: 3000 patterns with shape parameter uniformly distributed in $[0.4; 2.1]$.

In all matrixes, the effects associated to the propagation of the electromagnetic wave and the size variation of the ground resolution cell of the radar were compensated. After that, all patterns were normalized to have unit power. Training, validation and testing sets were built from the database using $1,200$, 300, and $1,500$ patterns of each class, respectively.

3.2 Classifier Design

MLP *and K-NN based* classifiers have been analyzed. A study of clutter statistical features has been carried out, to identify the most suitable set to be used as classifier inputs. This is the common strategy for controlling the computational cost and increasing the generalization capabilities of classifiers. The features defined in Table 1 have been considered, together with the correlation coefficient, ρ. Symmetry and kurtosis are complex, and each one gives rise to 2 real values; the entropy is estimated from the modulus and the real and imaginary parts, generating 3 real values. A total of 9 features have been studied.

Table 1. Analyzed statistical features

Features	Definition	Features	Definition
variance	$\sigma^2(x) = \sum_M \frac{(x-\mu(x))^2}{M}$	Entropy	$E(x) = -\sum_m p(x) \cdot log_2(p(x))$
symmetry	$S(x) = \sum_M \frac{(x-\mu(x))^3}{\sigma^3}$	kurtosis	$K(x) = \sum_M \frac{(x-\mu(x))^4}{\sigma^4}$

Symmetry and kurtosis are several orders of magnitude higher than the other features. As the classification problem conveys the calculus of distances, these two features can mask the information provided by others. Because of that, feature vectors have been normalized for having zero mean and unity variance.

4 Results

MLPs with different number of inputs (ranging from 1 to 9), one hidden layer and four inputs have been trained considering the 512 possible feature combinations. The Levenberg-Marquardt algorithm, [13], has been used in combination with a cross-validation technique, and MLPs have been initialized using the Nguyen-Widrow method [14]. For each feature combination, the training process has been repeated twenty times. Only the cases where the performances of all the trained networks were similar in average, have been considered to extract conclusions.

In Tables 2 and 3, a summary of the results is presented:

- The best classification performance has been obtained using $[\rho; \sigma^2; E(|\tilde{\mathbf{z}}|)]$.
- Table 2 shows the improvement associated to the variance.
- Although the input clutter matrixes have been normalized to have unity variance, the statistical properties of the variance estimator are different (Fig. 6).

Table 2. Miss-classification matrix for the most relevant features (%). Input patterns (rows), classifier outputs (columns).

	$[\rho;\sigma^2;E(\|\tilde{\mathbf{z}}\|)]$, L=6, NO=97				$[\rho;E(\|\tilde{\mathbf{z}}\|)]$, L=5, NO=72			
	G_{nc}	G_c	Weibull	K	G_{nc}	G_c	Weibull	K
G_{nc}	**100**	0	6.21	0	**100**	0	5.23	0
G_c	0	**100**	0	0	0	**100**	0	0
Weibull	0	0	**85.26**	14.98	0	0	**63.72**	10.25
k	0	0	8.53	**85.02**	0	0	31.05	**89.75**

Table 3. Miss-classification matrix for other features (%). Input patterns (rows), classifier outputs (columns).

	$[\sigma^2;E(\|\tilde{\mathbf{z}}\|)]$, L=5, NO=72				$[\rho;\sigma^2;S;K]$, L=9, NO=196			
	G_{nc}	G_c	Weibull	K	G_{nc}	G_c	Weibull	K
G_{nc}	**78.64**	43.64	1.43	0	**84.83**	0	57.69	11.29
G_c	21.36	**56.01**	3.58	0	0	**100**	0	0
Weibull	0	0.27	**87.50**	14.88	15.02	0	**23.64**	11.97
k	0	0.07	7.48	**85.12**	0.15	0	18.67	**76.74**

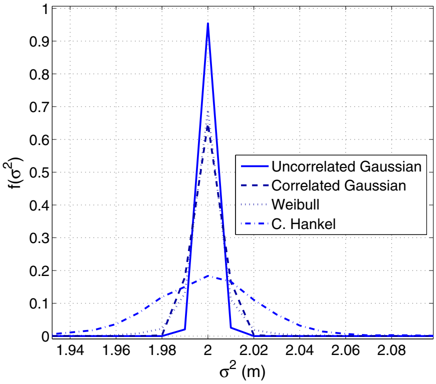

Fig. 6. Estimated fdp of the variance estimates

– Using $[\rho;\sigma^2;E(|\tilde{\mathbf{z}}|)]$, the 6.21% of the Weibull patterns are classified as Gaussian. This can be explained considering that the Weibull distribution for a shape parameter equal to 2 is the same as the Rayleigh, and the entropy of the modulus of the complex clutter samples is used. An analysis of the test set has revealed that the shape parameter for Weibull patterns classified as Gaussian is very close to 2.

In the first row of each table, the input feature vector, the number of hidden neurons, L, and the number of required operations NO, are provided. If P is the number of real inputs, NO is calculated as $(NO = (2P + 1)L + (2L + 1)4 + 3)$.

Solutions based on the KNN algorithm (euclidean distance) have been designed. The results obtained for $K = 10$ and 25 are presented in Table 4. The performances for both K values are very similar, so $K = 10$ is selected because of its lower computational cost (NO). For $N = 1,500$ training patterns for each class, $NO = 4N(3P+K+1)-0.5K(K-1)+3$. The performance of this solution is very similar to that provided by the MLP for the same input features, but the associated computational cost is significantly higher.

Table 4. Miss-classification matrix (%) for the K-NN classifier. Input patterns (rows), classifier outputs (columns)

	$[\rho;\sigma^2;E(\lvert\tilde{\mathbf{z}}\rvert)]$, k=10, NO=107,958				$[\rho;\sigma^2;E(\lvert\tilde{\mathbf{z}}\rvert)]$, k=25, NO=197,703			
	G_{nc}	G_c	Weibull	K	G_{nc}	G_c	Weibull	K
G_{nc}	**100**	0	3.27	0	**100**	0	5.33	0
G_c	0	**100**	0	0	0	**100**	0	0
Weibull	0	0	**86.80**	14.53	0	0	**87.13**	15.27
k	0	0	9.93	**85.47**	0	0	7.54	**84.73**

5 Conclusions

A neural network based pre-classification stage has been designed in order to improve the performance of radar detectors in sea clutter environments. The design of these detectors is really complex. If actual clutter statistics are different from those assumed in the design, performance losses can be very high. In order to design an adaptive detector capable of exploiting all the capabilities of solutions proposed in the bibliography, an automatic sea clutter classifier has been designed. The most extended sea clutter models have been used (Gaussian, Weibull and K-distributed), and the parameters of a commercial radar have been assumed for generating a synthetic data base from which training, validation and testing sets have been generated.

In order to reduce the complexity of the classifier and increase its generalization capabilities, a feature extraction stage has been designed for providing the classifier inputs. Variance, symmetry, entropy, kurtosis and correlation coefficient have been combined to determine the most suitable input vector. Results show that the best combination among those considered is the variance, the entropy of the modulus of the samples and the correlation coefficient. An MLP with 3 inputs, 6 hidden neurons and 4 outputs have been proposed. This classifier is able to provide a performance similar to the $K-NN$ algorithm with $K = 10$ with a significant reduction in computational cost, a very important feature in real time applications.

Acknowledgments. This work has been supported by the Spanish Ministerio de Economía y Competitividad, under Project TEC2012-38701

References

1. Neyman, Pearson: On the problem of the most efficient test of statistical hypotheses. Phil. Trans. Roy. Soc. A 231(9), 492–510 (1933)
2. Aloisio, V., di Vito, A., Galati, G.: Optimum detection of moderately fluctuating radar targets. IEE Proc. Radar, Sonar and Navigation 141(3), 164–170 (1994)
3. di Vito, A., Naldi, M.: Robustness of the likelihood ratio detector for moderately fluctuating radar targets. IEE Proceedings on Radar, Sonar and Navigation 146(2), 107–112 (1999)
4. Nayebi, M.M., Aref, M.R., Bastani, M.H.: Detection of coherent radar signals with unknown Doppler shift. IEE Proc. Radar, Sonar and Navigation. 143(2), 73–86 (1996)
5. Mata-Moya, D., Jarabo-Amores, P., Rosa-Zurera, M., Nieto-Borge, J.C., Lopez-Ferreras, F.: Combining MLPs and RBFNNs to Detect Signals With Unknown Parameters. IEEE Transactions on Instrumentation and Measurement 58(9), 2989–2995 (2009)
6. de la Mata-Moya, D., Jarabo-Amores, P., Vicen-Bueno, R., Rosa-Zurera, M., López-Ferreras, F.: Neural Network Detectors for Composite Hypothesis Test. In: Corchado, E., Yin, H., Botti, V., Fyfe, C. (eds.) IDEAL 2006. LNCS, vol. 4224, pp. 298–305. Springer, Heidelberg (2006)
7. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford Unviersity Press (1995)
8. Skolnik, M.: Radar Handbook, 3rd edn. Mc-Graw Hill (2008)
9. Conte, E., De Maio, A., Galdi, C.: Statistical analysis of real clutter at different range resolutions. IEEE Trans. Aerospace and Electronic Systems 40(3), 903–918 (2004)
10. Farina, A., et al.: High resolution sea clutter data: statistical analysis of recorded live data. IEE Proceedings Radar, Sonar and Navigation 144(3), 121–130 (1997)
11. Watts, S., Baker, C.J., Ward, K.D.: Maritime surveillance radar. Part 2: Detection performance prediction in sea clutter. IEE Proceedings Radar and Signal Processing 137(2), 63–72 (1990)
12. SCANTER 2001 Naval 2D Surface Surveillance, `http://www.terma.com`
13. Hagan, M.T., Menhaj, M.B.: Trainning feedfordward networks with the Marquardt alforithm. IEEE Trans. Neural Networks 6, 989–993 (1994)
14. Nguyen, D., Widrow, B.: Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights. In: Proc. of the Int. Conf. on Nerual Networks, IJCNN, pp. 21–26 (2009)