Automatic Object Classification with Active Sonar using Unsupervised Anomaly Detection

Pietro Stinco*
NATO STO CMRE
Centre for Maritime Research and Experimentation
La Spezia, Italy

Giovanni De Magistris
NATO STO CMRE
Centre for Maritime Research and Experimentation
La Spezia, Italy

Alessandra Tesei
NATO STO CMRE
Centre for Maritime Research and Experimentation
La Spezia, Italy

Kevin D. LePage
NATO STO CMRE
Centre for Maritime Research and Experimentation
La Spezia, Italy

***Abstract*—This work describes an unsupervised anomaly detection method for automatic contacts classification of an active sonar system. The proposed method refers to littoral, shallow water environments where there is a significant amount of clutter contacts from the seafloor and coastal reverberation. This huge amount of undesired contacts can be exploited to learn the "finger-print" of the clutter and then to identify the object related contacts as anomalies. The paper describes the proposed classification method and shows its performance with real data collected at sea using an echo-repeater as an artificial object.**

***Index Terms*—Unsupervised Learning, Anomaly Detection, Active Sonar, Sonar Contacts Classification.**

I. Introduction

In recent years, the need of systems for long-range detection of underwater objects has strongly increased. When dealing with very quiet objects, this operation is traditionally performed using Low-Frequency Active Sonar (LFAS) systems.

An LFAS consists of a powerful wideband source and a receiving hydrophone array. Both of them are towed systems, such that they are variable in depth and can be deployed in the most favourable acoustic layer.

Most of the work conducted by Centre for Maritime Research and Experimentation (CMRE) refers to littoral, shallow water environments where the contribution of false alarm echoes from the seafloor and coastal reverberation may be very significant, causing a deterioration of the sonar performance. In this case, real-time classification algorithms to discriminate object contacts from clutter contacts become crucial.

The task of classification is to select object related contacts (if present) from the large number of clutter contacts.

Traditionally, this task has often been left to a sonar operator who uses his expertise to classify contacts. Automatic classification can be helpful to reduce workload on sonar operators and it is a necessary feature when the receiver is towed by Autonomous Underwater Vehicles (AUVs), or in general for robots operating as receivers without human supervision.

Recent results at CMRE demonstrated how it is possible to get good classification performance using Convolutional Neural Networks (CNN) [1].

*pietro.stinco@cmre.nato.int

This work has been supported by the NATO Allied Command Transformation under the Cooperative Anti-Submarine Warfare research programme.

The challenges on the use of supervised learning techniques for Automatic Object Classification (AOC) for underwater surveillance are mainly these: data collection is costly, data labelling is time-consuming, it is difficult to generate accurate datasets and the datasets are often unbalanced.

As a matter of fact, the navigation accuracy of underwater vehicles is not very accurate and often the expected position of an object contact is far away from the effective one. Moreover, the object of interest is often sailing very close to cluttered regions (especially in shallow water environments) and this significantly increases the probability to erroneously label a clutter contact as an object contact (and vice-versa).

The training sets usually consist of a big amount of clutter data and very few object data. Moreover, being the object response very sensitive to the aspect angle and very difficult to collect sufficient object data for any possible aspect, it is hard to train a general model for AOC.

This paper introduces an unsupervised learning method based on anomaly detection for object classification. Anomaly detection is important in many applications, such as mine detection and classification [2], target detection in hyperspectral images [3] or mammographic image analysis [4]. There are many approaches to anomaly detection based on statistical models, machine learning, saliency based-methods, sparse representations, and more. The method presented here is based on machine learning since the normal model is learned by training the algorithm with a dataset of clutter samples.

Considering the challenges described above for supervised learning techniques applied to AOC, the proposed method exploits the clutter contacts to learn the clutter signature and then classifies the object contacts as anomalies if their signature is not similar to the learned one. The signature is a vector that collects several features obtained by processing the acoustic response of the sonar contact.

The algorithm is resumed in Figure 1 that shows the two phases of the anomaly detection approach for object classification. In the first phase, the algorithm learns the clutter signature by processing the features extracted by the acoustic signal of a training set with only clutter contacts. In the second phase, the features of each sonar contact are compared with the learned clutter signature and then an object is detected if its features are anomalous, i.e. not similar to the learned one.

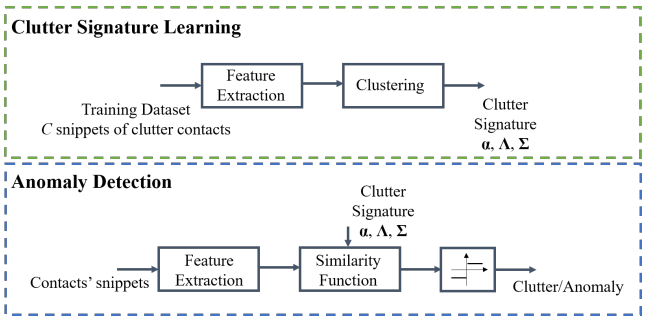

Fig. 1. Block diagram of the classification algorithm. *Clutter signature learning*: learn the clutter signature from a training set of clutter contacts. *Anomaly Detection*: contact classification by comparing the signature of each contact with the learned clutter signature.

Similarly to the approach proposed in [5], where the active sonar receiver increases its knowledge of the surrounding environment generating clutter maps based on the persistency of clutter contacts over a geographical area, in the proposed approach the receiver learns the clutter signature during the survey of the operational area. It is supposed that during this phase there is not any object; however, how will be discussed in next sections, when learning the clutter signature, the proposed method filters out the outliers that can affect the clutter signature estimate.

During a sea trial, it is possible to periodically estimate the clutter signature. This is very important when operating in shallow water scenario where the underwater environment rapidly changes and it is fundamental to adapt the receiver in order to reduce the false alarm rate.

The paper will show results with experimental data collected at sea during LCAS16 (Littoral Continuous Active Sonar 2016) sea-trial conducted in Italy in the Gulf of Taranto. The analysed data have been collected using an Echo-Repeater (ER) as an artificial object and the SLIm Cardioid Towed Array (SLICTA) as the sonar receiver. The SLICTA is a triplet array with port-starboard discrimination capability designed and developed at CMRE.

II. Clutter Signature Learning

Figure 2 shows the Beam Collapse Plot (BCP) of the data collected the 22 October 2016 from 09:10 to 10:50 UTC. In the analysed dataset, the receiving SLICTA triplet array and the acoustic source were towed by NRV Alliance while the ER was towed by CRV Leonardo. The data consists of 300 pings of 20 seconds collected when transmitting an LFM (Linear Frequency Modulated) pulse with a bandwidth of 800 Hz.

The BCP shows the output of the Adaptive Beamformer [6] for all the pings processed during the run. For any ping, the output of the beamformer is a time-bearing map where the time is the ping duration (fast time) and the bearing is the steering direction. The BPC collects, for any ping (slow time) and any fast time point, the maximum value of the beamformer output along the bearing direction. Figure 2 also shows in green the ground truth, i.e. the expected round trip delay between the object and the receiving sonar. From the BCP it is evident how the ER signal is embedded in a high

Algorithm 1 Clutter Signature Learning

Input: features of clutter contacts $\mathbf{F} = [\mathbf{f}_1 ... \mathbf{f}_C]$
Output: clutter signature α, $\mathbf{\Lambda}$, $\mathbf{\Sigma}$
randomly initialize N centroids θ_i from $\mathbf{F}$
randomly initialize $\mathbf{c}$ (c_i - index to which $\mathbf{f}_i$ is assigned)
initialize $\mathbf{d}$ = 0
Clustering
while ($\mathbf{d} \neq \mathbf{c}$) **do**
  $\mathbf{d} = \mathbf{c}$
  for ($i = 1 : C$) **do**
    $c_i \leftarrow$ closest to $\mathbf{f}_i$ centroid using dist($\mathbf{f}_i$,θ_i)
  end for
  for ($i = 1 : N$) **do**
    $\theta_i \leftarrow$ average of all points assigned to c_i
  end for
end while
Compression
k = 0
N_α = 0
for ($i = 1 : N$) **do**
  $n_i \leftarrow$ number of points with label i
  if ($n_i > n_{min}$) **then**
    $k \leftarrow k + 1$
    $\mathbf{X} \leftarrow$ set of points from $\mathbf{F}$ with label i
    $\mathbf{X} \leftarrow$ select 80% of closest points $\mathbf{x}$ to θ_i
    $\alpha_k \leftarrow$ number of points in $\mathbf{X}$
    $N_\alpha \leftarrow N_\alpha + \alpha_k$
    $\lambda_k \leftarrow$ centroid of $\mathbf{X}$
    $\sigma_k \leftarrow$ average value of $|\mathbf{x} - \lambda_k|$
  end if
end for
$\alpha \leftarrow \alpha$ / N_α
return α, $\mathbf{\Lambda}$, $\mathbf{\Sigma}$

cluttered environment. Very strong echoes are coming from the seafloor, such as from underwater canyons and rocks. The received signal is also affected by interfering ship traffic noise, i.e. continuous signals clearly visible in the BCP as vertical stripes. All these non-object signals are exploited to estimate the clutter signature.

Let consider a training dataset consisting of C clutter contacts. This dataset can be obtained in post-processing by randomly selecting C contacts at a sufficient distance from the ground truth. In real operations, the training contacts can be collected during an environmental characterization of the operating area, under the assumption that all the collected contacts are from clutter. The proposed algorithm works even if this assumption is violated, i.e. there are some unexpected object related contacts. As a matter of fact, the proposed method takes into account the presence of possible outliers to avoid that sporadic object contacts affect the clutter signature estimate. Let indicate with $s_c[n]$ the snippet on the contact c. The snippet is a finite time sequence (or a transformation of it) that collects the received acoustic samples of the contact, i.e. the output of the beamformer at the time-bearing coordinates of the detection. Let also indicate with $\mathbf{f}(s_c[n])$, in short $\mathbf{f}_c$, a vector that collects M real valued features on the contact c,

$$\mathbf{f}_c = \mathbf{f}\left(s_c\left[n\right]\right) \in \mathbb{R}^M \tag{1}$$

The features are obtained by processing the snippet $s_c[n]$, as an example by tacking its normalized moments, the number of peaks, etc.. Typically M is of the order of tens (30-40 or

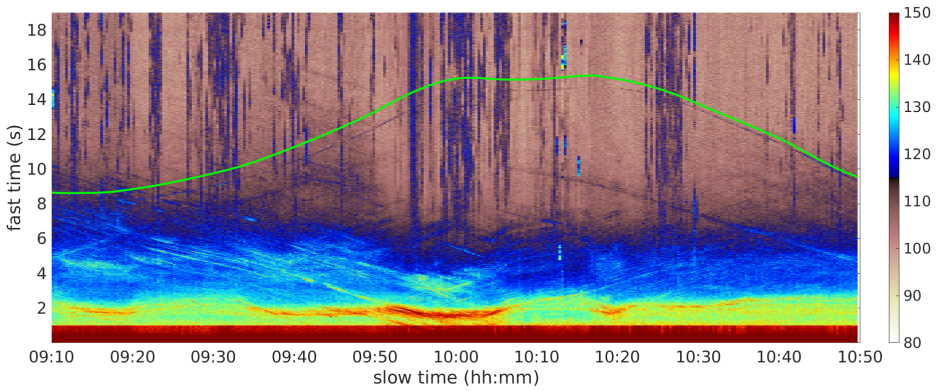

Fig. 2. Beam collapse plot at the output of the adaptive beamformer. Green: ground truth delayed of 0.5 s.

even more) and the features are scaled in order to have values of the same order of magnitude, i.e. with values from -1 to 1.

This paper focuses on how to process the features to estimate the clutter signature and then how to exploit this estimate for contacts classification. A good way to choose these features is to exploit those that might take on unusually large or small values in the event of an anomaly.

The pseudo-code in Table 1 describes the proposed algorithm. The input of the algorithm is matrix $\mathbf{F}$ that collects the feature vectors of the C clutter contacts of the training set. As described in the table, the algorithm is divided into two phases: a first *clustering* phase for an initial fit of the features model and a subsequent *compression* phase to remove clusters with few samples and to filter out possible outliers that can affect the clutter signature estimate.

Clustering is performed using the *K-means* method. The algorithm is initialised by randomly selecting N centroids from the input feature points, N is of the same order of magnitude of M. The clustering algorithm iteratively assign a label c_i to any feature point, considering the nearest centroid and using the Euclidean distance. After that a label has been assigned to any point, the values of the cluster centroids are updated with the average of all points with the same label. The algorithm is stopped when the labels no longer change.

The second step of the learning algorithm is the compression phase. In this phase, the clusters with few points are discarded. The minimum number of points is a small fraction (1-5%) of C/N. Moreover, for any of the remaining $K \leq N$ clusters, all the points far away from their centroid are discarded. For any cluster, only the 80% of points closest to the centroid are processed to estimate the clutter signature.

The clutter signature, is described by the three quantities

$$\begin{aligned} &\alpha \in \mathbb{R}^K \\ &\mathbf{\Lambda} = [\lambda_1 ... \lambda_k ... \lambda_K] \in \mathbb{R}^{M \times K} \\ &\mathbf{\Sigma} = [\sigma_1 ... \sigma_k ... \sigma_K] \in \mathbb{R}^{M \times K} \end{aligned} \tag{2}$$

where α is a vector collecting the normalized number of points (normalised with the total number of processed points) of each cluster, Λ is a $M \times K$ matrix that collects in each column the centroids of the clusters (after outliers removal), Σ is a $M \times K$ matrix whose k-th column collects the element by element mean distance of the cluster points from the k-th centroid. Next section will show how to process the learned clutter signature for anomaly detection.

III. Anomaly Detection

Once the clutter signature has been learned, the receiver evaluates, for each contact c, the features vector $\mathbf{f}_c$ as in (1) and then computes the similarity function ρ_c using

$$\rho_c = \sum_{k=1}^{K} \Theta_k(\mathbf{f}_c) \tag{3}$$

where $\Theta_k(\mathbf{f}_c)$ is the Gaussian kernel of cluster k on contact c, defined as

$$\Theta_k(\mathbf{f}_c) = \alpha_k \prod_{m=1}^{M} e^{-\frac{(\mathbf{f}_c(m) - \lambda_k(m))^2}{2\sigma_k^2(m)}}. \tag{4}$$

In words, for each of the K centroids that describe the clutter signature, the kernel function evaluates if the features vector is close to the centroid. A high value of the kernel indicates an high similarity between the features vector and the clutter cluster. The similarity value ρ_c is given by the sum of the K kernels, weighted with α_k. Meaning that the high populated clutter clusters have an high contribution in the similarity function. Note that the value of ρ_c is positive and lower than 1, $\rho_c \in [0, 1]$.

Note also from (4) that the feature components are treated as independent, the kernel is given by the product of the similarity between each component of the feature and the coordinates of the centroid. It is also possible to use multivariate Gaussian kernels to evaluate the similarity function by inferring the correlation between the clutter features during the learning phase. The use of multivariate Gaussian kernels automatically captures and exploits correlations between features but is computationally more expensive and does not scale with large number of features. This is because, for large number of features M it is possible that some of them are linearly dependent. In this case the resulting covariance matrix is rank deficient and its inversion generates numerical problem. On the other hand, the use of independent kernels is computationally

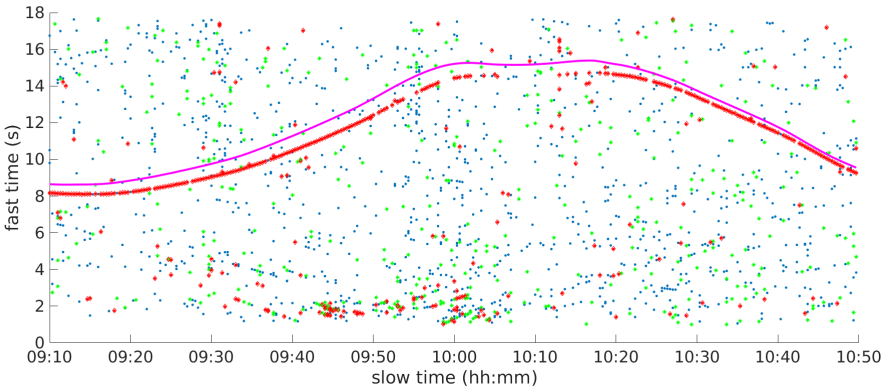

Fig. 3. Contacts at the output of the detection and clustering algorithm. Magenta: ground truth delayed of 0.5 s. Blue: all sonar contacts. Green: randomly selected clutter contacts exploited for clutter signature learning. Red: anomalous contacts, i.e. contacts classified as object.

cheaper, scales better with large features and provides the freedom to design very large vectors of features without taking care of possible dependencies between them.

The last step is contact classification, a contact is labelled as object (l_c = 1) if the similarity function is lower than a threshold ϵ, otherwise it is labelled as clutter (l_c = 0), i.e.

$$l_c = \begin{cases} 1 & \text{if } \rho_c < \epsilon \\ 0 & \text{if } \rho_c \geq \epsilon \end{cases} \tag{5}$$

The threshold ϵ decreases with increasing number of features, typical values are close to 10^{-M}.

IV. Results

Figure 3 shows with blue dots all the sonar contacts at the output of the detection and clustering algorithm compared with the object ground truth (magenta line). These contacts have been detected using the receiver described in [6]. The dataset analysed in this paper consists of 2098 contacts with a rate of almost 7 contacts per ping.

The green dots in figure are the clutter contacts used in the training phase to learn the clutter signature while the red dots are the anomalous contacts, i.e. those classified as object.

In this example, the green dots are 800 randomly selected clutter contacts representing different kinds of sonar clutter, since they are generated from compact clutter, diffuse reverberation and interfering ship noise. The anomalies have been obtained using the threshold $\epsilon = 10^{-40}$. The resulting percentage of true positives (object contacts classified as object) is close to 80% while the percentage of false positives (clutter contacts classified as object) is close to 10%.

The Receiver Operating Characteristic (ROC) of the proposed anomaly detection algorithm is shown in Figure 4 where the true positive rate is plotted as a function of the false positive rate. This plot has been obtained by Monte Carlo runs, averaging the performance obtained with different training sets of 800 clutters contacts and different values of the threshold.

The ROC of the proposed algorithm is compared with the one of the CNN described in [1]. As expected, the performance of the unsupervised classification algorithm are lower than those of the supervised algorithm. As discussed, the CNN

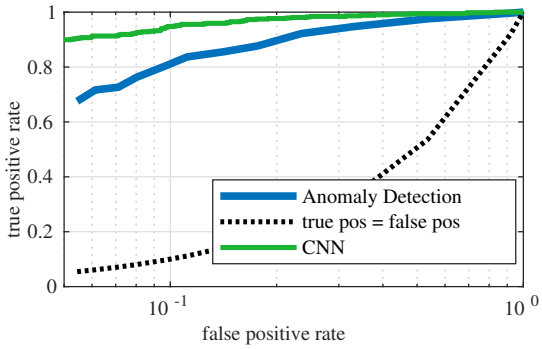

Fig. 4. ROC of the proposed algorithm (blue) and of the CNN in [1] (green).

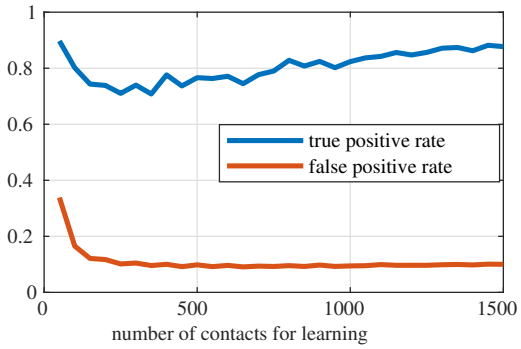

Fig. 5. True and false positive rates as a function of the training set cardinality.

is an excellent classification method but needs to be trained with very accurate datasets with related difficulty and cost in collecting and labelling the data. In particular, the CNN exploited to get the results in Fig. 4 has been trained with days of data collected in two sea trials using an ER as an artificial object [1]. On the other hand, the performance of the unsupervised anomaly detection algorithm are slightly lower than those of the CNN but the algorithm is trained online using only few hundreds of clutter samples that can be collected in less than one hour. For the analysed dataset, the contacts rate was of almost 20 contacts per minute. Then, to collect a training set of 800 clutter samples, the required survey duration is of almost 40 minutes.

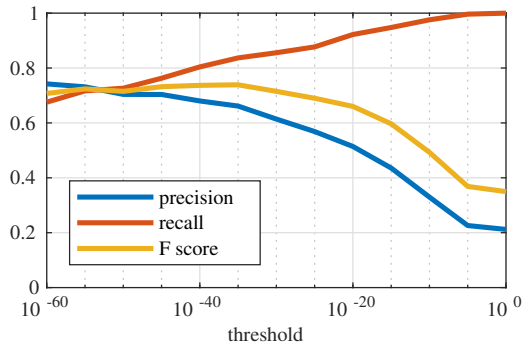

Fig. 6. Precision, recall and F score as a function of the threshold ϵ.

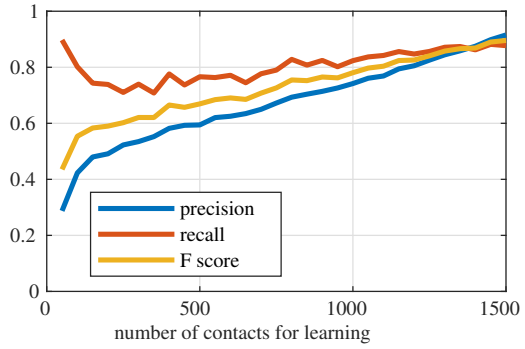

Fig. 7. Precision, recall and F score as a function of the training set cardinality.

Figure 5 shows the true positive and false positive rates as a function of the number of clutter samples used for training (training set cardinality). From this plot it is evident how, for this dataset, it is possible to get good performance even with few hundreds of training clutter samples. The higher is the number of training samples the higher is the true positive rate.

Figures 6-7 show the performance in terms of precision, recall and F score. Precision indicates how many positive contacts are relevant and is defined as the rate of true positive among the contacts classified as positive. On the other hand, recall indicates how many relevant contacts are selected and hence is the same of the true positive rate. The F score is the harmonic mean of precision and recall.

Figure 6 shows the performance as a function of the threshold ϵ. From this figure it is apparent that for increasing values of the threshold the recall is increasing while the precision is decreasing. This is obvious considering (5) and that a contact is classified as positive if its similarity is lower than the threshold. When the threshold tends to one all the contacts are classified as anomalies and hence the recall tends to one but the number of false alarms is maximum and the resulting precision tends to be very low. From the plot, precision and recall are similar when the threshold is close to 10^{-40} where 40 is the number of feature M of our classifier.

Figure 7 shows the performance as a function of the training set cardinality. As for the ROC in Figure 5, the higher is the number of contacts used to learn the clutter signature, the better are the classification performance. Clearly, the larger is the training set cardinality, the longer is the time required to collect the training set in a real operating environment.

The time required to learn on-line the clutter signature is less than one hour and hence, in real operations, it is possible to update/refine the clutter signature estimate several times during a day mission. This is very important when operating in shallow water environments where the clutter signature may rapidly change as a function of space and time.

V. Conclusions

This paper presented an unsupervised classification method for AOC based on an anomaly detection approach. The method is based on machine learning since the normal model is learned by training the algorithm only with clutter data. The contacts classified as object are those with anomalous features, i.e. not similar to the the learned normal model.

The main advantage of this algorithm with respect to conventional supervised learning technique is that there is no need to train the algorithm with object related contacts that, for underwater surveillance application, are very difficult and costly to be collected. On the other hand, especially in the challenging littoral, shallow water environment, there is an huge amount of clutter data and then the proposed method can be easily trained during a survey of the operational area.

The successful performance of the algorithm was demonstrated with real data collected at sea using an echo-repeater as an artificial target. The results show the capability of the proposed algorithm to cope with a variety of clutter contacts.

Acknowledgment

The collection of LCAS16 sea trial data used in this paper was made possible by the LCAS Multi-National Joint Research Project (MN-JRP), including as Participants the NATO Centre for Maritime Research and Experimentation, the Defence Science and Technology Organisation (AUS), the Department of National Defence of Canada Defence Research and Development Canada (CAN), the Defence Science and Technology Laboratory (GBR), Centro di Supporto e Sperimentazione Navale - Italian Navy (ITA), the Norwegian Defence Research Establishment (NOR), the Defence Technology Agency (NZL), and the Office of Naval Research (USA).

References

[1] G. De Magistris, P. Stinco, J.R. Bates, J.M. Topple, G. Canepa, G. Ferri, A. Tesei, and K. D. LePage, "Automatic Object Classification for Low-Frequency Active Sonar using Convolutional Neural Networks," *OCEANS 2019 MTS/IEEE SEATTLE*.

[2] G. Mishne and I. Cohen "Multiscale Anomaly Detection Using Diffusion Maps," *IEEE Journal of Selected Topics in Signal Processing*.

[3] Y. Chen, N. M. Nasrabadi and T. D. Tran "Sparse Representation for Target Detection in Hyperspectral Imagery," *IEEE Journal of Selected Topics in Signal Processing*.

[4] C. Spence, L. Parra, and P. Sajda "Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model," *Proc. IEEE Workshop Math. Meth. Biomed. ImageAnal. (MMBIA01), 2001, pp. 310.*

[5] M. Micheli, A. Tesei, G. Ferri, and P. Stinco, "Adaptive filter of seabed clutter onboard the AUVs of an active multistatic sonar network," *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans, OCEANS - Kobe 2018.*

[6] P. Stinco, A. Tesei, A. Maguer, A. Ferraioli, V. Latini, and L. Pesa, "Subbands beam-space adaptive beamformer for port-starboard rejection in triplet sonar arrays," *2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018 - Proceedings.*