# Adaptive Automatic Target Recognition with SVM Boosting for Outlier Detection

Kieron Messer[1], Josef Kittler[1], John Haddon[2], Graham Watson[2], and Sharon Watson[2]

[1] University of Surrey, Guildford, Surrey, UK. GU2 5XH.
K.Messer@surrey.ac.uk
Phone: (01483) 876049 ... Fax: (01483) 876031
[2] Defence Evaluation and Research Agency, Farnborough, Hampshire, UK.

**Abstract.** This paper is concerned with the detection of dim targets in cluttered image sequences. It is an extension of our previous work [7] in which we viewed target detection as an outlier detection problem. In that work the background was modelled by a uni-modal Gaussian. In this paper a Gaussian mixture-model is used to describe the background in which the the number of components is automatically selected. As an outlier does not automatically imply a target, a final stage has been added in which all points below a set density function value are passed to a support vector classifier to be identified as a target or background. This system is compared favourably to a baseline technique [12].

**Keywords:** Automatic target recognition, Mixture Modelling, Support Vector Machines, Outlier Detection.

## 1 Introduction

Automatic Target Recognition (ATR) is concerned with the detection, tracking and recognition of small targets using input data obtained from a multitude of sensor types such as forward looking infrared (FLIR), synthetic aperture radar (SAR) and laser radar (LADAR). Applications of ATR are numerous and include the assessment of battlefield situations, monitoring of possible targets over land, sea and air and the re-evaluation of target position during unmanned missiles weapon firing.

An ideal system will exhibit the properties of a low false positive rate (detection of a non-target as a target), whilst obtaining a high true positive rate (the detection of a true target). This performance should be invariant to the following parameters: sensor noise; time of day; weather types; target size/aspect and background scenery. It should be flexible such that it has the ability to detect previously unseen targets and be able to retrain itself if necessary. It is unlikely that one single system will cope well with all these possible scenarios [2]. The many challenges produced by ATR have been previously well documented in [3], [9] and [1].

In this paper an adaptive ATR system is proposed which is suitable for scenes with strong clutter which is spatially and temporally highly structured, such as sea glint and atmospheric scintillation. In the bootstrap phase a statistical Gaussian mixture-model of the background is built by using a set of texture filters. In operation, the same features are computed for each new pixel arriving at the sensor input. If the probability density

F.J. Ferri et al. (Eds.): SSPR&SPR 2000, LNCS 1876, pp. 104–113, 2000.
© Springer-Verlag Berlin Heidelberg 2000

value of the of this pixel feature vector falls below a set threshold it is considered as a potential target. A low probability density value does not necessarily imply a target, e.g. it could be sea glint. For this reason a final stage has been added to the system in which a support vector machine is used to classify all the outliers as a target or as clutter. This is consistent with realistic operational scenarios as the target objects required for training can easily be inserted in images synthetically.

Another novelty of this work is the technique applied to obtain a suitable set of filters which ensures that the background/target separation is maximised during training. In our previous work [7] we demonstrated that the use of a set of adaptive texture filters to model each background outperformed the more traditional Wavelet-based feature extractor. This set of filters was designed using Principal Component Analysis on randomly sampled image patches taken from a training image. This ensured that these filters had a mean response when presented with a similar looking texture. If an object with different texture, such as a target, is presented to the filter the resulting response should be non-mean, making its detection as an outlier easier. In this paper the filter design methodology is enhanced further to take into account the temporal dimension of the image data, i.e. the PCA is used to build 3-dimensional texture filters. Combining image data from different frames prior to detection is commonly known as the *track before detect* approach, TBD.

This method is compared to another TBD technique [12] in which targets are distinguished from the clutter by the analysis of the joint statistics of simple events such as glint flashes and regions of persistent brightness.

The rest of this paper is organised as follows: in the next section the DERA ATR system is briefly reviewed before our target detection algorithm is detailed in full. In section 4 experiments on two image sequences are performed. Finally, some conclusions are drawn.

## 2 Multivariate Conditional Probability

In [12] a target recognition approach was proposed in which the multivariate statistics of space-time structure was used to characterise spatially and temporally highly structured clutter, such as sea-glint and atmospheric scintillation. Targets were then recognised as unusual events.

Two three-dimensional filters were manually chosen and consisted of a constant-intensity blob filter and a filter tuned to sea glint flashes. A third feature was also used which was simply the vertical image co-ordinate. Dim targets were distinguished from the clutter by using the joint statistics of these three variables. A low joint-probability identified a possible target.

## 3 Adaptive Texture Representation

We also view the target detection as an outlier detection problem. That is, anything that does not normally occur in the background is viewed as a potential target. Our target detection algorithm has three basic steps:

**Model Generation** The background is described using a Gaussian mixture model.

**Model Optimisation** The model and model size are optimised using training data (if available).

**Target Detection** Outliers are found by deciding, per pixel, whether it is consistent with the model.

The background is represented by computing a feature vector, $\mathbf{f} = [y_0, y_1, \ldots, y_n]$, for every pixel in the training image. Each $y_k$ represents a measurement obtained by the $k^{th}$ filter. The distribution of these feature vectors is modelled by a mixture of Gaussians. Such a mixture model is defined by equation 1.

$$p(\mathbf{x}) = \sum_{j=1}^{M} p(\mathbf{x}|j)P(j) \tag{1}$$

The coefficients $P(j)$ are called the mixing parameters and are chosen such that

$$\sum_{j=1}^{M} P(j) = 1 \quad \text{and} \quad 0 \leq P(j) \leq 1 \tag{2}$$

Also note that the component functions satisfy the axiomatic properties of probability density functions

$$\int p(\mathbf{x}|j)d\mathbf{x} = 1 \tag{3}$$

In this work we used the normal distribution with a diagonal covariance matrix for the individual component density functions

$$p(\mathbf{x}|j) = \frac{1}{(2\pi\sigma_j^2)^{\frac{d}{2}}} \exp\left\{-\frac{\|\mathbf{x} - \mu_j\|^2}{2\sigma_j^2}\right\} \tag{4}$$

where $\mu_j$ is the mean of component $j$ and $\sigma_j$ is its standard deviation. The optimal values of the parameters $P(j)$, $\mu_j$ and $\sigma_j$ are estimated using the Expectation Maximisation algorithm, [4].

The EM algorithm requires, as an input, the number of components to be used for the data distribution modelling. This is achieved automatically using the model validation method proposed in [10]. This iterative algorithm systematically increases the model complexity until a model validation test is passed. This model selection strategy prevents both overfitting and underfitting.

To detect possible targets in test frames the same set of $n$ features is generated for every pixel in the image. Each feature vector, $\mathbf{f}_{test}$, is tested in turn to see whether it belongs to the same distribution as the background or is an outlier (i.e. possible target). This is done by computing the density function value for that pixel, based on the mixture model. If this value falls below a threshold, the pixel is considered an outlier and treated as a possible target. This threshold can be automatically determined from the training data.

There is also a problem of knowing which features to use to ensure the targets and background vectors are well separated in the feature space. For this reason a feature selection stage was added which selects features using the *sequential forward selection* algorithm [8].

### 3.1 Filter Design

The background regions of an image are described adaptively using Principal Component Analysis (PCA, also known as the Karhunen-Loeve transform). The representation adopted is an extension of an earlier method identified as the most promising in [7] in which we compared a PCA method against a standard Wavelet-based method and a method based on Independent Component Analysis. In our previous paper the filter design was two-dimensional. In this paper we incorporate the temporal dimension into the filter design.

Principal Component Analysis [5] finds a linear base to describe the dataset. It finds axes which retain the maximum amount of variance in the data. To construct a PCA base, firstly $N$ random rectangles of size $r \times c$ are taken from a set of training images. These rectangles are then packed into an $r \times c$-dimensional vector $\mathbf{x}_i$, usually in a row-by-row fashion. This results in a data set $\mathbf{X}$ containing $N$ samples. Assuming that the global mean of the vectors in $\mathbf{X}$ is zero, the principal components are the eigenvectors of the covariance matrix $\mathbf{XX^T}$. These are the columns of the matrix $\mathbf{E}$, satisfying

$$\mathbf{EDE}^{-1} = \mathbf{XX}^T \tag{5}$$

where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues corresponding to the eigenvectors in $\mathbf{E}$. The set of 2D filters is then generated by unpacking each row of $\mathbf{E}^T$ into a filter of size $r \times c$.

The design of 3D filters, used in all the following experiments, follows the same process as for the 2D design, however instead of extracting image rectangles from a single image, the rectangles are taken from $d$ consecutive images. This data is then unpacked to form a vector of $(r \times c \times d)$ dimensions. Typically, $d$ is set to 3.

## 4 Experiments

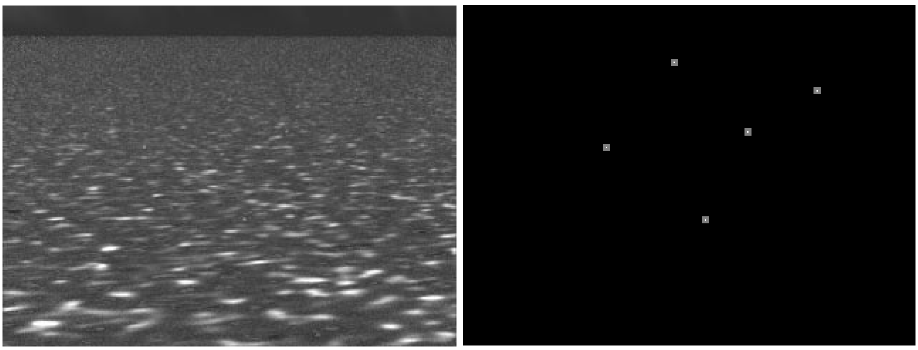
The proposed target detection technique has been applied to several sequences made available by DERA Farnborough and compared to the results obtained on the same sequence using the multivariate conditional probability (MCP) methods described in [12]. Typical results are shown in this section on a simulated sequence, SEASIM, and on a real sequence, AM.

### 4.1 On SEASIM

This sequence contains about twenty frames which have been artificially generated using a standard ray-tracing package. It represents the scenario of a sensor attached to a ship looking out over the ocean. Figure 1(a) shows the first frame of this sequence. Five targets have been inserted into this sequence; whose locations are given by the ground truth image of figure 1(b). These targets are very small (typically one pixel) and represent missiles moving towards the observer. The intensity of these targets are lower than the maximum intensity of the image and as the targets are moving slowly its pixel intensity will vary in time due to aliasing effects. A human observer will find it extremely difficult to identify all targets in this sequence. The two methods of target detection were then applied to this sequence.

| Method | Target Position |
|---|---|
| Reference | $[1, 2, 5, 7]$ |
| Proposed | $[1, 2, 3, 4, 5]$ |

**Table 1.** Probability ranking of real target

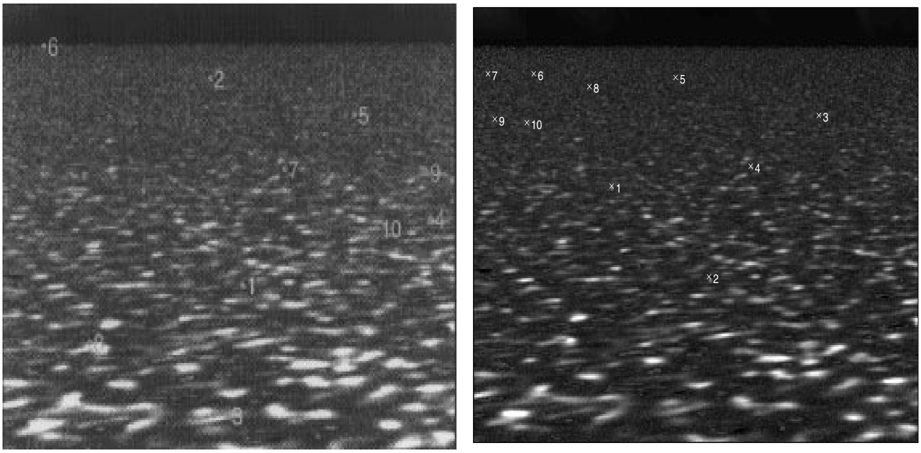


(a) First image

(b) Enhanced ground truth (the original size of each of the objects is 1 pixel).

**Fig. 1.** Sequence SEASIM.

The top ten most likely targets using multivariate conditional probability are shown in figure 2(a). The results obtained using the 3D-PCA and mixture modelling are shown in figure 2(b). The positions of the targets, for both methods, are shown in table 1. As one can see only four of the five targets have been recognised and two false positives have been identified in the top five for MCP. Using the proposed adaptive method of all five targets have been selected as the five most likely.




(a) Multivariate conditional probability (b) 3D-PCA and mixture modelling

**Fig. 2.** SEASIM: Top 10 detections using both methods

### 4.2 The AM sequence

A real infra-red sequence of just 8 frames was acquired. An artificial target was then placed in the sea area. The first frame of this sequence along with the ground truth image is shown in figure 3. Again manual identification of this target is extremely difficult.

The top ten most likely targets detected using the multivariate conditional probability method are shown in figure 4(a). The results obtained using the 3D-PCA and mixture modelling are shown in figure 4(b). For this sequence the MCP method has outperformed the proposed adaptive method. The single target was not found in the top 10 most likely candidates but is the most likely target identified by the MCP method. Our approach labels the real target as the 27th most likely.

## 5 Support Vector Machines

A target is identified if its density value is below a user set threshold. The system is very sensitive to this choice of threshold. If the threshold is set too low, targets are missed,

(a) First image (b) Enhanced ground truth

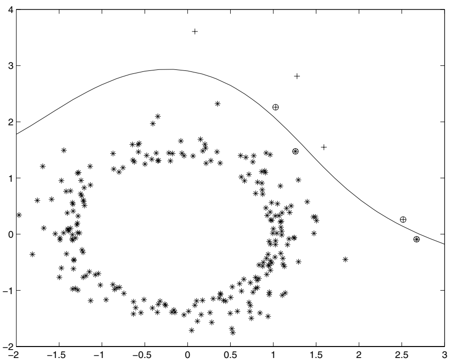
**Fig. 3.** Sequence AM.

(a) MCP (b) Proposed

**Fig. 4.** Top 10 detections on sequence `AM`

if it is set too high, many false positives are found. Also the targets may only occupy a corner of low density feature space, yet all points in the low density feature space are being identified as targets. This would explain the poor results for the AM sequence. The real target has been identified as an outlier but there are many other outliers also. To alleviate this problem a final stage has been added to the target recognition system which involves passing all the points below the threshold to a support vector classifier [11]. Most of these points will be false positives but still lie on the edge of the distribution. Using a classifier should eliminate some of these points.

Support vector machines have the major advantage that no density values are estimated. The classifier is designed on the principal of finding a boundary that optimally divides the two classes. The SVM boundary leaves the largest margin between the vectors of the classes. This makes SVM's highly insensitive to the curse of dimensionality and therefore do not require the large amounts of training data usually required to achieve a good general classification. Figure 5 demonstrates a typical decision surface generated for 2-dimensional training data.




(a) The best two selected features

**Fig. 5.** Example of decision surface formed by SVM classifier.

Labelled data is required to train the SVM. Typically the training target data set only contains a few vectors so all these must be used. However there are typically tens of thousands of known background vectors. It is not possible to use them all to train the SVM as the memory space required by the algorithm is quadratic in the number of training points. A representative set of points needs to be found. What we have found is that selecting random points on the edge of the background distribution (i.e. those with low density values) seems to give the best results. This corresponds to classifier boosting as advocated by Freund and Shapire [6].

A trained SVM was applied to the outliers obtained in the AM sequence. In this case the SVM only accepted five points as belonging to the target class, all the other outliers were classified as belonging to the background class. These five points were then ranked as a function of their distance from the decision boundary formed by the SVM. The points along with their corresponding ranks are shown in figure 6. As one can see the real target has been labelled as the most probable target in the scene.


(a) SVM

**Fig. 6.** Only 5 detections after using support vector classifier

## 6 Conclusion

In this paper we have demonstrated a system for detecting dim targets in a cluttered background, i.e. sea glint. This ATR system has been favourably compared to another leading edge technique used as a baseline in our study. Several improvements to our system have also been made to our original system presented in [7], namely:

**Clutter Model** A Gaussian mixture is used to model the feature distribution of the background. This allows for a more flexible representation of the clutter.

**Temporal Data** The temporal nature of the data is being incorporated into the design of the filters, making more robust filters.

**SVM** The application of support vector machines to aid the classification of the most outlying data points has significantly improved performance. ( It can also be argued that a similar performance level increase would of been observed if the outliers found by the MCP method were passed through an equivalent SVM).

An advantage of our approach is that the system is flexible which complies with realistic operational scenarios. We assume a model is available of what the sensor is looking for, i.e. the target. As a model of the current sensor input has been computed our system can be optimally tuned to distinguish between this target and the current background. If the background happens to change or the target model is modified the system can be adapted to this new environment.

In fact, both our method and the MCP method can be seen as complimentary. Both are looking for the same targets but each uses a different technique to obtain the posterior target probabilities. In theory, it should be possible to achieve a more robust ATR system by the combination of these probabilities.

**Acknowledgements** This research was funded by the MoD under the Corporate Research Program by Technology Group 10: Information Processing and Technology Group 3: Aerodynamics, Propulsion, Guidance and Control. ©British Crown copyright 2000. Published with the permission of the Defence Evaluation and Research Agency on behalf of the Controller of HMSO.

## References

1. B Bhanu. Automatic target recognition: State of the art survey. *IEEE Transactions on Aerospace and Electronic Systems*, 4:364–379, July 1986.
2. B Bhanu, D Dudgeon, E Zelnio, A Rosenfeld, D Casasent, and I Reed. Introduction to the special issue on automatic target detection and recognition. *IEEE Transactions on Image Processing*, 6:1:1–3, January 1997.
3. B Bhanu and T Jones. Image understanding research for automatic target recognition. *IEEE Transactions on Aerospace and Electronic Systems*, pages 15–22, October 1993.
4. A Dempster, N Laird, and D Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.
5. P Devijver and J Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall, 1982.
6. Y. Freund and R. E. Shapire. A decision-theoretic generalization of an on-line learning and an application to boosting in computational learning theory. In *EuroCOLT*, pages 23–37, 1995.
7. K Messer, D. de Ridder, and J Kittler. Adaptive texture representation methods for automatic target recognition. In *Proc British Machine Vision Conference BMVC99*, September 1999.
8. P Pudil, J Novovicova, and J Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119–1125, 1994.
9. M Roth. Survey of neural network technology for automatic target recognition. *IEEE Transactions on Neural Networks*, 1:28–43, March 1990.
10. L Sardo and J Kittler. Model complexity validation for pdf estimation using gaussian mixtures. In S Venkatesh A K Jain and B C Lovell, editors, *International Conference on Pattern Recognition*, pages 195–197, 1998.
11. V. Vapnik. *The nature of statistical learning*. Springer-Verlag, 1995.
12. G. Watson and S. Watson. Detection and clutter rejection in image sequences based on multivariate conditional probability. In *SPIE: Signal and Data Procssing of Small Targets*, volume 3809, pages 107–118, 1999.