MULTI-VIEW TARGET CLASSIFICATION IN SYNTHETIC APERTURE SONAR IMAGERY

David Williams[a], Johannes Groen[b]

[ab]NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, Italy

Contact Author: David Williams
NATO Undersea Research Centre, Viale San Bartolomeo 400, 19126 La Spezia, Italy.
Telephone: (+39) 0187.527.439. Fax: (+39) 0187.527.331. williams@nurc.nato.int

Abstract: *This work proposes an elegantly simple solution to the general task of classifying the shape of an object that has been viewed multiple times. Specifically, this problem is addressed in the context of underwater mine classification where the objective is to discriminate targets (i.e., mines) from benign clutter (e.g., rocks) when each object is observed in an arbitrary number of synthetic aperture sonar (SAS) images. The proposed multi-view classification algorithm is based on finding the single highest maximum correlation between (i) a set of views of a training shape of interest and (ii) a set of views of a given testing object. Classification is performed by using this measure of similarity, which we term the affinity, directly. This approach obviates the need for explicit feature extraction and classifier construction. Moreover, the framework induces no constraints on the number of views that each object can possess. Promising experimental results using real SAS imagery demonstrate the feasibility of the proposed approach for multi-view classification of underwater mines. In particular, it is shown that classification performance improves dramatically as the number of views of the objects increases.*

Keywords: *Target classification, mine detection, multi-view classification, data fusion, synthetic aperture sonar (SAS)*

1. INTRODUCTION

This work proposes a solution to the general task of classifying the shape of an object that has been viewed multiple times. Specifically, this problem is addressed in the context of underwater mine classification where the objective is to discriminate targets (*i.e.*, mines) from benign clutter (*e.g.*, rocks) when each object is observed in an arbitrary number of synthetic aperture sonar (SAS) images.

If a mine is observed from one orientation, it may be difficult to distinguish it from a rock; however, viewing it from a second orientation may reveal previously obscured characteristics that differentiate it from a rock. In general, the information accrued from multiple views of an object should translate into improved classification performance. This basic concept motivates the collection of multi-view data for the classification of underwater mines.

The nature of data collected for the underwater mine classification problem differs from most multi-view classification problems in that an *arbitrary* number of views of a given object will be an *unordered* set of observations. In this work, we propose a new classification method that combines information from multiple views, when each object can be viewed an arbitrary number of times. Specifically, the multi-view classification approach proposed here is based on finding the single highest maximum correlation between (i) a set of views (*i.e.*, images) of a training shape of interest and (ii) a set of views of a given testing object. Classification is performed by using this measure of similarity, which we term the *affinity*, directly.

This approach is particularly well-suited for the underwater mine classification problem for several reasons. For one, it fully exploits the recent advances in SAS systems by focusing on the detailed shape information of the objects that the high-resolution imagery provides. But more importantly, the approach also overcomes the unique challenges presented by the general underwater mine classification problem that cause standard (single-view) classification approaches to fail.

A fundamental assumption of machine-learning algorithms is that the underlying mechanisms (and hence statistics) that generate the training and testing data are the same [1]. In the underwater mine classification problem, however, this implicit assumption is often violated because different types of clutter objects can be encountered at different sites. As a result, a classifier learned from clutter training data collected at one site often will not generalize well on testing data collected at a different site.

Furthermore, in the underwater mine problem, the universe of target shapes that one is interested in detecting and classifying is typically small. However, within this small class of target shapes, marked differences exist among the different shapes. As a result, the features that are salient for certain shapes are not relevant for classifying other shapes. This fact induces a need for larger feature spaces, which in turn necessitates more training data to build a reliable classifier.

The fact that the image of a target is highly aspect-dependent further increases the need for more training data (namely, from multiple aspects). For example, the image of a broadside cylinder will look very different from the image of an end-fire cylinder. In general, the relative paucity of training data — of different targets, at different ranges, at different aspects, and in different site conditions — contributes to the difficulty of building a robust classifier.

The approach proposed in this work combats these challenges inherent to the underwater mine classification task by obviating explicit feature extraction and classifier construction. Moreover, in the proposed multi-view classification method, there is no constraint on the number of views that can be handled. Importantly, the elegantly simple approach is naturally suited to allow each object to be viewed an arbitrary number of times.

The remainder of this paper is organized in the following manner. The proposed multi-view classification approach is described briefly in Section 2. Experimental results using real

SAS imagery are presented in Section 3. Concluding comments are made in Section 4. Page constraints limit the detail into which the method can be described here; the interested reader can find a longer, more detailed presentation of this work in [2].

2. MULTI-VIEW CLASSIFICATION

The multi-view classification approach we propose is based on computing the maximum correlation between (i) a set of views of a known object of interest and (ii) a set of views of a test object of unknown identity.

We define the *affinity* to be the maximum over all of the correlation maxima obtained between every possible combination of a view of the training shape and a view of the testing object. That is, the affinity is a quantitative measure of the highest degree of similarity between the training shape and the testing object, considering all of the available views of each.

By construction, the affinity between any two objects will be monotonically increasing with increasing numbers of views (of one or both of the objects). Admittedly, there may be cases, for some limited number of views, where the affinity between two objects of different shapes will actually be greater than the affinity between two objects of the same shape. However, as the total number of views increases, the increase in affinity for the latter will in general be faster, and will therefore surpass the affinity of the former. That is, with enough views, the affinity between two objects of the same shape would be greater than the affinity between two objects of different shapes. This hypothesis relies on the belief that, eventually, the two objects of the same shape will be observed at similar enough orientations such that they appear nearly identical. It also assumes that two objects of different shapes will never appear *exactly* identical at any orientation.

It is this affinity — “the maximum of the maxima of the correlations” — with which classification decisions will be made. Specifically, if the affinity between a testing object and the given training object of interest is above a set threshold, the testing object is classified as the training object's shape. Otherwise, it is declared to be some other shape.

The proposed multi-view classification algorithm assumes that the target types (*i.e.*, shapes) of interest are known *a priori*. However, all supervised classification approaches always make this same assumption. If one does not know which targets are to be classified, devising a set of sensible features is not possible. Moreover, standard classification approaches implicitly assume that the targets in the training set will match the targets in the testing set. Therefore, our assumption about knowing the target types of interest is completely justified.

It should also be noted that the proposed approach does not require a training data set of both targets and clutter. Instead, training views of only the *targets* of interest are required. This aspect is important because significantly different types of clutter objects can be found in different areas (*e.g.*, at a training site and at a testing site). Thus, our approach circumvents the problem of training on clutter objects that are not representative of the type of clutter objects that one may encounter at a new testing site. Moreover, this approach also obviates the feature extraction and classifier learning processes, thereby avoiding the difficulties associated with them.

Finally, the proposed multi-view classification algorithm is predicated on the fact that *multiple* views of an object will be available. Nevertheless, if a given object is viewed only a single time, the proposed classification approach is still valid.

3. EXPERIMENTAL RESULTS

3.1. Data Set

In April-May 2008, the NATO Undersea Research Centre (NURC) conducted the Colossus II sea trial in the Baltic Sea off the coast of Latvia. During this trial, high-resolution sonar data was collected by the MUSCLE autonomous underwater vehicle (AUV), which is equipped with a 300 kHz sonar.

On 29 April, before the data collection was performed, six known targets — three cylinders and three truncated cones — were placed on the seabed. The vehicle then made multiple passes over the target area in different orientations. The data that resulted from this collection was then processed into SAS imagery.

A (contact) detection algorithm [3] that employs a matched filter to find highlight-shadow patterns characteristic of mine-like objects in the images was then applied. This detection algorithm generated a total of 2,395 detections. For each detection, a SAS image chip was then extracted.

Subsequently, all detections (*i.e.*, views) of a given object were grouped together by using vehicle-recorded latitude and longitude information. Based on this data-association process, it was established that a total of 317 unique objects comprised the 2,395 detections (for a mean of 7.55 views per object). The maximum number of views of any object was 27.

3.2. Experimental Procedure

The objective of these experiments is to assess the classification performance of the proposed affinity-based approach in terms of discriminating the six targets — three cylinders and three truncated cones — from clutter as a function of the number of views of each object. This goal is accomplished via the following experimental procedure.

Assume one of the six objects of interest is labeled training data. The remaining 316 objects are treated as unlabeled testing data that we wish to classify.

Randomly select *n* views of each object (or all of an object's views if fewer than *n* views are available). For each testing object, compute the affinity that the set of selected views of the object has with the set of selected views of the given training object.

To determine the performance for correctly classifying a testing object that is the same target shape as the training object, one can simply count the number of other testing objects that have a higher affinity than the affinity of the matched testing object. This number is exactly the number of false alarms, from which the probability of false alarm can be easily deduced.

In the above procedure, the affinity values that were obtained were between two objects when each was viewed no more than *n* times. However, each affinity value corresponds to only one possible pair of randomly selected views. Therefore, to increase the statistical strength of the experimental results, we repeat this entire process 1000 times — in lieu of considering every possible combination of views, which is computationally infeasible — where new sets of views for each object are randomly selected in each iteration.

Thus, for a given maximum number of views of each object, *n*, we have 1000 probability of false alarm values when trying to classify a testing object of the same target shape as the training object. The mean and variance of the probability of false alarm can then be readily calculated from the 1000 values.

All of the above is then performed for each possible value of *n*, from 1 to 27, to obtain performance as a function of the number of object views.

3.3. Classification Results

Fig.1 shows, for each maximum number of views of any object to be considered, the mean classification performance (averaged over the 1000 iterations, in which each iteration uses a random selection of each object's views) in terms of the probability of false alarm when each of the six objects of interest — three truncated cones and three cylinders — is treated as the training target. The error bars in the figures represent one variance (*i.e.*, square of the standard deviation) above and below this mean value.

As can be observed in the figure, the classification performance improves dramatically as the number of views of each object increases. Moreover, the benefit of obtaining additional views is most significant when only a limited number of views is possessed.

It should be noted that the targets of interest typically possessed more views than most of the clutter objects in this data set. This fact is because the only views (*i.e.*, images) that are considered in the classification stage are those that pass the detection stage's prescreener.

A view of a target from any aspect will usually appear mine-like, and hence pass the detection stage. In contrast, many clutter objects appear mine-like from only certain aspects, and hence pass the detection stage only occasionally. As a result, target objects will consistently be detected while clutter objects are detected only sometimes. Hence, target objects will have more views than clutter objects in the classification stage. In turn, target objects will have more opportunities to achieve a high affinity with a given training target of interest. This ostensible bias is actually desirable for the proposed approach because classification performance improves as the number of views of a target increases.

4. CONCLUSION

This work presented an elegantly simple solution for the general task of classifying the shape of an object that has been viewed multiple times. Specifically, this problem was addressed in the context of underwater mine classification where the objective is to discriminate targets (*i.e.*, mines) from benign clutter (*e.g.*, rocks) in SAS images.

The proposed approach was based on computing the maximum correlation between each of the multiple views of a known shape of interest with each of the multiple views of a test object of unknown shape. Specifically, the maximum of that set of correlation maxima, which we termed the *affinity*, was used directly to make classification decisions. A set of experiments with real SAS imagery demonstrated the feasibility of using the proposed method for multi-view classification of shapes.

A more detailed presentation of this work (in which the actual multi-view SAS imagery of the objects is shown) can be found in [2].

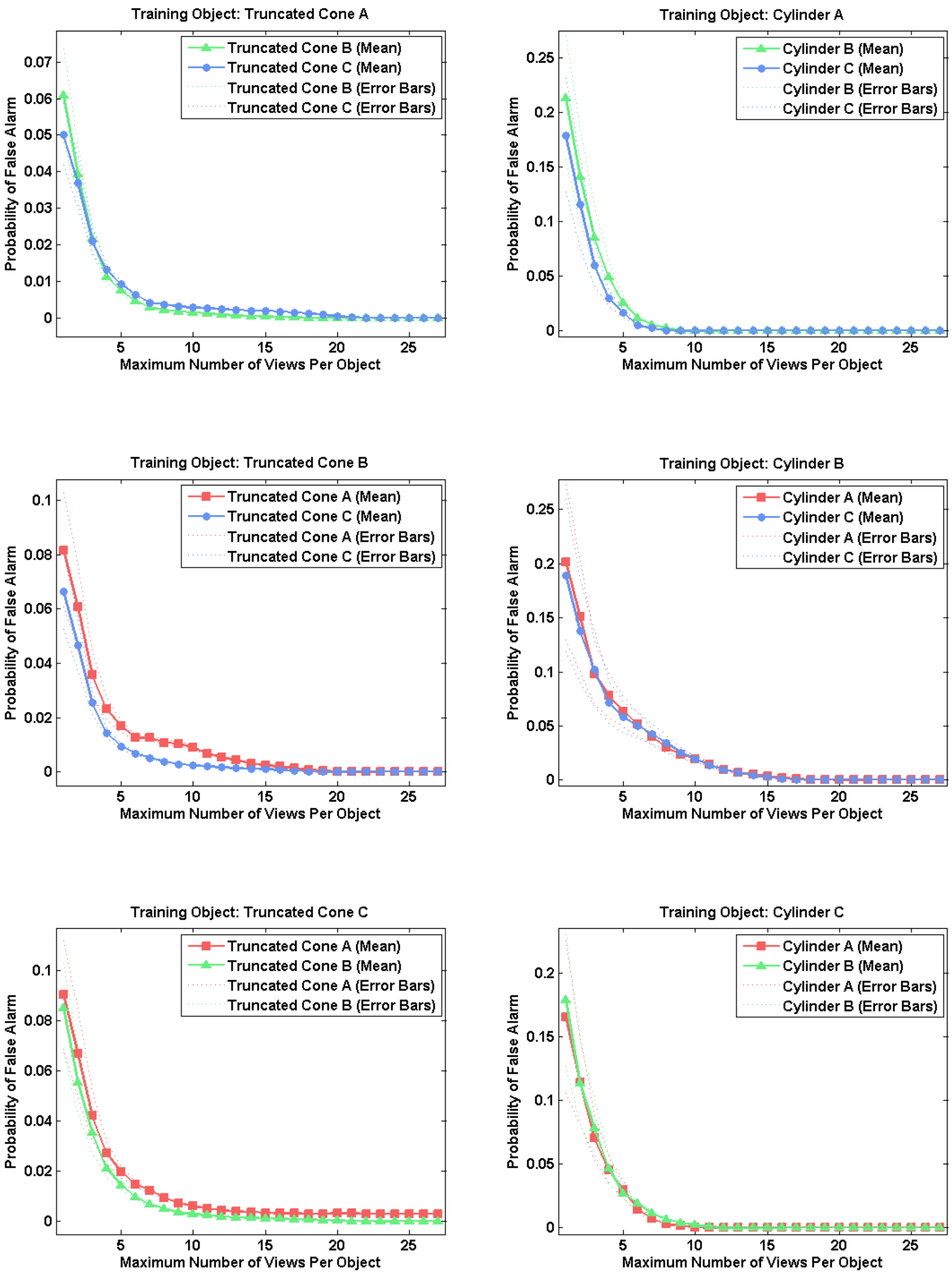

Fig.1: Performance of proposed approach for correctly classifying the objects noted in the legends, as a function of the number of views considered for each object in the data set, when the training object is that shown in the subfigure titles.

REFERENCES

[1] **V. Vapnik**, *Statistical Learning Theory*, Wiley, 1998.
[2] **D. Williams**, "On Multi-View Mine Classification with SAS Imagery," NATO Undersea Research Centre Technical Report, NURC-FR-2008-026, 2008.
[3] **J. Fawcett, A. Crawford, D. Hopkin, V. Myers, and B. Zerr**, "Computer-aided Detection of Targets from the CITADEL Trial Klein Sonar Data," Defence R&D Canada Technical Memo, DRDC Atlantic TM 2006-115, 2006.