SAS and Bathymetric Data Fusion for Improved Target Classification

David P. Williams

(1) Centre for Maritime Research and Experimentation (formerly NATO Undersea Research Centre), Viale San Bartolomeo 400, La Spezia 19126 (SP), Italy
E-mail: williams@cmre.nato.int

SHORT ABSTRACT: An algorithm is proposed for the fusion of multiple views of an object from each of two information sources – a synthetic aperture sonar (SAS) image and a bathymetric map. A parameter tied to the success of interferometric processing, and hence the reliability of the bathymetric estimates, automatically weights the relative contribution of each information source. The variation in fused images, measured by the Laplacian, is used to determine the image translation needed to align multiple views. The algorithm is completely model-free and requires no a priori knowledge about the types of objects that will be considered. As a result, the method has potential to be particularly useful for reducing false alarms generated by clutter objects, and in turn, for improving classification performance. The proposed fusion algorithm is demonstrated on three objects using real, measured data collected at sea.

Keywords: Data Fusion, Multi-View, Synthetic Aperture Sonar (SAS), Bathymetry, Classification.

1 INTRODUCTION

Mine countermeasures (MCM) operations depend on the ability to reliably classify objects on the seafloor as targets (i.e., mines) or benign clutter. The use of high-resolution synthetic aperture sonar (SAS) imagery [1] to accomplish this task has allowed a leap in performance over that which was previously possible with lower resolution side-scan sonar imagery.

However, it is well-known that an object’s appearance in a SAS image can depend heavily on the relative aspect at which the object is interrogated. The classic example of this aspect-dependence is that of a cylinder on a seafloor; a cylinder viewed at broadside will look markedly different than the same cylinder viewed at endfire. In fact, even with high-resolution imagery, it may not be possible to correctly classify an object with only a single (“unlucky”) view. This fact highlights the value of collecting multiple views of an object at different aspects. With view diversity, a more comprehensive understanding of the object can be obtained, and in turn, a more informed and confident classification decision can be made [2, 3].

In this work, we aim to aid classification performance by combining individual SAS views of an object into a single multi-view SAS image. Since this fusion is in fact the overlaying of multiple images of the same object, it can more properly be called image registration, which has a long history in many different fields and with many different sensor modalities (see [4] for a survey). However, the nature of SAS imagery – and the manner in which the sonar data is collected – presents several unique challenges that are not commonly encountered in other domains.

Sonar data for MCM operations is typically collected by an autonomous underwater vehicle (AUV). Because GPS cannot be used while underwater, the navigation accuracy of the AUV is limited [5].
Fig. 1: Motivating example: Unsatisfactory SAS image-fusion result of two views (at the same aspect) of the torus-like Object A when the fusion is performed using (a) navigation information, or (b) landmark (i.e., contact-location) information.

(This should be contrasted with the GPS-aided accuracy achievable in synthetic aperture radar (SAR) applications.) Consequently, registering objects based on navigation information is not feasible because the errors can be large relative to the size of the target.

As an example, Fig. 1(a) shows the SAS image-fusion result of two views of a torus-like object using only navigation information. The two views were collected by an AUV executing a survey composed of a series of parallel tracks; the views were obtained from different tracks (i.e., passes). It can be seen that the error between the two views of the object is about 2.5m, an unacceptable amount in this context.

One popular class of registration approaches, referred to as feature-based techniques [4], rely on the use of landmarks (i.e., prominent features) to align images. This approach is frequently employed to mosaic multiple large-scale sonar images [6, 7], but it cannot achieve the fine precision needed for aligning object images. An example of fusing the two views of the torus-like object based on the location of a landmark (in this case, the location of the object itself, obtained from the output of an automatic detection algorithm [8]) is shown in Fig. 1(b). Although the object alignment is better than in the navigation-based approach, it can readily be seen that the two views still do not coincide sufficiently.

Computer vision applications, such as object recognition in scenes [9], employ feature-matching methods with great success, but the non-repeatability of features owing to aspect dependence in SAS images (unlike the distinct, repeatable details that can be captured in photographs) disqualifies such approaches for our task.

The other major class of registration techniques are referred to as area-based methods [4]. This group is composed of correlation-based and template-matching approaches. Unfortunately, the nature of SAS images – namely, that the fundamental appearance of an object changes with aspect – makes image fusion via direct correlation-based methods infeasible. For example, the parts of an object that are characterized by shadows or highlights in a SAS image will be determined by the viewing aspect. As a result, the same object can look completely different in SAS images from different aspects, which contradicts the basic premise for employing a correlation-based approach.

This obstacle is partially circumvented in [10] by comparing a given SAS image to models of several known targets at numerous aspects. (The use of a model reference image for template matching is also common in many medical imaging applications such as brain scan registration [11].) However, such an approach is feasible only if the unknown object belongs to the set of targets for which a model is possessed. In practice, the unknown object may be something unexpected that has never previously been encountered. Hence, the approach is ill-suited for fusing multiple views of objects.
in the clutter class, to which the vast majority of detected objects typically belong [12]. Moreover, the computation required for this approach grows exponentially with the number of target-models considered.

The failure of common registration approaches and the inapplicability of others motivate the need for a new SAS-specific image-fusion algorithm. We contend that the ill-posed nature of the SAS image-fusion task – owing to the fact that the fundamental appearance of an object changes with aspect – suggests that in order to successfully fuse multiple views, it is necessary to have access to an auxiliary source of information. In this work, this second information source is bathymetric maps, which provide height estimates of the object, obtained from interferometric processing of the sonar data.

Bathymetric maps are a particularly valuable source of information in this fusion task because they measure intrinsic features of the object – the heights at specific locations – that should be invariant to viewing aspect [13]. (In contrast, we know a priori that the SAS image of an object will – and should – indeed vary with aspect; this means that features found in the SAS images will not be robust.) However, because the availability of the bathymetric maps is predicated on the ability to successfully perform interferometric processing, one cannot rely solely on the bathymetry to perform the image registration. (Successful interferometric processing cannot always be guaranteed for data collected at sea [14].) Therefore, in this work, we develop a (model-free) algorithm to determine the correct image registration – specifically, the correct translation – to align multiple SAS images of an object by exploiting both the SAS images themselves and, when they are reliable, bathymetric maps from those same views.

The remainder of this paper is organized as follows. In Sec. 2, the proposed multi-view fusion algorithm is described. Sec. 3 presents experimental results of this method on real, measured SAS data. Concluding remarks and directions for future work are given in Sec. 4.

2 MULTI-VIEW FUSION
Consider the case in which an object of unknown identity has been interrogated by each of two – not necessarily statistically independent – sensor “modalities” (or, more generically, information sources) multiple times. Assume spatial data (e.g., imagery) related to the object can be produced from a single view with either information source independently (i.e., without having any knowledge of the other source). In this work, the two sources of information representing the object are assumed to be (i) a SAS image and (ii) a bathymetric map obtained from interferometric processing. Our objective is to use the multiple views from these two data sources to create a single fused SAS image. The intent is that this fused image incorporating multiple views simultaneously will be more informative than the set of individual views, in terms of classifying the object.

Let \( I^s_i \) be the SAS image of the object obtained from the \( i \)th of \( N > 1 \) views. Let \( I^b_i \) be the estimated bathymetric map of the area centered around the object obtained from the \( i \)th view. We perform the desired multi-source, multi-view fusion in the following manner.

First, each image is rotated in order to transform the data from the \( N \) views into a common reference system. The rotation applied is determined from the heading recorded on the AUV during the data collection. Examples of this rotation step are shown in Fig. 2. After this rotation, the only remaining element in the fusion that is still unknown is the relative translation (i.e., shift) that should be applied to the images in order to align them.

To determine the appropriate translation, we exploit both the SAS images and the bathymetric maps. Specifically, we seek to minimize an objective function that expresses the collective variation that results after fusing translated images. The rationale for this approach is that low variation should correspond to a proper image alignment. The objective function is constructed to depend on the variation, rather than the original raw image, partially to circumvent the fact that SAS images of a given object will naturally look different when interrogated at different aspects.
The Laplacian operator, which is the second unmixed partial spatial derivative, has proven useful for detecting rapid intensity changes in an image [15], and therefore, we employ it in this work to measure the variation in the fused image. The Laplacian of an image with pixel values \( I(x, y) \) is given by

\[
\mathcal{L}(I) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2},
\]

with this calculation performed on a discrete image using a finite difference method. (For the sake of brevity and clarity, the \((x, y)\) argument of images are suppressed hereafter.)

For a given row and column translation, \( \tau \equiv \{r, c\} \), let \( I_s(\tau)_{i,j} \) be the fused SAS image that results from combining the SAS images from the \( i \)th view and \( j \)th (translated) view. The fusion is effected in this work by taking the maximum pixel value of the two views at each pixel location (after translation has been applied to the second image). This form of fusion using the maximum operator is chosen with consideration for the nature of SAS imagery, and in particular, the shadows that are created due to the object-sonar geometry. For example, averaging the pixel values would unfairly penalize portions of the object characterized by shadow in a view. Our choice also suggests that in the limit of an infinite number of views spanning all aspects, the resulting fused image would be similar to a circular SAS image [13].

Let \( V_s(\tau)_{i,j} \) denote the image-variation from fusing the SAS images from the \( i \)th and \( j \)th views (with translation \( \tau \)), defined by

\[
V_s(\tau)_{i,j} = \sum_x \sum_y |\mathcal{L}(I_s(\tau)_{i,j})|,
\]

with the summations over all pixels in the image.

The bathymetric maps are fused in a slightly different manner. For a given translation, \( \tau \), let \( I_b(\tau)_{i,j} \) be the fused bathymetric map that results from combining the bathymetric maps from the \( i \)th view and \( j \)th (translated) view. The fusion is effected in this work by taking the mean pixel value of the two views at each pixel location (after translation has been applied to the second image). This form of fusion is chosen with consideration for the nature of the bathymetric maps, which represent relative heights (with respect to an arbitrary reference point that is different for each view’s map) rather than absolute heights. It is for this reason that alternative choices, such as selecting at each pixel location the larger height or the height associated with the larger coherence, is not sensible.

Let \( V_b(\tau)_{i,j} \) denote the image-variation from fusing the bathymetric maps from the \( i \)th and \( j \)th views (with translation \( \tau \)), defined by

\[
V_b(\tau)_{i,j} = \sum_x \sum_y |\mathcal{L}(I_b(\tau)_{i,j})|,
\]

It should be noted that there will be one scalar \( V_s(\tau)_{i,j} \) and one scalar \( V_b(\tau)_{i,j} \) for each possible translation, \( \tau \). The minimum fused bathymetric map variation achieved by any translation will be denoted \( \tilde{V}_b^{i,j} \). The minimum fused SAS image variation achieved by any translation will be denoted \( \tilde{V}_s^{i,j} \).

The elegantly simple objective function we seek to minimize is then given by

\[
Q(\tau) = w \frac{V_b(\tau)_{i,j}}{\tilde{V}_b^{i,j}} + (1 - w) \frac{V_s(\tau)_{i,j}}{\tilde{V}_s^{i,j}},
\]

where \( w \in [0, 1] \) is a weight that controls the relative contribution of the SAS images and the bathymetric maps. This key weight is calculated as

\[
w = \min(f_i, f_j),
\]

where \( f_i \) is the fraction of pixels in the \( i \)th view for which the coherence (from interferometric processing) is above a threshold, \( \rho \). (In this work, \( \rho = 2/3 \).) As such, \( f_i \) represents a rough measure of
confidence about the bathymetric height estimates because a high coherence (above $\rho$) suggests the interferometric processing was successful. In (5), the minimum is selected because it makes sense to fuse two bathymetric maps only if both views are of good quality. That is, there is little (or no) benefit in attempting to fuse one accurate bathymetric map with one that is inaccurate (or unreliable).

In (4), the variations of the fused bathymetric map and fused SAS image are each normalized by the minimum variation achieved (by any translation) to ensure that the contribution of each term to the objective function is comparable in magnitude. We normalize by the minimum variation achieved instead of the maximum variation achieved because the latter approach would still have the potential to wildly skew the relative magnitudes.

The translation, $\tau$, that minimizes the objective function in (4) is selected as the correct translation needed to properly align the multiple views of the object and produce an accurate fused image. Upon determining this best translation, the two SAS image views are fused (keeping the maximum pixel value at each location, as above), the bathymetric maps are fused (averaging the height estimates at each location, as above), and the coherence maps are fused (taking the minimum value at each location, for reasons outlined above).

If an additional view is possessed (i.e., $N > 2$), all of the above procedure can be repeated treating the new fusion result as an input “view.” We choose to perform the fusion sequentially to ensure that the computational load scales only linearly with $N$; if the fusion were to be performed in parallel, the computation would instead scale exponentially with $N$.

3 EXPERIMENTAL RESULTS

The Centre for Maritime Research and Experimentation (formerly NATO Undersea Research Centre) has an AUV called MUSCLE that is equipped with an interferometric SAS (InSAS), which permits both synthetic aperture sonar processing and interferometric processing. As a result, multiple sources of information are available with which to aid classification for a given object detected on the seafloor. Using data collected by this AUV at sea, we demonstrate the proposed fusion algorithm outlined in Sec. 2 on three different clutter objects of unknown identity. Admittedly, in the absence of ground truth, it is difficult to rigorously evaluate the efficacy of the algorithm. The results here act merely as an initial proof-of-concept.

The first object, denoted Object A, has been viewed three times, twice at the same aspect and a third time at a different aspect. The aspects of the three views are $\theta_1 = 126^\circ$, $\theta_2 = 126^\circ$, and $\theta_3 = 37^\circ$. We first perform fusion using views 1 and 2. For this case, the interferometric processing was very successful, yielding high coherence values. Since the bathymetric maps are reliable in this case, they should play a strong role in the fusion process. This is precisely what occurs, as the weight controlling the relative contribution of each information source (SAS image and bathymetric map) in the fusion process was calculated to be $w = 0.84$. The result of the fusion procedure, as well as the data used to undertake it, is shown in Fig. 2. This fusion result should be compared to the (unsatisfactory) fusion results in Fig. 1 that would be achieved using navigation data or contact-location data.

After fusing these two views, the third view, which was at a different aspect, was also fused. This view was characterized by low coherence, and as a result, more importance was automatically placed on the SAS image during the fusion process via the weight, calculated to be $w = 0.37$. The result of this tri-view fusion is shown in Fig. 2(n).

Next, we demonstrate the fusion on an irregularly shaped object, denoted Object B. The aspects of the two views are $\theta_1 = 214^\circ$ and $\theta_2 = 306^\circ$. The fusion result of this object and the data used to achieve it are shown in Fig. 3. In this case, the coherence was again strong, so more importance was placed on the bathymetric fusion via the weight, calculated to be $w = 0.88$. This case is interesting because the SAS images from the two views are significantly different. The fusion result is also shown in a three-dimensional plot in Fig. 3(j) by draping the fused SAS image on top of the fused
Fig. 2: Object A data and fusion results. The aspects of the three views are $\theta_1 = 126^\circ$, $\theta_2 = 126^\circ$, and $\theta_3 = 37^\circ$. 
Fig. 3: Object B data and fusion results. The aspects of the two views are $\theta_1 = 214^\circ$ and $\theta_2 = 306^\circ$. 
Fig. 4: Object C data and fusion results. The aspects of the two views are $\theta_1 = 124^\circ$ and $\theta_2 = 254^\circ$. 
Lastly, we demonstrate the fusion on a third object, denoted Object C. The aspects of the two views are $\theta_1 = 124^\circ$ and $\theta_2 = 254^\circ$. In this case, the coherence was very weak, so more importance was placed on the SAS image fusion via the weight, calculated to be $w = 0.15$. Although plausible, the fusion result is difficult to quantify objectively without ground truth.

One appealing aspect of the proposed algorithm is that the data quality is automatically taken into account in the fusion procedure. If the interferometric processing is successful, the additional information provided by the bathymetric maps is exploited. If the interferometric processing is not successful, the algorithm knows to ignore that information source and instead base the fusion on the SAS images alone. However, it should be noted that relying solely on the SAS images may not produce a satisfactory fusion result in general. Indeed, it was for this very reason that the proposed algorithm was formulated to exploit the bathymetric maps.

4 CONCLUSION
An algorithm for the fusion of multiple views of an object from each of two information sources – a SAS image and a bathymetric map – was proposed. The technique is completely model-free and requires no a priori knowledge about the types of objects that will be encountered. Preliminary results demonstrated the promise of the technique, but the multi-view fusion problem is still far from solved.

Future work will seek to improve the robustness of the algorithm and see its application to larger data sets with known targets so fusion performance can be measured quantitatively. Additional research will aim to determine the best way to fuse data when $N > 2$ views are possessed. For example, it may make sense to fuse images from similar aspects first, or to first fuse images for which the interferometric processing was most successful.

References


