AIS-SAR Data Association Algorithms - Final Report

Report
R-14-111-1213

Prepared for:
Defence Research and Development Canada (DRDC)

Revision 1.2
April, 2015

Prepared by:
James Youden and Bing Yue

PWGSC Contract: W7714-09-1149

DRDC Technical Authority: Paris Vachon, Defence Scientist

The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

Captain Robert A. Bartlett Building
Morrissey Road
St. John's, NL
Canada A1B 3X5

T: (709) 864-8354
F: (709) 864-4706

Info@c-core.ca
www.c-core.ca

Contract Report
DRDC-RDDC-2015-C136
April 2015

Registered to ISO 9001:2008

This page is intentionally left blank

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2015
© Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2015

AIS-SAR Data Association Algorithms - Final Report

Prepared for:

Defence Research and Development Canada (DRDC)

Prepared by:

C-CORE

C-CORE Report Number:

R-14-111-1213

Revision 1.2

April, 2015

Smart Solutions for Challenging Environments

Captain Robert A. Bartlett Building
Morrissey Road
St. John's, NL
Canada A1B 3X5

T: (709) 864-8354
F: (709) 864-4706

Info@c-core.ca
www.c-core.ca

Registered to ISO 9001:2008

The correct citation for this report is:
C-CORE. (2015). *AIS-SAR Data Association Algorithms - Final Report*, C-CORE Report R-14-111-1213, Revision 1.2.

Project Team

James Youden (Project Manager)
Bing Yue

REVISION HISTORY

VERSION	SVN	NAME	COMPANY	DATE OF CHANGES	COMMENTS
1.0	34	James Youden Bing Yue	C-CORE	03/13/15	
1.1	40	James Youden	C-CORE	04/08/15	Added recommendations; minor edits
1.2	44	James Youden	C-CORE	04/28/15	Minor edits, header update, layout changes

DISTRIBUTION LIST

COMPANY	NAME	NUMBER OF COPIES
Defence Research and Development Canada (DRDC)	Paris Vachon	1 electronic

EXECUTIVE SUMMARY

Monitoring of Canada's territorial waters as well as other areas of interest is aided by the use of both Automatic Identification System (AIS) reports and the acquisition of synthetic aperture radar (SAR) imagery. The combination of information from these two systems provides the possibility of detecting vessels that have not otherwise been identified. The aim of the current review was to provide a summary of the AIS and SAR association algorithms that are available in the literature, including the present development level for the algorithms, and also to provide an evaluation of the performance of AIS-SAR association, in particular, for areas of high ship density, for coastal regions, and for time differences between AIS and SAR acquisitions of up to two hours. However, there are several challenges with a publication-based review of AIS-SAR association algorithms, especially in relation to comparisons of the relative performances, since these details for most current implementations appear not to be disclosed, especially as related to the performance under varying scenarios.

The current literature review therefore highlights the approaches that may be used to match AIS-identified targets with targets that have been detected in SAR imagery. The algorithms based on point-to-point association include point pattern matching, nearest neighbour and two-way nearest neighbour. The point-to-track algorithms are based on a multiple target tracking approach with association through nearest neighbour, global nearest neighbour and joint probabilistic data association methods, which include the use of the Munkres algorithm to find the optimal associations.

The AIS-SAR association algorithms are briefly evaluated in terms of their functionality and applicability, and in terms of their level of current development. The likely accuracy of AIS-SAR association as a function of ship density and the time difference between AIS and SAR acquisitions is also estimated. However, the actual performance of any AIS-SAR association algorithm will depend significantly on the scenarios under consideration, and is best evaluated for the specific areas of interest using actual shipping data. Furthermore, additional information such as the use of ground-based AIS receiving stations, target tracking algorithms, and ship-length correlation can all reduce the burden and increase the likely accuracy of any associations.

To date, simple nearest neighbour (NN) algorithms have been used most often in AIS-SAR association, with the assignment problem solved through the Munkres algorithm. An implementation of this approach in MATLAB could be easily accomplished for evaluation on specific scenarios of interest. Implementation of operational software in C++, including interfacing with currently existing image analysis and display software, would likewise be straightforward.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 ASSOCIATION ALGORITHM REVIEW

Canada possesses the world's longest shoreline. On any given day, Canada faces the laborious task of monitoring its marine activity, consisting of more than 250 ports and over 1,700 ships (Abielmona, Falcon, Vachon & Groza, 2014). It is crucially important to continue enhancing Maritime Domain Awareness (MDA) capabilities using the most advanced monitoring techniques. The high spatial resolution, large coverage, short revisit time and all-weather operation of Synthetic Aperture Radar (SAR) and the detailed information and coverage of Automatic Identification System (AIS) (including spaceborne AIS) have enabled these two sensors to provide valuable intelligence in maritime traffic surveillance.

One of the surveillance activities is to identify anomalous ships within territorial waters and nearby sea regions. Currently, the International Maritime Organization's International Convention for the Safety of Life at Sea requires AIS to be fitted aboard international voyaging ships of 300 gross tonnage or more, ships of 500 gross tonnage or more not engaged on international voyages, and all passenger ships regardless of size. When the system is on, an AIS device will automatically send the vessel information with high precision and fast response between AIS sensors on land, onboard ships and on satellites. The information includes vessel name, type, position, heading, velocity and so on. It is thus obvious that only a cooperative vessel can be tracked by AIS as only cooperative vessels use AIS properly. SAR imaging can detect ships, although there is obviously no identification. When using the two systems complementarily, an abnormal vessel can be identified if it is detected in a SAR image but it cannot be associated with a ship location or a ship track derived from AIS data.

In this literature survey, state-of-the-art algorithms are presented for the data association between different data sources in various application areas. The algorithms of interest are those that can be efficiently used to fuse ship detection results from SAR and AIS. The requirements of this study are also a timely response to Canada's Northern Strategy, since the country’s ability to exercise uninterrupted sovereignty over the Arctic region will largely impact its leadership in the economic and political realms for the years to come (Abielmona et al., 2014).

Although there exist various association algorithms in the literature, those specifically on AIS and SAR data are quite limited. The algorithms have been classified into different categories by researchers from different backgrounds. In a review paper of Abielmona et al. (2014) on SAR and AIS data fusion applications, the algorithms were grouped into three classes, computational intelligence (CI)-based, non-CI-based, and knowledge-based, which were used in five-level data fusion models. However, this complicated classification goes beyond the needs here and is not suitable as a guide for the current search of applied algorithms. This study therefore concentrates on the association algorithms in two main categories based on the types of objects to be associated, namely: Type 1: point to point; and Type 2: point to track. In both types, the association is from SAR to AIS, where the ships detected on SAR images can be represented as a point with or without attributes, and the ship positions derived from AIS can be represented either as independent points or a track (related point set), also with or without attributes.

A third type of algorithm also exists, which uses a track-to-track association approach. In this case, a time window needs to be allotted so that several ship locations on different SAR images can be used to construct a track. The output of the algorithm can answer questions such as whether the two tracks (one from SAR, one from AIS) are believed to belong to the same target (Liu & Shi, 2009; Suo & Liu, 2004). Since the ship track is not constructed by using just a single SAR image, these algorithms go beyond the simple data association problem and instead rely on tracking approaches that have to evaluate the reliability of previously acquired data and then update the most likely ship tracks with corresponding estimates of the uncertainty. This type of analysis is not considered further in this report.

1.1 Association Algorithm Type 1: Point to Point

In this class of algorithms, two point sets need to be constructed first, where one set includes the centroids of the detected ship pixels on the SAR image, and the other set contains the ship positions from AIS data. The points in each set are independent from each other. In the association process, the points' historical positions are not considered.

When projecting the AIS ship's position to a location on the SAR image for the association process, some algorithms compensate for the equivalent SAR Doppler displacement (Zhao, Ji, Zing & Zou, 2014). After position correction, the points in the two sets may present very similar locations, with only a small shift, as illustrated in Figure 1.

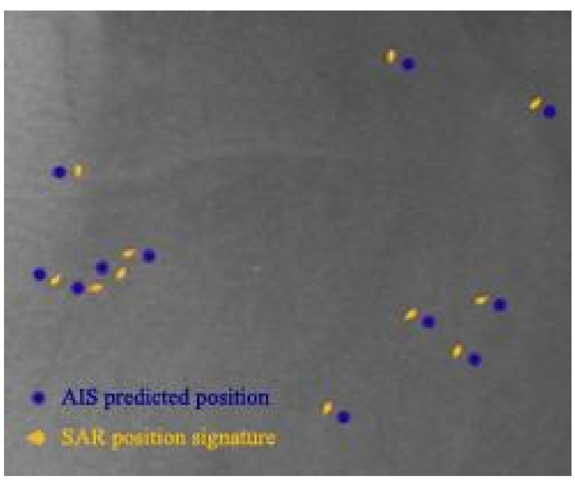

Figure 1. Example for ship positions detected from SAR image and predicted from AIS (adopted from Zhao et al., 2014).

However, the two point sets are not always related by a simple transformation, especially when the numbers of points in the two sets are not equal. The non-rigid Point Pattern Matching

(PPM) approach was adopted in Zhao et al. (2014) to solve this issue, with the non-linear geometrical transformation models used to solve the irregular distortion problem. The alignment of two point sets was considered as a probability density estimation problem. The Gaussian mixture model (GMM) centroids (representing the first point set) were fitted to the second point set by maximizing the likelihood. During the association process, the two point sets were forced to move coherently as a group to preserve the topological structure of the point sets.

The authors compared the association results of PPM and nearest neighbour (NN) for various conditions with simulated point sets. The tested conditions included low and high density shipping conditions, under- and over-estimated Doppler displacements, and with and without outliers (ships detected in SAR imagery with no AIS reports). The association results validated that the PPM method achieved better performance than the NN algorithm, especially under the high-density-shipping condition with outliers.

The authors also commented that SAR and AIS data association based on multiple feature information fusion could also benefit the results, where features such as size, speed and course could be jointly used to improve the accuracy of data association (Zhao et al., 2014).

SAR and AIS data were used in another point-to-point association algorithm (Ji, Zhang, Meng & Wang, 2014). In this paper, SAR and AIS data were acquired from the Bohai Bay area of China. Because of the busy maritime traffic in this area, the SAR and AIS point sets showed a large divergence in the distribution patterns, even when synchronized acquisitions of SAR and AIS were utilized. As shown in Figure 2, the SAR and AIS ship scenario contained more ships detected by SAR than reported by AIS so that there were a large number of ships with no AIS matching.

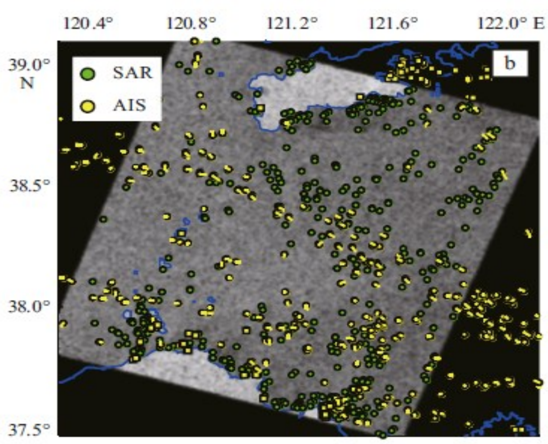

Figure 2. Example for RADARSAT-2 detected ship positions and AIS ship positions (adopted from Ji et al., 2014).

In the Ji et al. (2014) paper, the AIS velocity and heading information of the moving vessel target was used to calculate the ship's SAR Doppler offset. To simplify the point association, the SAR Doppler offset was applied to the reported AIS positions instead of the positions detected by the SAR. After the Doppler offset correction, the association algorithm conducted two-way searching. For point A in the first point set, the searching method tried to find its nearest point B in the second point set if the distance between them was within a distance threshold. For point B, the searching algorithm then tried to find its nearest point C in the first point set. After the two-way searching, if A and C were found to be the same point, the association between A and B was established, if A and C were not the same point, A was then considered as an outlier.

The algorithm is a two-way NN algorithm. Using the NN algorithm, good association can be obtained if the two point sets include similar points and the matched point pair are closest to each other. In ship associations using SAR and AIS data, this condition is most likely met when the two point sets are acquired at the same time or very close in time. This is also the limitation of the NN algorithm.

1.2 Association Algorithm Type 2: Point to Track

Another type of association is to assign a detected target (represented as a point) to a track. In SAR and AIS ship association, a detected ship in a SAR image is associated with a ship track derived from AIS data. In the end, if a ship cannot be assigned to any of the tracks, it should be marked as an outlier or an anomalous ship.

This is a typical multiple target tracking (MTT) problem involving multiple targets and multiple tracks, where each target can either belong to a single track or neither track. The generic MTT algorithm usually has two parts: an association algorithm and a tracking algorithm. The association method is the subject of interest in this report.

An association algorithm usually includes a cost function and an assignment method. The cost function is the metric to determine the likelihood of particular assignments, and the assignment method determines how to associate a target to a particular track with the requirement to minimize the overall cost. The literature publications on MTT are found mainly in military contexts. Several representative publications are discussed below.

In Uftring (2008), MTT algorithms were explored to support unknown target identification in applications such as anti-aircraft warfare and missile defense. Three common association methods were studied: nearest neighbour, global nearest neighbour and Joint Probabilistic Data Association (JPDA). The NN assignment here allowed a new measurement to be assigned to more than one track, while global nearest neighbour (GNN) assignment allowed each new measurement to be assigned to only one track. The statistical distance from the target to the track was used as a cost function for both of these NN methods.

The JPDA method handled the multi-target case based on their likelihood. It consists of two steps. The first is the calculation of the probabilities for the observation-to-track association, while the second phase is the hypothesis merging, where a weighted average of all the validated observations for a track is used to update its state.

In the ship association task, there is no need to update the track state as done in the second phase, however, the association probability can be used to make the target-to-track assignment based on the same principles as the GNN algorithms. Under these conditions, the combination of NN and JPDA methods, known as NNJPDA, was discussed in detail in Helmick (2000), and it was recommended and implemented as an Interacting Multiple Model (IMM)-JPDA, tracker association algorithm in DiFilippo, McAfee, Chen, Moore & Dawber (2006). The method was tested using simulation data and good results were achieved.

During the association process, the statistical distance from each target to a track can be obtained, so the statistical distances would be calculated from one target to the N tracks. If the statistical distance is less than the user-defined gate size, the target is said to be validated for that particular track. Usually, more than one observation is validated for a given track, or an observation is validated for more than one track, so that it is not clear which observation belongs to which track, especially when the number of targets is large. Searching the complete set of all possible associations for an optimal association means a heavy computation load. In Uftring (2008), the Munkres algorithm was recommended. The Munkres algorithm (also known as Hungarian algorithm) (Munkres, 1957) is an efficient algorithm to solve the assignment problem in polynomial time. The algorithm has many applications in combinatorial optimization. Comparing to other optimal searching methods such as Jonker-Volgenant-Castano (JVC) and Auction algorithms, Munkres is the easiest one to implement and optimality is guaranteed. In Uftring (2008), the use of statistical distance as the cost function with the GNN assignment via the Munkres algorithm worked well when both low and high level noise were added to the position measurements, based on results generated from simulated data.

It should be noted that the papers referred to in this section only studied the general tracking issue, and the methods were not specifically applied to AIS and SAR data association.

1.3 Conclusion

The state-of-the-art data association algorithms in the literature have been reviewed. In this report, we focus only on the ones that can be efficiently used to fuse ship detection results from SAR and AIS. Based on the types of objects to be associated, two types of algorithms are discussed: Type 1: point to point; and Type 2: point to track. In both types, the association is from SAR to AIS, where the ships detected in SAR imagery can be represented as a point, and the ship positions derived from AIS can be represented either as independent points or a track (related point set). The advantages and limitations of the reviewed algorithms are summarized below.

- NN is the most common association algorithm, which can be used in both types. It works well in the simple ship scenario case. However, it has limitations in the complex ship scenarios, such as when there are many outliers and a high ship density.
- Based on the results reported by the Zhao et al. (2014), PPM is a potential candidate for point-to-point association algorithms. It considers the divergence between the two point sets based on the probability density estimation, and searches the optimal matching between ship and AIS point sets by maximizing the likelihood.
- Among the point-to-track association algorithms, the combination of NN and JPDA is an attractive algorithm. It is similar to the PPM algorithm in the point-to-point association, where the association decision is based on a probability calculation. The method assesses the association probability for each point to all the tracks, and then a one-to-one association is made based on the highest association probability.
- However, no studies have been found to compare the association performance of the two algorithm types. The referred papers for the point-to point methods deal directly with AIS and SAR data association studies, whereas papers referred to for the point-to-track methods discussed the general tracking issue, and the methods were not specifically tested using AIS and SAR data.

2 AIS-SAR ASSOCIATION ALGORITHM EVALUATION

2.1 AIS-SAR ASSOCIATION ALGORITHM COMPARISONS

All the algorithms discussed in this report have the potential for the SAR-AIS association application in the RADARSAT Constellation Mission (RCM) context, where the best performance should be achieved under synchronized SAR-AIS acquisitions. When time differences exist between the SAR and AIS acquisitions, it is expected that the AIS ship positions would be propagated to the SAR acquisition time. A correction should also be applied for the SAR azimuth offset. The larger the time differences between the SAR and AIS data, the greater the uncertainty in the estimated positions.

In the low ship density areas and with limited outliers, the two-way NN should be a good candidate, which is easy to implement and can guarantee a one-one association result with high accuracy. As the time gap between AIS and SAR increases, the performance of NN-type algorithms may degrade dramatically, especially in the high ship density areas with a large number of outliers. In this case, the distance of a given ship between the SAR and AIS detections may change significantly, so that closest distance may not mean the same ship anymore. In this case, the more complicated non-rigid PPM algorithm should be considered. This algorithm considers the two point sets detected from SAR or AIS as two correlated topological structures that are distorted relative to each other. The distortion can be introduced by the uncertainty in the ship movements during the SAR and AIS time gap and by any outliers (i.e. points without a matching point in the SAR or AIS sets) in each point set. During the association process, the two point sets are forced to move coherently as a group to preserve the associated topological shape. In Zhao et al. (2014), the PPM surpassed the NN-type algorithms in the high ship density scenarios with some outliers in the SAR detections. However, more specific investigations on either actual or simulated SAR and AIS scenarios would be required to quantify the degree of distortion that the PPM algorithm can handle. A summary of the point-to-point algorithms is given in Table 1.

Table 1. SAR-AIS point-to-point association algorithms

Point to point (PP)	**Features**
Nearest Neighbour (NN)	• Easy to implement • Works well in simple scenarios: low-density-shipping area with few outliers, synchronized AIS and SAR acquisitions • One-multiple association
Two-Way Nearest Neighbour (TW-NN)	• Easy to implement • Simple relationships between SAR and AIS point sets are considered. • Works well in simple scenarios: low-density-shipping area, limited outliers, synchronized SAR and AIS acquisitions • One-one association
Point Pattern Matching (PPM)	• SAR and AIS point sets are considered as correlated topological structures, but distorted from each other • Works better in complex scenarios: high-density shipping area with some outliers and time gaps between SAR and AIS acquisitions • One-one association

Compared to the point-to-point algorithms, the point-to-track algorithms are more complicated to implement since an additional tracker algorithm needs to be included. The tracker algorithm estimates a ship's course based on the relationship derived from its available AIS detected positions. With this functionality, the ship movement during the time gap between SAR and AIS acquisitions can be estimated more accurately, thereby providing better information for the association process. In this respect, the overall performance of the point-to-track algorithms should be better than the point-to-point algorithms given the same ship density scenarios.

However, depending on the specific algorithm, the performance of the point-to-track approach may not be as robust as the point-to-point algorithm in areas of high ship density, and where the reported AIS positions are not adequate enough to derive reasonable estimates for the ship tracks, such that the statistical distance from the point to the track and the predicted ship positions cannot be determined correctly. This problem may typically occur in the high ship density areas when the AIS data are received from satellite-based receivers only. In high ship density areas, satellite AIS message reception is limited by the message collisions that occur at the receiver. Given the large field of view of satellite AIS, any areas that overlap high ship densities will particularly suffer from lost messages. Thus, AIS updates during a given pass may be infrequent, if they occur at all. There may be the need to keep track of AIS positions from previous RCM passes, which will further increase the time gaps.

The GNN and NNJPDA can both generate one-one associations. The GNN assigns a point to a track based on the statistical distance from the point to the track, and NNJPDA assigns a point to a track based on the Gaussian likelihood of the association. The NN-type algorithm is very simple, and does not perform well for multiple targets that are closely spaced together. The NNJPDA improved upon NN-type approaches in studies of attacking target tracking (Uftring, 2008; DiFilippo et al., 2006), however no direct studies using SAR and AIS ship detection data have been found in the literature. A summary of the point-to-track algorithms is given in Table 2.

Table 2. SAR-AIS point-to-track association algorithms

Point to Track (PT)	Features
Nearest Neighbour (NN)	• Works well in simple scenarios: low-density-shipping area with ships not closely spaced together • One-multiple association
Global Nearest Neighbour (GNN)	• Works well in simple scenarios: low-density-shipping area with ships not closely spaced together • One-one association
Nearest-Neighbour Joint Probabilistic Data Association (NNJPDA)	• One-one association • The associations between one point and every track are evaluated based on probability calculation

2.2 AIS-SAR Association Algorithm Implementation

The application of the association algorithms to the AIS-SAR problem may be evaluated in terms of the maturity of the algorithm along with its previous use in the AIS-SAR association specifically. The Technology Readiness Levels (TRL), as defined to assist in evaluating the maturity of goods and/or services for the Build in Canada Innovation Program (TRLs - Technology Readiness Levels, 2015), among other programs, is one measure of the maturity of the technology. A summary of the TRLs is given in Table 4. It should, however, be recognized that not all development cycles will necessarily follow these levels. The estimates of the TRLs for the association algorithms considered in this report are given in Table 4. Since these algorithms are not specific to the AIS-SAR association problem, Table 4 includes the TRL for both the algorithm development as well as its specific application to AIS-SAR association. The TRLs listed here are based on the available literature, or other known applications, and therefore may not represent the level of development of proprietary or unpublished implementations.

Also listed in Table 4 are the known intellectual property (IP) rights and the likely development effort required to produce working AIS-SAR association software.

Table 3. Definitions of TRL's

TRL Level	Definition
1	Basic principles of concept are observed and reported.
2	Technology concept and/or application formulated.
3	Analytical and experimental critical function and/or proof of concept.
4	Component and/or validation in a laboratory environment.
5	Component and/or validation in a simulated environment.
6	System/subsystem model or prototype demonstration in a simulated environment.
7	Prototype ready for demonstration in an appropriate operational environment.
8	Actual technology completed and qualified through tests and demonstrations.
9	Actual technology proven through successful deployment in an operational setting.

Table 4. Estimated TRL's for the AIS-SAR association algorithms considered

<table>
<tr><th></th><th>Algorithm</th><th>TRL General Application</th><th>TRL AIS-SAR Application</th><th>IP Rights</th><th>Development Effort</th></tr>
<tr><td rowspan="3">Point to Point</td><td>NN</td><td>9</td><td>9</td><td>Open source</td><td rowspan="3">1. Simulated or real SAR and AIS detection data with various time differences and ship densities need to be prepared as test data.
2. Algorithms can be initially implemented and tested in the MATLAB environment.
3. For real SAR and AIS data, synchronized ground truth information on the identifications of the SAR detected ships are necessary to calculate the association accuracy, which will usually require propagation of the AIS position to the time of the SAR acquisition.
4. PPM is more complicated in the aspects of concept and coding than NN and TW-NN algorithm.</td></tr>
<tr><td>TW-NN</td><td>9</td><td>6</td><td>Open Source</td></tr>
<tr><td>PPM</td><td>6</td><td>6</td><td>Open source</td></tr>
<tr><td rowspan="3">Point to Track</td><td>NN</td><td>9</td><td>Unknown</td><td>Open source</td><td rowspan="3">1. Simulated or real SAR and AIS detection data with various time differences and ship densities need to be prepared as test data.
2. Algorithms can be initially implemented, tested and integrated in the MATLAB environment.
3. For real SAR and AIS data, synchronized ground truth information on the identifications of the SAR detected ships are necessary to calculate the association accuracy, which will usually require propagation of the AIS position to the time of the SAR acquisition.
4. An independent tracker algorithm needs to be implemented to build ship tracks using AIS Data.
5. NNJPDA is more complicated in the aspects of concept and coding than GNN.</td></tr>
<tr><td>GNN</td><td>6</td><td>Unknown</td><td>(available to DRDC)</td></tr>
<tr><td>NNJPDA</td><td>6</td><td>Unknown</td><td>(available to DRDC)</td></tr>
</table>

2.3 AIS-SAR Association Accuracy

The application of AIS-SAR association algorithms would be best tested on either actual data or on simulated data derived from known distributions of shipping behavior. The success of a particular association algorithm will depend on the density of shipping within a given area and the time gap between the acquisitions of the AIS data and the SAR image. While these factors influence the difficulty of the association problem, the actual performance of any given algorithm should be confirmed under the scenarios of interest.

Off the East and West Coasts of Canada, the average shipping density varies from less than 1 ship per degree square to around 10 ships per degree square (degree of latitude by degree of longitude). Notable exceptions occur around larger ports, such as Victoria and Vancouver along with the Juan de Fuca and Georgia Straits, where the densities are around 130 and 225 ships per degree square, respectively. In general, on the east and west coasts of North America, the ship densities vary up to 300 to 400 ships per degree square, with a maximum of 1000 ships per degree square. For Europe, the densities vary up to about 500 ships per degree square, with a few values in the 1000 to 3000 ships per degree square range, while for Asia they vary up to 500 to 800 ships per degree square, with several values in the 1000 to 4000 ships per degree square range (C-CORE, 2012).

The higher ship densities generally occur along coastal shipping lanes and around major ports. In such areas, satellite AIS information may be limited due to message collisions. The use of ground-based AIS reception stations would therefore greatly simplify the association requirements of any algorithm, since more complete and timely AIS information would be available. The reliability could be further enhanced by using tracking algorithms to maintain and update known positional information, and thereby reducing the burden for any association function.

The accuracy of simple AIS and SAR association algorithms may be estimated from the density of shipping within a given area and the time difference between the AIS and SAR acquisitions. The uncertainty related to the dead reckoning of AIS positions over specified time gaps has been previously estimated (C-CORE, 2014), based on tracks obtained from AIS position, speed over ground and course over ground data. The procedure is outlined below.

- Obtain position reports for each ship track. (Here, a ship track is a continuous set of AIS messages from a ship where the time between each message was less than seven hours, based on the maximum dead-reckoning time considered of six hours).
- Calculate the dead-reckoned ship position for the desired time intervals for each ship position report.
- Interpolate the actual ship positions (from AIS data) at the dead-reckoning times for each ship position.
- Calculate the distance (deviation) between the dead-reckoned ship position and the interpolated point on the ship track for each dead-reckoning time.
- Bin the deviations on a 1 degree latitude and longitude grid.
- For each latitude and longitude grid cell, calculate the cumulative distribution function of the deviations for each dead-reckoned time.
- From the cumulative distribution function, determine the deviation at the desired confidence level.

The deviations, at a 90% confidence level, for a subset of the dead-reckoning times is given in Table 5 (C-CORE, 2014). Included in the table are the results for three regions, based on global

AIS data, and 3° latitude by 3° longitude grids of the East and West Coasts of Canada, centered on 47.5° N, 49.5° W and 51.5° N, 134.5° W, respectively.

Table 5. Deviations for the dead-reckoning times with a 90% confidence interval

Dead Reckoning Time (min)	Deviations (km)		
	Global	Canada East Coast (3x3 grid)	Canada West Coast (3x3 grid)
5	0.3	0.3	0.1
10	0.7	0.8	0.3
15	1.3	1.4	0.5
30	3.2	3.3	1.1
60	7.4	7.1	2.2
120	18.	15.	5.5

Assuming that the ships in the area of interest are distributed uniformly, then the probability that a ship may be within the circular area defined by the radius of the dead-reckoning deviation is simply the area of the deviation circle, at the specified confidence level, divided by the area of interest, specifically

$$p = \frac{\text{Area of deviation}}{\text{Area of interest}} .$$

Confusion in the association of AIS and SAR targets will occur if there is more than one target within the deviation circle. The probability of having k ships in a deviation circle is given by the binomial distribution as

$$P(k; N, p) = \binom{N}{k} p^k (1-p)^{N-k} ,$$

where N is the total number of ships in the area of interest. In the first approximation, one may assume that the probability of association is equally likely among any ships within the deviation circle, that is, no further consideration is given to the closest targets or the overall topology of the targets. In this case, the probability of correct association is then

$$P(\text{correct association}) = \sum_{m=1}^{N} \frac{1}{m} P'(m; N, p) ,$$

where P' is the conditional probability of there being m ships in the deviation area, given that there is at least the one ship under consideration there. The latter conditional probability is

$$P'(m;N,p) \equiv P(K=m;N,p|K \geq 1)$$

with

$$P(K=m;N,p|K>0) = \frac{P(m;N,p)}{1-P(0;N,p)}.$$

Using these equations, one may obtain an estimate of the probability of correct association based solely on the ship density, the time difference between the AIS and SAR acquisitions, and the uncertainty in the dead-reckoning of the AIS data.

The accuracy of AIS and SAR association could potentially be improved by considering additional ship characteristics. The heading of any SAR target that is elongated could be compared against the AIS-specified ship course over ground, albeit with a likely ambiguity of 180°. However, in many cases along standard shipping routes, the variations in the ship courses may not be sufficient to provide useful discrimination.

More promising is the comparison between ship lengths as obtained through AIS information and from the SAR signatures. For SAR resolutions of around 50 m or better, one might expect to be able to distinguish ships on roughly the same length scale, although the correspondence will not necessarily be one-to-one, due to SAR imaging effects. From the distribution of ship lengths for a given area of interest, one may define the probability that a ship is within a certain length class A_i as $P(A_i)$. The probability of distinguishing two ships within different length classes is then

$$P\left(A_i \& A_j, i \neq j\right) = \sum_{i=1}^{L} P(A_i)\left(1-P(A_i)\right),$$

where there are a total of L length classes. More generally, for m ships within an area of uncertainty, the probability of having a unique length class, that is, $P_m(\text{different})$, is

$$P_m(\text{different}) \equiv P_m\left(A_i \& A_j, i \neq j, \text{for } (m-1)\; j\right) = \sum_{i=1}^{L} P(A_i)\left(1-P(A_i)\right)^{m-1}.$$

Thus, if the distribution of ship lengths is included, then the probability of correct association becomes

$$P(\text{correct association}) = \sum_{m=1}^{N} \{P_m(\text{different}) + \frac{1}{m}(1-P_m(\text{different}))\} P'(m;N,p).$$

The length parameters have also been used to estimate the radar cross section of ships (Vachon, English & Wolfe, 2007), although at this time there has only been limited use of the radar cross section to identify certain classes of ships (e.g. Vachon, Kabatoff & Quinn, 2014).

For purpose of illustration, the potential overlap or confusion of separate AIS and SAR targets was considered for a few specific cases. The first example was taken from the area just to the west of the Cabot Strait at the entrance to the Gulf of St. Lawrence, within the region defined by 47° to 48°N and 60° to 61°W. Within this area, the AIS records indicate that there is an average of 6.3 ships, which have a corresponding speed over ground distribution as shown in Figure 3. The ship course over ground is shown in Figure 4, from which it is evident that the predominant ship courses are at 122° / 302° and 150° / 330°, thereby limiting the discrimination potential. The ship length distribution is shown in Figure 5, and illustrates that the ship lengths are spread throughout the range of lengths, with obvious peaks for particular length classes. The ship length would therefore appear to be a good discriminator among various ships, although the relation between the actual ship length and the ship length as determined from its SAR signature would need to be considered, as well as the associated uncertainty in the estimated ship length.

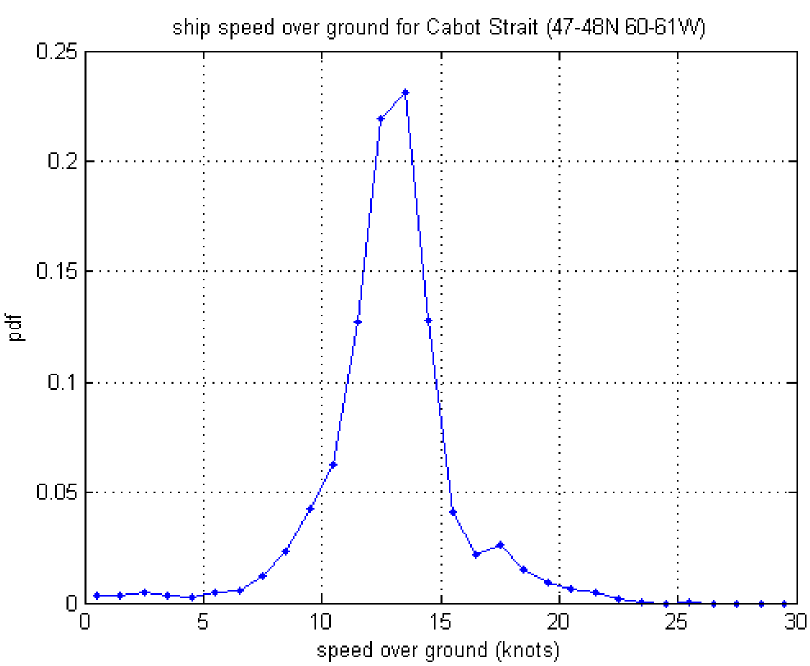

Figure 3. Probability density function of AIS transmitted speed over ground for ships near the west side of the Cabot Strait.

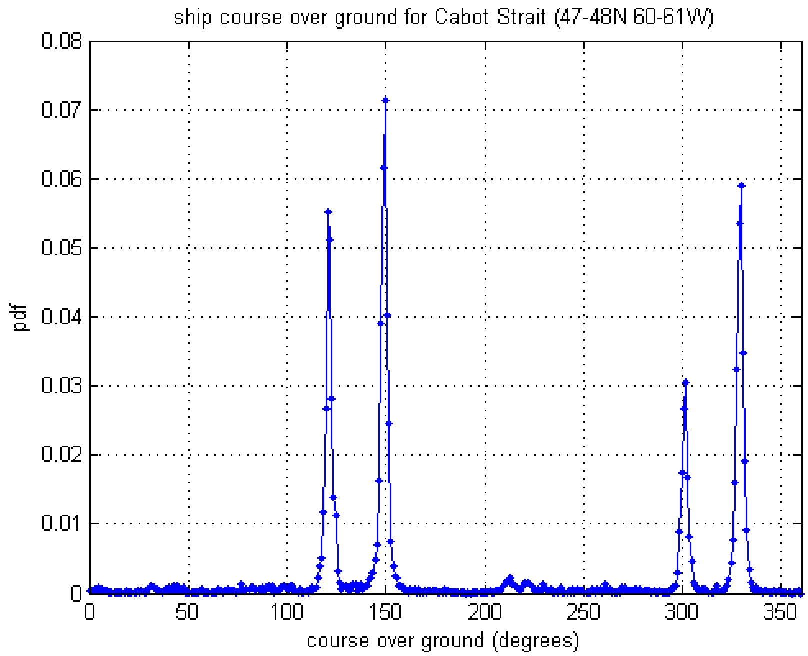

Figure 4. Probability density function of AIS transmitted course over ground for ships near the west side of the Cabot Strait.

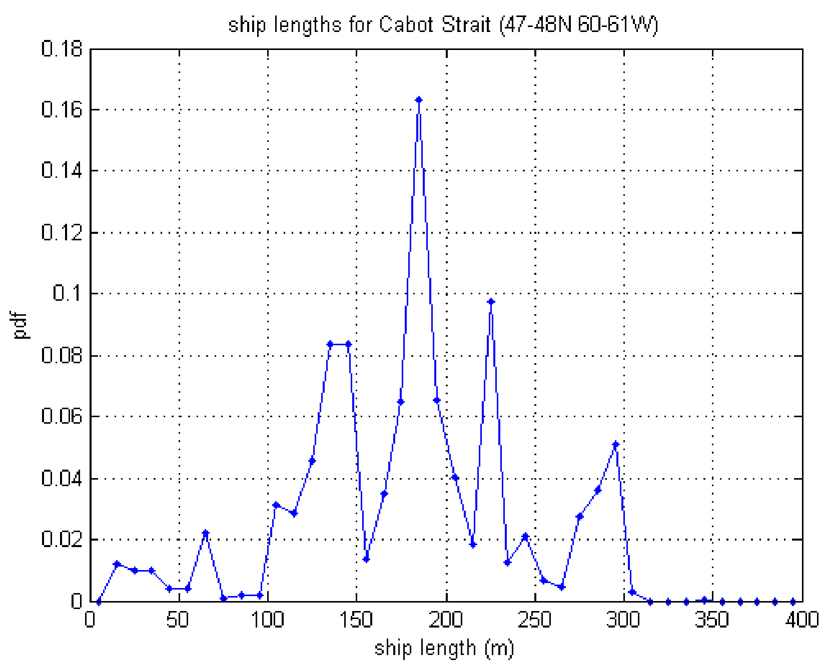

Figure 5. Probability density function of AIS transmitted ship lengths for ships near the west side of the Cabot Strait.

The probability of correct association for ships within this region can be estimated from the above equations, using the dead reckoning deviations based on global AIS data as given in Table 5, and copied in to the first two columns of Table 6. The resulting probabilities for the time differences from 5 minutes to 2 hours are listed in the remaining columns of Table 6. The probabilities are, respectively, the probability of finding one ship with an area equal to the uncertainty of the dead-reckoning deviation, the probability of correct association based on the ship density and the time differences, and the probability of correct association when including additional discrimination based on the ship lengths.

Table 6. Probability of correct association for a region of the Cabot Strait

Time Difference (minutes)	Deviation (km)	*p*(ship within deviation circle)	*P*(correct association)	*P*(correct association) (using ship length)
5	0.3	0.00003	1.00	1.00
10	0.7	0.00018	1.00	1.00
15	1.3	0.00063	1.00	1.00
30	3.2	0.0039	1.00	1.00
60	7.4	0.021	0.97	0.99
120	18	0.12	0.84	0.96

At higher ship densities, the time gap between AIS and SAR acquisitions must be reduced accordingly to avoid potential uncertainty in the AIS and SAR associations. In the Juan de Fuca Strait, for instance, the average ship density recorded by AIS is 25 ships within the degree square from 48° to 49°N and 124° to 125°W. The distributions for the speed over ground and the course over ground are given in Figure 6 and Figure 7, from which it is evident that of the ~80% of ships that are making way, the average speed is 12 knots and the predominant courses are 115° / 295° and 90° / 270°. The ship lengths are shown in Figure 8, and as above, they are likely to aid in the discrimination.

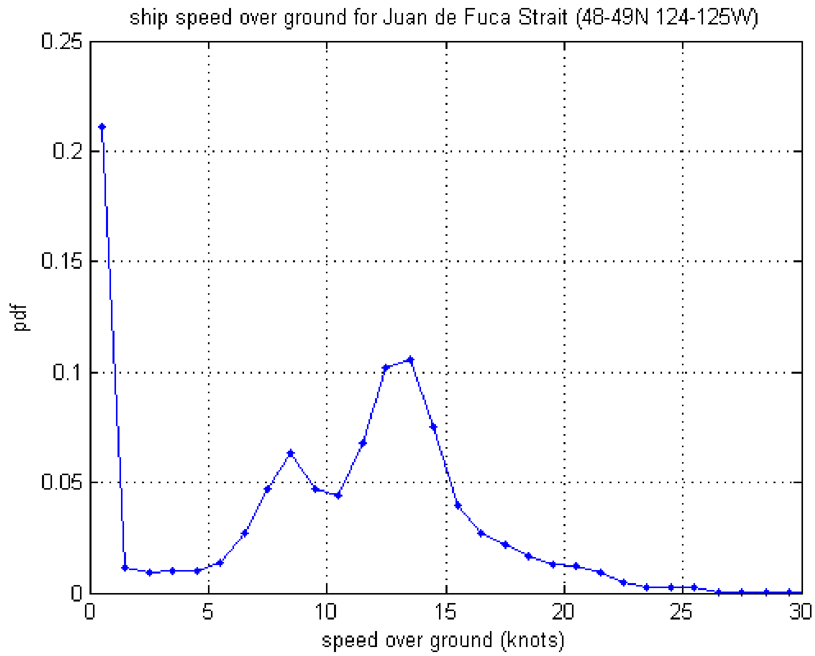

Figure 6. Probability density function of AIS transmitted speed over ground for ships near the entrance to the Juan de Fuca Strait.

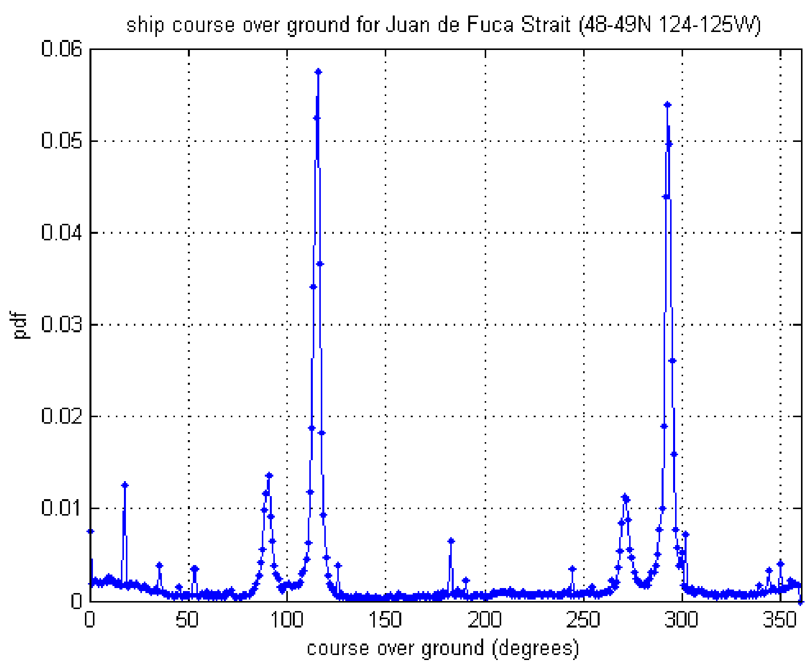

Figure 7. Probability density function of AIS transmitted course over ground for ships near the entrance to the Juan de Fuca Strait.

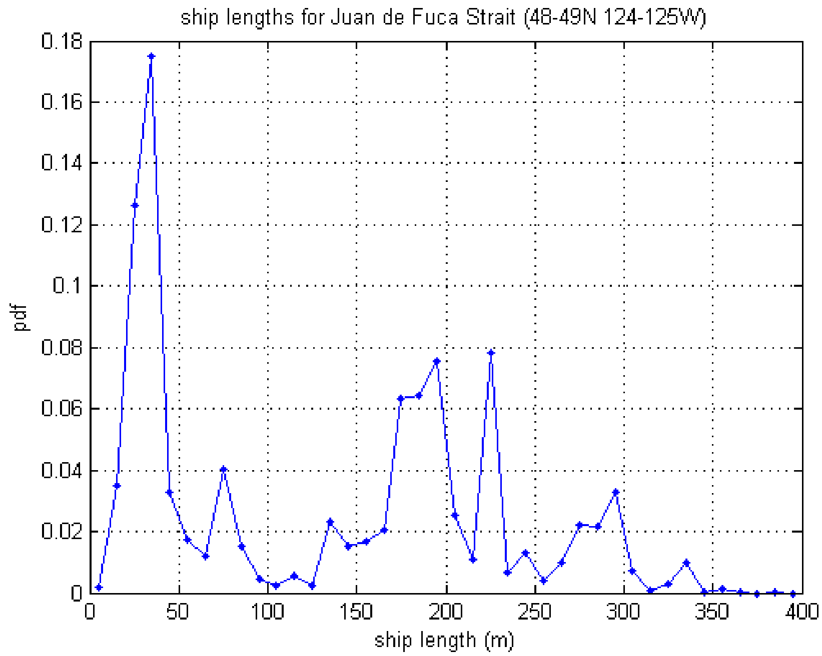

Figure 8. Probability density function of AIS transmitted ship lengths for ships near the entrance to the Juan de Fuca Strait.

The probabilities of correct association for ships within this region are estimated as above, and are listed in Table 7. Specifically, these are the probability of finding one ship with an area equal to the uncertainty of the dead-reckoning deviation, the probability of correct association based on the ship density and the time differences, and the probability of correct association when including additional discrimination based on the ship lengths.

Table 7. Probability of association for part of the Juan de Fuca Strait

Time Difference (minutes)	Deviation (km)	*p*(ship within deviation circle)	*P*(correct association)	*P*(correct association) (using ship length)
5	0.3	0.0001	1.00	1.00
10	0.7	0.0005	1.00	1.00
15	1.3	0.0019	0.99	1.00
30	3.2	0.011	0.93	0.98
60	7.4	0.061	0.66	0.88
120	18	0.36	0.12	0.32

From the above discussion, it is evident that the performance of any AIS-SAR association algorithm will depend significantly on the scenarios under consideration. The actual ship density in specific areas or along specific shipping lanes and the time differences between the AIS and SAR acquisitions are the main factors determining the demands on an association algorithm. However, additional information such as the use of ground-based AIS receiving stations, target tracking algorithms, and ship-length correlation can all reduce the burden and increase the likely accuracy of the association.

While the general behavior of specific point-to-point and point-to-track association algorithms has been given above, further evaluation of the algorithms for specific areas of interest would best be accomplished using actual shipping data.

2.4 RECOMMENDATIONS

The NN association algorithm is the simplest and the most computationally efficient, and has the highest TRL, especially in terms of its previous application and evaluation with respect to the AIS-SAR problem within the publically available literature. The assignment problem within the NN association is usually solved with the Munkres algorithm.

The confidence in the association may be estimated from the time difference between the AIS and SAR acquisitions and the number of targets under consideration in the region of interest. The uncertainty may be based solely on the propagation of the AIS position to the SAR acquisition time, or it may also include previously observed ship behaviour to aid in refining the uncertainties. The associations could also be further refined by including the correlation of ship lengths as obtained from the AIS messages and the SAR images.

Implementations of the NN algorithm are available through open sources. A MATLAB implementation would enable an evaluation of the algorithm for the scenarios of interest. Assuming the use of basic open source code, a MATLAB implementation could likely be developed and tested within about four weeks of effort. The algorithm could then be ported to a C++ implementation for operational software, which would integrate with existing image analysis and display software. Again, assuming the use of open source code from recognized sources, the C++ implementation could take from four to eight weeks of effort.

Other solutions would, in general, require additional time for development and evaluation. If desired, these could be developed and evaluated incrementally to the NN algorithm.

3 REFERENCES

Abielmona, R., Falcon, R., Vachon, P.W., & Groza, V. Z. (2014). Vessel Tracking and Anomaly Detection Using Level 0/1 and High-Level Information Fusion Techniques. *6th International Workshop on Soft Computing Applications*, Timisoara, Romania.

C-CORE. (2013). Satellite Automatic Identification System (SAIS) Performance Modelling and Simulation: Final Findings Report, DRDC Ottawa CR 2013-096, Defence R&D Canada - Ottawa.

C-CORE. (2014). AIS Dead Reckoning Deviations, C-CORE Report R-14-036-1160, Revision 1.0, July 2014. DRDC Contract Report DRDC-RDDC-2014-C177.

DiFilippo, D., McAfee, E., Chen, R., Moore, A., & Dawber, B. (2006). A Multifunction Radar Simulation for Adaptive Radar Control Studies. *TTCP Technical Report*, November, 2006.

Helmick, R. (2000). IMM Estimator with Nearest-Neighbor Joint Probabilistic Data Association. In Y. Bar-Shalom, & W.D. Blair (Eds.), *Multitarget-Multisensor Tracking: Applications and Advances*, Vol. 3, Ch. 3, Norwood, MA: Artech House.

Ji, Y., Zhang, J., Meng, J., & Wang, Y. (2014). Point Association Analysis of Vessel Target Detection with SAR, HFSWR and AIS. *Acta Oceanologica Sinica*, 33(9), 73-81.

Liu, C. & Shi, X. (2009). Study of Data Fusion of AIS and Radar. *Proceedings of 2009 International Conference of Soft Computing and Pattern Recognition*, 674-677.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1), 32-38.

Suo, J. & Liu, X. (2004). Fusion of Radar and AIS Data. Proceedings of 7th International Conference on Signal Processing, 2604-2607.

TRLs - Technology Readiness Levels. (2015). Public Works and Government Services Canada Document, Cited from: https://buyandsell.gc.ca/initiatives-and-programs/build-in-canada-innovation-program-bcip/program-specifics/technology-readiness-levels.

Uftring, L. (2008). A Multiple Target Range and Range-Rate Tracker Using and Extended Kalman Filter and a Multilayered Association Scheme. A Master of Science thesis submitted to Electrical Engineering Department, Tufts University, May 2008. Cited from: http://www.ece.tufts.edu/~elmiller/laisr/pdfs/ulftring_ms.pdf, Jan. 2015.

Vachon, P.W., English, R.A. & Wolfe, J. (2007). Ship Signatures in Synthetic Aperture Radar Imagery. *Geoscience and Remote Sensing Symposium*, IGARSS 2007. IEEE International. 1393-1396.

Vachon, P.W., Kabatoff, C. & Quinn, R. (2014). Operational Ship Detection in Canada Using RADARSAT. *Geoscience and Remote Sensing Symposium*, IGARSS 2014. IEEE International.

Zhao, Z., Ji, K.F., Xing, X.W., & Zou, H.X. (2014). Effective Association of SAR and AIS Data Using Non-Rigid Point Pattern Match. *IOP Conference Series: Earth and Environmental Science*, 17 012258.

LAST PAGE OF DOCUMENT