

# Behavioral Detection in the Maritime Domain

James W. Scrofani, Murali Tummala, Donna Miller, Deborah Shifflett, John C. McEachen

Electrical and Computer Engineering Department

Naval Postgraduate School

Monterey, CA, U.S.A.

[jwscrofa@nps.edu](mailto:jwscrofa@nps.edu), [tummala@nps.edu](mailto:tummala@nps.edu)

**Abstract** - *The maritime domain is important to the security, prosperity and vital interests of the global community. In order to protect these interests, governments require capabilities that provide situational awareness of the maritime domain. In [11] a spatiotemporal analysis approach is proposed that autonomously analyzes and classifies ship movement and possible intent at sea. The analysis focuses on detection of vessels of interest that exhibit one behavior, paralleling or following behavior. In this paper, we extend this approach by proposing a generalized semantic method that enables consideration of other behaviors of interest. Additionally we conduct a series of simulations using simulated and real AIS data to assess the performance of the algorithm to variation in behavior thresholds.*

**Keywords:** High-level data fusion; pattern detection; behavior detection; clustering, spatiotemporal analysis; maritime domain awareness; activity-based intelligence.

## 1 Introduction

The maritime domain is important to the security, prosperity and vital interests of the global community. It supports 90 percent of the world's trade and two-thirds of its petroleum are transported by sea [1]. In order to protect these vital national interests, governments require capabilities that provide situational awareness of the maritime domain, in other words, maritime domain awareness (MDA). MDA is the effective understanding of any activity associated with the maritime environment that could impact upon the security, safety, economy or environment [2].

Important contributions toward greater MDA have been realized by the information fusion community [3-8], but higher level fusion solutions remain elusive. Discovery and understanding of patterns, behaviors, and intentions of vessels of interest, and by extension, their command and control, is crucial to decision makers and maritime operators and analysts in developing courses of action or implementing strategic or tactical operations at sea.

Further, advances in this area will provide important contributions to the emerging area of activity-based intelligence (ABI) [9], a discipline of intelligence where the analysis and subsequent collection is focused on the

activity and transactions associated with an entity, a population, or an area of interest [10].

In [11], a spatiotemporal analysis approach is proposed that autonomously analyzes and classifies ship movement and possible intent at sea. The analysis focuses on detection of vessels of interest that exhibit one behavior, paralleling or following behavior. In this paper, we extend this approach by proposing a generalized semantic method that enables consideration of other behaviors of interest. Additionally we conduct a series of simulations using simulated and real AIS data to assess the performance of the algorithm to variation in behavior thresholds.

## 2 Methodology

### 2.1 Overview

Our approach, detailed in [11], can be viewed in terms of three major operations, localization of vessels; detection and identification of vessel behaviors of interest, and measurement of temporal congruence or persistence of vessel behaviors of interest over time. Localization partitions the input spatiotemporal database  $D$ , into sets or clusters of vessels within a region of interest that exhibit geographical similarity. Behavioral templating uses semantic expressions that are encoded to detect and identify vessel behaviors that are of interest. Temporal congruence measures the persistence of a particular behavior of interest over time and makes a declaration of behavior if the behavior is sufficiently persistent.

### 2.2 Spatiotemporal Database

Our method incorporates automated identification system (AIS) position reports that provide a temporal history of vessel position, course over ground, speed over ground and other vessel attributes. AIS was of particular interest for this analysis, but the approach is extensible to other sources of vessel information, including global positioning systems (GPS), maritime navigational radar (MNR) systems, and intelligence, surveillance, and reconnaissance (ISR) systems. The underlying requirement is to acquire spatial data and other vessel information of interest registered to temporal events, like sensor measurements. Fig. 1 portrays the concept of spatiotemporal data in a maritime setting. This can be

viewed as a time series representation of vessel position and other companion information (e.g., speed over ground, course over ground, etc.). The input spatiotemporal database used in this analysis is referred to as  $\mathbf{D}$ .

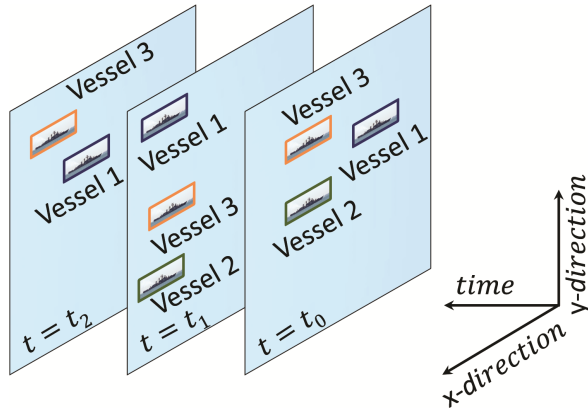


Figure 1. Spatiotemporal data is comprised of spatial information sampled in time.

## 2.3 Preprocessing

Preprocessing involves accepting data in the sensor output format and filtering and preparing that data for further processing. The available archive of data is filtered first by timestamp to a time period of interest and then by latitude and longitude to a geographic area of interest. Data are then further filtered, as necessary, for other details, such as vessel type or flag of origin, depending on user preferences.

## 2.4 Localization

Localization partitions the spatiotemporal database into clusters of vessels that exhibit geographical similarity.

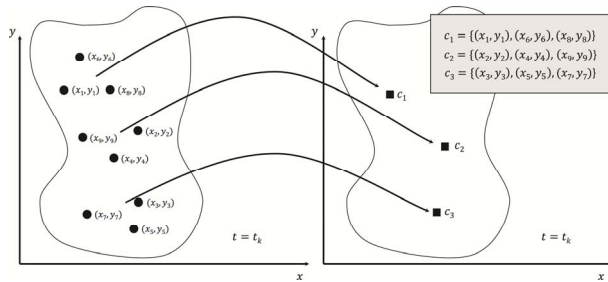


Figure 2. Localization of vessel positions to a set of representative clusters.

Given a set of vessel position reports, at a particular time  $t$ ,

$$\mathbf{x}_j \in \mathbb{R}^2, j = 1, \dots, N, \quad (1)$$

where  $\mathbf{x}_j$  contains position coordinate values  $(x_j, y_j)$  and  $N$  is the number of vessel reports in the database  $\mathbf{D}$  at time  $t$ , we

partition the data into  $k$  clusters, using the k-means clustering algorithm [12]. Fig. 2 illustrates the localization concept with a small number of vessels. Nine geographically disparate vessels represented by x-y coordinates are mapped to three representative clusters. Localization is performed at each time instance in the time period of interest and the resultant information is stored as a cell array

$$\xi^t = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k], \quad (2)$$

where  $t$  represents the current time and  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k$  are  $k$  column vectors that contain vessel identities localized to a particular cluster. A cell array construct is required as the columns of  $\xi^t$  may be of different lengths.

## 2.5 Behavioral Templating

Using a propositional logic-based approach, behavioral templating is used to algorithmically express sets of behaviors that are of interest and detect these behaviors from within the input database,  $\mathbf{D}$ . These behaviors are defined by the user and are multivariate, Boolean-valued functions of vessel attributes. For example, if an analyst is interested in understanding if particular vessels are under surveillance, an expression like

$$(s(i, j) < T_s) \wedge (c(i, j) < T_c) \wedge (d(i, j) < T_d) \rightarrow \gamma, \quad (3)$$

can be defined, where  $s(i, j)$ ,  $c(i, j)$ , and  $d(i, j)$  are measures of speed, course and distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  vessels in  $\mathbf{D}$ , respectively;  $T_s$ ,  $T_c$ , and  $T_d$  are thresholds of speed, course and distance, respectively; and  $\gamma$  is a binary variable representing the evaluated condition of the expression, i.e., TRUE or FALSE. This semantic expression reflects a query concerning vessels in an area of interest that are sufficiently close to other vessels and are transiting at similar speeds and courses, in other words, a possible surveillance condition.

Table 1. Behaviors of interest and their corresponding semantic expressions.

Behavior type	Natural language description	Semantic expression
Surveillance	vessel of interest (VOI) and target vessel (TV) transit at similar speed and similar course, but at a stand-off distance appropriate for visual	$(s(i, j) < T_s) \wedge (c(i, j) < T_c) \wedge (d(i, j) < T_d) \rightarrow$

	or sensor surveillance	
Harassment	VOI and TV transit at similar speed and similar course, but at a stand-off distance appropriate for weapons delivery or harassment.	$(s(i,j) < T_s) \wedge (c(i,j) < T_c) \wedge (d(i,j) < T_d) \rightarrow$
Rendezvous	VOI and TV rendezvous. Examine projection of VOI and target positions into the future and confirm intersection in time and place.	$(d(i,j) < T_d) \rightarrow \gamma, \text{ for } t > t_0$

Given a semantic expressions of interest, behavioral analysis is performed for each cluster contained in  $\xi^t$  for each time represented in  $\mathbf{D}$ . This analysis is implemented through employment of similarity measures. For the case of relative distance, the Euclidian distance is used. For relative speed and course, the  $L_1$  distance is used. Fig. 3 illustrates the application of behavioral analysis for a particular cluster,  $c_1$ , comprised of vessels  $i, j$  and  $k$  at time  $t = t_0$ . Appropriate distance measures are computed and attributes are compared to predefined thresholds. Behavioral templating is performed at each time instance in the time period of interest and the resultant information is represented as a cell array

$$\begin{array}{l}
 c_1 \\
 \left[ \begin{array}{l} \text{Vessel } i \\ \text{Vessel } j \\ \text{Vessel } k \end{array} \right] \\
 t = t_0
 \end{array}
 \begin{array}{l}
 \rightarrow \begin{bmatrix} d(i,i) & d(i,j) & d(i,k) \\ d(j,i) & d(j,j) & d(j,k) \\ d(k,i) & d(k,j) & d(k,k) \end{bmatrix} - T_d \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} < \epsilon_1 \\
 \rightarrow \begin{bmatrix} \Delta s(i,i) & \Delta s(i,j) & \Delta s(i,k) \\ \Delta s(j,i) & \Delta s(j,j) & \Delta s(j,k) \\ \Delta s(k,i) & \Delta s(k,j) & \Delta s(k,k) \end{bmatrix} - T_s \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} < \epsilon_2 \\
 \rightarrow \begin{bmatrix} \Delta attr(i,i) & \Delta attr(i,j) & \Delta attr(i,k) \\ \Delta attr(j,i) & \Delta attr(j,j) & \Delta attr(j,k) \\ \Delta attr(k,i) & \Delta attr(k,j) & \Delta attr(k,k) \end{bmatrix} - T_{attr} \times \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} < \epsilon_3
 \end{array}$$

Figure 3. Behavioral templating using semantic expressions to detect behaviors of interest.

$$\beta^t = [\phi_1, \phi_2, \dots, \phi_k], \quad (4)$$

where  $t$  represents the current time and  $\phi_1, \phi_2, \dots, \phi_k$  are column vectors that contain the identities of vessels that behaviorally conform to the criteria expressed by the threshold variables for a particular area of localization and time.

## 2.6 Temporal Congruence Analysis

Temporal congruence analysis (TCA) assesses the degree to which detected behavior persists over time. This is accomplished by evaluating the vectors in  $\beta^t$  over consecutive time steps. If vessel membership for a particular vector remains the same over the time of regard, then the behavior is said to persistent over this period of time. Membership is evaluated by comparing a particular vector in  $\beta^t$  with all vectors in  $\beta^{t+1}$  and computing the degree of intersection between them. The evaluation of

$$I = \max \left( \text{length} \left( \phi_i^t \cap \phi_v^{t+1} \right) \right) \text{ for } v=1,2,\dots,N_{\beta^{t+1}}, \quad (5)$$

yields a measure of similarity between this vector in  $\beta^t$  and all vectors in  $\beta^{t+1}$ . The maximum corresponds to vectors that exhibit greatest intersection over the time of regard. If no intersection is found, vectors in  $\beta^t$  are compared to vectors in  $\beta^{t+2}$  and this same analysis is performed. To assess the degree of similarity or intensity, the following congruence metric is computed

$$f = \frac{I}{\text{length}(\phi_i^t)}. \quad (6)$$

If a vector contained in  $\beta^t$  identically matches a vector contained in  $\beta^{t+1}$  then the congruence measure is identically 1 and detected behavior persists over the time of regard. Lesser values of  $f$  suggest diminishing conformance to behavior. This analysis is performed stepwise over the entire database  $\mathbf{D}$ . Fig. 4 depicts measurement of temporal congruence by comparing a reference vector at  $t = t_k$  to all vectors containing behaviorally-matched vessels at  $t = t_{k+1}$ . The reference vector exhibits maximum intersection with vector  $\phi_1^{k+1}$ . Fig. 5 considers three cases that demonstrate evaluation of the congruence metric. In panel a) the vectors or clusters that exhibit greatest membership intersection from  $t = t_0$  to  $t = t_1$ , contain identical vessels and therefore, the congruence metric is identically 1. In panel b) three of the four vessels are contained in the cluster that exhibits greatest membership intersection at  $t = t_1$  and therefore the congruence metric is 0.75. Finally in panel c) two of the four vessels are contained in the cluster that exhibits greatest membership intersection at  $t = t_1$  and therefore the congruence metric is 0.50.

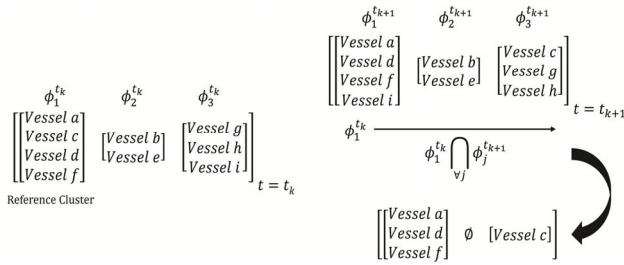


Figure 4. Temporal congruence is measured by comparing a reference vector at  $t = t_k$  to all vectors containing behaviorally-matched vessels at  $t = t_{k+1}$ .

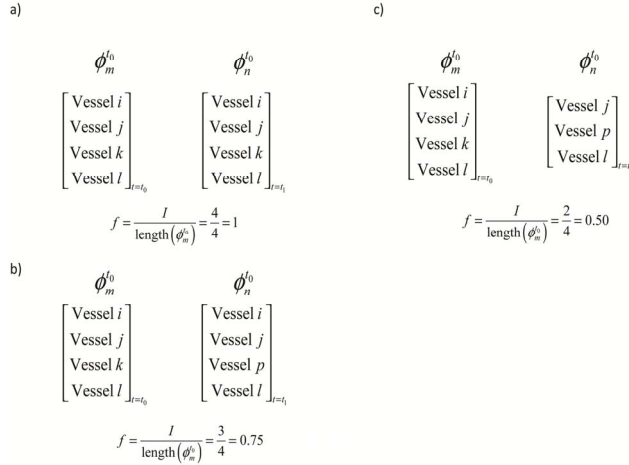


Figure 5. The degree of similarity between behaviorally-matched clusters is computed using a congruence metric.

### 3 Simulation Results

#### 3.1 Data Preparation

The behavior detection algorithm is applied to a data set comprised of both actual AIS vessel reports and synthetic data. The actual data set consists of AIS vessel position reports for a 10-hour period in the vicinity of the Strait of Gibraltar as shown in Fig. 6. The synthetic data, depicted in Fig. 7, is designed to simulate a convoy of five vessels transiting together from east to west. This convoy, referred to as the reference cluster path, is the blue line shown in Fig. 7. The synthetic data is combined with the actual data and this combined data set is used to conduct a series of simulations.

#### 3.2 Objectives

The objective of the simulations is to observe the formation of clusters centered on the synthetic convoy or reference cluster path and evaluate the influence of the actual AIS ship traffic on this formation. This provides a method to evaluate the performance of the algorithm. The performance metric used is given by

$$\rho = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \hat{x}_n)^2 + (y_n - \hat{y}_n)^2}, \quad (7)$$

where the  $(x_n, y_n)$  values are derived from the longitude and the latitude values of the synthetic ship cluster positions at each time step  $n$ , the  $(\hat{x}_n, \hat{y}_n)$  values are derived from the longitude and the latitude values of the combined AIS and synthetic ship cluster positions at each time step, and  $N$  is the total number of time steps.

#### 3.3 Simulations Runs

Default behavior thresholds for speed, course and distance are set as follows:  $T_s = 16$  knots (kts),  $T_c = 39^\circ$ , and  $T_d = 8$  nautical miles (nm). Simulations are run, keeping two of these parameters constant while varying the third.

In the first simulation, the distance threshold is varied from 1 to 30 nm in increments of 1 nm. The course and speed thresholds are held constant at  $39^\circ$  and 16 kts, respectively. The algorithm is run with these parameters for the 10-hour duration on the combined data set. The time step for the cluster centroid formation (fusion) is 12 minutes (which yields a total of 50 cluster center values for the 10-hour period) [13].

The second and third simulations are conducted in a similar manner. For the second simulation, the course threshold is varied from  $0^\circ$  to  $22^\circ$  in increments of  $0.5^\circ$  and for the third simulation the speed threshold is varied from 2 to 20 knots in increments of 2 knots.

#### 3.4 Observations

The simulation results are presented in Figures 8-10. Each plot shows the variation of the performance metric  $\rho$  versus particular behavior threshold. The following observations are made concerning these results: (a) the cluster path of the combined data set follows the synthetic ship cluster path closely, see Fig. 7; (b) for Figures 8-9, the plots indicate that the magnitude of the threshold beyond a certain value does not have an impact on performance. For example, the performance saturates for the distance threshold beyond 18 nm; (c) for Figure 10, the performance metric does not appear to saturate, indicating sensitivity to the magnitude of the speed threshold.

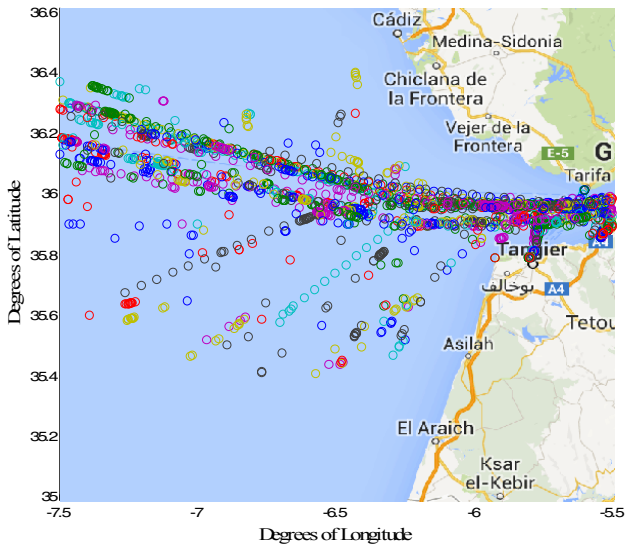


Figure 6. AIS vessel position reports collected in the Strait of Gibraltar for a 10-hour period.



Figure 7. Tracks of cluster centroids. The blue line indicates the cluster path of the ground truth convoy of ships; the dashed red line is the cluster path after processing the combined ship traffic. The two clusters merged or stayed very close to each other.

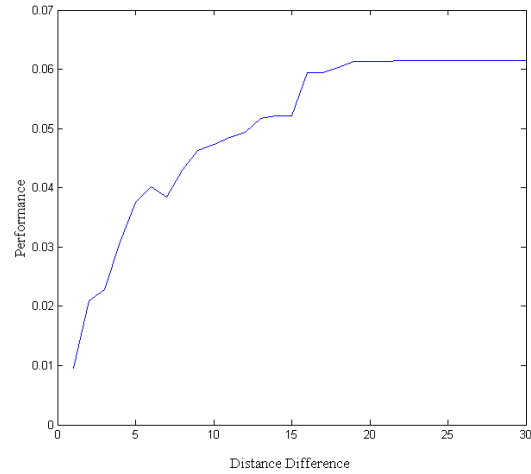


Figure 8. Algorithm performance as a function of the distance threshold with speed and course thresholds held constant at 16 knots and  $39^\circ$ , respectively.

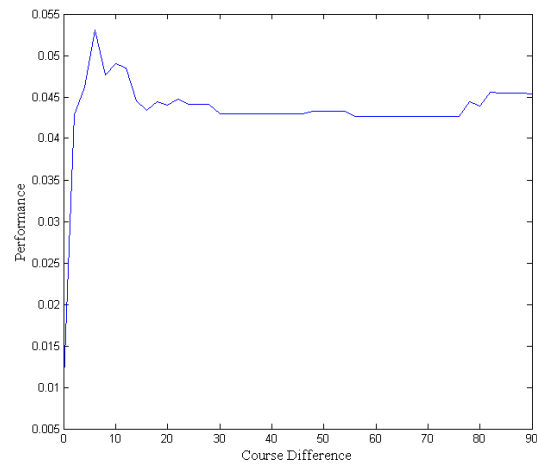


Figure 9. Algorithm performance as a function of the course threshold with speed and distance thresholds held constant at 16 knots and 8 nm, respectively.

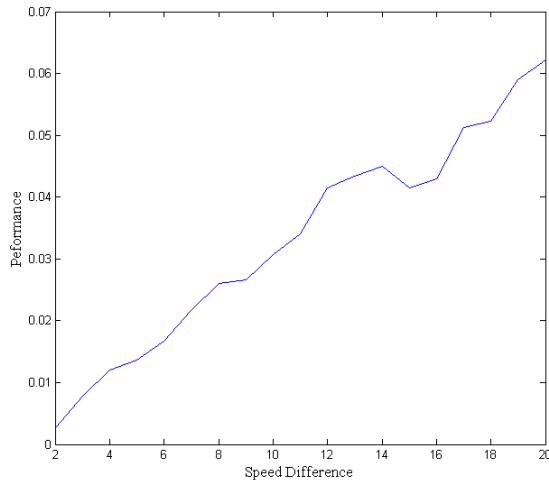


Figure 10. Algorithm performance as a function of the speed threshold with course and distance thresholds held constant at  $39^\circ$  and 8 nm, respectively.

## 4 Conclusions

A generalized behavior detection method for improving maritime domain awareness was presented as an extension to the spatiotemporal clustering algorithm presented in [11]. A propositional logic-based approach, was introduced which allows autonomous detection of vessel behaviors of interest based on natural language description. Additionally, a series of simulations were developed to assess the performance of the algorithm as a function of the distance, course and speed threshold values. The preliminary performance results suggest that variation in the threshold values affects the cluster formation and that performance is more sensitive to the speed threshold value than to the other two threshold values. We remark that the results will enable the user in selecting appropriate threshold values. Our eventual objective is to extend these results to derive optimal threshold values to maximize the algorithm's performance.

## References

- [1] J.T. Conway, G. Roughead, T.W. Allen, "A Cooperative Strategy for 21st Century Seapower," *Policy Paper*, Oct. 2007.
- [2] Department of Defense, "Dictionary of Military and Associated Terms, Joint Publication 1-02," Department of Defense, Arlington, VA, 08 Nov. 2010.
- [3] G. Pallotta, M. Vespe, and K. Bryan, "Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction," *Entropy*, vol. 15, no. 6, pp. 2218–2245, Jun. 2013.
- [4] E. Martineau and J. Roy, "Maritime anomaly detection: Domain introduction and review of selected literature," Tech. Memo. DRDC Valcartier TM 2010-460, October, 2011.
- [5] J. Hansen, G. Jacobs, L. Hsu, and J. Dykes, "Information domination: Dynamically coupling METOC and INTEL for improved guidance for piracy interdiction," *NRL Rev.*, 2011.
- [6] J. Roy, M. Davenport, Exploitation of maritime domain ontologies for anomaly detection and threat analysis, in: Proc. of the 2nd International on Waterside Security Conference (WSS 2010), Marina di Carrara, Italy, 2010, pp. 1–8.
- [7] A.C. van den Broek, R.M. Neef, P. Hanckmann, S.P. van Gosliga, D. vanHalsema, Improving maritime situational awareness by fusing sensor information and intelligence, in: Proc. of the 14th Int. Conf. on Information Fusion (Fusion 2011), Chicago, USA, 2011.
- [8] J. Gómez-Romero, M.A. Serrano, J. García, J. M. Molina, G. Rogova, Context-based multi-level information fusion for harbor surveillance, *Information Fusion*, Available online 19 February 2014, ISSN 1566-2535, <http://dx.doi.org/10.1016/j.inffus.2014.01.011>.
- [9] J. Llinas and J. Scrofani, "Foundational technologies for activity-based intelligence – a review of the literature," Naval Postgraduate School, Monterey, CA, Tech. Rep. NPS-EC-14-001, Feb. 2014.
- [10] M. Phillips. (2012) Trajectory Magazine, A brief overview of ABI and human domain analytics [Online]. Available: <http://trajectorymagazine.com/web-exclusives/item/1369-human-domain-analytics.html>.
- [11] K. Tester, J.W. Scrofani, M. Tummala, D. Garren, J. McEachen, "A spatiotemporal clustering approach to maritime domain awareness," *Signal Processing and Communication Systems (ICSPCS)*, 2013 7th International Conference on , vol., no., pp.1-9, 16-18 Dec. 2013.
- [12] J.B. MacQueen, "Some methods for classification and analysis of multivariate observations". Proc. 5th Berkeley Sym. on Mathematical Statistics and Probability, University of California Press. pp. 281–297.
- [13] K. Tester, "A spatiotemporal clustering approach to maritime domain awareness" M.S.E.E. Thesis, Department of Electrical and Computer Engineering, Naval Postgraduate School, Monterey, CA, Sep. 2013.