Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER 516582
UNCLASSIFIED	

TITLE

Fuzzy clustering for data fusion in a recognized maritime picture

System Number: 516582

Patron Number:

Requester:

Notes:

DSIS Use Only:

Deliver to:

This page is left blank

This page is left blank

Fuzzy Clustering For Data Fusion In A Recognized Maritime Picture

Christopher Helleur
Defence Research Establishment Ottawa
3701 Carling Avenue,
Ottawa, Canada K1A 0Z4
chris helleur@dreo dnd ca

Abstract - *The generation and maintenance of a Recognize Maritime Picture consists in part of associating contact reports from sensor sources with existing tracks or initiating new tracks. This decision-making process often takes place in an environment in which the goal and the constraints are not known precisely To deal quantitatively with imprecision, we usually employ the concepts and techniques of probability theory The use of a probabilistic approach requires that the imprecision can be equated with randomness and that the characteristics of this randomness are reasonable well known This is generally not the case for the generation of the RMP An alternative approach is to view the problem as fuzzy decision making and to employ the concepts and techniques of fuzzy sets The approach proposed in this paper makes use of the non-real-time nature of the problem to make maximum use of the available data. The proposed approach makes use of the reverse Cuthill-McKee ordering technique to establish an initial estimate of the number of clusters and the c-mean clustering techniques refine the cluster and to establish fuzzy membership functions The effectiveness of the approach is demonstrated on data acquired off the East Coast of Canada*

Keywords: Fuzzy reasoning, fuzzy set, Recognize Maritime Picture, data fusion.

1 Introduction

The Canadian Forces are responsible for the production and maintenance of the Recognized Maritime Picture (RMP) of Canadian waters for other departments, such as Fisheries and Oceans, and the Coast Guard, as well as for defense purposes It is expected that one of the key sources of surveillance of the 200NM Exclusive Economic Zone (EEZ) will be a system of High Frequency Surface Wave Radar (HFSWR).

There presently exist two radar sites, at Cape Race and Cape Bonavista, Newfoundland, that are being used to demonstrate the capability of the HFSWR technology [1] One of the more difficult platforms to track are smaller fishing vessels These vessels, which have small radar cross-sectional area, are constantly maneuvering while fishing. As a result they often produce broken tracks Since there are usually a number of fishing vessels in close proximity in any actively fished area, it is often difficult to associate these contacts unambiguously A possible means of associating the HFSWR track segments is by performing track-to-track fusion between the HFSWR and radar Electronic Support Measures (ESM) data.

In the military context, emitter characteristics are compared with a library of parametric descriptions, which yields a list of likely candidates The results of the analysis are reported as an ELINT report [2]. The recipient of these reports may track these vessels using an ELINT Correlator that sequentially associates incoming reports to existing tracks or initiates new track. The association process makes use of the positional measurement error to determine if two successive reports are Geo-feasible. It uses known variations in the values of the radio frequency (RF), pulse repetition interval (PRI) and the scan rate (SCAN) for each emitter type to determine whether the reported values can be feasibly associated with previous reports.

In a civilian context such a library does not exist As a result precise values for the emitter parameter variation are not available Also, it is not unusual to find a cluster of small vessels with emitter characteristics that would be very difficult to distinguish. Depending on the ESM data collection process, the report update rate may be very low resulting in a vessel maneuver being under sampled to maintain an accurate track on the target. This makes unambiguous association of ESM reports difficult in this scenario.

This paper proposes the use of an adaptive association technique based on a fuzzy c-mean clustering to produce fuzzy ESM clusters A track-to-cluster correlation between the HFSWR tracks and the resulting fuzzy ESM clusters provide a means of associating the HFSWR track segments

Reprinted from Proceedings of the fourth International Conference on Information Fusion Fusion 2001, August 7-10 Montreal, QC Volume II, pp ThB2-11-ThB2-17

The techniques used are discussed and an evaluation of the procedure is presented using recent trial data from the HFSWR site at Cape Race, Newfoundland and simulated ESM contact reports based on known positions of the vessels used in the trial.

2 HFSWR track and ESM cluster association

The approach is divided into three steps The first is to estimate the number of emitters, 'c', in the area of interest using a feasibility matrix. The second step is to use this value of c to generate a fuzzy clustering of the ESM contact reports The next step is the Fuzzy track-to-cluster association Finally, it is then possible to correlate the HFSWR track segments based on their membership in the ESM clusters An overview of the process is shown below in Fig 1

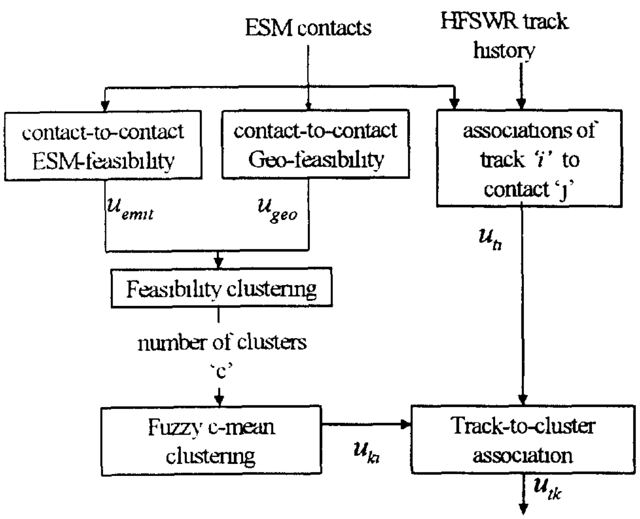

Figure 1 - Overview of ESM contacts and HFSWR track fusion

2.1 Estimation of number of emitter using feasibility matrix

An ESM contact report will provide the reporting time, the estimated position of the emitter, an Area of Uncertainty (AOU) ellipse for that position (expressed as a major and minor axis and the ellipse orientation), radio frequency (RF), the pulse repetition interval (PRI), the radar scan rate (SR) and the pulse duration (PD). The determination of whether two contacts could feasibly be from the same platform based on the location, size and orientation of the Area of Uncertainty (AOU) ellipse is referred to here as Geo-feasibility The determination of whether two contacts could feasibly be from the same emitter based on the similarity of the emitter parameters is referred to as Emit-feasibility [3-4]. For the purposes of this paper, the evaluation of the Emit-feasibility is limited to using the RF and the PRI.

Geo-Feasibility – Two contacts are considered to be Geo-feasible if it is feasible for a surface vessel to travel the distance between the two reported positions in the interval between the two reported times. This can be determined by expanding the AOU ellipse of the earlier contact, AOU_i, by the distance that the vessel can travel at its maximum velocity (e.g. 18 knots) during the time elapsed between the contacts to form $AOU_{j,i}$. If this ellipse overlaps the later ellipse AOU_j, we would conclude that the two contacts are Geo-feasible. A more complete discussion of testing confidence ellipses is presented in [5] We can express this feasibility as:

$$\mu_{Geo}(x_i, x_j) = \begin{Bmatrix} 1; & AUO_{i,j} \cap AOU_j \succ 0 \\ 0; & otherwise \end{Bmatrix} \quad (1)$$

where:

AUO_j = Area of uncertainty ellipse of contact j.

$AUO_{i,j}$ = Area of uncertainty ellipse of contact I at the reporting time of contact j.

Emit-Feasibility – A contact is considered to be Emit-feasible to another contact if the difference in the emitter parameters are within the known variations (i.e. tolerances) In the case of commercial navigational radar these tolerance values are not general known This makes a precise determination of whether an emitter parameter is sufficiently different from the parameters assigned to a track to warrant not associating with that track a subjective process. We can express this feasibility as:

$$\mu_{emit}(x_i, x_j) = \begin{Bmatrix} 1; & |x_i - x_j| \le tolerances \\ 0; & otherwise \end{Bmatrix} \quad (2)$$

Feasibility Cluster - If X is a set of n items, X = {x_1, x_2...x_n}, to cluster in X means the identification of an integer c, 2 ≤ c < n, and a partitioning of X by c mutually exclusive, collectively exhaustive subsets (clusters) of X. For this problem the condition for clustering is that members of each cluster, A_k, must be Geo-feasible and Emit-feasible with all other members of the cluster

$$A_k = \{x \in X \mid \mu_{geo}(x_i, x_j) \cap \mu_{emit}(x_i, x_j) = 1\} \quad (3)$$

The objective is to find the minimum 'c' that meets this criterion.

To illustrate this, Table 1 presents sample ESM contact reports with the latitude and longitude and Fig. 2 provides a plot of these contact reports. A contact feasibility visualization matrix is shown in Fig 3 For each contact pair the Emit/Geo-feasibility ($\mu_{ij} = \mu_{emit}(x_i,x_j) \cap \mu_{geo}(x_i,x_j)$) is computed The contacts have been renumbered to produce a matrix with a reduced bandwidth using a Reverse Cuthill-McKee ordering [6]

A border has been drawn around clusters that could potentially have come from the same emitter In the case of the first cluster (contact 1-6), the clustering is unambiguous since all the contacts in the grouping are feasible with and only with each other In the second cluster this is not the case since contact 7 cannot be feasible associated with contact 13 and contact 12 cannot be feasible associated with contacts 10 and 11 We have grouped contacts 7-11 together and 12-13 to form two additional clusters This result clearly indicates that there are at least three emitters in this region during this period. However, this grouping is not the only one possible There are three ways the last seven contact reports can be formed into two grouped Therefore we refer to the grouping as ambiguous

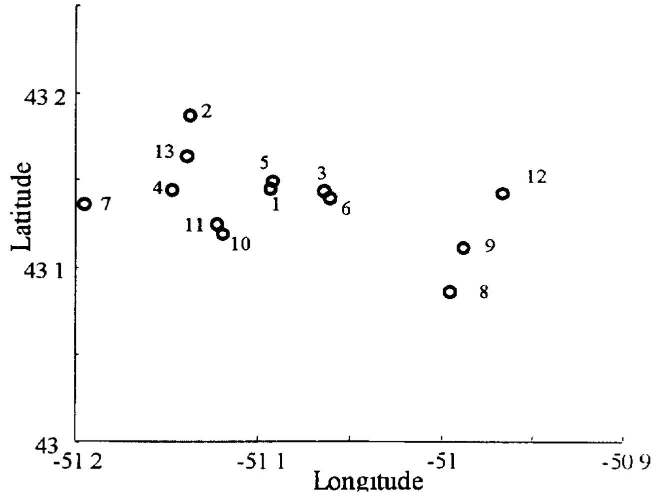

Figure 2- ESM sample data

Table 1 – Sample ESM contact reports

Report no	Lat (deg)	Lon (deg)	RF (Mhz)	PRI (μsec)	Time (hr)
1	43 15	-51 10	9435 00	1327 66	16 52
2	43 19	-51 14	9435 63	1327 66	17 10
3	43 14	-51 07	9434 38	1327.66	15 07
4	43 14	-51 15	9435 00	1327 66	16 90
5	43 15	-51 09	9435 63	1327 66	16 32
6	43 14	-51 06	9434 38	1327 66	15 22
7	43 14	-51 20	9404 38	1690 95	15 23
8	43 09	-51 00	9406 25	1691 90	17 10
9	43 11	-50 99	9407 50	1694 90	16 88
10	43 12	-51 12	9405 63	1690 90	16 53
11	43 13	-51 12	9407 50	1690 90	16 33
12	43 14	-50 97	9406 25	1689 20	16 52
13	43 16	-51 14	9405 63	1690 90	15 25

ESM Contact Reports

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	0	0	0	0	0	0	0
2	1	1	1	1	1	1	0	0	0	0	0	0	0
3	1	1	1	1	1	1	0	0	0	0	0	0	0
4	1	1	1	1	1	1	0	0	0	0	0	0	0
5	1	1	1	1	1	1	0	0	0	0	0	0	0
6	1	1	1	1	1	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	1	1	1	1	1	0
8	0	0	0	0	0	0	1	1	1	1	1	1	1
9	0	0	0	0	0	0	1	1	1	1	1	1	1
10	0	0	0	0	0	0	1	1	1	1	1	0	1
11	0	0	0	0	0	0	1	1	1	1	1	0	1
12	0	0	0	0	0	0	1	1	1	0	0	1	1
13	0	0	0	0	0	0	0	1	1	1	1	1	1

Figure 3 - Geo/Emit-feasibility of contact report pairs
1 = TRUE 0 = FALSE

2.2 Fuzzy Clustering of ESM Contacts

A fuzzy set is a class of objects in which there is no sharp boundary between those objects that belong to the class and those that do not. Let $X = \{x\}$ denote a collection of objects (points) denoted generically by x. Then a fuzzy set A in X is a set of ordered pairs:

$$A = \{(x, u_A(x))\}, \quad x \in X$$

where $u_A(x)$ is termed the membership of x in A. When u_A can only have values of 0 and 1, A is non-fuzzy and its membership function becomes identical with the characteristic function of a non-fuzzy set. In what follows, u_A is the interval [0, 1] with 0 and 1 representing, respectively, the lowest and highest grades of membership. Thus a fuzzy set A can be defined by associating with each object x a number between 0 and 1 that represents its grade of membership in A

The process of establishing membership grades $u_A(x)$ is referred to as abstraction The two abstraction methods described below are used in this work for the fuzzy ESM contact clustering and the fuzzy track to contact association.

The ambiguity in clustering some of the ESM contacts can be addressed by the use of a fuzzy clustering approach. The membership of an ESM data points in a cluster, abstraction, is achieved here using the Fuzzy c-Mean Clustering technique C-clustering is the process of partitioning a data set X into c mutually exclusive and collectively exhaustive subsets (clusters) of X Hard c-mean clustering is a technique by where each data point belongs to a cluster by a degree of 1 or 0 Fuzzy c-mean clustering is a technique by where each data point belongs to a cluster by a membership grade of between1 or 0 The technique used in this paper is based on the approach described in [7] The algorithm used is taken from [8]

The algorithm is based on the optimization of an objective function J(U) incorporating the distance of the data point, x_i, from the estimated center, x_j^* of the cluster and the fuzzy c-partitions U Using a gradient descent method, the algorithm produces estimates of the cluster centers and the partition matrix U The resulting partition matrix u_{ij} is used as a measure of the data point, x_i, membership in cluster j

2.3 Fuzzy track-to-cluster association

The fuzzy set concept of intersection and the corresponding notation is presented below A more complete discussion of fuzzy set concepts is presented in [9]

Intersection The intersection of A and B is denoted by $A \cap B$ and is defined as the largest fuzzy set contained in both A and B The membership is given by:

$$u_{A \cap B}(x) = u_A \wedge u_B = Min(u_A(x), u_B(x)), \quad x \in X \qquad (5a)$$

Union The union of A and B is denoted by $A \cup B$ and is defined as the smallest fuzzy set containing both A and B The membership is given by

$$u_{A \cup B}(x) = u_A \vee u_B = Max(u_A(x), u_B(x)), \; x \in X \qquad (5b)$$

The membership of the ESM contacts in HFSWR track is based solely on Geo-feasibility evaluation. However, unlike the crisp memberships used in section 2.1, it is necessary to establish a fuzzy membership of the EMS contact in the HFSWR track The HFSWR radar track, which is the output of a Kalman filter, provides estimates of the position, course and speed. It is possible to estimates of the location the target at the time of an ESM contact reports and to establish an AOU ellipse for this position (Fig. 4). The AOU ellipses can be expressed as a covariance matrix and a fuzzy membership function can be established as follows

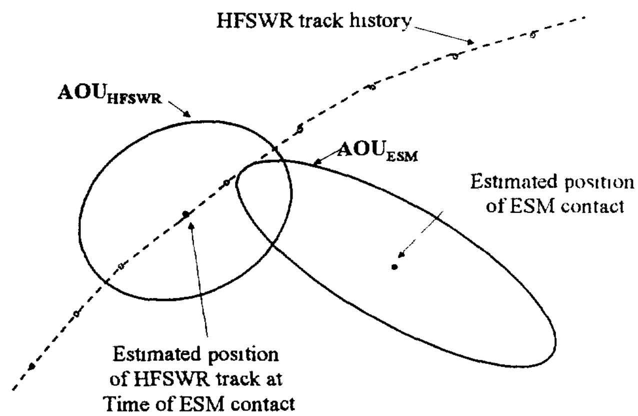

Figure 4 – Fuzzy correlation of ESM contact with HFSWR track based on AOU ellipses

The points in an ellipse E_i can be denoted by $\{X \mid (X - X_i)^T S_i^{-1}(X - X_i) \le 1\}$ The points in an ellipse D_i that is similar to E_i can be expressed as $\{X \mid (X - X_i)^T S_i^{-1}(X - X_i) \le k\}$.

where·

- X denotes the coordinates of an arbitrary point.
- X_i denotes the coordinates of the centre of the i^{th} ellipse (i.e. the estimate of the i^{th} contact report), and
- S_i denotes the covariance matrix of the i^{th} ellipse

If we have two ellipses E_1 and E_2, the space can be transformed so that E_1 is a unit circle at the origin and the axis of the ellipse E_2 are parallel to the coordinate axis

$$X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; X_2 = \begin{pmatrix} a \\ b \end{pmatrix}, \; S_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; S_2 = \begin{bmatrix} 1/c & 0 \\ 0 & 1/d \end{bmatrix}.$$

If the center of E_2 lies within E_1 then the two points are considered geo-feasible [5]. If the center of E_2 lies outside of E_1 then there exists an ellipse D_2 that is similar to E_2 that intersects with E_1 at only one point P. The normal vector to E_1 at P is $S_1^{-1}(X_1 - P) = I(0 - P) = -P$. The normal vector to D_2 at P is $S_2^{-1}(X_2 - P)$ Since E_1 and D_2 are tangent at P, the normal vectors of D_2 proportional to the normal vectors of E_1 at P Hence.

$$S_2^{-1}(X_2 - P) = tP$$

$$P = X_2[1 + tS_2]^{-1} \tag{6}$$

Substituting in values for X_2 and S_2 into the above equation yields

$$P = \begin{pmatrix} ac/(t+c) \\ bd/(t+d) \end{pmatrix} \tag{7}$$

From the equation for E_1 we have $P^T S_1^{-1} P = 1$. Substituting in our values for P and S_1, we have

$$\begin{pmatrix} ac/(t+c) & bd/(t+d) \end{pmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} ac/(t+c) \\ bd/(t+d) \end{pmatrix} = 1$$

$$\left(\frac{ac}{(t+c)}\right)^2 + \left(\frac{bd}{(t+d)}\right)^2 = 1 \tag{8}$$

This results in a fourth order equation in t. The real positive root corresponds to the point of tangency between E_1 and D_2 From the equation of D_2 we have $tP^T S_2^{-1} tP = k$ Substituting in our values for t, P and S_2^{-1}, we can solve for k If k is equal to 1 then the ellipse D_2 is identical to E_2 and we can conclude that E_2 intersects E_1 at only one point If k is less than 1, then E_2 intersect E_1 at more than one point and there is therefore an overlap between E_1 and E_2 If k is greater than 1, then E_2 does not intersect E_1 and there is therefore no overlap between E_1 and E_2

A fuzzy membership function u_{ti} for an ESM contact 'i' in HFSWR track 't' can be expressed as·

$$\mu_{ti} = \begin{Bmatrix} 1, & X_{ESM} \text{ inside } AOU_{HFSW} \\ 1-k; & X_{ESM} \text{ outside } AOU_{HFSW} \text{ and } k < 1 \\ 0, & \text{otherwise} \end{Bmatrix} \tag{9}$$

where X_{ESM} is the center of the ESM AOU ellipse.

The membership, u_{tk}, of the track 't' in an ESM cluster 'k' can be determined by determining the ESM contact with the highest membership in track 't' and associating with it the cluster in which the ESM contact has the highest membership. This can be expressed mathematically as follows:

$$u_{tk} = \bigvee_{i=1}^{\text{no of contacts}} \left(u_{ti} \wedge u_{ik}\right) \tag{10}$$

where u_{ti} is the membership of ESM contact 'i' in HFSWR track 't' and u_{ik} is the membership of contact 'i' in cluster 'k'

Finally, it is then possible to correlate the HFSWR track segments based on their membership in the ESM clusters Tracks are assumed to the members of the cluster to which it has maximum membership and tracks that are member of the same cluster are associated

3 Application

3.1 Trial data

Trials conducted of the East Coast of Newfoundland provided us with some HFSWR data and corresponding ground truth. The trials involve three vessels that were tracked over an area of 30 km^2 for a period of 5 hours Figure 5 shows the GPS data for the three vessels and the HFSWR tracks. The radar tracked vessel 1 intermittently, vessel 2 was tracked for only a very short period at the end of the 5-hour period and vessel 3 was not tracked at all Since no ESM is available for this analysis, simulated data is generated based on the GPS data and assumed characteristics of the ESM equipment and the navigational radar

3.2 Simulated ESM data

The objective is to simulate the ESM data that might be collected from three vessels with similar emitter parameter The emitters were assigned RF values of 9 402 MHz, 9 404 MHz, 9.406 MHz and PRI values of 1002 µsec, 1004 µsec, and 1006 µsec respectively The AOU ellipses where assigned major and minor axis of 2 NM and 1 NM respectively and the ellipse orientation was assigned a random value between 0 and 360 degrees Randomly selecting the desired number of contacts from the GPS positions of the vessels and adding a random position error assigned the locations of the detections. We assumed a normally distributed position error with a standard deviation equal to half the major and minor axis of the AOU ellipse Figure 6 illustrates the resulting ESM contact reports along with the HFSWR tracks. In this example a total of 25 ESM contact reports were generated for the three vessels

An association is made between an ESM contact and a HFSWR track if the AOU ellipse of the ESM contact overlaps the AOU associated with the HFSWR track. In the case of the example shown in Fig.5, each of the HFSWR tracks has at least one ESM contact sufficiently close to it to make an association

3.3 ESM Contact Clustering and Track to Cluster Association

Using the method described above, it is possible to generate the feasibility visualization matrix shown in Fig. 7. A minimum of three clusters in which each contact is Emit-feasible and Geo-feasible with all the other contacts in the cluster can be identified in this matrix. The solid lines in this figure illustrate a possible clustering. Applying the Fuzzy c-Mean clustering technique to this sample data with c=3, we obtain the memberships shown in Table 2

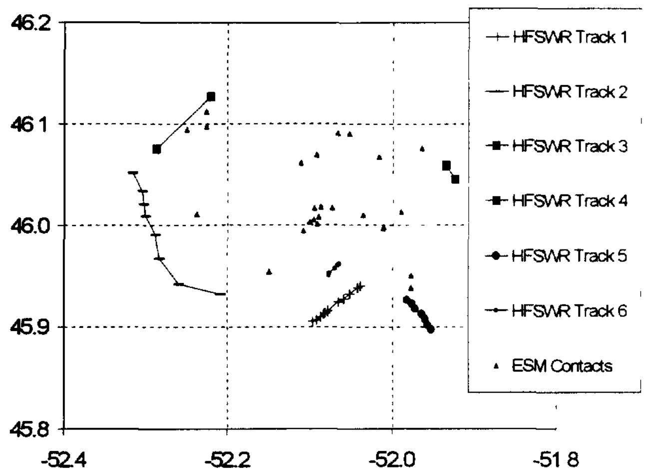

Figure 6 - HFSWR tracking of these vessels and simulated ESM contact reports

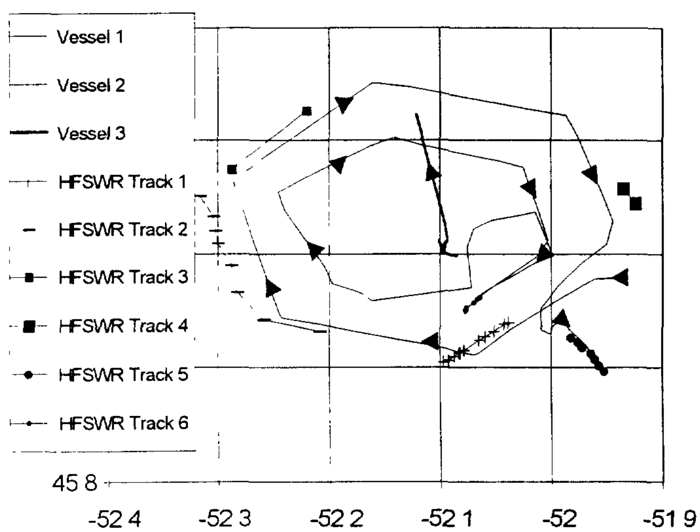

Figure 5 - Ground truth for three vessels over a 5 hour period and HFSWR tracking of these vessels

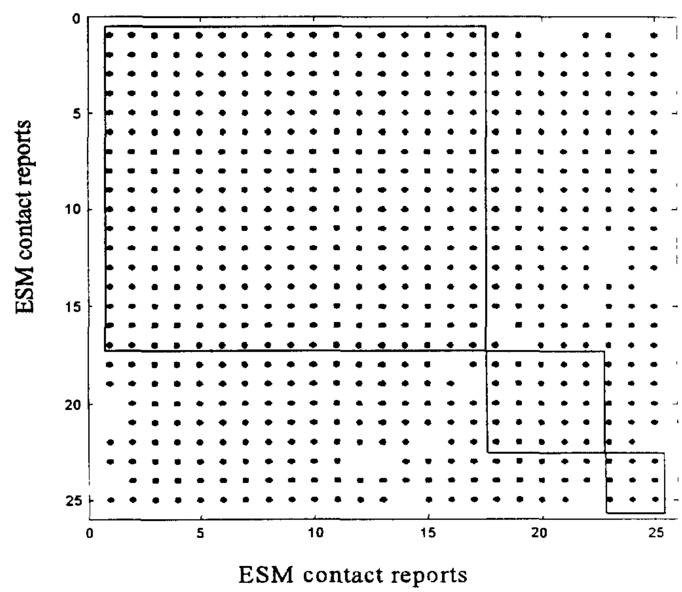

Figure 7 – Feasibility visualization matrix A "•" indicates contact pair feasibility

Table 2 – Contact membership (U_{ij}) in the clusters and contact-to-track association

Contact	Platform	Cluster 1	Cluster 2	Cluster 3	Contact-Track Association
1	1	0 13	0 02	0 85	
2	1	0 08	0 02	0 90	HFSW Track1
3	3	0 05	0 94	0 01	
4	1	0 04	0 01	0.96	
5	2	0 66	0 26	0 08	
6	3	0 06	0 92	0 02	
7	3	0 07	0 91	0 02	
8	3	0 05	0 94	0 02	
9	2	0 99	0 00	0 01	
10	1	0 05	0 01	0 94	HFSW Track2
11	1	0 07	0 02	0 91	HFSW Track3
12	1	0 62	0 04	0.34	
13	3	0 05	0 93	0 02	
14	2	0 66	0 27	0 07	
15	2	0 97	0 01	0 02	
16	3	0 07	0 91	0 03	
17	2	0 99	0 00	0 00	
18	3	0 08	0 90	0.02	
19	1	0 22	0 03	0.75	HFSW Track4
20	1	0 08	0 02	0 89	
21	2	0 83	0 04	0 13	HFSW Track6
22	3	0 03	0 96	0 01	
23	1	0 13	0 03	0.84	HFSW Track5
24	1	0 02	0 00	0 98	
25	1	0 10	0 02	0 88	

The first two column of Table 2 provide the contact report number and the vessel to which the ESM contact corresponds. The following three columns provide the membership value of each ESM contact in the three clusters The final column indicates which ESM contacts have the maximum membership in the HFSWR tracks segments The HFSWR track segments 1-5, which correspond to vessel 1, are associated with ESM contacts that have high membership values in cluster 3 HFSWR track segment 6, which correspond to vessel 1, is associated with an ESM contact with a high membership value in cluster 1

3.4 Results

Simulated EMS contacts were generated for a large number of scenarios in which the number of contacts generated during the 5-hour period varied from 20 to 50. The results indicate that the track-to-contact association is correct approximately 85% of the time and the contact-to-emitter association is correct approximately 90% of the time when the clustering technique correctly identifies there being three emitters present. However, the frequency with which three emitters is identified is dependant on the number of ESM contacts reported during the 5-hour period. The simulated results indicate that for this scenario, approximately 35 contacts need to be reported to have a greater than 50% probability of correctly identifying 3 emitters

4 Discussion

The use of the feasibility matrix to estimate the number of emitters, the use of the c-mean fuzzy clustering technique for fuzzy membership abstraction and the fuzzy HFSWR track to ESM contact association have been shown to be an effective means of correlating HFSWR track segments. The primary limitation on the approach is its ability to correctly identify the number of emitters with a small number of ESM contact reports. In a routine surveillance mode of operation, it is not reasonable to expect a high rate of reporting. This scenario is particularly challenging due the proximity and movement of the platforms, the large error ellipses associated with the ESM contact reports and the similarity of the emitter parameter chosen for the evaluation.

5 References

[1] Moutray, R.E., Ponsford, A.M., *Integrated Maritime Surveillance for the Grand Banks*, Oceans '97 MTS/IEEE Conference Proceedings, p981-6 vol.2 1997, Halifax, Canada

[2] Schleher, D.C , *Electronic Warfare in the Information Age*, Norwood, MA: Artech House, 1999.

[3] Lefebvre, E., Simard, M., Helleur C. *A Multisource Information Adaptive Fuzzy Logic Correlator for Recognized Maritime Picture*, Applied informatics 2001 - Symposium 1 Artificial Intelligence and Applications, Innsbruck, Austria February 2001.

[4] Simard, M., Lefebvre, E, Helleur, C. , *Multisource Information Fusion Applied to Ship Identification for the Recognised Maritime Picture*, SPIE Conference, Orlando, April 2000.

[5] Myhre, J.,Rennie, M, Duquette, W., *Testing and Combination of Confidence Ellipse: A Geometric Analysis*, Technical Memo No 2, Aug. 1985, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California.

[6] Gilbert, J.R., Moler, C , Schreiber, R., *Sparse Matrices In Matlab: Design and Implementation*, SIAM J Matrix Anal. Appl., Vol. 13, No.1, pp 333-356, January 1992

[7] James C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[8] Jang, J.S R., Gulley, N., *Matlab: Fuzzy Logic Toolbox*, The Math Works Inc , Natick, MA, 1995.

[9] Dubois, D , Prade, H., *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980.

516582

CA011983