Automated Association of Track Information from Sensor Sources with Non-Sensor Information in the Context of Maritime Surveillance

Eric Lefebvre
Lockheed Martin Canada
6111 Royalmount Avenue
Montreal, Quebec, H4P 1K6
Canada
eric.lefebvre@lmco.com

Christopher Helleur
DRDC-O
3701 Carling Avenue
Ottawa, Ontario, K1A 0Z4
Canada
chris.helleur@drdc-rddc.gc.ca

Abstract - *Maritime surveillance of coastal regions requires the processing of a large amount of data. Therefore, limited availability of human or material resources may prevent full use of information and knowledge. A software prototype that incorporates automated association of sensor information with non-sensor information has been developed. Sensor information, which provides contact reports, is modelled as tracks while non-sensor information, or user's knowledge that represents arbitrary ship behaviour, is modelled as templates. For the fusion process elements of fuzzy logic are introduced through membership functions to take into account uncertainties from track and template information. These membership functions are used to evaluate the possibility of track and template association. The knowledge-based system used for this level 2 data fusion incorporates a genetic algorithm for the track modelling and Boolean combination of parent templates for the modelling of user's knowledge. This paper also discusses database management and user interface.*

Keywords: Maritime surveillance, association, sensor, non-sensor, information.

1 Introduction

Maritime surveillance not only involves management of sensor information but also infrastructure, systems, plans and strategies from a maritime perspective. For national sovereignty purposes, areas under surveillance can include the 200 NM Exclusive Economic Zone (EEZ) and for defence purposes, extend well beyond [1]. Civilian and Military maritime organisations may have access to a number of surveillance sources. A country's ability to make full use of these systems is limited by its ability to fuse the data and information from all sources in a timely, accurate, and complete manner. Considering the amount of information that needs to be processed, available human or material resources may not be sufficient to perform all the required data analysis. Automated processing and information management then become an essential capability for maritime surveillance systems.

A system performing level 1 fusion, as defined by the US Joint Directors of Laboratories Data fusion Group, for maritime surveillance has been presented in references [2,3]. This paper presents additional functionality for this software prototype to support the fusion of information from sensor sources with non-sensor information. In reference [4] a method for fusion of radar tracks with ship reports and plans has been proposed. Here a method to fuse non-sensor information that is not coming from a specific vessel is presented.

Sensor information, which is reported in the form of contact reports, is modelled as tracks while non-sensor information is modelled as templates. Templates represent information known about ship behaviour or it represents behaviour that decision makers are looking for. Both sensor and non-sensor information have uncertainties that should be taken into account when performing fusion. Methods to manage and represent these uncertainties are presented herein.

2 Track modelling

Tracks are based on a collection of contact reports provided by the sensors. However, to perform data fusion at track level a track model must be computed. Usually, when contact reports are received in real time by a surveillance system, this model consists of a state estimate for the track generated by a Kalman filter. In our case, much of the data is characterised by delays of several hours and received in batches [3]. As a result the modelling process is performed in batch mode and tracks are modelled by a collection of parametrised line segments. This approach provides track models that are better suited to track-template fusion. Ships usually navigate using waypoints performing leg segments between each waypoint. The line segments model is therefore appropriate for transiting ships. Fishing ships, however, will perform short segments at low speeds in a small area. Due to their size and motion they may be difficult to track for coastal radar and will be often represented by sensor output as stationary vessels.

Coastal radar will provide near real-time tracking with position updates every 3 minutes on average for each track. Surveillance aircraft flying with a predefined path may update ship position only one time per day.

The algorithm used for track modelling is a hybrid optimisation algorithm based on both a genetic algorithm (GA) and simulated annealing (SA), SAGACIA [5]. The advantage of SAGACIA over conventional GA or SA are that it works from a population, which takes the advantage from GA. It can also easily escape from local optimum and converge rapidly. The different states of the track are represented with chromosomes. Chromosomes are made up from contacts ordered in time. Each gene represents a track's contact and has two possible values, 0 or 1. The value 1 indicates the beginning of a segment. The track's first and last contacts are always represented by 1. Moreover, each segment must have at least four contacts. Fig. 1 shows a possible chromosome's representation for a 20-contact track. This track model has three segments, and despite the last gene being equal to one, its corresponding contact belongs to Segment 3.

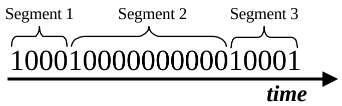

Fig. 1 Chromosome representation of a track.

The initial state of the population is randomly generated with the only constraint being that each segment must have at least four contacts. The state with only one segment is always present in the initial population.
The fitness or cost function is the sum of the chi-square, c , per degree of freedom, u , of all segments:

$$Cost = \sum_{s=1}^{ns} \left(\frac{c^2}{n} \right)_s \tag{1}$$

Where s is the segment number and ns is the number of segments. The chi-square per degree of freedom is defined [6] by

$$\frac{c^2}{n} = \frac{1}{2n-4} \left\{ \sum_{q=x,y} \sum_{i}^{n} \left(\frac{q_i - \left(a_q + b_q t_i\right)}{s_i} \right)^2 \right\} \tag{2}$$

Where n is the number of contacts in a segment, q is one of the two orthogonal referential axes (here q is either along the longitude or along the latitude at each contact position $(x,y)_i$), t_i is the time of the i^{th} contact, and a_q and b_q are the adjusted parameters given by:

$$a_q = \frac{\sum_i^n \frac{t_i^2}{s_i^2} \sum_i^n \frac{q_i}{s_i^2} - \sum_i^n \frac{t_i}{s_i^2} \sum_i^n \frac{q_i t_i}{s_i^2}}{\sum_i^n \frac{t_i^2}{s_i^2} \sum_i^n \frac{1}{s_i^2} - \left(\sum_i^n \frac{t_i}{s_i^2} \right)^2}$$
$$b_q = \frac{\sum_i^n \frac{q_i}{s_i^2} - a_q \sum_i^n \frac{1}{s_i^2}}{\sum_i^n \frac{t_i}{s_i^2}} \tag{3}$$

With s_i defined by the maximum error projection on the orthogonal referential axes[1]

$$\begin{aligned} s_x &= \mathrm{MAX}\left(a \sin q, b \cos q\right) \\ s_y &= \mathrm{MAX}\left(a \cos q, b \sin q\right) \end{aligned} \tag{4}$$

Where a and b are respectively the semi-major and semi-minor axes and q the error ellipse orientation.

The selection method for each generation is elitist and cost-proportionate (roulette-wheel sampling). The algorithm ends when the state with the lower cost is the same for 10 generations in a row. This state defines the number and the distribution of segments in the track model.

For this optimisation situation, genetic algorithms provide good results. Usual genetic algorithm downsides, like high cpu time consumption, possibility of being trapped in a local minimum, and difficult transposition in genes, do not apply here. Transposition into binary chromosomes is straightforward, and track segments are sufficiently defined to address small populations (around 30 individuals) and to avoid many local minima in the fitness function.

Results of the modelling can be seen in Fig. 2 where track models (red lines) are overlaying two sets of fused real contact reports (black crosses and blue circles). Two coastal radar coverage regions and positions of three oilrigs (orange triangles) are also represented.

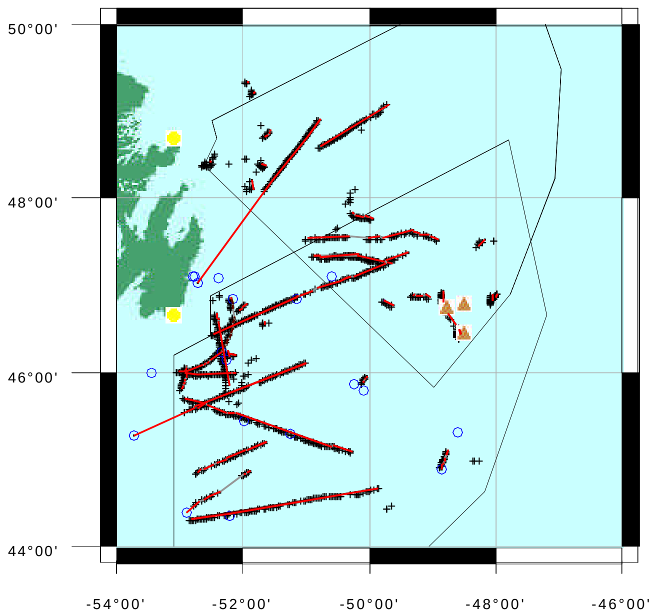

Fig. 2 Track models overlaying contact reports.

[1] The general formalism to perform this fit involves the moment matrix of measurment correlation. However, this matrix is not completly known and since measurment errors are of the same magnitude for most of the contact reports, our definition of s_i remains a good approximation.

3 Templates

In this paper non-sensor information corresponds to the knowledge the user has on vessels of interest. In order to be associate with tracks, this information must also be modelled. This modelling is done through templates, which then become the mathematical representation of the user's knowledge. In reference [7], maritime vignettes have been elaborated to represent relevant ship behaviours for Intelligence Surveillance and Reconnaissance (ISR). General parameters can be extracted from these vignettes to develop parent templates that can be combined in different ways to represent the user's knowledge. The three parent templates that have been selected are as follows:

1. Origin/Destination - Origin or destination of ship is known
2. Intercept Region - Area of intercept is known
3. Path - Ship's course or speed is known

Examples of the first template, Origin/Destination, are the case where information about a self-reporting vessel that has announced it is entering territorial water or the case where a vessel carrying contraband that has been identified by a platform in the region (origin and destination may be entry and exit points of a detection area and not necessarily the actual ports of origin and destination.). Examples of the second template, Intercept Region, are the cases where a pollution slick has been sighted or informations about fishing activities are known. The third template, Path, may represent Search and Rescue (SAR) information on a vessel with a known route, but which has been reported missing with no specific location provided.

A user interface has been developed to encode these types of information. They can be input with numerical values, when precise information is required, or with fuzzy values. The choice of numerical or fuzzy values will automatically select the level of information uncertainty. Possible inputs corresponding to the three parent templates are:

1. Latitude and longitude of origin/destination, time at the origin/destination.
2. Ellipse parameters delimiting the intercept region (latitude and longitude of the centre, semi-major and semi-minor axes, orientation), time interval during which the vessel intercepts the region (time after, time before.)
3. Ship course and speed.

Partial information can be input (e.g., origin is known but not the destination, intercept area is known but not the time interval, course is known but not the speed.) Combination of parent templates is performed with the Boolean operators: OR, AND, OR NOT, AND NOT.

"OR" ("OR NOT") signifies that the (non-)occurrence of the parent template is facultative, and "AND" ("AND NOT") signifies that the (non-)occurrence of the parent template is mandatory in the resulting template. For example, in the case of illegal fishing monitoring, a region is defined in order to monitor vessels entering a specific fishing area, but ships coming from an acceptable fishing port can be excluded. The template representing this situation can be a combination of parent template 1 (AND NOT origin) with parent template 2 (AND intercept region).

Templates can also be deactivated. When deactivated templates are still in the database, they can be saved in the configuration file (see section 6), but will not be used in the association process.

4 Track-Template Association

In Eq. (2) mathematical models have been defined for tracks with parametrised line segments:

$$\begin{aligned} x &= a_x + b_x t \\ y &= a_y + b_y t \end{aligned} \tag{5}$$

Where (a_x, a_y) is the location at time t=0 and (b_x, b_y) the velocity along each axis (x,y). The general model for parent templates 1 and 2 is the ellipse:

$$\begin{aligned} &(x-h)^2 A + (y-k)^2 B + (x-h)(y-k)C = 1 \\ &A = \frac{\cos^2 q}{a^2} + \frac{\sin^2 q}{b^2} \\ &B = \frac{\sin^2 q}{a^2} + \frac{\cos^2 q}{b^2} \\ &C = \frac{\sin 2q}{b^2} - \frac{\sin 2q}{a^2} \end{aligned} \tag{6}$$

Where (h,k) is the centre of the ellipse, a and b the semi-major and semi-minor axes and q its orientation. For origin and destination, circles are used (a=b=r). Values of a, b and r will be discussed in the next section.

A track will be associated with a template when an intersection occurs between one of the track segments and one of the template ellipses. Times of this intersection, t_a and t_b, are given by

$$\begin{aligned} &t_a^{(+)}, t_b^{(-)} = \frac{-Q \pm \sqrt{Q^2 - 4PR}}{2P} \\ &P = b_x^2 A + b_y^2 B + b_x b_y C \\ &Q = 2b_x(a_x - h)A + 2b_y(a_y - k)B + \\ &\qquad \left[b_y(a_x - h) + b_x(a_y - k)\right]C \\ &R = (a_x - h)^2 A + (a_y - k)^2 B + (a_x - h)(a_y - k)C - 1 \end{aligned} \tag{7}$$

Possibility of association between tracks and templates is expressed by a membership function M. The values of this function include uncertainties and will be defined in the next section.

For the first parent template, Origin/Destination, the condition for an association between a track and the origin circle is that the start time of the first segment of the track t_0^{first} must be greater than the time of intersection t_b of the extrapolation of this segment with the origin circle. Similarly, the end time of the last segment t_1^{last} must be smaller than the time of intersection t_a of the extrapolation of this segment with the circle of the destination (see Fig. 3):

$$
\begin{aligned}
membership_{origin} &= M \Rightarrow t_0^{first} \geq t_b \\
membership_{destination} &= M \Rightarrow t_1^{last} \leq t_a \\
membership &= 0 \Rightarrow else
\end{aligned}
\tag{8}
$$

If a time T is defined for the origin or destination the resulting membership will be

$$
membership = membership \times \begin{cases} 1 \Rightarrow t_b \leq T + d \text{ and} \\ \quad t_a \geq T - d \\ 0 \Rightarrow else \end{cases}
\tag{9}
$$

$$
d = \begin{cases} 0 \Rightarrow numerical \\ 3hrs \Rightarrow fuzzy \end{cases}
$$

Where d takes into account whether time has been input with fuzzy or numerical precision.

For parent template 2 (intercept region) association, a track may have intercepted a region before sensor measurements occur, during the measurement or after the sensor acquisition stopped. This is expressed by the following rules for the membership of association between track and intercept region template.

$$
\begin{aligned}
membership_{intercept} &= M \Rightarrow \begin{cases} t_1^{first} \geq t_b \text{ or } t_0^{last} \leq t_a \text{ or} \\ t_0^{n} \leq t_a \text{ and } t_1^{n} \geq t_b \end{cases} \\
membership &= 0 \Rightarrow else
\end{aligned}
\tag{10}
$$

If a time slot is defined $[T_{after}, T_{before}]$ for the intersection, the following rules for the membership also apply.

$$
membership = membership \times \begin{cases} 1 \Rightarrow t_b \leq T_{before} + d \text{ and} \\ \quad t_a \geq T_{after} - d \\ 0 \Rightarrow else \end{cases}
\tag{11}
$$

$$
d = \begin{cases} 0 \Rightarrow numerical \\ 3hrs \Rightarrow fuzzy \end{cases}
$$

Fig. 3 shows some examples of possible track-template intersections with the time parameter identified. The plot is in a two dimensional space (x,y) with the time increasing from the bottom up.

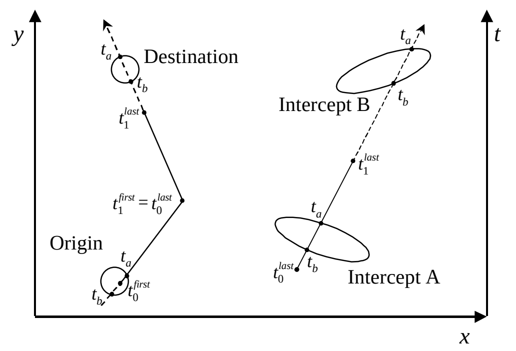

Fig. 3 Examples of tracks-templates possible intersections.

For parent template 3 (course and speed), a comparison between the course and the speed of each segment n and those defined in the template is performed.

$$
\begin{aligned}
membership_{course} &= 1 \Rightarrow course(n) = Course \pm s_{course} \\
membership_{speed} &= 1 \Rightarrow speed(n) = Speed \pm s_{speed}
\end{aligned}
\tag{12}
$$

Depending on whether the inputs are fuzzy or numerical a tolerance is added. Hence

$$
s_{course} = \begin{cases} 5^{\circ} \Rightarrow numerical \\ 22^{\circ} \Rightarrow fuzzy \end{cases} \quad s_{speed} = \begin{cases} 2\text{Kts} \Rightarrow numerical \\ 5\text{Kts} \Rightarrow fuzzy \end{cases}
\tag{13}
$$

Fig. 4 shows results of associations of tracks from Fig. 2 to two templates. The green and red circles represent, respectively, the origin where a vessel is expected to enter the area of detection and the port of destination. The yellow ellipse represents a fishing area.

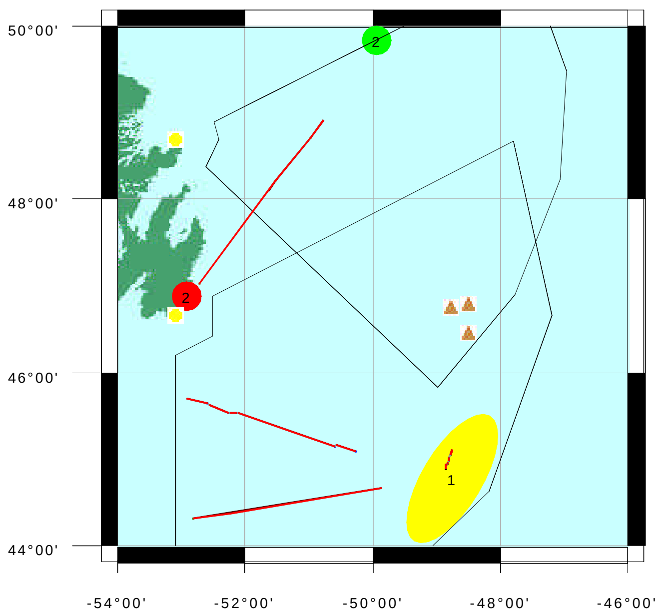

Fig. 4 Templates representation and associated tracks.

When the association is performed the software interface filters tracks not associated with active templates, showing only associated tracks. As can be seen here, there is one track associated with the first template (origin and destination) and three tracks associated with the second template (fishing area). For the second template, two tracks are going to or coming from the fishing area while one is in the area.

When parent templates are combined the resulting membership depends on the Boolean operators used. If all parent templates are combined with the OR (OR NOT) the resulting membership will be the maximum membership of parent templates involved:

$$
membership_{\text{OR}} = \text{Max}(M_1, M_2, M_3)
\tag{14}
$$

If one or more parent templates are combined with AND (AND NOT), the resulting membership will be the minimum membership of parent templates combined with AND (AND NOT):

$$membership_{\mathrm{AND}} = \mathrm{Min}\left(M_{\mathrm{i}},...\right) \qquad (15)$$

5 Uncertainties

Sensor measurements of any kind are always subject to uncertainties or errors. It is important to distinguish between two types of errors: systematic and random. Random errors can be handled by the theory of statistics. Systematic errors, however, are uncertainties in the bias of the data and they are related to a sensor's calibration. Long-range sensor measurements may be affected by several factors (ionosphere, atmospheric condition, sea state…) and these factors may vary in time. Their effects are usually not precisely known. The estimation of systematic errors becomes important when considering data fusion from different sensors, since all sensors will not bias the data in the same way. Unfortunately, systematic errors cannot be calculated the way statistical errors can. They are not output by the sensor nor do they appear in the filter's covariance matrix, which includes only random errors. Therefore, their estimation always involves a certain level of subjectivity. It should be noted that even if a sensor has significant systematic error it can remain reliable. Here we distinguish systematic error from sensor reliability, the former is related to a bias that can be estimated while the later is related to the consistency of the sensor output.

Fig. 5 presents the random and systematic errors for a coastal radar tracking a stationary oil platform. The radar (not visible) is located at the same latitude as the Hibernia oil platform (orange triangle) but about 4 degrees West (~165 NM). The sensor's radial resolution is better than the angular one. This can be seen by the spread of the contact reports (black crosses), which are more important in latitude than in longitude. The contact reports are actual measurements collected over a 6-hour period. The random error is related to the spread of the contact reports and the systematic error is related to the offset between the centre of the contact reports' distribution and the real position of the platform. The offset may not be the same for all positions on the sensor coverage area and will change depending on the atmospheric and sea conditions.

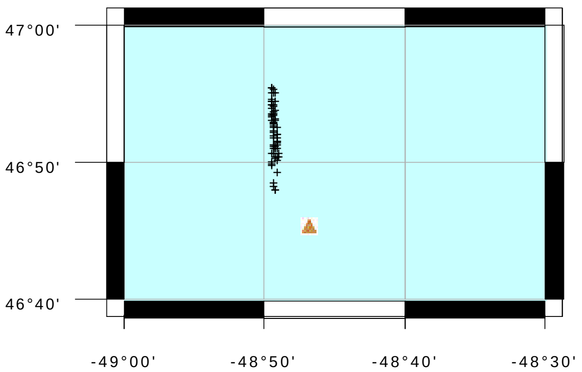

Fig. 5 Sensor's random and systematic errors.

The software prototype presented here allows users to input a *track bias*, which can be added to the random errors when fusing tracks from different sensors or when fusing tracks with templates. Based on trials with available GPS data, an error of about 2 NM seems appropriate for a high frequency surface wave radar with a detection range of 50 NM to 200 NM. This bias should be considered as an average value since it may vary depending on the range and other conditions already cited. For the data shown in this paper, a track's random errors are negligible compared with the *track bias*. Therefore, only track bias will be considered in track to template fusion.

If *track bias* represents the uncertainty of the track position, template uncertainty should also be taken into account. When a user inputs template positions with fuzzy values instead of numerical values, the track to template association process will automatically vary parameters of template ellipses to represent the template's uncertainty (treatment of time uncertainty has already been presented in previous section). Semi-axes length of ellipses will be increased and the membership value *M* will correspond to the dimension of the ellipses where the intersection with the track model occurred first. For the origin and destination template, a circle of radius *r* (ellipse with *a*=*b*=*r*) is centred on the origin and destination locations. The value of *M* in Eq. (8) is

$$M = \begin{cases} 1.0 \Rightarrow r = 2 \cdot track\_bias \\ .75 \Rightarrow r = 4 \cdot track\_bias \text{ (fuzzy only)} \\ .50 \Rightarrow r = 8 \cdot track\_bias \text{ (fuzzy only)} \end{cases} \qquad (16)$$

If, using the smallest value of *r*, no intersection occurs between the track model and the template (t_a and t_b not real in Eq. (7)) the value of *r* is increased and will be increased again if there is no intersection with this new value. The note,"(fuzzy only)", indicates that *r* will be increased only if the origin or destination location is input with fuzzy values.

For intercept region templates, the semi-axes lengths input by the user (*a'* and *b'*) will be successively increased by a factor of *track bias*.

$$M = \begin{cases} 1.0 \Rightarrow a,b = a',b' + track\_bias \\ .75 \Rightarrow a,b = a',b' + 2 \cdot track\_bias \text{ (fuzzy only)} \\ .50 \Rightarrow a,b = a',b' + 4 \cdot track\_bias \text{ (fuzzy only)} \end{cases} \qquad (17)$$

Association between track and template will be made for any of the three possible values of *M* (1.0, 0.75, 0.5). On the template editor interface, the user is able to see which track is associated with which template. Membership values are also presented, informing the user on the degree of possibility of the association between a specific track and template.

6 Data management

When the software prototype performs fusion level 1 as described in references [2,3], resulting tracks are modelled (as described in Section 2) and stored in track files.

Templates can be stored in the software prototype configuration file or in track files when track to template fusion occurs. Fig. 6 presents the data processing for track-to-template fusion. In step 1, track models are loaded from track files. Templates can also be loaded from track files if fusion has been performed previously or loaded from the configuration file if they have been saved from a previous session. In step 2, the user can use the template editor interface to create, modify or delete templates. Modifications are reflected on the template representations on the software prototype interactive map. When template edition is completed, the user can start the track-to-template fusion process in step 3. All track models loaded in step 1 and all activated templates are sent to the Cortex knowledge base system [8] on which the software prototype is built. Cortex's agent will perform the association between track and template and will update track files by writing or erasing templates to complete the fusion in step 4. If a template has been fused with a track in a previous run and, is no longer associated with that track after it has been modified, it will be erased from the track file.

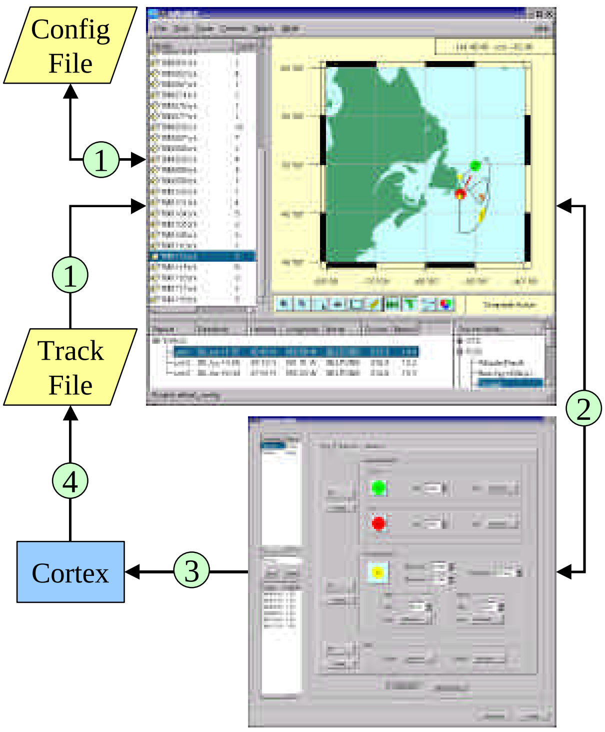

Fig. 6 Data processing for track to template fusion.

7 Conclusions

In this paper we have proposed an automated means to associate sensor based tracks with non-sensor based information and the means to manage these associations. This method has been integrated to a software prototype for maritime surveillance and tested with real data coming from coastal radar and surveillance aircraft.

This paper also demonstrated how templates representing ship behaviour could be created and modified by the user through a template editor interface.

Tracks were modelled with a genetic algorithm using sensor contact reports to produce parametrized line segments. Track-to-template association is performed by computing the intersections between both track and template mathematical models. Membership functions are introduced to indicate the level of uncertainty of templates when they are associated to tracks. For track uncertainties that are not of statistical nature, user can input an estimation of the bias. This bias, not usually reported by sensors, corresponds to a sensor's systematic errors and is an important factor in data fusion involving information from different sources. Especially for long range sensors, which are more sensitive to external conditions (atmosphere, sea state…) systematic errors may become larger than random errors (or statistical errors).

The template editor interface allows users to filter the information for a quick identification of track to template association. Graphical representation of tracks and templates can be viewed on an interactive map.

Automated association of sensor information with non-sensor information is an important functionality for surveillance systems, which can help them accomplish search and rescue, monitor specific regions and identify ship activities that may threaten environment or national security.

References

[1] Peter Avis. Surveillance and Canadian maritime domestic security. *Canadian Military Journal*, 4(1):9–14, 2003.

[2] Eric Lefebvre and Christopher Helleur. Multisource information adaptive fuzzy logic correlator. *Proc. 4th Annual Conf. Information Fusion*, volume II, ThB2-19 – ThB2-24, Montréal, Canada, 7-10 August 2001.

[3] Eric Lefebvre. Data fusion testbed for recognized maritime picture. *Proc. NATO ASI on Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management* , Tsaghkadzor, Armenia, 18–29 August 2003. (to be published).

[4] O. M. Mevassvik, A. Lokka. Fusion of Radar Tracks, Reports and Plans. *Proc. 3th Annual Conf. Information Fusion*, volume I, TuD1-10 – TuD1-15, Paris, France, 10-13 July 2000.

[5] B. Li, W. Jiang. A novel stochastic optimization algorithm. *IEEE Trans. Sys., Man Cyber.- Part B: Cybernetics*, 30(1), 193-198, 2000.

[6] Byron P. Roe. *Probability and Statistics in Experimental Physics.* Springer-Verlag, New-York, 1992.

[7] Steve Doré, Christopher Helleur and C.W. Van Fong. Scenario-based assessment of sensors for the Canadian recognized maritime picture. *Proc. Of the 7th ICCRTS*, Québec, Canada, 16-20 September 2002.

[8] Elisa Shahbazian, Jean-Rémi Duquet, Maciej Macieszczak and Pierre Valin. A generic expert system infrastructure for fusion and imaging decision aids. In Mark Bedworth and Jane O'Brien editors, *Proc. Int. Conf. Data Fusion*, pages 167-174, Great Malvern, UK, 6-7 October 1998.