Multi-Sensor Fusion using an Adaptive Multi-Hypothesis Tracking Algorithm

Leon J.H.M. Kester

TNO Physics and Electronics Laboratory,
P.O. Box 96864, 2509 JG The Hague,
The Netherlands

ABSTRACT

The purpose of a tracking algorithm is to associate data measured by one or more (moving) sensors to moving objects in the environment. The state of these objects that can be estimated with the tracking process depends on the type of data that is provided by these sensors. It is discussed how the tracking algorithm can adapt itself, depending on the provided data, to improve data association. The core of the tracking algorithm is an extended Kalman filter using multiple hypotheses for contact to track association. Examples of various sensor suites of radars, electro-optic sensors and acoustic sensors are presented.

Keywords: Sensor fusion, tracking, data association, MHT.

1. INTRODUCTION

For fusion of dissimilar and distributed sensors a general and adaptive approach is desirable. This approach is examined at different levels. In chapter 2 the architecture is discussed. Here it is argued that in most cases it is preferable to incorporate fusion in the tracking process. In chapter 3 the implementation of the 'fuse while track' or contact fusion algorithm is described. Using multiple hypotheses for contact to track association improves the performance, especially for dissimilar sensors. Also, for dissimilar or distributed sensors, the benefits are discussed of an autonomously adaptable track estimation state. This is illustrated with two examples in chapter 4. One example concerns the fusion of a co-located radar and camera. Another example concerns fusion of radars and acoustic sensors in a network.

2. ARCHITECTURE

2.1 Single sensor signal processing scheme

In the customary signal-processing scheme, a number of processes is identified: pre-processing, detection, tracking and classification. The result of the last process is classified tracks that compose the recognized surface picture. In figure 1 the signal-processing scheme is presented.

The sequence of the processes is similar for different types of sensors. If all processes are applied, however, depends on the application. There may be no interest in recognizing the objects or there may not be any need to estimate the properties of the objects with a tracking process.

The reason that the tracking process is performed before the classification process is because in tracks the information of a number of contacts is integrated. This will lead to a better classification. Each track, for example, can contain the minimum and maximum velocity of the object, which excludes certain classes or the behavior of the object in time reveals features that improve classification results.

One might argue that if the classification process precedes the tracking process, classification results could improve tracking. If, for example, a contact is classified as a sailboat it can only be associated to the track of a sailboat (if there are no ambiguous classes). It is important to realize, however, that the task of each process is limited to reduce and transform the data in such a way that it is optimal for the next process. The process preceding the tracking process should therefore provide optimal information for the tracking process. This information may be a certain class but it is

favorable to use the features on which this classification result is based. A feature 'red' measured with a camera or 'modulation frequency' measured with a radar will be more useful to the tracking process than a classification result derived from these features. This is because a classification process is usually not a one to one mapping from features to classes and therefore destroys information. It is, however, conceivable that some kind of classification is performed with contact features. This is regarded as feature extraction, which is part of the detection process.
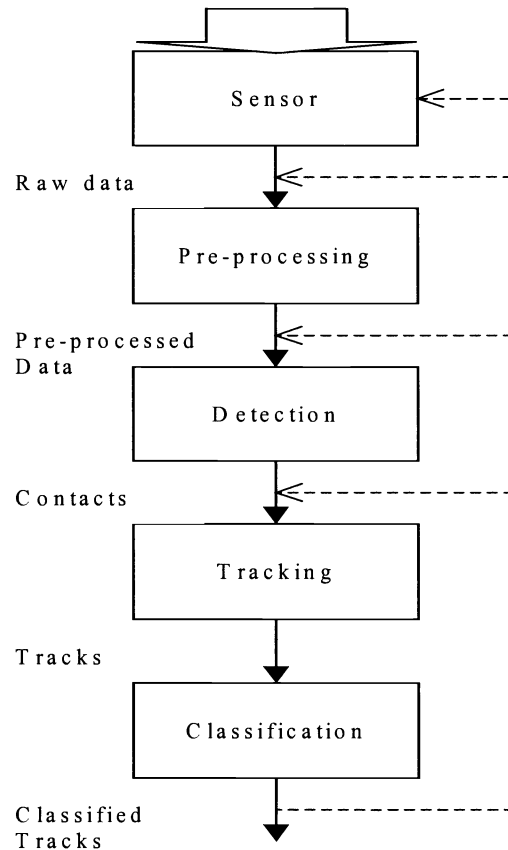

Figure 1: Scheme of the processes in the single sensor signal processing chain, including sensor and processing management.

Also shown in figure 1 is the sensor and processing management performed by feeding back the results from the processes to earlier processes and to the sensor itself. The more the system manages itself, the more intelligent the system must be since it must be aware of the implications of changing the algorithm or the sensor settings.

2.2 Multi sensor signal processing scheme

In the case of more than one sensor there are several data streams. The object is, however, to have a single recognized surface picture. Therefore, somewhere in the signal-processing scheme, the data streams have to be fused. In principle fusion can be performed at more than one level. In figure 2 a general signal-processing scheme is shown for more than one sensor, thus including fusion at every level.

In the single sensor scheme processes are optional. This also applies to the fusion processes. The dotted line between a fusion process and the subsequent process indicates that both processes can not always be separated. If the fusion process can be separated, however, it precedes the other process. The gradually smaller arrows indicate the decreasing number of independent data streams.

Similar sensors such as camera's that are co-located can be fused at any level since already the raw or pre-processed data are represented in the same data space. For dissimilar sensors like a camera and radar, fusion of raw data does not seem feasible. Trying to associate pre-processed data of both sensors will, apart from the profit, certainly demand a

large amount of processing power. Besides that, the sensors have to operate synchronously. The demand of processing power is diminished at the detection level since the detection process reduces the amount of data to only the regions of interest and subsequently the contacts. Fusion of contacts offers the possibility to associate data using features extracted in the detection process. The features that can be extracted depend of course on how sophisticated the sensors are. It is obvious that contacts of a multi-spectral EO-system provides more information and therefore its contacts contains more features than contacts obtained with a black and white camera. Likewise, contacts obtained with a coherent polarimetric radar will contain more features than contacts obtained with a non-coherent radar. As mentioned before, for an optimal performance of the multi-sensor system, each sensor must be sensitive to one or more common features to be able to associate the others. Although fusion of contacts before the tracking process seems feasible and useful, there are three disadvantages. The sensors have to operate synchronously or nearly synchronously, the fusion process does not use the measurement history and the process is sensitive to sensor failure.
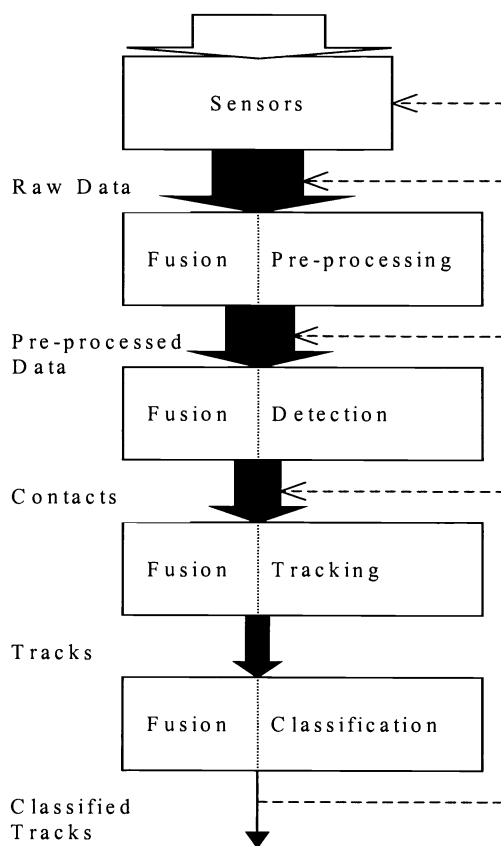

Figure 2: Scheme of the processes in the multi-sensor signal processing chain including sensor and process management.

A solution to all three problems is to integrate the fusion process in the tracking process. This process must be able to handle contacts from all sensor types that arrive asynchronously and integrate that to tracks represented in a global space. The process must be able to predict new measurements for each type of sensor. In this way the contacts are fused in tracks while the track history improves the association.

One step further down the chain is fusion of tracks. Although history can be used in the fusion process, the individual track processes of each sensor are less accurate since they were not aware of the information provided by other sensors. Moreover, track fusion is far from trivial since it has to deal with lost, broken, mixed up, ambiguous and false tracks from the individual track processes. Especially for dissimilar sensors, the results of these processes can be quite different.

It can be concluded that fusion integrated in the tracking process offers the best prospects. In the next paragraph we discuss how this can be implemented.

3. FUSE WHILE TRACK

Before examining how the fusion process can be incorporated in the tracking process, the tracking process itself is discussed.

The geometrical and dynamic features of the contacts are usually measured in the polar coordinate system of the sensor (measurement space). The tracks however, are usually defined in the Cartesian or geodetic space (state space), which is more suitable for situation assessment of maneuvering objects and which is independent of the sensor. If there is more than one sensor and the sensors are dissimilar or distributed the state space will differ from the individual sensor measurement spaces.

Note that the measurement space and state space in which the contacts and tracks are defined is not limited to the 3-D space in which the kinematical quantities are defined. Also features like temperature, signal modulation or radar cross section can be dimensions in this space. Conditions are that transformations between both spaces are well defined and that there is a model that is able to make predictions.

A general scheme for the tracking process is shown in figure 3.
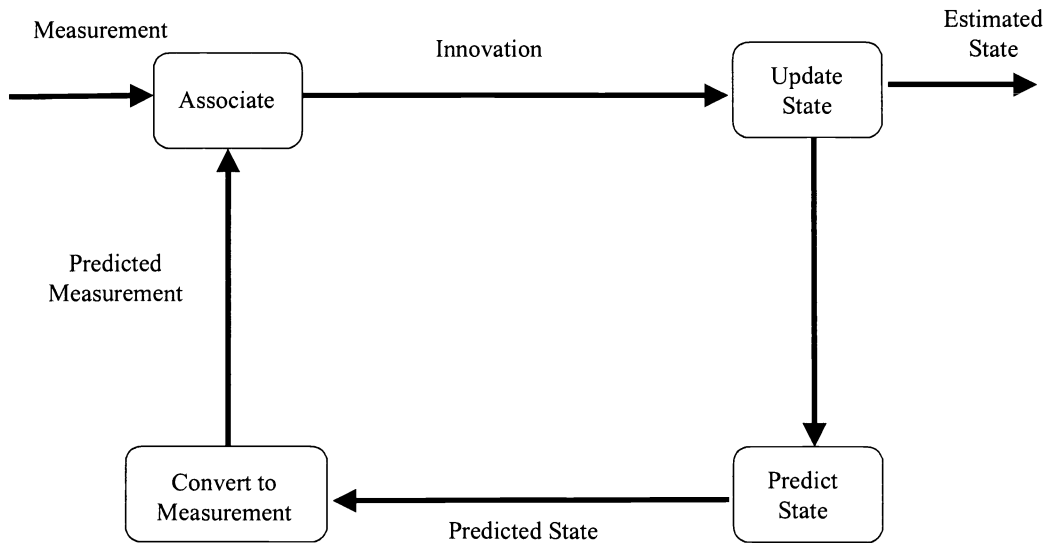

Figure 3: General scheme for the tracking process.

In this scheme contacts are associated to predictions from existing tracks in the measurement space. The result of this association is used to update existing tracks and possibly to create new tracks in the state space. In most single sensor applications, the time interval between measurements is fixed. Therefore predictions can be made as soon as the tracks are updated. Then the predictions are transformed to the measurement space. At the next point in time, when new contacts are measured, the process starts again.

How can this tracking process be used to fuse data from dissimilar or distributed sensors? In this case there are contacts defined in a sensor specific space while the results are global tracks, associated to objects in the state space. The prediction process does not have to be altered. It must however be supplied with the point in time at which the next measurement (by any of the sensors) has been performed. The transformation process must know to which measurement space the predictions have to be transformed.

Not only the state space usually differs from the measurement space, it may also differ from the prediction space since the state space is not necessarily the most suitable state to predict.

Figure 3 represents the situation where there is already an estimated state from which to a prediction can be made. However, when the track is initiated the estimated state can not contain more information than the measurement. Usually, when more measurements are associated to the track, the estimated state adapts itself to the desired state.

In figure 4, a scheme is shown that is analogue to the scheme of figure 3 but now specific for the extended Kalman filter. In this scheme also the conversions that may be performed are indicated.
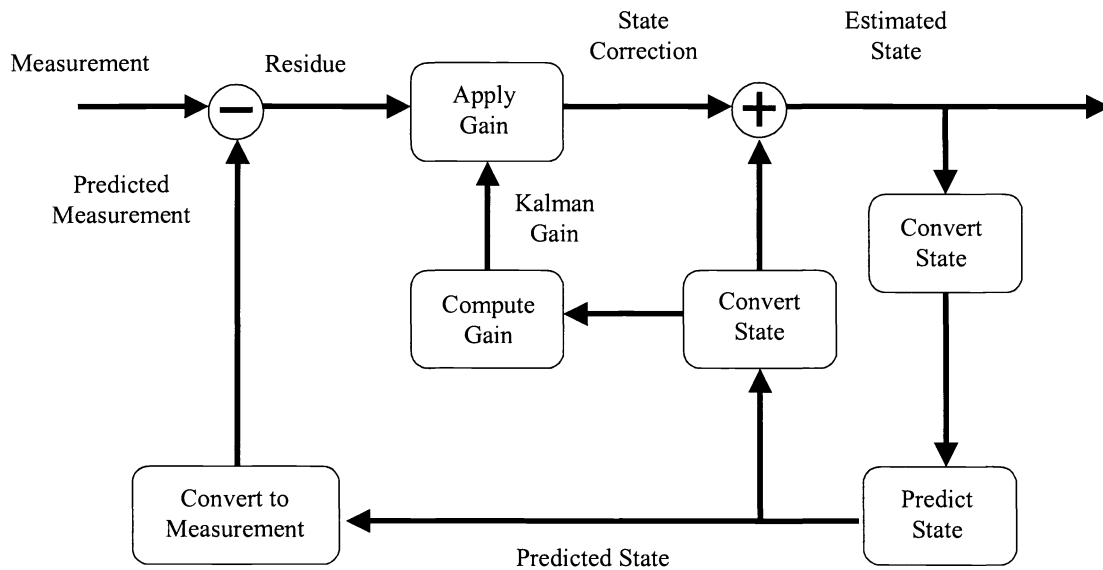

Figure 4: Scheme for the Kalman filter process with contacts originating from possibly different sensors.

The main reason for fusion of various sensors is to exploit the complementary data. However, to exploit the complementary data, the association of this data should be unambiguous. Therefore, the best performance is achieved if there is a balance between complementary data that provides more information about the environment and data common to all types of sensors to ensure that the information is from the same physical entity.

Another issue that needs to be addressed, is the association of contacts to predictions. Three possibilities are considered:
- Nearest Neighbor
  Every contact is uniquely associated to a track such that the overall likelihood is largest.
- Joint Probability Density Association (JPDA) [2]
  All contacts within the track gate are, weighted with their individual association probabilities, combined.
- Multi Hypothesis Testing or Tracking (MHT) [3]
  Decisions are delayed by retaining several association hypotheses. During the tracking process the number of hypotheses increases dramatically. At a certain stage the number of hypotheses has to be reduced by merging them and/or disregarding less probable hypotheses.

Data association is often uncertain due to the difference in measurement space, therefore delaying decisions by retaining more than one association hypothesis is favorable. The fuse while track algorithm should be able to process asynchronously arriving contacts. Since the time interval between successive measurements is variable, also for this reason a MHT approach seems more suitable. This is because the shorter the time interval, the more the association problems for all measurements are correlated. The two other methods do not take this into account.

To be able to calculate, on arrival of a measured contact, the hypotheses probabilities, the probability is needed of not detecting the object with any sensor in the time between the last and current measured contact. Accurate *a priori*

knowledge of detection probabilities in space and time are needed to estimate the probability of a missed detection. The probability that the measured contact is a false alarm might be derived from the measurements in which clutter is present.

To be able to evaluate the tracking algorithms a sensor fusion test bed (M6T, figure 5) is developed which considers multiple sensors from multiple platforms, multiple targets, multiple prediction models and multiple measured features and supports multiple hypotheses reasoning.
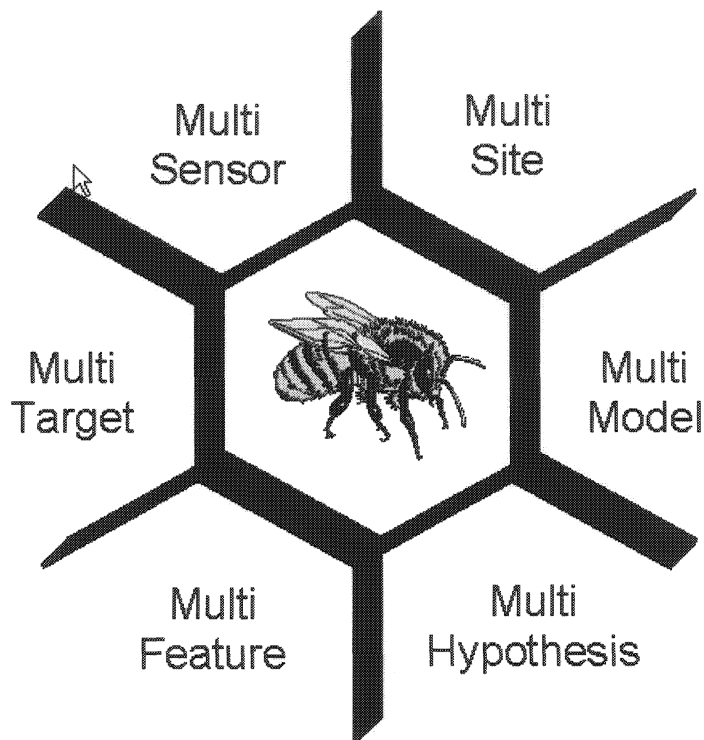

Figure 5: M6T test bed.

With the increase of complexity of the general tracking algorithm the flexibility or adaptability should also increase. The examples discussed in the next chapter focus on multi hypothesis tracking using an adaptive estimation state.

4. EXAMPLES

4.1 Radar and camera for surface surveillance

Measurements were performed with a radar and camera one the sea surface. The camera measures elevation and azimuth; the radar however measures range. Since the radar was staring no azimuth was determined. In figure 6 measurements and tracking results are shown for a vessel. In the top left panel the measurements of the radar are visible on a straight line at 2530 m. To show the measurements of the camera in the same plot, the range is derived from the measured elevation. Due to the much larger error the measurements of the camera are scattered around the measurements of the radar. At the bottom left panel elevation is shown versus time. Here, the elevation of the radar measurements is derived from the range. The picture is essentially the same, but the radar measurements are shifted with respect to the camera measurements indicating the estimated systematic error. In the top right panel the azimuth is shown as a function of time. To be able to plot the radar measurements the azimuth is assumed zero with a very large error.
The solution that is chosen in this case is track initiation on measurements of the radar since the number of false alarms from this sensor is small. The estimated state of the track therefore only includes range. After more than one

measurement radial velocity can also be included. Multiple hypotheses are generated for the associations of camera measurements whereby the estimated state changes to range, radial velocity and azimuth. After more than one camera measurement also velocity in azimuth is included.
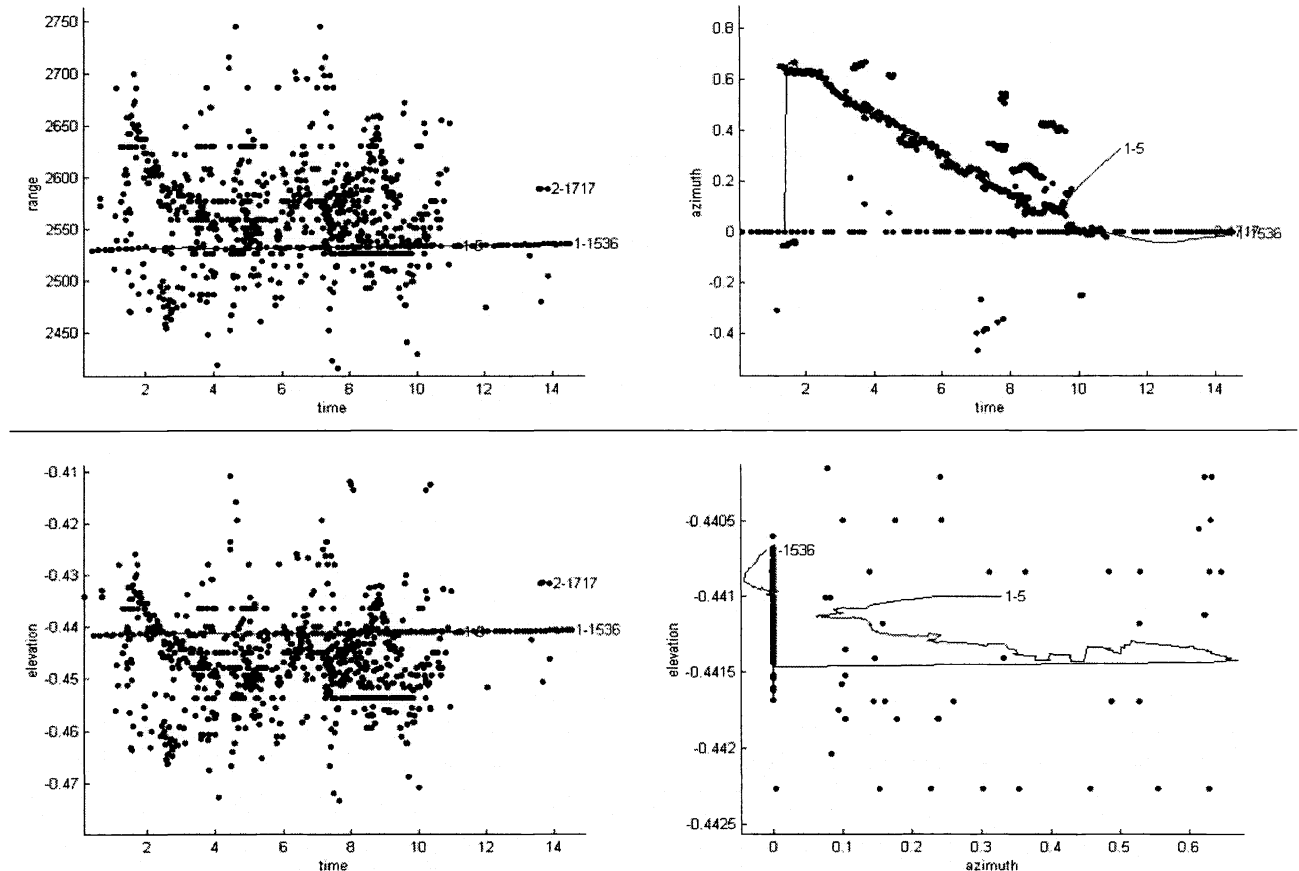

Figure 6: Measurements and tracking results of radar and infrared camera. At the top left panel range versus time, at the bottom left panel elevation versus time, at the top right panel azimuth versus time and finally at the bottom right panel elevation versus azimuth.

This is illustrated in figure 6 where tracks are shown which belong to the best hypothesis. In the upper left panel a track is started on the radar measurements at a range of 2530 m. In the upper right panel the azimuth value of the estimated state is determined by the azimuth values of the measurements from the camera. In the estimated state the range is hardly effected by the camera measurements since the accuracy of the predicted elevation is much better than the accuracy of the elevation measured with the camera.

If there is more than one object or clutter in the same area the risk is large that the association is ambiguous since there is no feature that is measured accurately with both sensors. Indeed at about 10 seconds the camera measurements of wake are associated to the track (1-5) instead of the measurements from the vessel. By the time the detections from wake stop the hypothesis is lost that associated the measurements of the vessel to the track. The original track is ultimately lost while a new track (1-1536) is started on the vessel measurements.

In the lower right panel of figure 6 elevation versus azimuth is shown for the measurements and tracks. Since the track accuracy in elevation (range) is determined by the radar measurements and the accuracy in azimuth by the camera measurements, the measurements of both sensors can hardly be related to the tracks. Also in this plot it is clear that the track is unjustly broken.

The association may be improved by retaining more hypotheses or including features in the estimated state which are correlated to features measured with the camera and features measured with the radar.

4.2 Network of Radars and acoustic sensors for ground surveillance

Simulations are performed for a network of cheap radars that only measure range combined with acoustic sensors that are able to determine the azimuth of a detected object. The situation is depicted in figure 7. How can the state be estimated of the object in the Cartesian coordinated system? A single measurement only gives either range or azimuth. If all options are kept open two hypotheses are generated. Either the measurement starts a new track or is a false alarm. When a subsequent measurement arrives from a different sensor this can be interpreted as the start of another track, again a false alarm or a measurement from the same object and therefore be associated to the track. If the contacts are both from a radar or from a radar and an acoustic sensor there are usually two solutions of the estimated state in the Cartesian coordinate system (see figure 7). One can continue tracking with the estimated state in this coordinate system, which is suitable for prediction, and generate two hypotheses for the two solutions or one can opt for a composite estimated state of two ranges (or one range and an azimuth value). Which choice is better depends on the uncertainties of the measurements and the a priori object track prediction models. Usually it is not a good idea to transform to the Cartesian coordinate system if the estimated state is non-linear over the uncertainty region. However, if the prediction models highly dependent on the direction in the Cartesian coordinate system, e.g. because a road is present within which the object motion is more probable, it may still be worthwhile to estimate the state in this coordinate system.
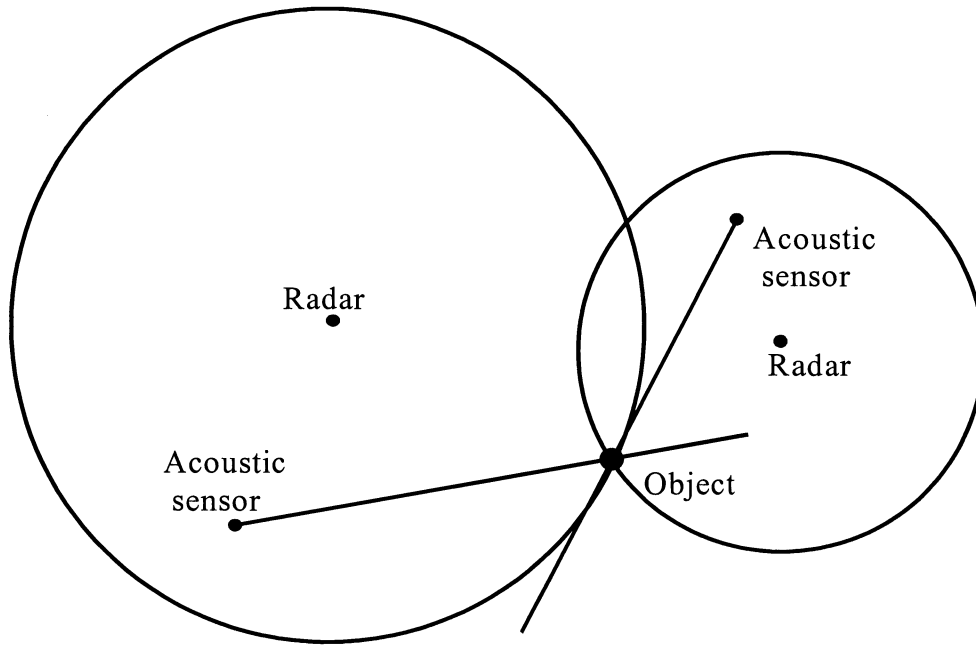

Figure 7: An object detected by a net of radars and acoustic sensors.

If the probability of the state of the object is not dependent on its position in the Cartesian system, the ambiguity in the estimated state will not be resolved by associating more measurements from the same sensors to the track. For this, a measurement of a third sensor is needed. This was demonstrated with tracking of simulated measurements. When two sensors were used two tracks with equal probability were generated in two hypotheses. If measurements from a third sensor were processed one of the hypotheses was eventually lost.

5. CONCLUSIONS

For possibly asynchronously operating (dissimilar) sensors, fusion before the tracking process is not attractive since the tracking process is the optimal process for interpolation and extrapolation in time. If fusion is performed after the tracking process, the sensor specific tracking processes are not aware of the information provided by other sensors. Besides that, a complicated track to track association problem has to be solved. Most appropriate therefore is integration of the fusion into the tracking process.

This fuse while track algorithm processes contacts on arrival. Predictions and transformation of the predictions to the sensor specific measurement space must be delayed until a contact arrives. Due to the variable time interval between successive measurements, an association algorithm that retains a number of hypotheses seems most suitable. Since the fuse while track process does not depend on the presence of measurements of specific sensors, it is flexible and not prone to sensor failure.

The combination of an adaptive estimation state and the option of multiple association hypotheses leads to various solutions. The analysis for the two examples shows that the estimation state should only contain information that is really measured. The elements of the estimation state should be chosen such that the correlation of the elements does not change too much over the uncertainty region. This may conflict with the interest to constrain the errors in the prediction process. The estimation and prediction state should be chosen such that the tracking process is most discriminative between genuine measurements and false alarms.

Multiple hypotheses can not only be retained for contact to track association but also for more than one possible discrete solution of the estimated state, derived from measurements of distributed sensors. Furthermore, the combination of multiple hypotheses and autonomous adaptive states leads to hypothesis clusters where these states may be defined in different dimensions.

6. REFERENCES

1. R.E.Kalman, A new approach to Linear Filtering and Prediction Problems, Trans. ASME, J. Basic Engineering, 82:34--45, March 1960.
2. Y. Bar-Shalom and X.R. Li, Estimations and Tracking: Principles, Techniques and Software, Artech House, Boston, 1993.
3. S. S. Blackman and R. Popoli, Design and analysis of Modern Tracking Systems, Artech House, Norwood, MA, 1999.