Annals of Operation Research manuscript No.
(will be inserted by the editor)

A bi-criteria approach for the data association problem

Hadrien Hugot[1], **Daniel Vanderpooten**[1], **Jean Michel Vanpeperstraete**[2]

[1] LAMSADE, Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75 775 Paris Cedex 16, France, e-mail: `{hugot,vdp}@lamsade.dauphine.fr`

[2] Thales Airborne System, 1, Boulevard Jean Moulin, 78 852 Élancourt Cedex, France, e-mail: `jean-michel.vanpeperstraete@fr.thalesgroup.com`

The date of receipt and acceptance will be inserted by the editor

Abstract The data association problem consists of associating pieces of information emanating from different sources in order to obtain a better description of the situation under study. This problem arises, in particular, when, considering several sensors, we aim at associating the measures corresponding to a same target. This problem, widely studied in the literature, is often stated as a multidimensional assignment problem where a state criterion is optimized. While this approach seems satisfactory in simple situations where the risk of confusing targets is relatively low, it is much more difficult to get a correct description in denser situations. This is why, we propose, for the first time to our knowledge, to address this problem in a multiple criteria framework using a second complementary criterion, based on the identification of the targets. Due to the specificities of the problem, simple and efficient approaches can be used to generate non-dominated solutions. Moreover, we show that the accuracy of the proposed solutions is greatly increased when considering a second criterion. A bi-criteria interactive procedure is also introduced to assist an operator in solving conflicting situations.

Key words data association – bi-criteria optimization – multidimensional assignment problem – interactive procedure.

Introduction

The data association problem consists of associating pieces of information emanating from different sources in order to obtain a better description of the situation under study. Typically, the situation to be described consists of different objects or targets that must be discriminated or localized. Sources are sensors which analyze independently the situation from several points of view or locations and provide information. Typical application domains are multiple: in military domains such as ground target tracking (Kirubarajan et al., 2000), radar tracking in the presence of electronic countermeasures (Blair et al., 1998), space surveillance (Bar-Shalom and Li, 1995), air traffic controls (Wang et al., 1999), coastal monitoring (Sevgi et al., 2001), or in civil domains such as air traffic controls, medical diagnosis (Kirubarajan et al., 2001) or environmental modelling (Hall and Llinas, 1997) to name only a few.

In most contexts, the data association problem consists of determining from which target a particular detection (sensor measurement) originated. Sensors in these systems (radars, ESM, infrared, etc) do not provide perfect information about the targets. In general, detections are imprecise mainly due to measurement noise. The problem of data association is the central problem of tracking multiple targets with a multisensor system. This problem, which is the original motivation of this work, has been studied extensively in the literature (see, *e.g.*, Bar-Shalom and Fortmann (1988), Bar-Shalom (1992) and Bar-Shalom and Li (1995)).

The data association problem was first stated as a set packing problem by Morefield (1977). Recent works, using the construction of Morefield, formulated the data association problem as a single criterion multidimensional assignment problem (see, *e.g.*, Pattipati et al. (1992), Poore (1994) and Popp et al. (2001)), where the criterion is based on state or kinematic estimation, such as position, velocity, range or angle measurements, and each dimension corresponds to a sensor. In many cases, however, this unique criterion does not allow a clear discrimination of the targets. This is mainly due to imprecision of the sensors. Obviously, confusion, corresponding to inaccurate association of targets, tends to increase when the density of targets increases (Poore and Rijavec, 1994). In such cases, an operator should intervene to solve the conflicting associations.

Generally, when there is ambiguity in using state information alone, the operator bases his/her judgement on identification information of targets. Identification information of targets, usually refers to a classification of targets according to their characteristics like length, type of radar used by the target or any relevant feature. Recently, Bar-Shalom et al. (2003) proposed to aggregate into a single criterion state and identification information using a combined likelihood ratio. This a priori aggregation between heterogeneous criteria into one overall criterion requires specific assumptions about the relative importance of each criterion. Moreover, it does not allow the operator to appreciate the conflicting nature of the criteria and to explore the possible tradeoffs between the criteria.

Our work differs in that we formulate, for the first time to our knowledge, the data association problem as a bi-criteria multidimensional assignment problem. We propose to use identification information separately in a second criterion. The inclusion of an additional complementary and uncorrelated criterion aims at reducing risks of confusion.

As confirmed by our experiments, situations of simultaneous confusion on both criteria appear only in very few cases. Moreover, instead of concealing this difficulty by providing only the solution optimizing one criterion, our approach allows the operator to evaluate the degree of conflict and possible tradeoffs between these two main criteria. An interactive procedure is proposed to find a compromise solution and to support the operator in his/her choice. The operator can use different sources of information to establish his/her judgement such as information from sensors and context-dependent information.

Our paper is organized as follows. In section 1, we formally define a bi-criteria model for the data association problem. Then in section 2, we introduce a simple and efficient method for generating the set of efficient solutions, and a bi-criteria interactive procedure making use of additional information. Finally, in section 3, we present computational results and illustrate the interactive procedure with a realistic example. Conclusions are provided in a final section.

1 The model

As indicated previously, the data association problem is usually formulated as a multidimensional assignment problem. In this paper, we propose to study the data association problem as a bi-criteria assignment problem based on two criteria: the classical state criterion and an identification criterion (see section 1.2). First, we need to introduce some notations. We consider a planar region where a target p is described by its cartesian coordinates $X_p = (x_p, y_p)$ and its identification class. We work with s sensors, and assume that there are n targets in the detection zone. The problem, which is made more complex because of noisy measures, consists of associating the observations from s lists of n measurements obtained from sensor 1 to sensor s as shown in Figure 1. An association of measurements from each sensor is called *elementary association*. A *global*

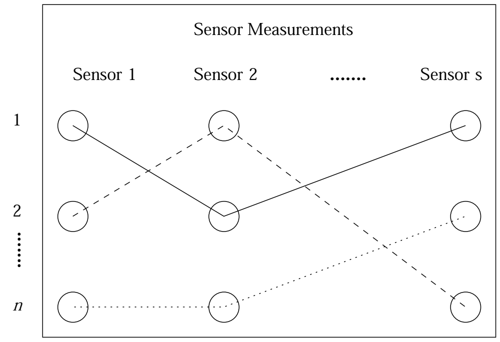


Figure 1 A data association problem

association is the set of all the elementary associations. Finally a gating technique, which drastically reduces the size of the model by eliminating unlikely elementary associations, is described in section 1.3.

1.1 Decision variables and constraints

An elementary association is defined as the association of one detection from each sensor, and is supposed to describe the same target. Formally, an elementary association can be represented by the following variables:

$$x_{i_1,\ldots,i_s} = \begin{cases} 1 \text{ if detections } i_1,\ldots,i_s \text{ coming respectively} \\ \quad \text{from sensors } 1,\ldots,s \text{ are associated} \\ 0 \text{ otherwise} \quad (i_k = 1,\ldots,n)\ ,\ (k = 1,\ldots,s) \end{cases} \tag{1}$$

In the data association problem, we require that each detection is used once and only once in elementary associations, which gives rise to the following constraints:

$$\begin{aligned} \sum_{i_2=1}^{n} \cdots \sum_{i_s=1}^{n} x_{i_1,\ldots,i_s} &= 1 \qquad (i_1 = 1,\ldots,n) \\ \sum_{i_1=1}^{n}\sum_{i_3=1}^{n} \cdots \sum_{i_s=1}^{n} x_{i_1,\ldots,i_s} &= 1 \qquad (i_2 = 1,\ldots,n) \\ \vdots \qquad\qquad & \vdots \qquad\qquad \vdots \\ \sum_{i_1=1}^{n} \cdots \sum_{i_{s-1}=1}^{n} x_{i_1,\ldots,i_s} &= 1 \qquad (i_s = 1,\ldots,n) \end{aligned} \tag{2}$$

1.2 Criteria

We define and formulate two criteria: the classical state criterion and an identification criterion. It should be pointed out that these criteria which refer respectively to the position and the type of targets, are very different in essence and provide complementary information to characterize the situation. In the following, we consider that sensors give for each detection a measurement consisting of state and identification data.

State criterion Measurements for the state criterion are in the form of kinematic information, such as azimuth, range, elevation, or position (Pattipati et al., 1992; Poore, 1994; Deb et al., 1997; Popp et al., 2001). Whatever the underlying information, the state criterion is based on the measurement of a distance between detections. In our case, we formulated the state criterion using positions only. Clearly, the bi-criteria model can be applied without any restriction to any type of kinematic information.

We consider here that each sensor provides state measurements in the form of two-dimensional vectors of cartesian coordinates. More precisely, we denote by $z^k_{i_k}$ the state measurement vector associated with detection i_k of sensor k. Due to imprecision of the sensors, state measurement vectors are noise-contaminated. Error measurement of sensor k is represented by covariance matrix Σ_k. Let $Z^k = \{z^k_{i_k}\}^n_{i_k=1}$ denote the set of all measurements from sensor k and $Z = Z^1 \times \ldots \times Z^s$ the set of all possible elementary associations in the entire detection zone. Consider the s-tuple of measurements $Z_{i_1,\ldots,i_s} \in Z$, corresponding to an elementary association. In our formulation presented in section 1.1, selecting a s-tuple $Z_{i_1,\ldots,i_s}$ amounts to set $x_{i_1,\ldots,i_s} = 1$. Therefore, our problem is to find the most likely feasible partition γ of the set Z into s-tuples $Z_{i_1,\ldots,i_s}$. Let Γ denote the set of all feasible partitions. Finally, the most likely partition is obtained by maximizing $L(\gamma)$, the joint likelihood function of all the measurements in partition γ:

$$\max_{\gamma\in\Gamma} \qquad L(\gamma) = \prod_{Z_{i_1,\ldots,i_s}\in\gamma} \Lambda\big(Z_{i_1,\ldots,i_s}|X_p\big) \tag{3}$$

where $\Lambda\big(Z_{i_1,\ldots,i_s}|X_p\big)$ is the likelihood that an s-tuple of measurements originated from detected target p with known state $X_p = (x_p, y_p)$. Thus we have:

$$\Lambda\big(Z_{i_1,\ldots,i_s}|X_p\big) = \prod_{k=1}^{s} p(z^k_{i_k}|X_p) \tag{4}$$

Probability $p(z^k_{i_k}|X_p)$ is assumed to follow a bi-dimensional Gaussian distribution $\mathcal{N}_2\left(X_p, \Sigma_k\right)$. In (4), state $X_p = (x_p, y_p)$ is in reality unknown, and hence it is replaced by its maximum likelihood estimate:

$$\widetilde{X_p} = \arg\max_{X_p} \Lambda(Z_{i_1,\ldots,i_s}|X_p) \tag{5}$$

Therefore, equation (4) becomes:

$$\widetilde{\Lambda}\big(Z_{i_1,\ldots,i_s}|\widetilde{X_p}\big) = \prod_{k=1}^{s} \frac{1}{|2\pi\Sigma_k|^{1/2}} \exp^{-\frac{1}{2}\left[z^k_{i_k}-\widetilde{X_p}\right]^T \Sigma_k^{-1}\left[z^k_{i_k}-\widetilde{X_p}\right]} \tag{6}$$

Furthermore, the maximization problem (3) is equivalent to the minimization of the negative log-likelihood and, using equation (6), the problem becomes:

$$\min_{\gamma\in\Gamma} \qquad -\ln\widetilde{L}(\gamma) = -\sum_{Z_{i_1,\ldots,i_s}\in\gamma} \ln\left(\widetilde{\Lambda}\big(Z_{i_1,\ldots,i_s}|\widetilde{X_p}\big)\right) \tag{7}$$

Hence, using equations (6,7), problem (3) can be stated as:

$$\min_{\gamma\in\Gamma} \sum_{Z_{i_1,\ldots,i_s}\in\gamma} c^1_{i_1,\ldots,i_s} \tag{8}$$

where

$$c^1_{i_1,\ldots,i_s} = \sum_{k=1}^{s} \quad \ln|2\pi\Sigma_k|^{1/2} + \frac{1}{2}\left[z^k_{i_k} - \widetilde{X_p}\right]^T \Sigma_k^{-1}\left[z^k_{i_k} - \widetilde{X_p}\right] \tag{9}$$

This can be reformulated, under the constraints that the partition is feasible (see subsection 1.1), as the following objective function:

$$\min \mathcal{Z}_1(\mathbf{x}) = \sum_{i_1=0}^{n}\sum_{i_2=0}^{n}\ldots\sum_{i_s=0}^{n} c^1_{i_1,\ldots,i_s}\; x_{i_1,\ldots,i_s} \tag{10}$$

Identification criterion Before defining this criterion, it is important to distinguish clearly recognition and identification. While object recognition deals with recognizing the precise object identity, object identification deals with assigning objects to classes. In our problem, we use identification information and not recognition information to build up the criterion.

It follows that using an identification criterion requires to pre-define identification classes. In this application, we used classes corresponding to reference lengths. Thus, the criterion evaluates the likelihood that the detected lengths of an elementary association correspond to a same reference length. Other more sophisticated classifications could be used.

Similarly to the state criterion, a generalized likelihood which involves the target identification estimates for the candidate association is used to assign costs to each feasible s-tuple of measurements. If we consider the same notation as for the formulation of the state criterion, the most likely partition of the measurement set into s-tuples is obtained by maximizing the joint likelihood function of all the measurements in partition γ. The likelihood that an s-tuple of measurements $Z_{i_1,\dots,i_s}$ originated from target p with identification length l_p is:

$$\Lambda\big(Z_{i_1,\dots,i_s}|\ l_p\big) = \prod_{k=1}^{s} p(z_{i_k}^k|\ l_p) \tag{11}$$

Whereas in the formulation of the state criterion probabilities $p(z_{i_k}^k|X_p)$ are known, for the identification criterion it is not the case. We should consider a probability density function that approximates the sensor measurements. Hence estimation $\widetilde{l_p}$ of true length l_p is not a very accurate approximation. We can estimate in (11) the unknown identification l_p by its maximum likelihood estimate:

$$\widetilde{l_p} = \arg\max_{l_p} \Lambda(Z_{i_1,\dots,i_s}|\ l_p) \tag{12}$$

As for the state criterion, we formulate the problem as the global optimization of the costs:

$$\min \mathcal{Z}_2(\mathbf{x}) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\dots\sum_{i_s=1}^{n} c^2_{i_1,\dots,i_s}\ x_{i_1,\dots,i_s} \tag{13}$$

Where $c^2_{i_1,\dots,i_s}$ represents the identification cost of selecting s-tuple $Z_{i_1,\dots,i_s}$ as an association.

1.3 Gating technique

The number of variables $x_{i_1,\dots,i_s}$ can be reduced drastically by forbidding implausible elementary associations. In fact, some elementary associations are so unlikely that the corresponding decision variables are not even defined in the assignment problems to be solved. In practice, operators would immediately reject such associations. Therefore, a procedure, called gating technique (Blackman and Popoli, 1999), is applied on each elementary association, or equivalently on each variable, before formulating and solving the assignment problem. The gating technique performs a validation test on each variable in order to eliminate unrealistic associations. This means that the gating technique should be seen, primarily, as a way of ensuring that all proposed elementary associations are acceptable, rather than as a means of reducing the combinatorial nature of the problem, even if it has clearly a strong impact on this aspect.

For the state criterion, the validation test is based on the probability density function $p(z_{i_k}^k|X_p)$. Since $p(z_{i_k}^k|X_p)$ is normally distributed and follows $\mathcal{N}_2(X_p, \Sigma_k)$, the quadratic term $\left[z_{i_k}^k - X_p\right]^T \Sigma_k^{-1} \left[z_{i_k}^k - X_p\right]$ follows a χ^2 distribution. Hence, the validation test can be viewed as a binary hypothesis problem with the two hypotheses:

H_0 : the s-tuple $Z_{i_1,\dots,i_s}$ can be associated (only if it fits the state estimation well enough),

H_1 : the s-tuple $Z_{i_1,\ldots,i_s}$ cannot be associated.

We reject H_0 and accept H_1 according to the validation criterion:

$$\sum_{k=1}^{s}\left(\left[z_{i_k}^{k}-\widetilde{X_p}\right]^{T}\Sigma_k^{-1}\left[z_{i_k}^{k}-\widetilde{X_p}\right]\right)>\mu \tag{14}$$

where $\left[z_{i_k}^{k}-\widetilde{X_p}\right]$ is the difference between the measurement and its estimate and Σ_k is the covariance matrix representing error measurement of sensor k. The parameter μ is obtained from tables of χ^2 distribution. For example, in a problem with 2 sensors (implying 4 degrees of freedom), $\mu = 13.277$ corresponds to a validation region with a confidence of 99%. All the variables whose validation cost is higher than μ are eliminated. In some cases, the resulting assignment problem contains only 5% of the initial variables. A gating technique could also be envisaged on the identification criterion.

2 Solution of the model

We propose in this section a methodology assisting the operator in his/her characterization of the situation under study, based on the model presented in the previous section. In a bi-criteria context, the most simple and informative way of presenting candidate solutions and possible tradeoffs between these solutions is to generate and display the efficient frontier. We use a simple technique, based on a classical ε-constraint approach, that performs well in our case due to some specificities of the data association problem (see section 2.1). Using information and graphical displays of the efficient frontier, at the level of global and elementary associations, but also additional context-dependent information, the operator can react on the current proposal by providing different types of modifications on the model, which can be solved iteratively. This interactive procedure is described in section 2.2.

2.1 Generation of efficient solutions

Solutions are evaluated over two criteria. Due to the conflicting nature of the objectives, an optimal solution that simultaneously minimizes all criteria is usually not obtainable. This is why we are interested in efficient solutions, that have the property, that no improvement on any objective is possible without sacrificing on the other objectives. Formally, a global association a is efficient (or non-dominated) if and only if there is no other feasible global association b which is at least as good as a on all the criteria and strictly better than a on at least one criterion. Therefore, we would like to generate only the set of efficient global associations. Unfortunately, the number of efficient solutions in a combinatorial problem tends to increase rapidly with the problem size and the number of criteria ; it can even be exponential in the size of each of these parameters (see, e.g., Serafini (1986) and Ehrgott (2000) about the intractability of multi-objective combinatorial optimization problems). Also the single objective assignment problem with at least 3 dimensions is NP-complete (Garey and Johnson, 1979). In the multiple objective case, NP-completeness holds, even in the case of 2-dimensional assignments problems, for the problem of deciding if a given solution is efficient (Serafini, 1986). The associated counting problem is #P-complete (Neumayer, 1994). To handle these problems, metaheuristic techniques, which generate an approximate set of efficient solutions, have been developed for bi-criteria assignment problems in Tuyttens et al. (2000) using simulated annealing or in Gandibleux et al. (2003) using a genetic algorithm. Exact algorithms have been presented in Bhatia et al. (1982) and Ulungu and Teghem (1994) to generate the whole set of efficient solutions. Other methods are presented by Ehrgott and Gandibleux (2000) in their survey.

Table 1 Generation of the efficient frontier by an ε-constraint approach

Initialization: $k \leftarrow 1$, generate $\mathbf{x}^1$, solution of:

$$\begin{array}{ll} \min & \mathcal{Z}_1(\mathbf{x}) + \alpha \mathcal{Z}_2(\mathbf{x}) \\ \text{s.t.} & (2') \end{array} \tag{15}$$

where α small positive value,and $(2')$ corresponds to assignment constraints (2) and, possibly, additional constraints derived from the interactive procedure.

Repeat

$k \leftarrow k + 1$

Generate $\mathbf{x}^k$, solution of:

$$\begin{array}{ll} \min & \mathcal{Z}_1(\mathbf{x}) + \alpha \mathcal{Z}_2(\mathbf{x}) \\ \text{s.t.} & (2') \\ & \mathcal{Z}_2(\mathbf{x}) \leq \epsilon^k \end{array} \tag{16}$$

where $\epsilon^k < (1 - \beta)\mathcal{Z}_2(\mathbf{x^{k-1}})$, and β small positive value.

Until problem (16) does not admit a feasible solution.

In order to generate the set of efficient global associations, we use a simple algorithm for bi-criteria problems based on the ε-constraint approach (Chankong and Haimes, 1983). The main idea of the algorithm, described in Table 1, is to solve the model optimizing on one criterion while adding a constraint on the other criterion which expresses a requirement level and to iterate the process with an higher level. A small positive value α is used to discriminate weakly efficient solutions and guarantee that all generated solutions are efficient. This parameter α as well as the way of setting ε^k makes the approach quasi-exact in the sense that some efficient solutions, very close to the generated solutions, might be omitted.

With such a simple approach, computation times, for a specific instance, depend on two factors:

1. the number of efficient solutions in the frontier,
2. the time for generating a specific solution of the frontier.

In our case, the gating technique described in section 1.3 has a strong impact on both factors. Clearly, gating on a criterion reduces the number of values that can be taken by this criterion, which also reduces the number of efficient solutions. Moreover, on the second factor, each problem to be solved for generating a new solution has a limited number of variables (between 5 and 20 % of the initial number), which considerably reduces computation times. Our experiments, presented in section 3, will illustrate these points.

We just point out here that the situation may be rather different for general instances of bi-criteria 2-dimensional assignment problems. For comparison purposes, we solved with our approach the instances used in Ulungu and Teghem (1994) (instances between 20×20 and 50×50 variables) and Gandibleux et al. (2003) (instances between 60×60 and 100×100 variables)[1].

The results are reported in Table 2. Computation times were obtained using a Centrino 1.4GHz with 512Mb RAM using Opl Studio and Ilog Cplex 8.1 with default MIP options.

A first observation is that this approach seems practical for small instances up to 50×50 variables. The most striking point is that time for finding an efficient solution may widely vary, above all for the largest instances. In our case, as will be shown in our experiments, this time remains very stable and very low even for the largest instances. Indeed, for 100×100 data association problems the maximum time for generating one

[1] These instances are available at the following address:
`http://www.univ-valenciennes.fr/ROAD/MCDM/ListMOAP.html`

Table 2 Instances used by Ulungu and Teghem (1994) and Gandibleux et al. (2003)

$n \times n$		20	30	40	50	60	70	80	90	100
Total Time		2.27	8.86	21.17	49.93	68.72	185.48	308.70	2876.40	5089.82
Time for finding	Min	0.01	0.05	0.09	0.13	0.02	0.03	0.39	0.64	0.62
an efficient	Avg	0.04	0.10	0.17	0.31	0.54	1.07	1.58	15.06	22.82
solution	Max	0.10	0.23	0.40	1.78	1.99	9.41	22.95	352.02	1444.75
Size of efficient set		55	88	127	163	128	174	195	191	223

solution is less than half a second - see Table 5 - vs 1444.75 seconds for the 100×100 instance of Gandibleux et al. (2003).

Such results convinced us that, for data association problems, such a simple and quasi-exact approach was preferable to metaheuristic techniques generating approximations of the frontier.

2.2 A bi-criteria interactive procedure

We propose a bi-criteria interactive procedure to explore the different data association schemes and support the operator in determining satisfactory global association(s). Such a procedure alternates at each iteration a calculation phase, corresponding to the generation of efficient global associations, and a dialogue phase where the operator reacts on the current proposal and provides information on the tactical situation. This information is translated into various types of constraints which modify the problem to be solved at the next iteration. This procedure is iterated until the operator is satisfied with the final global association(s).

Since the calculation phase has been described in the previous section, we focus now on the dialogue phase. Concerning the presentation of the results it is important to recall that associations are described at two levels.

At the global level, the relevant display is the bi-criteria representation of the efficient frontier which gives a clear picture of the tradeoffs between the two criteria (see Figure 3). Given that the state criterion is often considered the most important one, the reference solution is the current global association optimal on the state criterion. The basic question for the operator is to evaluate whether improvements on the identification criterion can compensate deteriorations on the state criterion. While improvements to the current reference solution seem interesting, the operator is encouraged to continue his/her exploration.

At a finest level, it is important for the operator to evaluate the plausibility of elementary associations. A very useful result of the calculation phase is to indicate the frequency of each elementary association in the efficient frontier. Clearly, variables corresponding to elementary associations with high (resp. low) frequency could be set to 1 (resp. 0). This type of reasoning is particularly relevant in the very first iterations of the interactive procedure, *i.e.* when the size of the efficient frontier is large. Obviously, setting to 1 (resp. 0) only the variables corresponding to elementary associations which are always (resp. never) present in the efficient frontier will not give rise to a new frontier at the next iteration of the procedure (this may, however, reduce substantially computational times). It is thus necessary to consider that above (resp. below) a given threshold, e.g. 90% (resp. 10%), a variable could be set to 1 (resp. 0). The operator can validate this by consulting graphical displays showing the associations (see Figure 4).

When the size of the efficient frontier is relatively small, the operator should perform a finer analysis by considering confusion groups, i.e. groups of detections located in a small area for which several alternative associations could be reasonably considered. While the state criterion will be poorly discriminant in such cases, information on the identification can be extremely helpful, at least in order to forbid some elementary

Table 3 100 runs for scenarios with 2 sensors, 20 targets and 6 identifications classes (2D-20T6I)

Average number of confusions :		3		5		7		10	
Gating :		with	without	with	without	with	without	with	without
Total Time	Min	0.00	0.02	0.00	0.04	0.03	0.09	0.06	0.16
	Avg	0.03	0.51	0.07	0.56	0.19	0.68	0.48	0.97
	Max	0.14	2.30	0.26	2.69	0.53	2.59	1.35	2.85
Time to find an efficient solution	Avg of Min	0.004	0.023	0.007	0.022	0.010	0.022	0.012	0.023
	Avg	0.008	0.041	0.011	0.038	0.016	0.038	0.023	0.039
	Avg of Max	0.010	0.060	0.016	0.057	0.025	0.059	0.038	0.066
Size of the efficient set	Min	1.00	1.00	1.00	2.00	2.00	3.00	4.00	6.00
	Avg	3.41	12.66	6.59	14.72	11.82	17.86	20.46	24.74
	Max	14.00	52.00	21.00	55.00	32.00	50.00	56.00	66.00
Accuracy of single criterion solutions	Min %	50.00	50.00	40.00	40.00	20.00	20.00	20.00	20.00
	Avg %	86.65	86.65	73.00	73.00	59.65	59.65	49.25	49.25
	Max %	100.00	100.00	100.00	100.00	100.00	100.00	80.00	80.00
	Exact	14.00	14.00	2.00	2.00	2.00	2.00	0.00	0.00
Accuracy of bi-criteria solutions	Avg of Min %	83.10	58.40	67.90	53.00	54.15	47.20	44.05	40.15
	Avg %	87.36	71.14	77.25	67.86	69.10	65.21	61.06	59.11
	Avg of Max %	97.25	97.25	92.15	91.95	88.40	88.60	84.50	84.55
	Exact	77.00	77.00	48.00	48.00	27.00	27.00	18.00	18.00

associations. Graphical displays indicating identification class information for each detection (see Figure 4) support the operator in this procedure.

In addition to this information directly resulting from the calculation phase, the operator may also make use of context-dependent information and knowledge of the zone under study. For instance, physical or technical constraints should be respected (e.g. a 30-meter long target cannot be in a forest), or partial knowledge about the situation could be included (e.g. there should be an alignment of targets with a same identification). A priori information could also be used to forbid specific types of associations (e.g. associations between specific types of identification classes).

It should be pointed out that fixing some variables may involve the fixation of other variables (constraint propagation due to the assignment constraints) and, above all, will largely influence the frequency of the associations corresponding to the remaining free variables.

Finally it is important to observe that these modes of interaction do not require quantitative and complex information. The operator can thus easily add information to the model.

3 Experimental results and illustration

We first present extensive computational results on the use of the ε-constraint approach for generating the efficient set of bi-criteria assignment problems resulting of data association problems (section 3.1). Then we illustrate the interactive procedure on a typical example (section 3.2).

3.1 Experimental results for the generation of the efficient set

Experimental design Because of the gating technique, two instances with the same number of targets may vary in difficulty. In order to evaluate this difficulty, one classical measure is the density d. Density is given

Table 4 100 runs for scenarios with 2 sensors, 50 targets and 6 identifications classes (2D-50T6I)

Average number of confusions :		3		5		7		10	
Gating :		with	without	with	without	with	without	with	without
Total Time	Min	0.02	9.30	0.10	12.36	0.41	10.77	1.94	15.63
	Avg	0.21	19.60	0.77	23.55	2.00	25.33	6.23	35.43
	Max	0.85	37.56	2.84	48.20	6.34	47.71	16.72	70.96
Time to find an efficient solution	Avg of Min	0.021	0.147	0.029	0.158	0.037	0.158	0.047	0.152
	Avg	0.028	0.351	0.042	0.340	0.058	0.325	0.081	0.297
	Avg of Max	0.037	1.058	0.057	0.995	0.089	1.008	0.157	0.950
Size of the efficient set	Min	1.00	27.00	3.00	32.00	9.00	31.00	25.00	60.00
	Avg	7.31	55.85	18.56	69.30	34.55	77.90	76.93	119.30
	Max	29.00	125.00	68.00	125.00	90.00	140.00	227.00	248.00
Accuracy of single criterion solutions	Min %	70.00	70.00	56.00	56.00	44.00	44.00	28.00	28.00
	Avg %	87.58	87.58	76.42	76.42	64.70	64.70	51.74	51.74
	Max %	100.00	100.00	92.00	92.00	88.00	88.00	72.00	72.00
	Exact	5.00	5.00	0.00	0.00	0.00	0.00	0.00	0.00
Accuracy of bi-criteria solutions	Avg of Min %	86.52	48.20	75.40	46.70	63.40	44.80	49.96	41.30
	Avg %	90.45	73.19	83.77	71.41	76.65	68.59	67.31	65.48
	Avg of Max %	97.34	97.20	93.78	93.60	90.16	89.20	85.46	86.00
	Exact	48.00	45.00	20.00	25.00	5.00	5.00	2.00	0.00

by $d = N/n^s$ where N is the number of feasible elementary associations after using the gating technique and n^s is the number of all possible elementary associations (s denotes the number of sensors). Nevertheless, density presents the disadvantage to be incomparable for situations with different numbers of targets. For example, in a 100 targets scenario, with 2 sensors, a density of 0.1 expresses that the problem has 1000 variables, whereas in a 5 targets scenario a density of 0.1 is impossible to achieve since the minimal density for this problem is 0.2 corresponding a trivial problem with just one feasible global association. Thus, for computational experiments, we defined a more practical indicator, the Average Number of Confusions (ANC). The ANC represents the average number of associations to which a given detection can belong. This indicator is measured after the application of the gating technique. For instance, in a 100 targets scenario with 2 sensors, an ANC of 10 expresses the fact that each sensor detection can potentially be associated with in average 10 other sensor detections and therefore that the assignment problem has 1000 variables and 200 constraints. This ANC is very large since in realistic situations a detection can be confused with about 5 other sensors detections. Therefore, in our computational results we study scenarios with different ANC around 5: $3, 5, 7$ and 10. Realistic situations involve at most 100 targets ; scenarios involving 20 or 50 targets are more reasonable. In the following, we denote by sD-nTiI-c a scenario with s sensors, n targets, i identification classes and c confusions in average. When creating a scenario, we first locate the targets and then generate detections for each sensor, so as to satisfy the required ANC. Computational experiments were performed on a Centrino 1.4GHz with 512Mb RAM using Opl Studio and Ilog Cplex 8.1 with default MIP options. We tested three types of scenarios: 2D-20T6I, 2D-50T6I, and 2D-100T6I using four different ANC ($3, 5, 7$ and 10). Computational results for each three types of scenario are presented respectively in Tables 3, 4 and 5. Each scenario is solved over 100 randomly generated instances. Moreover, in order to appreciate the impact of the gating technique, each instance is solved with and without gating. Obviously, the results without gating are of no interest for solving realistic situations. In each table, we present three main types of information: computation times, size of the efficient set and accuracy of the solutions.

Table 5 100 runs for scenarios with 2 sensors, 100 targets and 6 identifications classes (2D-100T6I)

Average number of confusions :		3		5		7		10	
Gating :		with	without	with	without	with	without	with	without
Total Time	Min	0.30	143.18	1.15	446.06	6.00	335.62	17.67	466.23
	Avg	2.38	485.41	7.14	647.51	18.30	716.49	45.54	892.34
	Max	46.76	803.85	24.48	976.91	42.84	1795.24	111.69	1466.38
Time to find an efficient solution	Avg of Min	0.072	0.710	0.086	0.764	0.106	0.757	0.131	0.719
	Avg	0.119	2.185	0.113	2.094	0.146	2.000	0.197	1.803
	Avg of Max	0.403	10.050	0.277	12.353	0.240	11.630	0.397	12.944
Size of the efficient set	Min	3.00	79.000	10.00	217.00	41.00	176.00	72.00	249.00
	Avg	20.06	222.20	63.02	309.30	125.78	358.30	231.04	494.90
	Max	67.00	355.00	216.00	390.00	249.00	849.00	630.00	814.00
Accuracy of single criterion solutions	Min %	76.00	78.00	61.00	61.00	46.00	46.00	39.00	39.00
	Avg %	87.15	86.30	74.59	74.59	64.41	64.41	51.72	51.72
	Max %	98.00	95.00	89.00	89.00	79.00	79.00	66.00	66.00
	Exact	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Accuracy of bi-criteria solutions	Avg of Min %	86.38	43.90	74.19	44.40	64.01	41.20	51.21	38.60
	Avg %	90.69	72.00	83.69	73.21	77.91	69.59	69.23	63.77
	Avg of Max %	96.59	96.20	93.82	95.80	91.49	91.10	85.19	85.20
	Exact	12.00	20.00	4.00	0.00	0.00	0.00	0.00	0.00

Computation times We present the *(minimum/average/maximum)* total time, in seconds, required to generate the efficient set. As expected, the average total time increases with the ANC. But, for each scenario computation times remain very small, even in the worst case. The largest realistic scenario solved here has 1000 variables (scenario 2D-100T6I-10) and is solved in average in 45.54 seconds and in 111.69 seconds in the worst case. In order to evaluate the intrinsic difficulty of each problem (16), we present the *average (and average minimum/maximum)* time to find an efficient solution. It is noteworthy that the difference between the average maximum time and the average minimum time is very small, above all when using gating. This shows that, in our problem, the time to find an efficient solution is very stable unlike in arbitrary instances of assignment problems as underlined in Table 2. Observe also that without gating, total time, which depends on the size of the efficient set and also on the time to find an efficient solution, is considerably increased.

Size of the efficient set We present the *(minimum/average/maximum)* numbers of efficient solutions. Tables 3 to 5 show that increasing the ANC always increases the size of the efficient set. This is not surprising, since conflicting situations increase with the number of confusions. Minimum and maximum numbers of efficient solutions show that the size of the efficient set may vary a lot for a same scenario. As confirmed in comparing instances with and without gating, the size of the efficient set is reduced by the gating technique. This reduction of the efficient set implies a reduction of the total time, since, in our computational experiments, the average time to find an efficient solution remains relatively stable.

Accuracy of solutions We compare the accuracy of the global associations of the detections between the single criterion formulation and the bi-criteria formulation. Accuracy is evaluated using as a reference the *exact* global association which consists of all the *correct* elementary associations, where an elementary association is correct when it regroups all detections describing the same target. The single criterion solution is the solution of problem (15), which corresponds to an optimal solution for the state criterion (more precisely in case of alternative optima, it is the solution with the best value on the identification criterion). For the single

Table 6 100 runs for scenarios with 3 sensors, 50 targets and 6 identifications classes (3D-50T6I) with gating

Average number of confusions :		3	5	7	10
Total Time	Min	1.28	31.87	16.30	138.68
	Avg	18.25	73.75	126.97	304.60
	Max	31.92	214.67	274.81	467.75
Time to find an efficient solution	Avg of Min	1.159	1.220	1.347	1.594
	Avg	1.342	1.287	1.496	1.928
	Avg of Max	3.555	2.502	5.235	10.657
Size of efficient set	Min	1.00	25.00	11.00	74.00
	Avg	13.60	57.30	84.90	158.00
	Max	27.00	165.00	184.00	240.00
Accuracy of single criterion solutions	Min %	74.00	58.00	46.00	30.00
	Avg %	89.40	77.31	69.80	57.23
	Max %	100.00	90.00	88.00	73.00
	Exact	10.00	0.00	0.00	0.00
Accuracy of bi-criteria solutions	Avg of Min %	87.20	73.40	67.40	55.93
	Avg %	91.32	88.45	81.58	69.43
	Avg of Max %	98.37	95.45	92.28	87.40
	Exact	50.00	20.00	7.00	3.00

criterion formulation, we present the *(minimum/average/maximum)* percentages of exact global associations and the number of times that the exact global association coincides with the optimal solution. For the bi-criteria formulation, we obtain at each run a set of efficient solutions. For each of these solutions, we compute the percentage of exact associations. The accuracy of a given efficient set is then evaluated by considering the *minimum, average, and maximum* percentages over all solutions belonging to this efficient set. The presented results correspond to an average of these three values over 100 runs. We also give the number of times when the exact global association is present in the efficient set for the 100 runs.

First, we can observe that the exact global association is generally not the single criterion solution. This only happens for scenarios with few targets and small ANC. However, the bi-criteria model often includes the exact global association in the efficient set. Also the average number of correct associations in the bi-criteria case is always better than in the single criterion case, and the difference tends to increase with the increase of the ANC. The bi-criteria model is more robust when the ANC varies. In fact, the average percentage of correct elementary associations in the single criterion case is very close to the average minimum of correct associations of the bi-criteria formulation. This clearly expresses that the single criterion solution is more or less the worst solution of the efficient set, in terms of correct associations. Moreover in the bi-criteria case, the efficient set contains very good associations, since the maximum average percentage is at least 84% of correct associations, even in very confused scenarios.

Experiments in the 3 *dimensional case* For 3 sensors, we present computational results in Table 6 for 100 randomly generated instances of scenario 3D-50T6I for $3, 5, 7$ and 10 ANC. All instances are solved after using the gating technique. First, remark that the accuracy of the solutions is slightly better than previously, but computation times increase considerably. It is true that we could envisage using other sophisticated and more efficient methods, but the weak improvement of the accuracy of the solutions, does not counterbalance substantially increasing computation times. Observe also that the time to find an efficient solution increases compared with the 2-dimensional case but remains also quite stable.

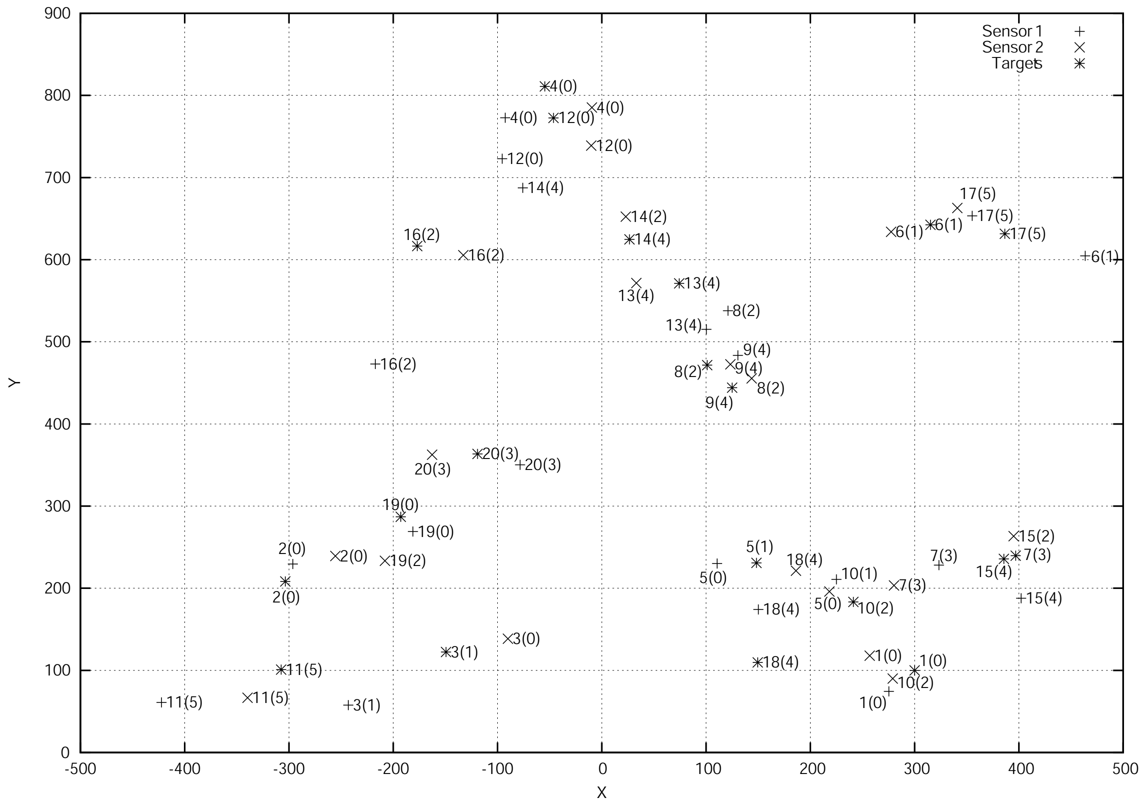
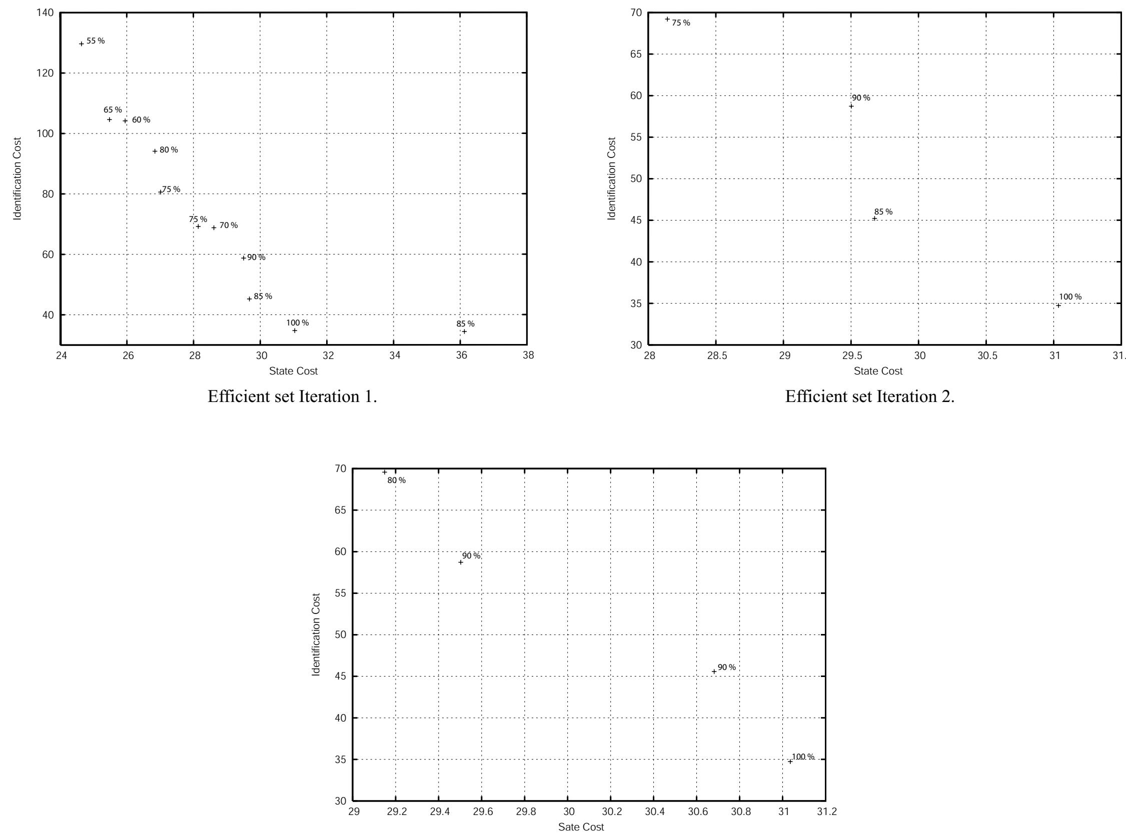


Figure 2 Location and identification of targets and sensor detections

3.2 An illustrative example

In this section we present various types of interaction on a realistic situation. We consider a scenario with two sensors, 20 targets, an average number of confusions of 7 and 6 identification classes (class 0: objects of length 1 meter, class 1: 2 meters, class 2: 6 meters, class 3: 10 meters, class 4: 15 meters and class 5: 20 meters). This leads to a 2D-20T6I-7 scenario, corresponding to an assignment problem with 140 variables and 40 constraints due to the required ANC. Figure 2 represents the locations of the targets and detections of the two sensors. Targets are represented by a $\ast$, sensor 1 detections are represented by a $+$, and sensor 2 detections by a $\times$. For each plot, the first number represents, for the targets, the target number and, for the sensors detections, the detection number; the second number, in brackets, represents in both cases the identification class. For more intelligibility when reading the results, detections take the same number as the target from which they are coming. We denote by (i, j) the elementary association of detection i from sensor 1 with detection j from sensor 2. This scenario is a representative scenario among 2D-20T6I-7 scenarios: the initial efficient set is generated in 0.18 seconds (the average time is 0.19 seconds) and contains 11 efficient solutions (the average number of efficient solutions is 11.82). Figure 3 gives a graphical display of the efficient set, according to the state cost and the identification cost. For each efficient solution, we report the percentage of correct elementary associations (obviously this information is not available to the operator). The reference solution is the optimal solution for the state criterion. We can remark, with respect to the reference solution, that a small increase in the state cost implies a significant improvement of the identification cost.

As indicated in section 2.2, the frequency of elementary associations in the efficient set is a useful information. As shown by the graphical display presented in Figure 4, 7 elementary associations (represented by continuous lines) are always present in the efficient set while 6 elementary associations (represented by



Efficient set Iteration 1. Efficient set Iteration 2.

Efficient set Iteration 3.

Figure 3 Efficient sets of the different iterations

dashed lines) appear with a frequency higher than 90%. A first observation is that all the high frequency elementary associations are correct associations. More interestingly, they correspond either to trivial cases (like association $(16, 16)$ corresponding to an isolated target) or slightly more complex cases involving simple groups of confusion, with few targets and rather consistent identification information. Our approach allows the operator to focus on these cases that he/she should be able to solve rather easily. A typical example is the group of confusion involving targets 6 and 17, in the upper right corner of Figure 4. The only global association of the efficient frontier containing elementary associations $(6, 17)$ and $(17, 6)$ is the solution optimizing the state criterion, whereas the identification criterion favors correct associations $(6, 6)$ and $(17, 17)$, which are present in all other solutions of the efficient frontier. Clearly, the operator could arbitrate easily, considering also that the difference in terms of state cost is negligible, whereas it is not on the identification cost. We assume that the operator accepts all the associations with a frequency higher than 90% and that corresponding variables are set to 1.

We compute the new efficient set, which is presented in Figure 3 (iteration 2). This new efficient set, consisting of 4 global associations, is contained in the previous efficient set. After this iteration, the remaining efficient solutions have much better values on the identification criterion, while values on the state criterion are nearly the same. On a new graphical display, where associations which have been accepted are removed, the operator might focus on more complex confusion groups. In our case, we are left with two confusion groups: a first group involving targets 1, 10 and 18 and a second group involving targets 8, 9 and 13 (see Figure 4). Regarding the first group, the acceptation of association $(5, 5)$ at previous iteration, implies a

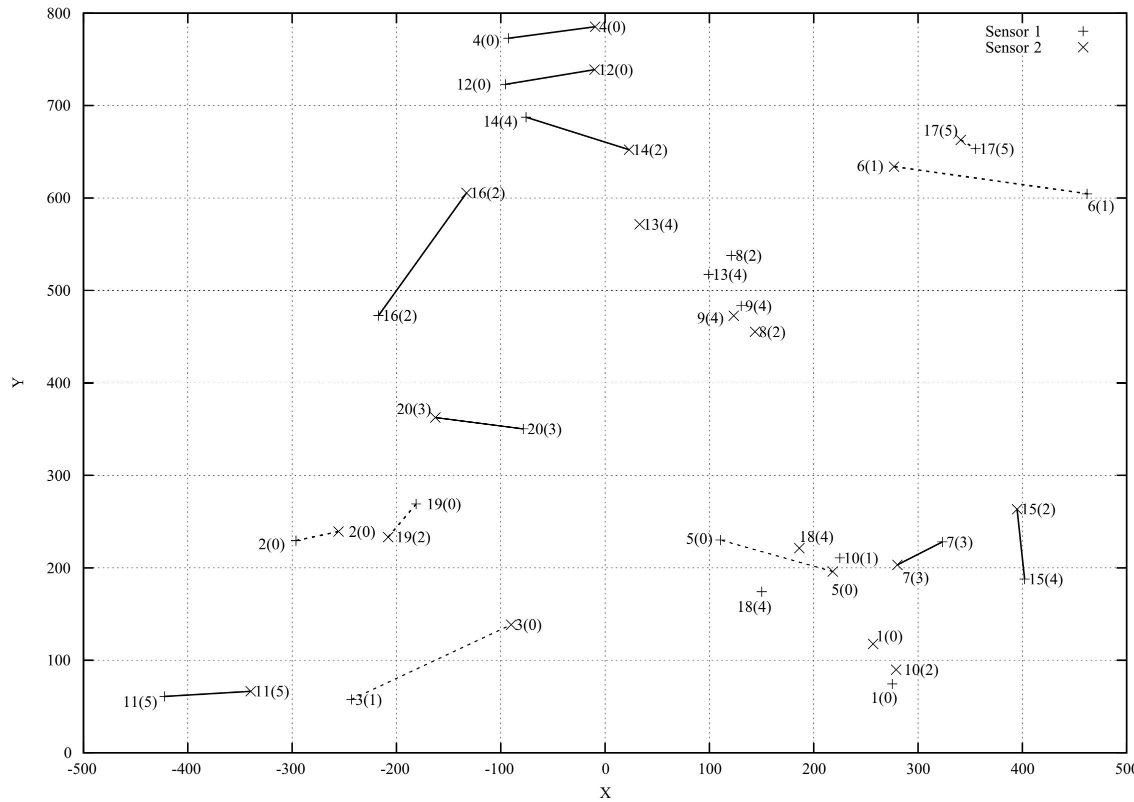


Figure 4 The high frequency elementary associations in the efficient set at iteration 1

frequency of 0.75 for elementary association $(18, 18)$. In addition, the identification class for these detections is, in both case, class 4. The confusion results from the existence of potential associations $(10, 18)$ and $(18, 1)$. But in both cases, identification does not fit. Therefore, the operator might accept elementary association $(18, 18)$. As to the second group, he/she might consider that detection 13 of sensor 1 and detection 13 of sensor 2 could be associated since their identification measure is identical. He/she might also consider that detection 8 of sensor 1 with identification class 2 (6 meters length) cannot be associated with detection 13 of sensor 2 with identification class 4 (15 meters length). Therefore, he/she discards elementary association $(8, 13)$. The new efficient set is presented on Figure 3 (iteration 3). This efficient set contains new efficient solutions which were not present in the previous iteration. The operator might continue examining residual confusion groups. Even if he/she were to stop at this stage, he/she would get a global association containing at least 80% of the correct elementary associations, a real improvement compared with the optimal solution on the state criterion which contains only 55% of correct elementary associations.

4 Conclusions

The goal of this work has been to formulate the data association problem as a bi-criteria problem. We showed that the use of a complementary criterion, based on identification, besides the classical state criterion really improves the quality of the solutions by revealing clearly the conflicting situations and providing the operator with relevant information. Computation times using a simple ε-constraint method, are quite satisfactory, at least in the 2-dimensional case. It seems also that using an additional sensor does not enhance significantly the quality of the solutions and requires substantial computational effort.

References

Bar-Shalom, Y., editor (1992). *Multitarget-multisensor tracking: advance application applications.* Artech house, Norwood, Massachusetts.

Bar-Shalom, Y. and Fortmann, T. E. (1988). *Tracking and data association.* Academic Press, New-York.

Bar-Shalom, Y., Kirubarajan, T., and Gokberk, C. (2003). Tracking with classification-aided multiframe data association. *Proceedings of SPIE, Signal and Data Processing of Small Targets*, 5204:131–142.

Bar-Shalom, Y. and Li, X. R. (1995). *Multitarget-multisensor tracking: principles and techniques.* Y. Bar-Shalom.

Bhatia, H. L., Malhotra, R., and Puri, M. (1982). Bi-criteria assignment problem. *Operations Research*, 19(2):84–96.

Blackman, S. and Popoli, R. (1999). *Design and Analysis of Modern Tracking Systems.* Artech House Radar Library Series.

Blair, W. D., Watson, G. A., Kirubarajan, T., and Bar-Shalom, Y. (1998). Benchmark for radar allocation and tracking in ECM. *IEEE Transactions on Aerospace and Electronic Systems*, 34(4):1097–1114.

Chankong, V. and Haimes, Y. Y. (1983). *Multiobjective decision making.* Elsevier Science Publishing, New York.

Deb, S., Yeddanapudi, M., Pattipati, K. R., and Bar-Shalom, Y. (1997). A generalized S-D assignment algorithm for multisensor-multitarget state estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 33(2):523–537.

Ehrgott, M. (2000). *Multicriteria optimization.* Lecture Notes in Economics and Mathematical Systems. Springer-Verlag.

Ehrgott, M. and Gandibleux, X. (2000). A survey and annoted bibliography of multiobjective combinatorial optimization. *OR Spektrum*, 22:425–460.

Gandibleux, X., Morita, H., and Katoh, N. (2003). Population based-heuristic for solving the assignment problems with two objectives. Technical Report 7, LAMIH-ROI.

Garey, M. and Johnson, D. (1979). *Computers and intractability: a guide to the theory of NP-completeness.* Freeman and Co., New York, San Fransico.

Hall, D. L. and Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6–23.

Kirubarajan, T., Bar-Shalom, Y., and Pattipati, K. R. (2001). Multiassignment for tracking a large number of overlapping objects. *IEEE Transactions on Aerospace and Electronic Systems*, 37(1):2–21.

Kirubarajan, T., Bar-Shalom, Y., Pattipati, K. R., and Kadar, I. (2000). Ground target tracking with variable structure IMM estimator. *IEEE Transactions on Aerospace and Electronic Systems*, 36(1):2–21.

Morefield, C. R. (1977). Application of 0-1 integer programming to multitarget tracking problems. *IEEE Transactions on Automatic Control*, 22(3):302–312.

Neumayer, P. (1994). Complexity of optimization on vectorweighted graphs. In Bachem, A., Derigs, U., Jünger, M., and Schrader, R., editors, *Operations Research*, volume 93, pages 359–361, Heidelberg. Physica Verlag.

Pattipati, K. R., Deb, S., Bar-Shalom, Y., and Washburn, R. B. (1992). A new relaxation algorithm and passive sensor data association. *IEEE Transactions on Automatic Control*, 37(2):198–213.

Poore, A. B. (1994). Multidimensional assignment formulation of data association problems arising from multitarget and multisensor tracking. *Computational Optimization and Applications*, 3:27–57.

Poore, A. B. and Rijavec, N. (1994). A numerical study of some data association problems arising in multi-target tracking. In Hager, W. W., Hearn, D. W., and Pardalos, P. M., editors, *Large scale optimization: state of the art*, pages 339–361.

Popp, R. L., Pattipati, K. R., and Bar-Shalom, Y. (2001). m-best S-D assignment algorithm with application to multitarget tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 37(1):22–37.

Serafini, P. (1986). Some considerations about computational complexity for multiobjective combinatorial problems. In Jahn, J. and Krabs, W., editors, *Recent advances and historical development of vector optimization*, volume 294 of *Lecture Notes in Economics and Mathematical Systems*, pages 222–232, Berlin. Springer-Verlag.

Sevgi, L., Ponsford, A., and Chan, H. C. (2001). An integrated maritime surveillance system based on high-frequency surface-wave radars. *IEEE Antennas and Propagation Magazine*, 43(4):52–63.

Tuyttens, D., Teghem, J., Fortemps, P., and Nieuwenhuyze, K. V. (2000). Performance of the MOSA method for the bicriteria assignment problem. *Journal of Heuristics*, 6(3):295–310.

Ulungu, E. and Teghem, J. (1994). The two-phases method: an efficient procedure to solve bi-objective combinatorial optimization problems. *Foundations of Computing and Decision Sciences*, 20(2):149–165.

Wang, H., Kirubarajan, T., and Bar-Shalom, Y. (1999). Precision large scale air traffic surveillance using IMM with assignment estimators. *IEEE Transactions on Aerospace and Electronic Systems*, 35(1):225–266.