

# Comparison of Belief Functions and Voting Method for Fusion of Mine Detection Sensors

Nada Milisavljević<sup>a,b</sup>, Sebastiaan P. van den Broek<sup>c</sup>, Isabelle Bloch<sup>b</sup>,

Piet B.W. Schwering<sup>c</sup>, Henk A. Lensen<sup>c</sup> and Marc Acheroy<sup>a</sup>

<sup>a</sup>Signal and Image Centre - Royal Military Academy,  
Av. de la Renaissance 30, 1000 Brussels, Belgium

<sup>b</sup>École Nationale Supérieure des Télécommunications,  
TSI – CNRS URA 820, 46 rue Barrault, 75013 Paris, France

<sup>c</sup>TNO Physics and Electronics Laboratory,  
P.O. Box 96864, 2509 JG The Hague, The Netherlands

## ABSTRACT

In this paper, two methods for fusion of mine detection sensors are presented, based on belief functions and on voting procedures, respectively. Their application is illustrated and compared on a real multisensor data set collected at the TNO test facilities under the HOM-2000 project. This set contains data acquired by metal detector, infrared camera and ground penetrating radar. The data acquisition and preprocessing are briefly described. For some typical cases presented in this data set, the characteristics extracted and used by both methods are discussed, as well as the answers given by each method and possible causes of potential differences in results. Also, it is shown how the different voting schemes compare to belief functions modeling in various situations, based on the knowledge that is put into the belief functions. Since the roots of the two methods are different, i.e. belief functions involve expert knowledge while voting is a simple approach, the explanations involve these differences. Problems that arise when comparing and evaluating different methods are also addressed. Finally, it is shown that both of the methods have their advantages and drawbacks, depending on the measurement and operational conditions. This paper is a result of a joint work at three European institutions towards a common goal: humanitarian demining.

**Keywords:** Humanitarian mine detection, fusion, belief functions, voting

## 1. INTRODUCTION

Due to the great diversity of types of AP mines as well as of situations where they can be found, there is no single sensor solution of the overall humanitarian mine detection problem. A sensor that works well in one situation and for one type of mines fails to detect them in another situation. Therefore, a promising way to go is to combine several complementary sensors. We discuss here possibilities to fuse three sensors often used in this domain of application: an infrared camera (IR), a metal detector (MD) and a ground penetrating radar (GPR). One of the main goals of our work is to analyze how much can be gained in the final fusion results once the knowledge about the problem is taken into account in comparison to a simple method. To this purpose, we analyze two fusion methods, voting and belief functions. In brief, voting is a simple method that does not ask for inclusion of the existing knowledge (about the types of AP mines, mine laying principles, sensors, etc.) The belief function method fuses sensors on feature level. It is based on selecting and extracting discriminative measures coming from the three sensors and modeling them, taking into account the knowledge about the problem and information contained in each of the measures. Because of these fundamental differences between the two methods, comparison of their fusion results must be performed carefully and loosely. Each method is applied on a multisensor data set collected at the TNO test facilities under the HOM-2000 project. For comparing the methods, the belief function and voting fusion start from the same alarms. It can be expected that their results are the same if there is no knowledge included in the belief function model.

The paper is organized as follows. Some characteristics of the used data set and of the applied preprocessing are given in Section 2. The voting fusion is discussed and the results are shown in Section 3, while Section 4 contains a description of our belief function method as well as the obtained results. In Section 5, results obtained by the two methods are compared and some specific cases discussed in detail. Finally, conclusions are given in Section 6.

The paper is a result of a joint work at three European institutions within both the Dutch HOM-2000 and the Belgian HUDEM projects.

## 2. DATA ACQUISITION AND PREPROCESSING

Here, we briefly discuss only those characteristics of the data set as well as data acquisition and preprocessing that are important from the point of view of the further analysis and fusion. The preprocessed data, resulting from this step, are used as starting points for the two fusion methods. For the comparison between the methods, the same settings for obtaining alarms from the preprocessed data are used as well.

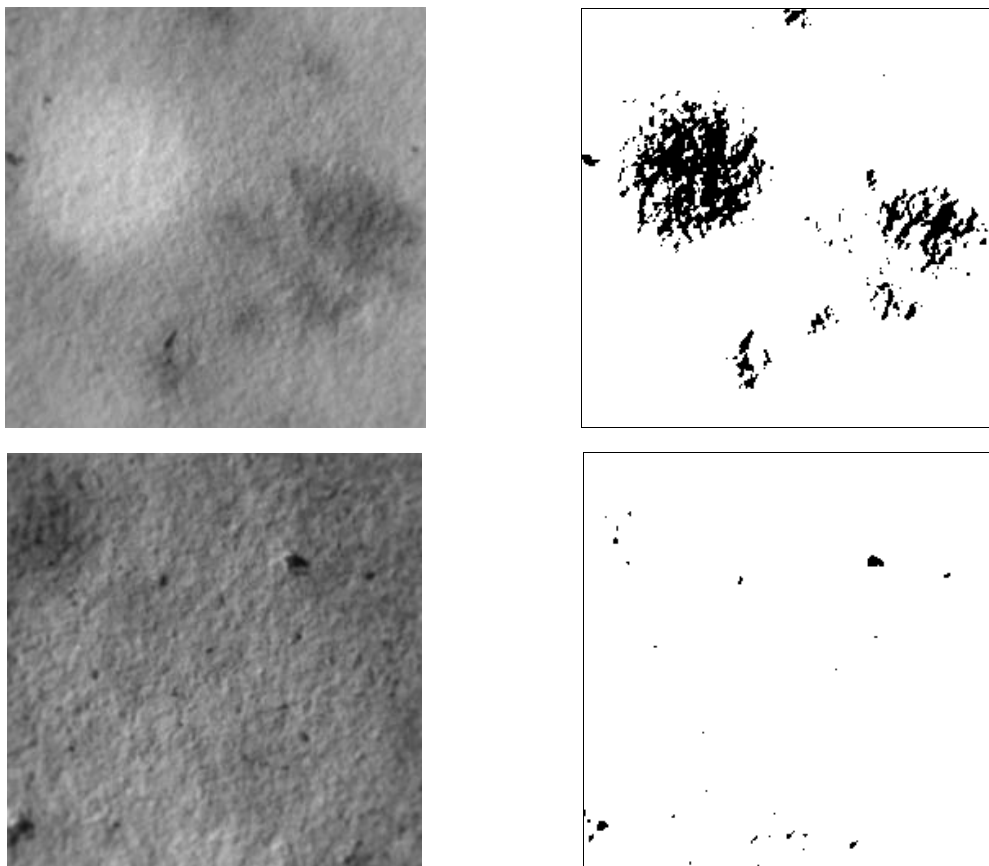


Figure 1: Top: A raw IR image of an NR24 with ring, just below the surface (left), result of region selection (alarming) for  $f=3$  (right); bottom: a raw IR image of an NR24 without ring, 10 cm below the surface (left), resulting alarm for  $f=3$  (right).

### Description of the test lane

The used data were acquired in a part of the lane filled with sand at the TNO test facilities<sup>1</sup>. This lane contains both anti-tank (AT) and anti-personnel (AP) mines, placed false alarms and empty areas. The mines are either dummies or mines from which all explosives are removed. The area that is analyzed contains several types of AP mines buried at various depths. The AP mines range from small mines containing no metal at all to larger mines and mines with metal detection rings. The area reserved for one object is approximately  $50 \times 45 \text{ cm}^2$ . Objects are placed in lines. Because no GPR data are

available for the area of the lane containing the surface-laid objects, a smaller part of the lane is analyzed, that includes 14 buried mines and one surface-laid mine, and two iron parts buried at two different depths.

## IR

The acquired IR data consist of separate images of size  $31 \times 31 \text{ cm}^2$ , that partly overlap and cover the whole sand-lane. The size of one pixel is  $1.21 \times 1.21 \text{ mm}^2$ . We analyze images one by one. Since the acquired IR data are of very good quality, and since we are looking for fast, realistic, automatic and simple solutions, the alarming or the selection of regions that possibly contain mines (i.e. which should be further analyzed) is performed directly on the raw IR data. IR region selection has to take into account the daily evolution of the IR signature<sup>2</sup>, meaning both the inversion of contrast as well as various levels of contrast between a mine and the background. A simple and reliable solution is to estimate the background for each image separately, from some preset number of the bordering pixels around an image, small enough not to include pixels belonging to a mine, and still large enough to be a statistically reliable estimation. In case of daylight measurements, the thermal clutter is quite strong, so the background can significantly change from one IR image to another, making the idea of background estimation from bordering pixels even more attractive. Therefore, we find the mean of the background area,  $m_{back}$ , as well as the standard deviation of the background,  $s_{back}$ , and keep only the pixels that fulfill this condition:

$$|v - m_{back}| > f \cdot s_{back}, \quad (1)$$

where  $f$  is a factor that can be varied, depending on the situation. Two examples, obtained for  $f=3$ , are given in Figure 1.

## GPR

The direction of collecting B-scans was along the lines where objects were placed. The acquisition step in the cross-track direction was 10 cm, while it was around 9 mm in the scanning direction. In depth, representing wave travel-time proportional to the depth, 256 samples were gathered, with sampling time of 50 ps.

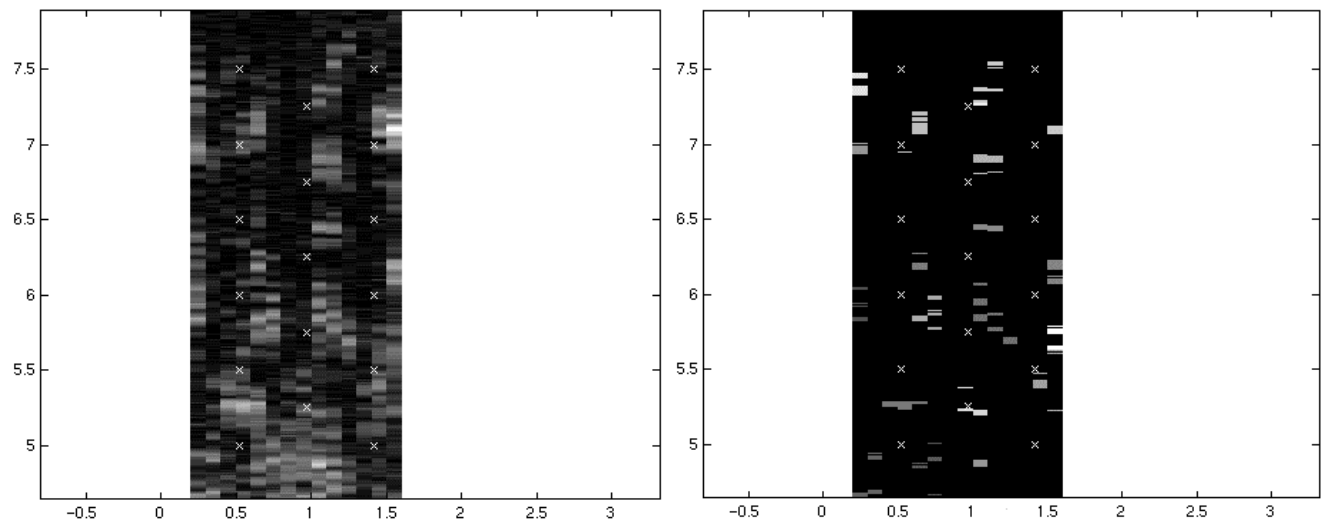


Figure 2: A 2D map related to the energy projection of the GPR data (left), result of the proposed region selection method for  $t=30$  (right); the ground truth is indicated by "x" marks.

The preprocessing of the data<sup>3</sup> consists of:

- 1) removing background (sliding window within which the mean of A-scans was calculated and subtracted from the central A-scan);
- 2) taking the absolute values for each point in the A-scan;
- 3) applying a unique threshold on each A-scan value, and counting how many samples exceed it.

These steps result in a 2D image presenting a map of a value related to the energy projection to the surface (see the left side of Figure 2). Still, deeper buried objects have weaker signals and, accordingly, weaker energy<sup>2</sup>, so it could happen that they would not be detected if the GPR region selection were a simple thresholding. In order to preserve weaker signals, without causing a strong increase in false alarming, we propose a new way for the selection of alarms<sup>2</sup>. It consists of finding the local maxima exceeding a threshold  $t$  and grouping together all the points around each of them, that have a value close to

that local maximum. By that, we obtain a region or “blob” around every local maximum, as shown in the right side of Figure 2, for  $t=30$ .

## MD

This data set was collected using a rigid MD array<sup>2</sup>, covering a wide area, hence improving the speed of data acquisition. The step in the scanning direction (along the lines with objects, as for GPR) was 4 mm, while the step in the cross-track direction was 12.5 cm. We face some problems with these data<sup>2</sup>:

- responses of a few objects are joined, due to the fact that they are large metallic and too closely spaced;
- one large metallic object may hide a smaller one in its neighborhood or make it appear that a non-metallic object is seen;
- saturation of the sensor, which often occurs with MDs, due to the fact that it cannot be known for sure which sensitivity to choose if there is no certain knowledge on types of objects, their depths, etc.;
- negative signals, appearing near a large metallic object as a result of the principle of operation of the applied MD array; due to the arrangement of the lane, this signal may be combined with signals of other objects, making the data hard to interpret; as a result, signals from other objects may become invisible, and the shape of the signal may appear to indicate the presence of metal where there is none, as it introduces more local maxima.

Although all these effects may also be seen in a real minefield, the presented data are a rare case, which provides some problems in drawing conclusions for more common situations.

Within the preprocessing part, a metal-free region in the lane is used for background estimation. Its mean is taken, as well as the maximum value of the difference between pixel values in that region and the mean, representing noise. From the whole image, this mean is subtracted. Then, just pixels exceeding noise by some preset factor  $p$  are preserved, and the result (see the left side of Figure 3,  $p=3$ ) was the input of the region selection step. Due to the specificities of the acquired data, mentioned above, a local maximum method<sup>2</sup> is used with  $r$  being a threshold on local maxima value. This step results again in blobs which present alarmed regions (see the right side of Figure 3, where  $r=0$ ), that will be further analyzed in each of the two fusion frameworks, together with the regions of the other two sensors. To solve the problem of large blobs near local maxima due to saturation, a window is used to limit the size of blobs.

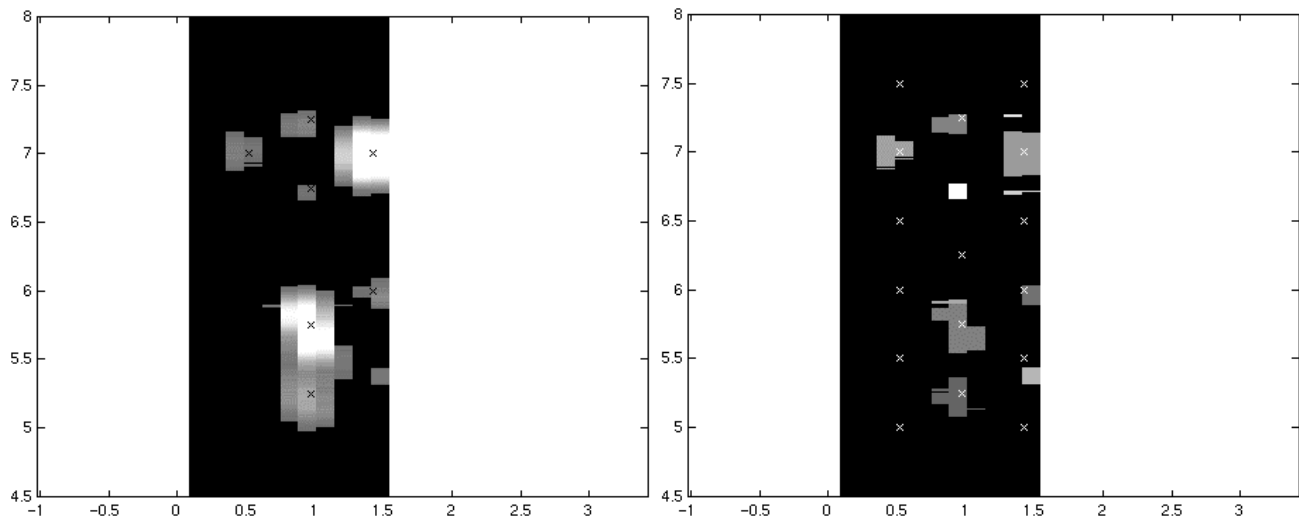


Figure 3: Preprocessed MD data (left) for  $p=3$ , result of the proposed region selection method for  $r=0$  (right). The ground truth is indicated by “x” marks.

## 3. VOTING

### On voting methods

Voting is a simple method that can be used to fuse detections from different sensors, even if no knowledge about the objects to be found is available. In addition to the preprocessing described in Section 2, for the IR processing, the results obtained

by thresholding are clustered, and clusters made up of less than 250 pixels are discarded, to avoid detections due to noise (as seen in the bottom part of Figure 1). This could be regarded as the only difference in preprocessing with the belief function fusion as described in Section 4. Alarms are given where the data exceeds the threshold ( $f$ ,  $t$  and  $p$ , as described in Section 2). In order to be able to compute area of alarms, processing for voting is done on a common grid. For all sensors, the alarm maps, which are at the resolution of the sensor data, are transformed to this grid with  $2.5 \times 2.5 \text{ cm}^2$  grid cells. In this mapping it is made sure that no detections are lost, i.e. a cell in the common grid shows an alarm if any data point covered by it shows an alarm, or if a lower resolution data point overlapping with the cell does. Three voting schemes are used:

- OR voting, where voting results in an alarm if for a grid cell any of the three sensors indicates an alarm;
- majority voting, where voting results in an alarm if at least two of the three sensors indicate an alarm;
- AND voting, where voting results in an alarm if all three sensors indicate an alarm.

Since the voting schemes are used here as a reference that includes no knowledge, they are kept simple. It means that decisions of single sensors are not weighted (see for examples in<sup>4,5</sup>) and that no more information is used in sensor specific detection algorithms (see<sup>3</sup>).

The results are scored against the ground-truth in the following way. An area of 5 cm around each object location is mapped onto the common grid. The 5 cm radius allows for some error in placement of the mines, drift of mines, and positional errors in the acquisition of the data. An object is scored as being detected when any grid cell in the ground truth area shows an alarm. As a cost function related to clearing an area of mines, based on the given alarms, the alarm area that has to be searched to remove the object is estimated, by counting the area of all grid cells within 10 cm of any alarm. In this way, the value for the x-coordinate in the ROC-figure is more related to operational cost than when only counting alarms.

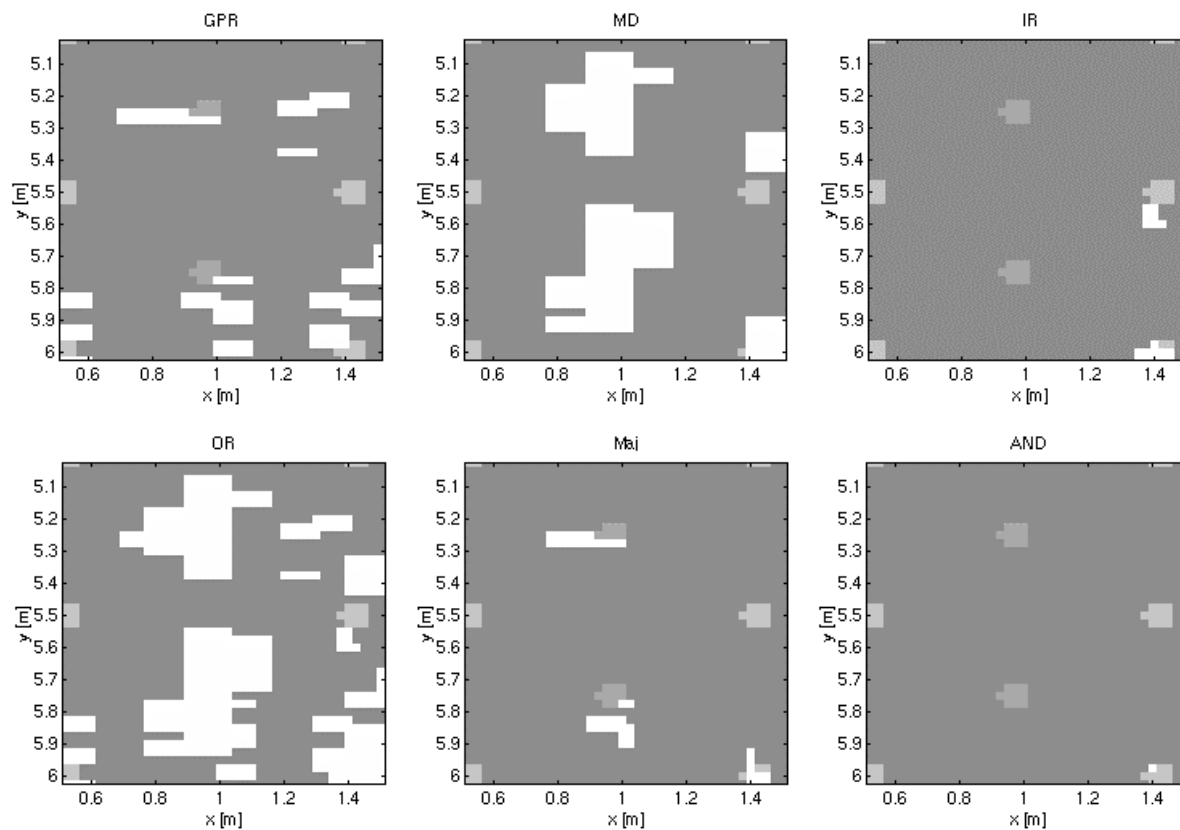


Figure 4: Parts of images of each of the sensors used as inputs of voting, as well as results for the three voting methods, for  $f=3$ ,  $t=30$  and  $p=3$ . White regions are classified as  $M$  for voting, while light grey regions indicate ground-truth positions.

## Results

An example of the voting procedure is given in Figure 4. The number of detections and alarm area are computed for a large number of threshold settings (values  $f$ ,  $t$  and  $p$ ) and plotted as ROC-curves. Since the possible values for number of detections are discrete, this leads to a stair-like function. As a more realistic description of the curves, relevant thresholds are chosen, minimizing the alarm area for a given number of detections.

The results are presented in ROC-curves in Figure 5. The lines indicate the results for the single sensors. Except for the IR, the sensors were not able to detect all mines, even with lowest threshold setting, due to the preprocessing (see Section 2). The circles show the IR results. It is interesting to see that above 8 detections, the behavior of the ROC-curve changes. In that region, detections are made more and more by chance due to alarms caused by clutter. The GPR (diamonds) and MD (stars) partly follow the same line through the origin, except for low number of detections, where with each detection immediately quite a large area appears, and for the MD at higher number of detections.

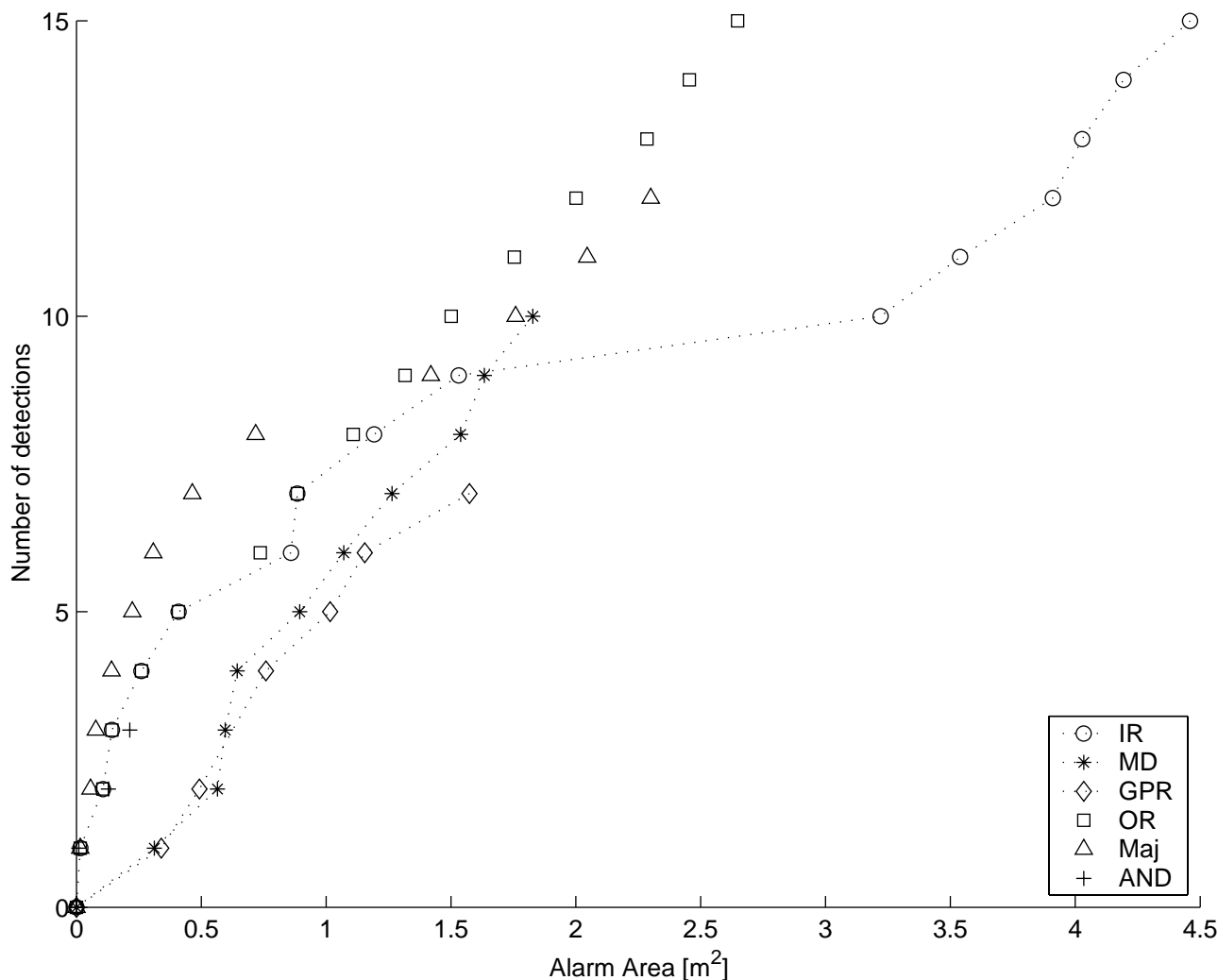


Figure 5: ROC results for single sensors, and voting. The lines give the results for the single sensors, the other points are the best results obtained for OR, majority and AND voting.

Detection maps obtained for the single sensors, corresponding to the different lines in the figure, are fused using the three voting schemes. All possible combinations are computed, for example low threshold IR with high threshold MD and GPR. This results in over a thousand points in the ROC space, for each of the schemes. In Figure 5, the best results are shown, being the points with lowest alarm area for each number of detections. The actual “clouds” of points extend to below the single sensor curves. The results show that improvements with simple voting schemes are possible. However, not all combinations result in an improvement, as compared to the best sensor. One can obtain curves within the cloud by changing a single parameter, for example by increasing thresholds for each sensor in the same way. It is not possible to find a curve that shows improvements by voting in a large region. This proves that settings should be chosen carefully, for each scenario. Indications on how to do this are found when looking at which combinations of settings result in the points presented in the figure.

To obtain a small alarm area with medium number of detections, majority voting (indicated by triangles in the figure) gives good results. These results are achieved by taking an IR setting resulting in 8 or 9 detections, a high threshold GPR setting and discarding MD. In other words, the IR results are improved by reducing alarm area with the GPR. The majority voting results for higher numbers of detections, which are close to those of OR voting, are obtained by using MD instead of GPR. Higher number of detections, with better alarm area than using IR alone, are reached by using OR voting (the squares in the figure), by taking a setting near the change in the IR curve, with settings for MD and GPR that yield only a few detections. The data set is not suitable to judge AND voting (crosses in the figure), since the number of detections is really low, as, clearly, the sensors detect different mines when high thresholds are used.

The results for voting are very scenario specific. Indications to use certain settings or a combination of sensors depend on which types of mines are expected to be found, as well as on background and environment. The case presented here shows that general indications for settings can be found for a scenario, that improve results over a single sensor. This may be the case even more when the number of different types of mines expected to be encountered in a given minefield is lower, or the types less diverse. Depending on scenario, operational requirements and environmental aspects at time of data acquisition, it may well be that a certain sensor should be ignored, or only used with very high settings. These findings can be obtained for a scenario, after scoring voting results for a minefield with other methods, and use ROC curves to find the thresholds.

#### 4. FUSION IN TERMS OF BELIEF FUNCTIONS

##### On belief functions

*Brief overview of the developed model.* The general idea of this fusion method we have developed<sup>2,6,7</sup> is to model each piece of information extracted from the sensors regarding the identity of the analyzed region (such as its area, its shape, etc.) as a belief function. These models are based on a reasoning that includes knowledge about the sensors, about the situations and possible ambiguities. For example, we know that if the extracted area is very small or very large, such an object is most possibly not a mine, so in such cases we give a high belief that this is a non-dangerous (friendly) object. Otherwise, if the area is such that the object could be a mine, it could be anything else as well, so a high belief would be given to the union of mines and friendly objects. This point illustrates one of the advantages of this framework, which is its ability to easily handle unions of classes, not only single classes. This aspect is crucial for us, since neither of the used sensors is a mine detector, they are all anomaly detectors; in other words, whenever they detect something that can be a mine, it can be a non-dangerous object as well. The choice and the width of the area interval within which we believe that such an object could be an AP mine depend on the existing knowledge about expected types of mines in the region: the higher the knowledge, the narrower the interval will be.

After the modeling, we combine all these functions using Dempster's rule<sup>8</sup> in unnormalized form<sup>9,10</sup> to decide if the sensors detect a mine or a friendly object. Since we extract several pieces of information from each of the sensors and since the combination rule is associative<sup>10</sup>, we firstly combine these functions per sensor and then we combine the sensors, in order to have decisions of each sensor as well as of their fusion. Due to the fact that the three sensors detect different phenomena, our model allows for a possibility that not all sensors detect the same object, e.g., the IR sensor detects something plastic on the surface, while GPR and MD detect some metallic object buried deep below. In that case, based on the depth information extracted from GPR, the sensors are regrouped so that the ones in the same subgroup refer to the same object, and fusion takes place only within each subgroup. The chosen measures, their modeling, the combination procedure as well as the decision rule are discussed in detail in our previous work<sup>2,6,7</sup>.

*Flexibility of the model to adjust to real data.* For IR, due to the good quality of the data, all measures foreseen by our theoretical model<sup>2,6,7</sup> are useful, which are: shape elongation, shape ellipticity and area. On the other hand, due to some specificities of the acquired GPR and MD data pointed out in Section 2, not all theoretically envisaged pieces of information of these two sensors are useful, such as shape information (because of a very poor resolution in the cross-track direction). In other words, only some of their measures make sense, and they are:

- for MD: width of the analyzed region in the scanning direction;
- for GPR, we analyze two separate sets of measures<sup>2</sup>:
  - set A: depth of the object (from A-scan), width of the region in the scanning direction, and the energy of the local maximum of the region;
  - set B: depth of the object, propagation velocity of electromagnetic waves in the medium above the object, and the size of the object (all extracted from hyperbolae detected on B-scans<sup>2</sup>).

It should be noted that this aspect is a good means to test the flexibility of the developed model to fit a real situation. For each alarmed region, after extracting useful pieces of information, modeling them and fusing for each of the sensors separately, the decision is made regarding the true object identity, i.e. whether it is a mine (M) or a non-dangerous (friendly) object (F). Our tests on the other part of the lane, which is not analyzed further in this paper and which contains mainly AT mines, show that the used IR sensor could not reach the depths larger than 5 cm. Therefore, before fusing the sensors, if the depth detected by GPR is:

- larger than 5 cm, then the two sensors are clustered separately, meaning that they detected different objects;
- smaller than or equal to 5 cm, then the GPR depth information is used to partly weaken IR measures so that the deeper the object, the weaker the relevance of IR information.

After this step, fusion of the sensors is performed.

## Results

In this section, we show the results of applying the belief function fusion method to the alarmed regions corresponding to a chosen setting of the three sensors. Namely, the IR region selection factor  $f$  is chosen to be 3 (as illustrated in Figure 1), in order to preserve responses of shallowly buried targets without significantly increasing the clutter in the case of surface-laid objects. Following similar reasoning,  $t$  is 30 (as in Figure 2), while  $p=3$  and  $r=0$  (as in Figure 3).

The obtained results for each of the sensors as well as after their fusion are summarized in the following.

For IR, all surface-laid mines and all except one shallowly buried mine are correctly classified. The detected area of the rejected shallowly buried mine is very small, so it is classified as F. In case of the remaining five mines, buried 10 cm below the soil surface, IR did not sense them at all, but only some noise on the surface above them, hence all of them are classified as F as well. In other words, the results worsen as the burial depth grows, which is in agreement with IR principles of operation. Also, one placed false alarm (iron fragment) is correctly classified, i.e. as F.

In case of MD, all alarmed regions containing metallic objects (six mines and two placed false alarms) are classified as M. Three false alarms caused by clutter and the chosen region selection method are correctly classified as F. There are no false rejections.

For GPR, set A rejects one out of twelve mines in alarmed regions, while there are no rejections of mines in case of set B. Set A performs better for placed false alarms, while set B is superior in removing alarms caused by clutter (only one out of seven is classified as M), which is often the largest problem for GPR. It can be noted that we analyze the two GPR sets of measures separately. There are several reasons for that, amongst which:

- combination of the two sets would lead to increasing belief for F, i.e. worsening the results regarding the correct classification of mines, due to some aspects of the theory and the model<sup>2,6,7</sup>;
- in reality, it makes sense to use only one of the sets, since they imply different feature extraction methods;
- their combination would have to be done carefully, because only independent (in the cognitive sense<sup>9</sup>) measures can be combined.

After fusion of the three sensors, of the total of 15 mines placed in the lane, two and one for A- and B- set respectively, are rejected. The difference in one rejected mine is due to the difference in results between the two GPR sets of measures. The other mine is in both cases rejected due to IR. It is a non-metallic mine placed 10 cm below the surface, for which only IR gave an alarm, which was actually clutter on the surface (see the bottom part of Figure 1). In other words, this rejection is due to this very difficult situation for the used sensors and not due to our model or the chosen belief function framework. The only fusion method that would avoid this rejection in an automatic way is one where all alarmed regions of all sensors would be classified as mines. This is similar to OR voting. It must be pointed out that all results given here are obtained in an automatic way, although our belief function model allows for a possibility that the deminer influences the reasoning process by giving his opinion about the reliability of each of the sensors for each situation. For example, once we serve the deminer with the alarmed IR regions, it can be expected that, seeing only a few randomly spaced points in the questionable case (bottom part of Figure 1), he would give a low importance to the IR information. That would result in the correct classification of the mine, thanks to the cautious aspect of our model, i.e. whenever the highest belief resulting from fusion is given to the union of M and F, the final decision is M. It should be noted that this example is only an illustration of how the deminer's opinion could influence (improve) detection in difficult cases for the three sensors. It is out of the scope of this paper to analyze how the inclusion of the deminer in the reasoning process would affect the overall results.

After the fusion, regardless the GPR data set used, one of the two placed false alarms is misclassified. Also, from the total of nine clutter-caused alarms, five or eight are correctly classified as F, depending whether GPR A-set or B-set is used, respectively. Finally, it should be noted that in case of the three mines buried at 10 cm alarmed both by GPR and IR, the rejection due to IR is corrected thanks to the depth information of GPR, which asked for grouping the two sensors separately. In other words, in each of these three regions, two objects are declared, a surface-laid F (seen by IR) and a deeper buried M (seen by GPR).



As a conclusion, for the chosen setting, the belief function method results in 14 out of 15 placed mines detected, where the remaining mine is rejected due to the detection limitations of the three sensors themselves. Based on the included knowledge about the problem, this method allows for discriminating between the mines and non-dangerous objects, leading to significant improvements in false alarm reduction. For example, in case of set B, only 2 out of 11 false alarms are classified as mines. This aspect shows the real power of the method.

## 5. COMPARISON

With this comparison of the two methods, our main goal is to analyze what the improvement in false alarm reduction will be once knowledge is introduced and whether that inclusion affects the number of false rejections of mines. Therefore, the settings used for the belief function fusion in Section 4 are also used to compute corresponding detection maps in the voting fusion. In order to be able to compare the results, from these maps blobs are taken and visually related to the regions used in the belief function fusion. The only remaining differences are:

- in voting, IR alarms smaller than 250 pixels are removed; as a result, some IR alarms analyzed in case of belief functions, most of which are declared as friendly object, are not seen for voting (such as region 12, illustrated in the bottom part of Figure 1);
- for belief function fusion, some imprecision is allowed while associating alarms of the sensors, i.e. regions are linked if the distance of their centers is small enough (less than 20 cm); for voting, regions are linked if they overlap in the common grid; in majority or AND voting, this may result in a smaller region (that does not overlap with the ground truth area as used in Section 3, but in this comparison is still linked to the object).

Although the above differences could have been avoided, the ideas behind introducing them are linked to the methods themselves. For example, for voting, the only way to discard alarms caused by noise is to remove small alarms, while in case of belief functions, such alarms can remain and be discarded based on their area measure (so, they do the same in different ways). In other words, the mentioned differences do not influence the overall comparison and conclusions.

Table 1 gives an overview of results for the different objects that are processed using both voting and belief function methods. As can be noticed, only belief function results using GPR B-set are shown. Namely, the results of the two sets are already compared in Section 4, on the one hand, and on the other hand, the comparison of results of belief function and of voting methods would lead to the same conclusions regardless the chosen set.

In the processing for voting, all declarations are for possible mines. The fact that scoring is done differently in Section 3, explains that, for example, more mines are detected for GPR than the maximum, as shown in Figure 5. These are due to alarms more than 5 cm from the ground-truth location, but with their center within 20 cm.

From the table it can be seen that belief function fusion improves results very much as compared to the single sensors. OR voting also indicates most mines, but it also indicates all clutter-caused as well as placed false alarms as possible mines. A summary of the results given in Table 1 is shown in Table 2, containing the number of correct and wrong classifications of the alarmed regions (alarms), depending whether they contain mines, friendly objects or nothing (clutter-caused alarms), by each sensor and by each fusion method. Note that the number of detected mines is the same for belief functions and OR voting. A main difference is in the number of false alarms, that can be explained by the following. As in voting no difference is made between friendly and minelike objects, any alarm is a possible mine. Therefore, using measures, a single sensor may discard an alarm as related to a friendly object (e.g., false alarm region 18 in the table), or be persuaded to do so by other sensors (e.g., the iron fragment at 10 cm depth, region 9). For voting, the latter will only happen in majority or AND voting, if enough sensors do not give an alarm.

For voting, small clusters of alarms, including detections covering a larger area, but with no continuous alarm, are discarded, to avoid noise. For the belief functions, these areas are taken into account, and discarded on other grounds, as seen in regions 10, 14 and 15 where, due to the depth information of GPR, IR is grouped separately, and two objects are declared in each region, one F on the surface, and one M below it.

An interesting example is region 3. This mine contains no metal at all, hence it should not have been seen by MD. Indeed, as can be seen in Table 1, majority voting shows that although two sensors give a detection in this region, there is no overlap between them that can be related to the object. This is also illustrated in Figure 4, where region 3 corresponds to the ground-truth position centered at  $x=1.42$  m,  $y=5.5$  m. A strange and non-realistic MD detection of such a mine is caused by the signal of a neighboring large metallic object (as discussed in Section 2). Therefore, it can be expected that with other arrangement of objects only the IR alarm would be present, and it would remain correctly classified both in case of belief function fusion and of OR voting.

Observing results for regions 2 and 9 obtained by belief function fusion, another interesting conclusion can be drawn. For region 2, two sensors have alarms, IR and GPR. IR claims it is F, GPR claims it is M, and the fusion result is M. In case of region 9, all three sensors have alarms, and although two (MD and GPR) of them decide it is M, the final result is F. The reason behind is that for region 2, GPR is more confident in its decision than IR, so GPR decision prevails. For region 9, on the contrary, confidence of IR is far higher than the ones of GPR and MD, and as a result, the decision of IR overrules the decisions of the other two sensors and after the fusion, the object is classified as F.

Region number	Object	Depth [cm]	Belief functions				Voting					
			IR	MD	GPR (B-set)	Fusion (B-set)	IR	MD	GPR	OR	Maj	AND
1	Iron fragment	0		<i>M</i>	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	
2	M14	0	<b>F</b>		<i>M</i>	<i>M</i>			<i>M</i>	<i>M</i>		
3	NR22, no ring	0	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>	<i>M</i>		<i>M</i>		
4	NR24 with ring	0	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
5	PFM-1 (butterfly)	0		<i>M</i>		<i>M</i>		<i>M</i>		<i>M</i>		
6	PMN	0			<i>M</i>	<i>M</i>			<i>M</i>	<i>M</i>		
7	PMN2	0	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
8	Wooden AP	0	<i>M</i>		<i>M</i>	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>		
9	Iron fragment	10	<i>F</i>	<i>M</i>	<i>M</i>	<i>F</i>		<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	
10	M14	10	<b>F</b>		<i>M</i>	<i>M+F</i>			<i>M</i>	<i>M</i>		
11	NR24 with ring	10		<i>M</i>	<i>M</i>	<i>M</i>		<i>M</i>	<i>M</i>	<i>M</i>		
12	NR24, no ring	10	<b>F</b>			<b>F</b>						
13	PFM-1 butterfly	10			<i>M</i>	<i>M</i>			<i>M</i>	<i>M</i>		
14	PMN	10	<b>F</b>		<i>M</i>	<i>M+F</i>			<i>M</i>	<i>M</i>		
15	PMN2	10	<b>F</b>		<i>M</i>	<i>M+F</i>			<i>M</i>	<i>M</i>		
16	Wooden AP	10			<i>M</i>	<i>M</i>			<i>M</i>	<i>M</i>		
17	PFM-1 butterfly	+	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
18	Nothing			<i>F</i>		<i>F</i>		<i>M</i>		<i>M</i>		
19	Nothing			<i>F</i>		<i>F</i>		<i>M</i>		<i>M</i>		
20	Nothing			<i>F</i>	<i>F</i>	<i>F</i>		<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	
21	Nothing				<i>F</i>	<i>F</i>			<i>M</i>	<i>M</i>		
22	Nothing				<i>F</i>	<i>F</i>			<i>M</i>	<i>M</i>		
23	Nothing				<i>F</i>	<i>F</i>			<i>M</i>	<i>M</i>		
24	Nothing				<i>F</i>	<i>F</i>			<i>M</i>	<i>M</i>		
25	Nothing				<i>M</i>	<i>M</i>			<i>M</i>	<i>M</i>		
26	Nothing				<i>F</i>	<i>F</i>			<i>M</i>	<i>M</i>		

Table 1: Detection results for the different sensors and fusion methods. Depth + means on the surface, 0 means just below the surface. Empty cells indicate that the region selection did not give an alarm for the corresponding sensor and the object. False alarms are written in italic. Grey cells indicate missed or misclassified (bold F) mines.

It can be concluded that in case of the voting schemes, a large number of detections can only be reached by OR voting, resulting in many false alarms due to GPR. On the other hand, the same number of detections is obtained using belief functions, but, once measures are extracted and knowledge about them is included in the belief function model, many of these false alarms are discarded. Finally, with voting, mines that are seen by only one sensor (for example, mines at 10 cm depth, which are mainly found by GPR), can only be detected using OR voting, which, again, leads to many false alarms. Using belief functions, a single sensor may still yield good results, since false alarms may be lost thanks using measures.

Object type	Classification	Belief functions				Voting					
		IR	MD	GPR (B-set)	Fusion (B-set)	IR	MD	GPR	OR	Maj	AND
Mine (total: 15)	Correct	5	6	12	14	5	6	12	14	4	3
	Wrong	5			1						
Placed false alarms (total: 2)	Correct	1			1						
	Wrong		2	2	1		2	2	2	2	
Clutter-caused alarms (total: 9)	Correct		3	6	8						
	Wrong			1	1		3	7	9	1	

Table 2: Summary of the results from Table 1 regarding classification of the alarmed regions.

## 6. CONCLUSION

We have presented some experimental results on mine detection with the aim of investigating the influence of knowledge inclusion in fusion methods. Several conclusions can be drawn, concerning both the methods and the data we used.

Two methods, voting and belief functions, are analyzed and compared. While voting is a simple method that only combines sensors, either taking all their alarms, including false ones, or losing many alarms, belief function fusion can use all alarms and reduce number of false alarms using measures, and combining them for each sensor prior to combining sensors. In other words, in voting schemes, a crisp decision is taken for each sensor, without accounting for quality of this decision. This is avoided in the belief function framework, and intervals based on knowledge are used for summarizing the results of each measure or each sensor and for fusion. The differences in obtained results between the methods are due to the fundamental differences between the two methods.

Another important difference between the two methods is in the use of the depth information in case of belief functions. By that, it is possible to determine that not all sensors see the same object so that there is no point to fuse them, which may be extremely important in some cases, e.g., when IR detects only some noise on the top of a deeper buried mine.

Although results given here are obtained in an automatic way for both methods, a possibility of including the deminer's opinion in the reasoning process is envisaged in case of belief functions, which could be, in some very difficult cases, the only way to avoid mine rejection.

A main stress of the work presented here is on general aspects and differences of the two fusion methods. From that point of view, these multisensor data of different types of mines are very well-suited. Still, it has to be noted that the used data introduce several problems to any further exploitation. Firstly, the objects were placed close to each other in the lane, which caused mixing of signals coming from neighboring ones, for MD especially and for GPR to some extent. Also, the diversity of the objects in the lane was too high in comparison to their number to allow for any statistically relevant estimation of performance of each of the sensors as well as of each of the fusion methods (or their comparison). Related to this, statistical methods were excluded because of lack of data for a reliable learning<sup>4,11,12,13</sup>.

Voting is a simple method and with tuning of threshold settings, the area to be searched to find all mines can be reduced. This requires some knowledge on the performance of the sensors in encountered situation. If reliable knowledge about the problem is available, the belief functions can take this into account to differentiate between mines and friendly objects, and as a main advantage, significantly decrease the number of false alarms. If such knowledge does not exist, meaning that our ignorance about any of the measures extracted from our sensors is maximum, all alarms would be classified as mines for belief functions as well. In other words, in such cases, the differences in results of the belief function and voting methods (i.e. OR voting) would be diminished.

## REFERENCES

1. W. de Jong, H.A. Lensen, and Y.H.L. Janssen, "Sophisticated Test Facility to Detect Landmines", *Proc. SPIE* **3710**, pp.1409-1418, Orlando, USA, 1999.
2. N. Milisavljević, "Analysis and Fusion Using Belief Function Theory of Multisensor Data for Close-Range Humanitarian Mine Detection", PhD thesis, ENST, Paris, France, 2001.

3. M.G.J. Breuers, P.B.W. Schwering, and S.P. van den Broek, "Sensor Fusion Algorithms for the Detection of the Land Mines", *Proc. SPIE* **3710**, pp. 1160-1166, Orlando, USA, 1999.
4. F. Cremer, K. Schutte, J.G.M. Schavemaker, and E. den Breejen, "A Comparison of Decision-Level Sensor-Fusion Methods for Anti-Personnel Land-Mine Detection", *Information Fusion*, accepted for publication.
5. A.L. Klein, "*Sensor and Data Fusion Concepts and Applications*", SPIE Optical Engineering Presss, Bellingham, 1993.
6. N. Milisavljević, I. Bloch, and M. Acheroy, "Modeling, Combining and Discounting Mine Detection Sensors within Dempster-Shafer Framework", *Proc. SPIE* **4038**, pp. 1461-1472, Orlando, USA, 2000.
7. N. Milisavljević, I. Bloch, and M. Acheroy, "Characterization of Mine Detection Sensors in Terms of Belief Functions and their Fusion, First Results", *Proc. 3rd International Conference on Information Fusion (FUSION 2000) II*, pp. ThC3.15-ThC3.22, Paris, France, 2000.
8. G. Shafer, "*A Mathematical Theory of Evidence*", Princeton University Press, 1976.
9. P. Smets, "Belief Functions: the Disjunctive Rule of Combination and the Generalized Bayesian Theorem", *International Journal of Approximate Reasoning* **9**, pp. 1-35, 1993.
10. I. Bloch, "Some Aspects of Dempster-Shafer Evidence Theory for Classification of Multi-Modality Medical Images Taking Partial Volume Effect into Account", *Pattern Recognition Letters* **17**, pp. 905-919, 1996.
11. M. Roughan and D. McMichael, "Data Fusion for Land-Mine Detection: Data Requirements for Experimental Design", Technical Report to LSOD, DSTO, Cooperative Research Centre for Sensor, Signal and Information Processing, Adelaide, Australia, 1996.
12. N. Milisavljević and M. Acheroy, "An Approach to the Use of the Bayesian Rule in Decision Level Fusion for Multisensor Mine Detection", *Proc. Physics in Signal and Image Processing (PSIP'99) Conference*, pp. 261-266, Paris, France, 1999.
13. R. Voles, "Confidence in the Assessment and Use of Mine Detection Systems", *Proc. IEE Second International Conference on the Detection of Abandoned Landmines*, pp. 28-30, Edinburgh, UK, 1998.