Abstract—We present an approach of automatic seabed recognition from multiple views of side-scan sonar in the context of the project ASEMAR of the pole Mer Bretagne. We integrate detailed knowledge about each view: the nature of the seabed, the position and the uncertainty and the imprecision related to each information. To exploit information from multiple views, a fusion strategy for seabed recognition has been developed. It is based on the theory of belief functions, that deals with the imperfection of information, computed over tiles of seabed. We show the application of our method on simulated sonar data given by an autonomous underwater vehicle. This application illustrates the interest of a belief fusion approach and the analysis of the final results show the benefits.

I. INTRODUCTION

Seabed recognition are among the various applications that need view-specific inputs. Autonomous Underwater Vehicles (AUV) are used to explore underwater seabed using specific sensors such as sonar. A typical mission of an AUV is to map an area to determine the nature of seabed or to determine if there are any sort of mines. Each mission is controlled by a trajectory defined by an expert and gives a map of seabed of the explored area. Our objective is to define automatically this trajectory based on previous missions (previous maps) to explore areas with risks or with a poor map. For this aim, we use the multi-view information (resulting from multiple viewpoints of a single mission or more) that we combine to give a more accurate and an optimized map of the observed area. Figure 1 gives an example of a mission trajectory and the corresponding simulated seabed.

The information gathered, in most cases, is corrupted by imperfections generally related to the studied environment and the used sensors. In this work, the information is described by the kind of seabed: for a given position of the sonar, we have the information about seabed in the perpendicular direction of the side-scan sonar. Thus, the information depends on the position which is imprecise and not well known. Several choices are suitable to tackle this kind of imprecise information: either we try to remove these imperfections, which requires a comprehension of the physics which led to them; or we try to develop a robust system to cope with these imperfections; or we try to model them. A precise modeling of uncertain and imprecise data can be carried out using the theories of uncertainty like the fuzzy sets theory [18], the possibilities theory [19], [7] or the theory of belief functions [4], [15]. Our choice fell on the theory of belief functions which, in the same theoretical framework, model uncertainty and imprecision and also offers advantages to model the lack of information from each viewpoint.

The organization of this paper is as follows: In Section 2, we present the multi-view fusion for the seabed recognition with the theory of the belief functions and the voting fusion. Therefore we give a description of the simulated data used to validate the approach of multi-view fusion. We present and discuss classification results in Section III-C.

II. MULTI-VIEW FUSION FOR SEABED RECOGNITION

The multi-view fusion for seabed recognition can be expressed as a problem of combining classification results (seabed classes) from each viewpoint. The conventional methods from the theory of uncertainty (fusion by weighted voting, Bayesian fusion, belief functions fusion, etc.) can be used to achieve this fusion [10], [11]. In these theories of uncertainty, two different concepts are required to fully model data imperfections: the uncertainty and imprecision. The uncertainty characterizes a degree of conformity with reality (qualitative default of information), while imprecision measures a lack of quantitative information (for example an error measurement) [11].

Several studies in the literature use the information extracted from several viewpoints to detect objects, such as in [3]; they studied the contribution of the use of multiple sonar viewpoints.
to improve classification rates of objects. In the same context, [2] use multi-view classification for mines detection. For sonar images, multi-view classification has seen several interests such as the positioning of an autonomous underwater robot, and generally requires an image registration [6]. Multi-view fusion was also used for classification of the gender based on a walk sequence [8]. In the field of remote sensing, [13] have used an approach based on the theory of the belief functions and the theory of possibilities to detect land-mines.

The general principle of an information fusion approach is described in Figure 2. Thus we consider the information from different viewpoints to combine, as well as additional information and knowledge related to external application to assist the combination.

![Fig. 2. Representation of the information fusion](image)

The process for information fusion is described by four steps: the modeling, the estimation, the combination and the decision (cf. Fig 3). The model defines the choice of the used formalism, which will be in our case the theory of belief functions. The estimation defines the used functions, depending on the application, in the modeling stage. The combination is the information consolidation phase. The last step gives a decision over the combination result.

![Fig. 3. Fusion node](image)

### A. Multi-view fusion with belief functions

We propose here the use of the theory of the belief functions for multi-view fusion for a better seabed recognition.

The theory of belief functions is based on the use of mass functions. The mass functions are defined over $2^Θ$ the set of all possible disjunctions realized using elements from the frame of discernment $Θ = \{C_1, \ldots, C_N\}$ and of values in $[0,1]$, where $C_q$ represents the assumption “the tile belongs to the class $q$”. Moreover, we add a condition of normalization, given by:

$$\sum_{A \in 2^Θ} m(A) = 1,$$

where $m(.)$ represents the mass function. The first difficulty is to define these mass functions according to the problem. Other belief functions can be defined from these mass functions, such as the functions of credibility, representing the intensity that all sources believe in an element, and as the functions of plausibility representing the degree with which we believe on an element.

In order to estimate the mass functions to be combined, [1] suggests two models addressing three axioms which involve the use of $N$ mass functions, each having only three focal elements $\{C_q, C_q^c, \Theta\}$. Such a function is called dichotomous mass function. Furthermore, an axiom guarantees the equivalence with the Bayesian approach where the reality is perfectly known. Both models are substantially equivalent for our data, we use in this article the model (denoted model $I$) given by:

$$
\begin{align*}
    m_{iq}(C_q)(x) &= \frac{\alpha_{iq} R_i p(V_i(x)/C_q)}{1 + R_i p(V_i(x)/C_q)} \\
    m_{iq}(C_q^c)(x) &= \frac{\alpha_{iq}}{1 + R_i p(V_i(x)/C_q)} \\
    m_{iq}(\Theta)(x) &= 1 - \alpha_{iq}
\end{align*}
$$

where $p$ is a probability, $R_i = (\max_{q \in q} p(V_i(x)/C_q))^{-1}$ is a normalization factor, and $\alpha_{iq} \in [0,1]$ is a discounting factor. $\alpha_{iq}$ controls the reliability of the information provided by the view $i$: $V_i(x)$ for a class $C_q$ that we choose equal to 0.95 without more information. The difficulty for this model is the estimation of probabilities $p(V_i(x)/C_q)$. Since the given $V_i(x)$ of the view $i$ is the response of a classifier given in the form of class (symbolic data), the estimation of these probabilities can be realized using the confusion matrices during the learning step.

Denœux [5] proposes another estimation of the mass functions using a distance based model (we denote it model 2), defined for each nearest tile $x_j^ε$ by:

$$
\begin{align*}
    m_{iq}^ε(C_i|x_j^ε)(x) &= \alpha_i \gamma_i d^2(x,x_j^ε) \\
    m_{iq}^ε(\Theta|x_j^ε)(x) &= 1 - \alpha_i \gamma_i d^2(x,x_j^ε)
\end{align*}
$$

where $C_i$ is the associated class to $x_j^ε$, and $\{x_j^ε\}_{j=1,...,k}$ are the $ε$-nearest tiles to $x$ (i.e. $d^2(x,x_j^ε) \leq ε$). The distance $d$ used is the euclidean metric. $\alpha_i$ and $\gamma_i$ are coefficients of discounting, and of normalization which can be optimized [20]. The mass functions are computed in a different from the one proposed in [5] to compute mass functions. In [5], the authors fix the number of neighbors used to compute the mass functions. Here, we look for tiles in the sphere of radius $ε$ and centered in $x$. This estimation approach allows to integrate the imprecision of the position.

The combination of $N$ (number of views) mass functions can be a real problem if this number is high. Indeed, the first
combination rule in the belief function framework proposed by Dempster is the normalized conjunctive combination rule given for two mass functions \(m_1\) and \(m_2\) and for all \(X \in 2^\emptyset\), \(X \neq \emptyset\) by:

\[
m_{\text{ID}}(X) = \frac{1}{1-k} \sum_{A \cap B = X} m_1(A)m_2(B),
\]

where \(k = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)\) is the inconsistency of the combination.

Smet [16] proposes to consider an open world, therefore the conjunctive rule is non-normalized and we have for two mass functions \(m_1\) and \(m_2\) and for all \(X \in 2^\emptyset\) by:

\[
m_{\text{Conj}}(X) = \sum_{A \cap B = X} m_1(A)m_2(B).
\]

\(m_{\text{Conj}}(\emptyset)\) can be interpreted as a non-expected solution. In the Transferable Belief Model of Smet, the repartition of the inconsistence is done in the decision step by the pignistic probability (6).

These two rules (4) and (5) are not idempotent. So the combination of \(n\)-times the same mass function \(m\), \(m_{\text{Conj}}(\emptyset)\) and \(m_{\text{ID}}(\emptyset)\) tend to 1 when \(n\) tends toward \(\infty\), that is what we call the auto-conflict in [14]. Hence the normalization by \(1-k\) in the combination rule (4) or in the pignistic probability can be problematic. Many other rules have been proposed, a brief state of the art as well as new rules for coping with the conflict within the combination are given by [12].

In order to preserve maximum of information, it is preferable to stay on a credal level (i.e. to handle belief functions) during the information combination stage to make the decision on the belief functions resulting from this combination. If the decision taken by the maximum of credibility is too pessimistic, the decision obtained by the maximum of plausibility is too optimistic. The maximum of the pignistic probability, introduced by [17], is the most used compromise. The pignistic probability is given for all \(X \in 2^\emptyset\), with \(X \neq \emptyset\) by:

\[
betP(X) = \sum_{Y \in 2^\emptyset, Y \neq \emptyset} \frac{|X \cap Y|}{|Y|} \frac{m(Y)}{1 - m(\emptyset)}.
\]

**B. Weighted voting fusion**

The simplest approach for fusion is the majority voting approach based on the combination of classes of classified tiles. The fusion is done using the principle of majority voting by taking the maximum on the number of times that the tile \(x\) is assigned to a given class [9]. We calculate the normalized histogram \(p_x\) of number of times that tiles \(x_i, i = 1, \ldots, N_x\), are classified to a class \(C_q, q = 1, \ldots, N\):

\[
p_x(C_q) = \frac{\text{card}\{i = 1, \ldots, N_x ; x_i \in C_q\}}{N_x}.
\]

Using this approach, the class \(C_x\) of tile \(x\) is the maximum of \(p_x\):

\[
C_x = \arg\max_{q=1,\ldots,N} p_x(C_q).
\]

If the maximum is reached for several values of \(q\), we can choose, for example, the class of the nearest tile to \(x\). Let \(M_x = p_x(C_x)\) be the value of the maximum of \(p_x\). Calculating the value of \(M_x\) of all tiles in the vicinity of \(x\), we form a matrix of class of tiles \(x\) noted \(I_x\) and a matrix \(I_c\) containing the values \(M_x\). This matrix (with a maximum of 1) indicates a sort of “certainty” on the classification of each tile: a value close to 1 indicates that the classifier is “sure” about the class devoted to this zone.

We did not make a difference, in terms of distance between the tiles in the vicinity of the tile \(x\) for the decision of its class. We can use a weighting of these tiles using a “high” weight for the neighborhood tiles \(x_{i_0}\) and a “low” weight for the other tiles. This weight is a function \(\psi\), decreasing with distance between the tile \(x\) and the other neighborhood tiles. For example:

\[
\psi_d(x) = e^{-\rho d_2(x_0, x)}, \rho \geq 0,
\]

In this case, to find \(C_x\) the class of the tile \(x\), we use a weighted vote by weighting the value of the histogram \(h_x\) by the summation of weights of tiles in vicinity of \(x\) of the same class:

\[
p_w(q) = p_x(C_q) \sum_{i \in C_q} \psi_{d_2}(x, x_i),
\]

where \(x_i\) are the neighborhood tiles of \(x\).

\[
C_x = \arg\max_{q=1,\ldots,N} p_w(q).
\]

**III. Experiments**

We describe in this section the simulated data to validate the approach of multi-view belief fusion for seabed recognition. We give in Section III-C the obtained results using the belief approach based on the confusion matrix model and the distance model to compute mass functions. Results are compared to the classical vote approach.

**A. Simulated data**

The data are obtained from two simulated missions of an AUV on a same area but at different times and with different trajectories. The simulated trajectories are designed to present a strong sonar overlapping. The data and trajectories were simulated and may not be fully representative of the difficulty of a real operational context. They were simulated for the benefit of an initial stage of the study. Data are expressed by the kind of seabed found in the transition of the AUV. For a given position of the AUV, we have the nature of the seabed for 13 tiles to the right of the side-scan sonar and for the 13 tiles to the left of the side-scan sonar. Each tile is characterized by the nature of the seabed (from 4 classes: silt (class 1), sand (class 2), gravel (class 3) and rock (class 4)), the size, the distance model to compute mass functions. Results are compared to the classical vote approach.
two AUV missions and the corresponding simulated seabed. Each tile is of size 10 m × 10 m. The performance of the sonar is binary (1 for a good reliability and 0 for a poor one). The coverage is also a binary number: a value equal to 0 means that we have no information about the seabed of the tile (represented by a red color in Figure 5). We can note a difference of simulated seabed between the mission 1 and the mission 2. Indeed, the mission 2 was conducted later than the mission 1, and the sand and silt move and free up a rock area.

The coverage is also a binary number: a value equal to 0 means that we have no information about the seabed of the tile (represented by a red color in Figure 5). We can note a difference of simulated seabed between the mission 1 and the mission 2. Indeed, the mission 2 was conducted later than the mission 1, and the sand and silt move and free up a rock area.

The coverage is also a binary number: a value equal to 0 means that we have no information about the seabed of the tile (represented by a red color in Figure 5). We can note a difference of simulated seabed between the mission 1 and the mission 2. Indeed, the mission 2 was conducted later than the mission 1, and the sand and silt move and free up a rock area.

### B. Estimation of mass functions

We give here the mass functions used for data described in Section III-A. As noted before, each measure is associated with imprecision that we will use in the mass functions modeling. A value of the coverage equal to 0 shows that the nature of the seabed may be silt or sand. So we can discard the mass associated to the silt class (class 1) by: ⎧⎨⎩\( m'(C_1) = \beta_1 m(C_1) \) and the associated mass with the sand class (class 2) by: ⎧⎨⎩\( m'(C_2) = \beta_2 m(C_2) \) and the ignorance by ⎧⎨⎩\( m'(\Theta) = 1 - \beta_1 - \beta_2 + (\beta_1 + \beta_2)m(\Theta) \). Where \( \beta_1 \) and \( \beta_2 \) are two real chosen between 0 and 1. The performance of the sonar controls the reliability of the source (view). The mass in this case is updated using: \( m'(C_q) = \beta_3 m(C_q) \) and \( m'(\Theta) = 1 - \beta_3 + \beta_3 m(\Theta) \).

Here we give the steps used to calculate the mass functions for a given tile \( x_0 \). Firstly, we enlarge each tile with its longitude and latitude imprecision, we look then for tiles \( x^\pm_i \) that intersect with the tile \( x_0 \). Finally, we use the Equation (3) to compute the mass functions considering only tiles that intersect with \( x_0 \). Figure 6 illustrates how to find the tiles used for combination for a given tile \( x_0 \). In this example, the tile \( x_j \), enlarged using its longitude and latitude imprecisions, intersects with the tile \( x_0 \), however, the tile \( x_i \) does not intersect with \( x_0 \) and thus it is not used for the fusion. The vote approach uses the number of tiles that intersect with \( x_0 \) for each class. It does not take into account the distance between tiles in contrast to the belief approaches, where the distance is taken into account.

### C. Results

We give in this section results for seabed recognition using multi-view belief fusion. We use model 1 and model 2 for multi-view fusion described in section II-A. The model 1 based approach uses the confusion matrix given by Equation (12) to calculate the mass functions. This confusion matrix is specific to the classifier of the AUV, used to generate the data.

\[
\begin{bmatrix}
88.18 & 11.81 & 0.01 & 0.00 \\
10.56 & 78.87 & 10.56 & 0.01 \\
0.01 & 10.56 & 78.87 & 10.56 \\
0.00 & 0.01 & 11.81 & 88.18
\end{bmatrix}
\]  

The parameters used by the model 2 based approach are given by: \( \beta_1 = 0.7, \beta_2 = 0.3 \) and \( \beta_3 = 0.4 \). The value of \( \epsilon \) (for both models) is equal to the tile size (10 m in our example). These parameters are tuned manually. We give in Figure 7 results found by distance based model approach, confusion
matrix based approach and the vote approach. Table I gives the associated confusion matrices (CM) and recognition rates (RR). The final maps, of the explored area, are represented in Figure 8.

The first row in Figure 7 represents the expected seabed of each tile. The second row gives the seabed points found after the AUV simulated mission; we can observe the overlapping between classes that gives an idea about the difficulty of the problem. The last three rows give, respectively, the distributions of seabed tiles using belief fusion models (model 1, model 2) and the vote approach. The model 2 approach and the vote approach give a good recognition of seabed, indeed, the classes are well separated in contrast to model 1 approach where class 1 (silt), class 2 (sand) and class 3 (gravel) are overlapped.

The numerical results, in terms of confusion matrices (CM) and recognition rate (RR) represented in Table I, show the significant gain using fusion approaches. Indeed, we obtained a high recognition rate (93.29%) with the model 1 approach compared to that found without fusion (54.32%). The model 2 based approach gives also a best recognition rate (92.19%). The vote approach gives a recognition rate of 90.93% smaller than these found with belief approaches. In terms of confusion matrices, the model 1 approach gives the best recognition rate of class 2 (the majority class in this study) and class 3. However, the vote approach gives the best recognition rates for class 1 (silt) and class 4 (gravel). The CPU time used by the model 2 based approach is about 183.86 seconds for 35880 tiles (a mean of about 5.1 ms to compute the class of a single tile). The vote approach gives the best CPU time with 65.82 seconds and the model 1 approach gives the worst CPU time with 1,008.5 seconds. We note that the time changes from a tile to another which depends on the number of neighborhood tiles used in the combination.

We have done fusion for the second mission, we give in Figure 9 the final maps found by the different approaches and in Table II the associated recognition rates and confusion matrices. The final maps show the improvement of seabed recognition using model 1 and model 2 approaches.

Here again, the model 1 belief approach gives the best recognition rate (92.33%) followed by the model 1 approach. The weighted voting approach gives the small recognition rate with 88.01%.

Figure 10 shows the maximum of belief and the ignorance
for the AUV missions using the model 2 approach. We note that the maximum of belief is small for tiles on the boundaries between two different seabeds. Moreover the ignorance is important in these locations and is low in homogeneous seabed areas. We note also that the ignorance and the maximum of belief is too small for the external tiles (limits of the explored area), this can be explained by the use of small number of tiles for fusion. For application these kind of maps is interesting to plan other missions.

Until now, we used the fusion of single mission independently to the other missions. The use of multiple missions can improve seabed recognition if the ground truth does not change over time which is not the case for areas of sand or silt (we can see the difference between the two missions used in this paper). Figure 11 gives the final map of the fusion of the two missions, using the model 2 belief approach by discounting the calculated masses over tiles of the first mission, and the corresponding maximum of belief and the ignorance. The recognition rate is 89.01% which is smaller than that found using a single mission. This can be explained by the variation of the simulated ground truth between both

<table>
<thead>
<tr>
<th>Approach</th>
<th>RR (%)</th>
<th>CM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>54.46</td>
<td>65.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>22.44</td>
<td>5.34</td>
</tr>
<tr>
<td></td>
<td>5.19</td>
<td>19.77</td>
</tr>
<tr>
<td></td>
<td>64.30</td>
<td>14.85</td>
</tr>
<tr>
<td>Model 1</td>
<td>92.33</td>
<td>97.33</td>
</tr>
<tr>
<td></td>
<td>2.67</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>25.66</td>
<td>74.34</td>
</tr>
<tr>
<td>Model 2</td>
<td>90.08</td>
<td>97.97</td>
</tr>
<tr>
<td></td>
<td>2.03</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>18.55</td>
</tr>
<tr>
<td></td>
<td>81.45</td>
<td>85.76</td>
</tr>
<tr>
<td></td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>Vote</td>
<td>88.01</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>5.62</td>
<td>94.38</td>
</tr>
</tbody>
</table>

**Fig. 8.** Final maps using belief and vote approaches for the first mission

**Fig. 9.** Final maps using belief and vote approaches for the second mission

**TABLE I**

**Belief Fusion Performance using model 1 and model 2**

<table>
<thead>
<tr>
<th>Approach</th>
<th>RR (%)</th>
<th>CM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>54.32</td>
<td>64.34</td>
</tr>
<tr>
<td></td>
<td>30.38</td>
<td>4.62</td>
</tr>
<tr>
<td></td>
<td>0.66</td>
<td>0.33</td>
</tr>
<tr>
<td>Model 1</td>
<td>93.29</td>
<td>89.27</td>
</tr>
<tr>
<td></td>
<td>10.73</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Model 2</td>
<td>92.19</td>
<td>94.80</td>
</tr>
<tr>
<td></td>
<td>5.20</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Vote</td>
<td>90.93</td>
<td>99.92</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**TABLE II**

**Belief Fusion Performance, for the second mission, using model 1 and model 2**
missions. The confusion matrix is given by:

\[
\begin{pmatrix}
95.97 & 4.03 & 0 & 0 \\
7.19 & 92.53 & 0.28 & 0 \\
0 & 17.76 & 81.39 & 0.85 \\
0 & 0 & 24.56 & 75.44
\end{pmatrix}
\]  \hspace{1cm} (13)

We note that the recognition rate of the sand class has been enhanced in contrast to the other classes where the single mission gives the best recognition. The maximum of belief shows the border between areas of seabed and the ignorance is high for all tiles except tiles of the limits of the explored area.

IV. CONCLUSIONS

In this paper, we studied the contribution of the use of multi-view fusion to improve seabed recognition. Indeed, areas swept by an AUV can be overlapped: a single zone can be viewed more than one time. We combine then information (seabed found by the AUV) from different viewpoints to improve the quality of the map (seabed of the explored area). The belief approaches used here have led to a significant improvement of recognition rate for multi-view seabed recognition compared to voting approach. Indeed we have obtained recognition rates (93% for the first mission and 92% for the second mission) greater than the obtained rates using only one view (around 54% for both missions). The confusion matrix based model gives the best recognition rate compared to the distance based model and to the classical vote approach. The belief approach allows the drawing of the ignorance and the maximum of belief maps. We have shown that the overall recognition rate for the fusion of both missions is small in comparison to the results of a single mission. This can be explained by the fact that the ground truth has changed. Thus, the use of multiple missions depends on the interval time between missions and the nature of the studied area. We note that these results are preliminary and further work on this subject remains.

ACKNOWLEDGMENT

The authors would like to thank THALES UNDERWATER SYSTEMS for providing the simulated data (with special thanks to Julien Ferrand and Jean-Philippe Malkasse).

REFERENCES


