



Fusion of Local Statistical Parameters for Buried Underwater Mine Detection in Sonar Imaging

Frederic Maussang, Michèle Rombaut, Jocelyn Chanussot, Alain Hétet, Maud Amate

► To cite this version:

Frederic Maussang, Michèle Rombaut, Jocelyn Chanussot, Alain Hétet, Maud Amate. Fusion of Local Statistical Parameters for Buried Underwater Mine Detection in Sonar Imaging. EURASIP Journal on Advances in Signal Processing, 2008, 2008 (ID 876092), 19 p. 10.1155/2008/876092 . hal-00276968

HAL Id: hal-00276968

<https://hal.science/hal-00276968>

Submitted on 5 May 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research Article

Fusion of Local Statistical Parameters for Buried Underwater Mine Detection in Sonar Imaging

F. Maussang,¹ M. Rombaut,² J. Chanussot,² A. Hétet,³ and M. Amate³

¹ *Institut TELECOM; TELECOM Bretagne, UEB; CNRS Lab-STICC/CID, Image and Information Processing Department, Technopôle Brest-Iroise - CS 83818, 29238 Brest Cedex 3, France*

² *GIPSA-Lab, Signals and Images Department, Grenoble INP, INPG - 46 Avenue Félix Viallet, 38031 Grenoble Cedex, France*

³ *Groupe d'Etudes Sous-Marines de l'Atlantique, DGA/DET/GESMA, BP 42, 29240 Brest Armées, France*

Correspondence should be addressed to F. Maussang, frederic.maussang@telecom-bretagne.eu

Received 30 May 2007; Revised 6 December 2007; Accepted 11 January 2008

Recommended by Ati Baskurt

Detection of buried underwater objects, and especially mines, is a current crucial strategic task. Images provided by sonar systems allowing to penetrate in the sea floor, such as the synthetic aperture sonars (SASs), are of great interest for the detection and classification of such objects. However, the signal-to-noise ratio is fairly low and advanced information processing is required for a correct and reliable detection of the echoes generated by the objects. The detection method proposed in this paper is based on a data-fusion architecture using the belief theory. The input data of this architecture are local statistical characteristics extracted from SAS data corresponding to the first-, second-, third-, and fourth-order statistical properties of the sonar images, respectively. The interest of these parameters is derived from a statistical model of the sonar data. Numerical criteria are also proposed to estimate the detection performances and to validate the method.

Copyright © 2008 F. Maussang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Detection and classification of different kinds of underwater mines are crucial strategic tasks. For this purpose, high-level technique systems are required, especially for partially or completely buried object detection. The objective of these processes is to decrease, as much as possible, the number of false alarms (harmless objects detected as mines) while all the mines are actually detected.

Many approaches have been proposed in underwater mine detection and classification using sonar images. Most of them use the characteristics of the shadows cast by the objects on the seabed [1] (Figure 1(a)). These methods fail in case of buried objects, since no shadow is cast (Figure 1(b)). That is why this last case has been less studied. In such cases, the echoes (high-intensity reflection of the wave on the objects) are the only hint suggesting the presence of the objects. Their small size and the similarity of their amplitude with the background make the detection more complex.

Starting from a synthetic aperture image, a complete detection and classification process would be composed of three main parts as follows.

- (1) *Pixel level*: the decision consists in deciding whether a pixel belongs to an object or to the background.
- (2) *Object level*: the decision concerns the segmented object which is “real” or not: are these objects interesting (mines) or simple rocks, wastes? Shape parameters (size, ...) and position information can be used to answer this question.
- (3) *Classification of object*: the decision concerns the type of object and its identification (type of mine).

This paper deals with the first step of this process. The goal is to evaluate a confidence that a pixel belongs to a sought object or to the seabed. In the following, considering the object characteristics (size, reflectivity), we will always assume that the detected objects are actual mines. However, only the second step of the process previously described, which is not addressed in this paper, would give the final answer.

Different approaches have been proposed to solve the problem of object detection at the pixel level. Some methods tend to enhance the echoes by filtering, or to isolate them by segmentation [2]. Such processes are unsatisfactory in

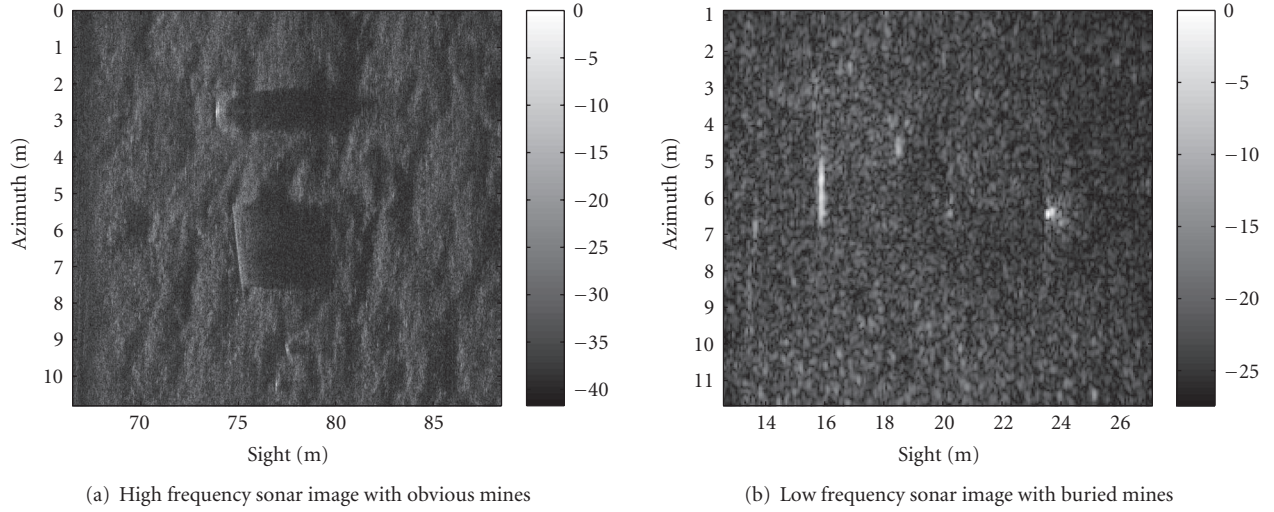


FIGURE 1: Example of sonar images (SAS, normalized dB scale).

case of low signal-to-noise ratio (case of buried objects). The most frequent approaches can be divided into two steps. The first step consists in selecting various sized regions of interest likely to contain echoes. Many methods have been proposed to reach this goal: filters enhancing echoes before a selection of the pixels of high value [3], making use of several parameters (mean, variance, lacunarity, etc.) that are locally estimated on the image [4, 5], or searching for outlines including one or several echoes [6, 7].

After this location, the second step consists in extracting several statistical or morphological parameters from the selected regions. These parameters are eventually fused. The fusion process can use likelihood, potentially after orthogonalization [3, 8–10] or neural networks [5].

Most of the times, these parameters are directly extracted from the sonar image [5]. They can also result from the combination of different sources [5] or different algorithms [9]. A learning phase, based on a well-adapted and manually labeled learning set, is required in many cases. In such cases, the general architecture is fixed and adding a new parameter implies a new learning phase. Finally, most of the methods provide a binary result (mine/not mine), leaving the expert with no flexibility for the final decision. As a matter of fact, beyond a binary decision, mine hunting experts rather expect the system to give clues and hints but leave the final decision to them. The belief function theory is well suited for this framework. It has already been used in detection of buried objects, but for a different application (terrestrial mines buried in the ground [11]) or for very different application as nondestructive testing [12].

In this paper, we propose a detection method structured as a data fusion system. This type of architecture is a smart and adaptive structure: the addition or removal of parameters is easily taken into account, without any modification of the global structure. The *inputs* of the proposed system are the parameters extracted from a synthetic aperture sonar (SAS) image. SAS systems are increasingly popular in seabed imagery thanks to the high resolution they provide for

inspection and detection [13]. The SAS principle consists in an active sonar moving along a straight trajectory. For the data presented in this paper, the sonars were guided by a rail in the direction of the azimuthal axis of the figures. The sound wave is directed perpendicularly to the rail. The grey levels on the figures represent the signal intensity given by the receiver at each position. Such a system gives a resolution of $1\text{ cm} \times 1\text{ cm}$ with a frequency of about 160 kHz and a band of 60 kHz (Figure 1(a)). These frequencies do not penetrate in the sea floor and buried objects thus remain invisible. To overcome this shortcoming, the frequency can be decreased (about 17 kHz) while preserving a sufficient spatial resolution (about 6 cm). With such a configuration, buried object generate small echoes constituted of a few pixels (Figure 1(b)): the sought objects are underwater mines about 30 cm large and from 30 cm to 1 m (cylindrical mine) long, the generated echoes approximately correspond to 5×5 pixels areas in the image. Figure 3(a) represents, with white boxes, the location of the mines.

Considering the statistical properties of such SAS images, the use of statistical parameters will be discussed. These parameters are estimated locally, using a square estimation window sliding over the image and correspond to the first-, second-, third-, and fourth-order statistical moments, respectively (Section 2).

The *outputs* of the system are the areas detected as potentially including an object. The fusion system, based on the belief theory, also provides the confidence that the detected potential objects belong to the “real object” or the “nonobject” class (Section 3). In order to assess the performances of the proposed classification system, the results are evaluated visually and compared to a manually labeled ground truth using a standard methodology (receiver operating characteristic (ROC) curves). This is described in Section 4. Results obtained with various real SAS data, containing underwater mines lying on the seabed, buried or partially buried in the sea floor, are presented with both a qualitative and a quantitative evaluation in Section 5.

2. EXTRACTION OF LOCAL STATISTICAL PARAMETERS

The key issue when designing a classification algorithm is to choose the right parameters discriminating the classes of interest. The two main approaches are (i) use of statistical knowledge about the process, (ii) use of expert knowledge, eventually derived from a physical model of the process. In this application, the statistical characteristics of the seabed pixels are well known and follow statistical laws (Rayleigh and Weibull distributions for instance). As a consequence, the data fusion process is based on the comparison of the statistical characteristics locally extracted for each pixel and these laws.

2.1. Statistical properties of the sonar images

The sonar images, as any image formed by a coherent system (radar imagery is another example), are seriously corrupted by the speckle effect. They thus have a strong granular aspect. This noise comes from the presence of a large number of elements (sand, rocks, etc.) that are smaller than the wavelength and randomly distributed over the seabed. The sensor receives the result of the interference of all the waves reflected by these small scatterers within a resolution cell [14].

The speckle effect can be described by different statistical models. The most usual model of the received amplitude A , assuming the real and the imaginary parts of the response on a resolution cell have a Gaussian distribution, is based on the Rayleigh distribution [14]. However, this model is not satisfactory when the number of elementary scatterers within a resolution cell significantly decreases. The assumption of Gaussianity, based on the central limit theorem, is no longer valid and the Rayleigh approximation falls. This case is frequently observed in high-resolution images [1] such as SAS images. In this case, the received amplitude A is better described by a Weibull law (Figure 2) whose probability density function is

$$p_W(A) = \frac{\delta}{\alpha} \left(\frac{A}{\alpha} \right)^{\delta-1} \exp \left\{ - \left(\frac{A}{\alpha} \right)^{\delta} \right\}, \quad A \geq 0 \quad (1)$$

with α a scale parameter and δ a shape parameter, strictly positive.

The following relation of proportionality between the standard deviation σ_A and the mean μ_A of $p_W(A)$ holds for this probability density function:

$$\mu_A = k_W(\delta) \cdot \sigma_A \quad (2)$$

with a coefficient $k_W(\delta)$ defined by

$$k_W(\delta) = \frac{\Gamma(1 + 1/\delta)}{\sqrt{\Gamma(1 + 2/\delta) - \Gamma(1 + 1/\delta)^2}}, \quad (3)$$

where Γ is the Gamma function: $\Gamma(z + 1) = z! = \int_0^{\infty} e^{-t} t^z dt$.

In this paper, the Weibull law is chosen to statistically model the intensity of seabed pixels. Other statistical

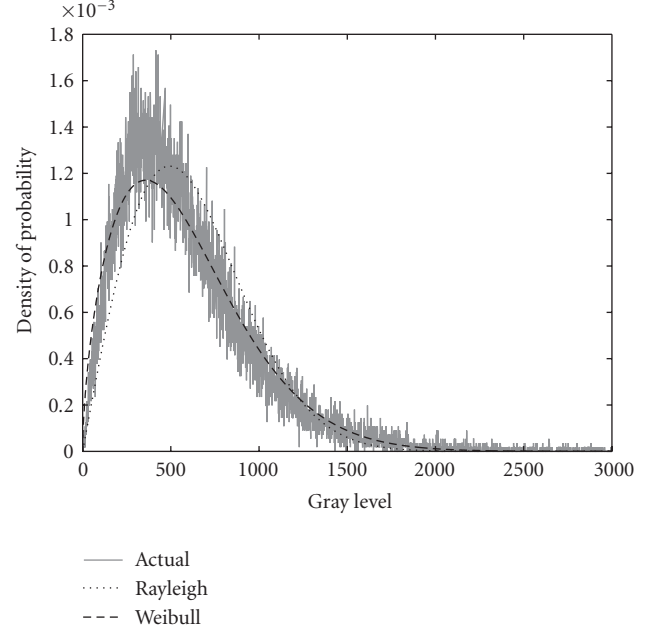


FIGURE 2: Comparison of two statistical models (Rayleigh and Weibull) with actual data: example of the image (Figure 1(b)).

models exist in the literature (K laws, ...) but the Weibull model offers a good trade off between the accuracy of modeling and the complexity of the parameters estimation. The limit of this model is reached when one resolution cell does not include any scatterer, but it never happens in the considered practical cases.

The echoes generated on the SAS images by the objects are considered as deterministic elements surrounded by the noisy background. These elements do not follow the Weibull law. This is true if we assume that the echoes are larger than the resolution, which is easily ensured for the objects of interest.

2.2. First- and second-order parameters

The proportional relation of the statistical model describing the seabed sonar data (see (2)) is used to extract the two first parameters. The local mean and standard deviation are estimated on the SAS image using a square sliding window. These values become the coordinates of the processed pixel in the mean standard deviation plane. Figure 3(a) presents an example of SAS image where the sought objects are marked by white boxes. Figure 3(c) presents the corresponding mean standard deviation representation (every cross on this figure corresponds to one pixel of the original image). The dashed line features the proportionality relation estimated from (3). As explained in [15], the choice of the size of the computation window is chosen as slightly larger than the spatial extension of the echoes, this size depends on the resolution of the sonar image and the quality of the preprocessing chain.

An automatic segmentation of the sonar image can be performed in order to identify the regions likely to

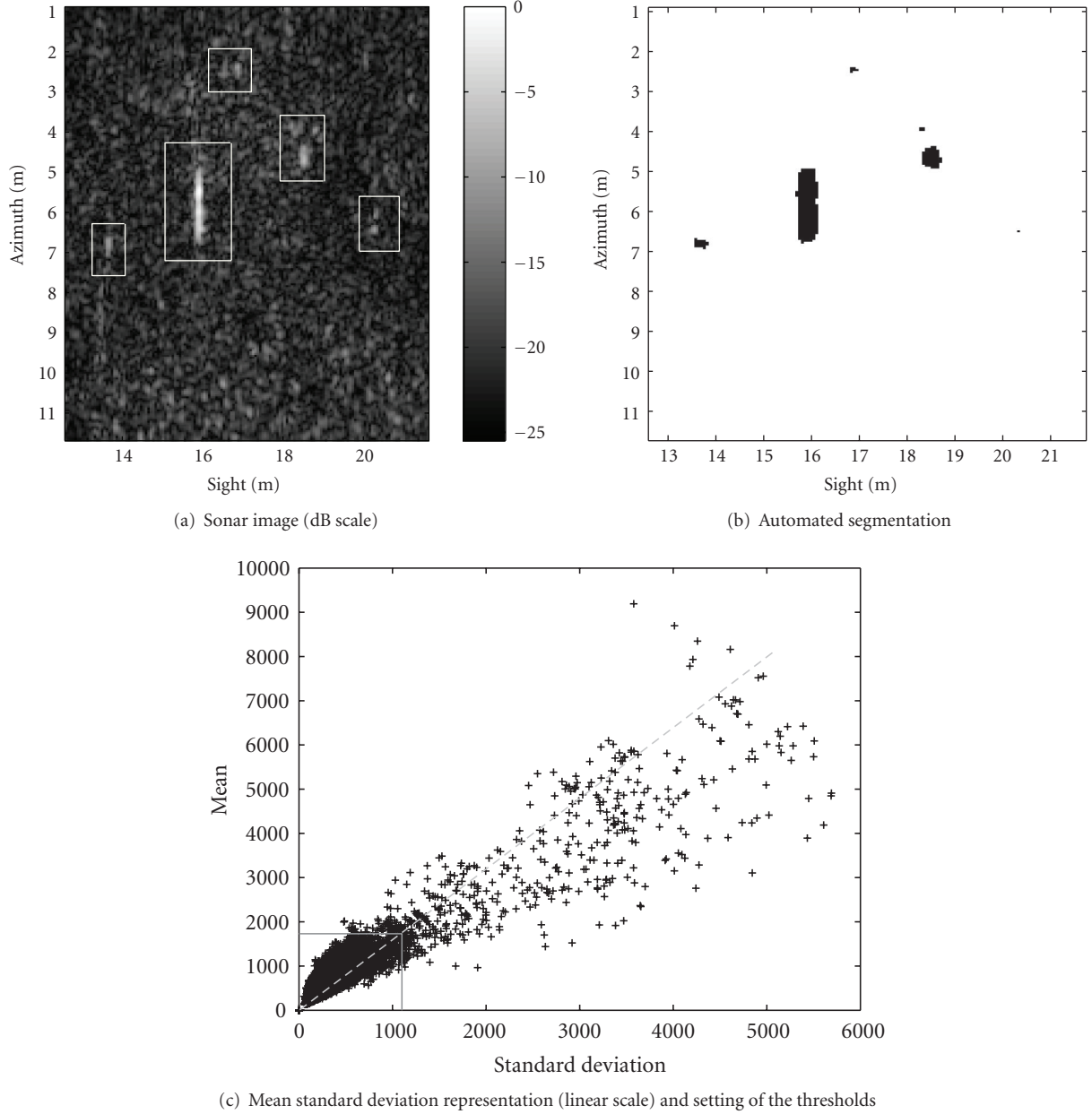


FIGURE 3: Automated segmentation of a sonar image.

be echoes reflected by objects. Such pixels are isolated by a simple threshold in the mean standard deviation plane (see the rectangle in Figure 3(c)). The threshold in the standard deviation dimension and the threshold in the mean dimension are linked by the proportionality relation. The optimal setting is found automatically by optimizing the spatial distribution of the resulting segmentation. The setting of these threshold and further details about this segmentation technique are given in [2, 15]. The resulting segmentation is presented in Figure 3(b): the echoes of interest are fairly accurately detected, and a few false alarms remain.

Note that the computed threshold value is used in the following for the fusion process.

2.3. Higher-order statistics

Pertinent information regarding SAS data can also be extracted from higher-order statistics (HOSs). In particular, the relevance of the third-order (skewness) and the fourth-order (kurtosis) statistical moments for the detection of statistically abnormal pixels in a noisy background is discussed in [16, 17]. In this previous work, an algorithm aiming at detecting echoes in SAS images using HOS is described. It basically consists in locally estimating the HOS on a square sliding window (Figure 4(b) where all the objects of interests are framed by high values of the kurtosis, the size of the frame being linked to the size of the computation window). A theoretical model of these frames is used to perform a

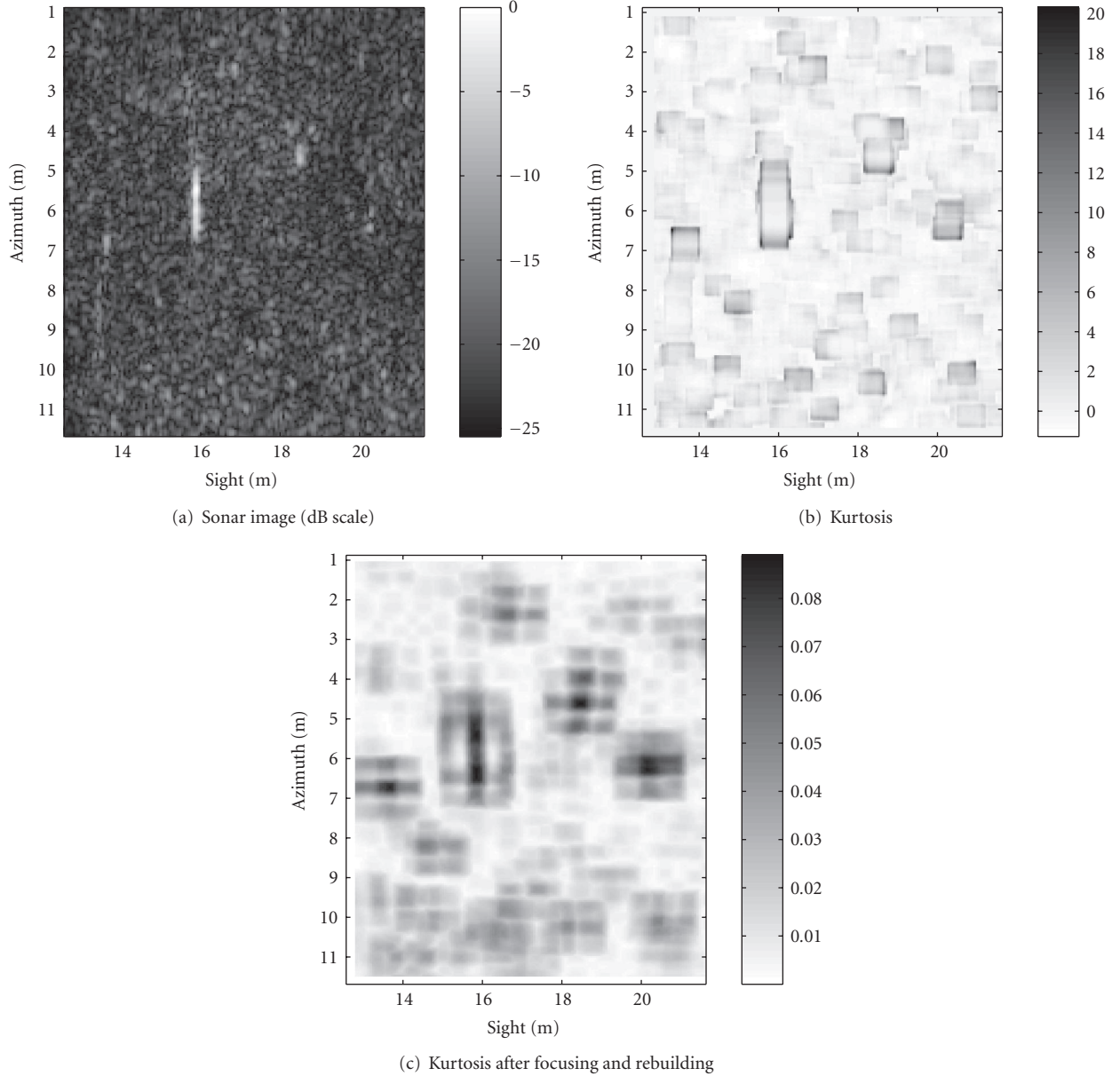


FIGURE 4: Estimation of kurtosis on the sonar image (Figure 3(a)).

matched filtering and thus refocus the detection precisely at the center of the objects of interest. The last step consists in rebuilding the objects using a morphological dilation. The corresponding detection result is presented in Figure 4(c): all the objects of interest are marked by high values, thus providing a good detection. However, some false alarms remain, and the detection is not as accurate as with the first algorithm (see Section 2.2). This will be taken into account for the fusion process.

3. DETECTION ON SONAR IMAGES USING BELIEF FUNCTION THEORY

In the previous section, we have briefly presented two algorithms aiming at detecting echoes in SAS images. In

order to further improve the detection performances, this paper presents a fusion scheme taking advantage of the different extracted parameters.

The combination of parameters in a fusion process can be addressed using probabilities. This popular framework has a solid mathematical background [18]. Numerous papers have been written on this theory using modeling tools (parametric laws with well-studied properties) and model learning. However, these methods are affected by some shortcomings. Firstly, they do not clearly differentiate doubt from conflict between sources of information. Single hypothesis being considered, the doubt between two hypotheses is not explicitly handled and the corresponding hypotheses are usually considered as equiprobable. Conflict is handled in the same way. Moreover, probabilities-based fusion methods

usually need a learning step using a large amount of data, which is not necessarily available for an accurate estimation.

Another solution consists in working within the belief function theory [19, 20]. The main advantage of this theory is the possibility to deal with subsets of hypotheses, called propositions, and not only with single hypothesis. It allows to easily model uncertainty, inaccuracy, and ignorance. It can also handle and estimate the conflict between different parameters. This theory has been used in a large number of applications, from medical imaging [21], to remote sensing [22–24], just to name a few. Regarding the problem of detection, this theory enables the combination of parameters with different scales and physical dimensions. Finally, the inclusion of doubt in the process is extremely valuable for the expert who can incorporate this information for the final decision.

As a conclusion, the belief function theory is selected to address the considered application. The proposed fusion scheme is described in the next subsection.

3.1. Fusion scheme

For the detection of echoes in SAS images, the frame of discernment Ω defined for each pixel is composed of the two following hypotheses:

- (i) “object” (O) if the pixel belongs to an echo reflected by an object,
- (ii) “nonobject” (NO) if it belongs to the noisy background or a shadow cast on the seabed.

The set of propositions 2^Ω is thus composed of four elements: the two single hypothesis, also called singletons, O and NO, the set $\Omega = \{O, NO\}$, noted $O \cup NO$ ($\cup$ means logical OR) and called “doubt,” and the empty set called “conflict.” In this application, the world is obvious closed (Ω contains all the possible hypotheses).

The proposed fusion process uses the local statistical parameters extracted from the SAS image, as presented in Section 2. These parameters are fused as illustrated on Figure 5: the relationship between the first two statistical orders is taken into account by using the thresholds in standard deviation and mean estimated by the automatic segmentation, the third and fourth statistical moments are used after focusing and rebuilding operations.

3.2. Definition of the mass functions

The mass of belief is the main tool of the belief function theory as the probability for the probability theory. The definition of the mass functions enables to model the knowledge provided by a source on the frame Ω . In this application, every parameter is used as a source of information. For one given source i , a mass distribution m_i^t on 2^Ω is associated to each value t of the parameter. This type of functions verifies the following property: $\sum_{a \in 2^\Omega} m(a) = 1$. We propose to define each mass function by trapezes or semitrapezes. In the considered application, only the three propositions (O), (NO), and $(O \cup NO)$ are concerned. Four thresholds must thus be defined, namely, t_1^1 , t_1^2 , t_1^3 , and t_1^4 (see Figure 6). They

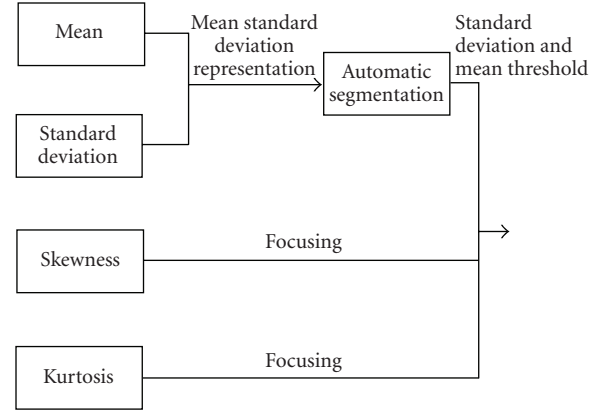


FIGURE 5: Main structure of the proposed detection system.

are set using knowledge on local and global statistics of sonar images. They also take into account the minimization of the conflicts while preserving the detection performances (no *nondetection*).

The first mass function concerns the two first statistical orders simultaneously because they are linked by the proportional relationship. In order to build the trapezes, we consider the pair (*mean; standard deviation*) as used for the automatic segmentation. Figure 7 illustrates the design of the corresponding mass function, based on the mean standard deviation representation. We first describe this function in the general case, the setting of the parameters being described afterward.

Pixels with a local standard deviation below t_1^1 are assigned a mass equal to one for the proposition “nonobject” and a mass equal to zero for the others. Pixels with a local standard deviation between t_1^1 and t_1^2 are assigned a decreasing mass (from one to zero) for the proposition “nonobject,” an increasing mass (from zero to one) for the proposition “doubt,” meaning “object OR nonobject” ($O \cup NO$). These variations are linear in function of the standard deviation. The construction of the mass functions goes in a similar way for t_1^3 and t_1^4 . This mass function is function of the standard deviation, but, considering the proportional relation holding between the mean and the standard deviation, an equivalent mass function can easily be designed for the mean. Then the mass function corresponding to the mean being redundant with the standard deviation, is not computed.

We propose to set the different parameters of these mass functions using the following expressions:

$$\begin{aligned}
 t_1^1 &= M_W(\hat{\sigma}_B), \\
 t_1^2 &= M_W(\hat{\sigma}_B) + \sqrt{V_W(\hat{\sigma}_B)}, \\
 t_1^3 &= \sigma_s - \frac{1}{2}\sqrt{V_W(\hat{\sigma}_B)}, \\
 t_1^4 &= \sigma_s + \frac{1}{2}\sqrt{V_W(\hat{\sigma}_B)},
 \end{aligned} \tag{4}$$

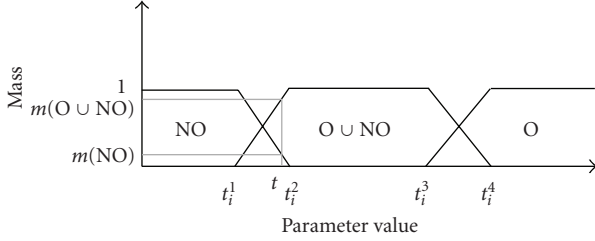


FIGURE 6: Definition of the mass functions.

where $\hat{\sigma}_B$ stands for the background standard deviation estimated, using the Weibull model previously computed, on a region of the image without any echo. σ_s is the threshold in standard deviation fixed by the algorithm described in Section 2.2. M_W and V_W are the mean and variance of the standard estimators [25] applied on σ_s considering the size of the computation window used for mean standard deviation building. This allows to take into account the uncertainty in the statistical parameters estimation by the fuzziness of the mass distributions.

The two other mass functions concern the HOSs: the skewness and the kurtosis, respectively. As mentioned in Section 2.3, the corresponding detector provides less accurate results, which prevent a precise definition of the areas of interest. Furthermore, some artifacts generate false alarms. As a consequence, the information provided by these parameters will only be considered to assess the certainty of belonging to the background. A null mass is thus systematically assigned to the proposition “object,” whatever are the values of the HOS. The mass is distributed over the two remaining propositions: “nonobject” and “doubt.” This is illustrated on Figure 8: only two parameters remain, t_2^1 and t_2^2 .

Parameters t_2^1 and t_2^2 (skewness) are set by considering the normalized cumulative histogram, noted $H(t)$, of the HOS values over the whole SAS image. This is illustrated in Figure 9. Considering that pixels with low HOS values necessarily belong to the noisy background and that pixels with high values (that might belong to an echo of interest) are extremely rare, the following expressions are used:

$$\begin{aligned} t_2^1 &= H^{-1}(0.75), \\ t_2^2 &= H^{-1}(0.90). \end{aligned} \quad (5)$$

These equations are valid for t_3^1 and t_3^2 (kurtosis). This assumes that at least 75% of the image belong to the background, which is easily fulfilled. Similarly, the 10% pixels with the highest HOS values are considered as potential objects (doubt has a mass equal to one).

3.3. Combination of the mass functions

Based on the local statistical moments of the data, three mass functions have been defined. The data fusion aims at improving the detection performances and ease the final

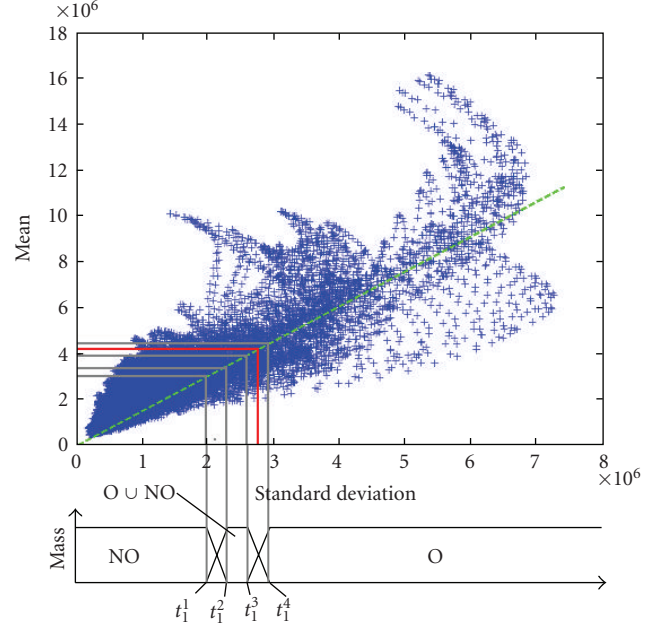


FIGURE 7: Definition of the mass functions for the first two-order statistical parameters: $t_i^2 - t_i^1 = t_i^4 - t_i^3 = \sqrt{V(\hat{\sigma}_W)}$ (the thresholds obtained from the automatic segmentation are in red. The mean standard deviation graph given on this figure has been calculated on the image presented on Figure 20(a)).

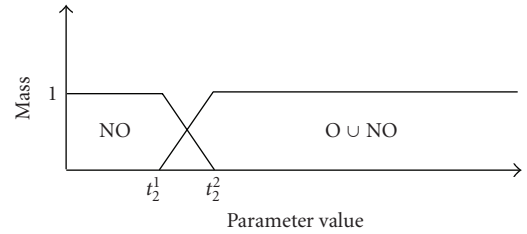


FIGURE 8: Definition of the mass functions for the higher-order statistics (the graphic is valid for definition of t_3^1 and t_3^2).

decision by the expert. It is performed using the following conjunctive rule:

$$m_{1,2,3} = m_1 \oplus m_2 \oplus m_3, \quad (6)$$

where $\oplus$ is the conjunctive sum: $(m_1 \oplus m_2)(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$ with A a proposition.

Note that, for the sake of simplicity, the superscript t of the mass m_t^i , corresponding to the parameter value, is removed from the notations.

A conflict between the different sources can appear during this combination phase. This information is preserved as it is valuable to assess the adequacy of the fused parameters. If one of the fused parameters provides irrelevant information, the conflict is high. Further investigations are then required to determine the cause of this situation (bad estimation of the parameter, limits of the data, etc.).

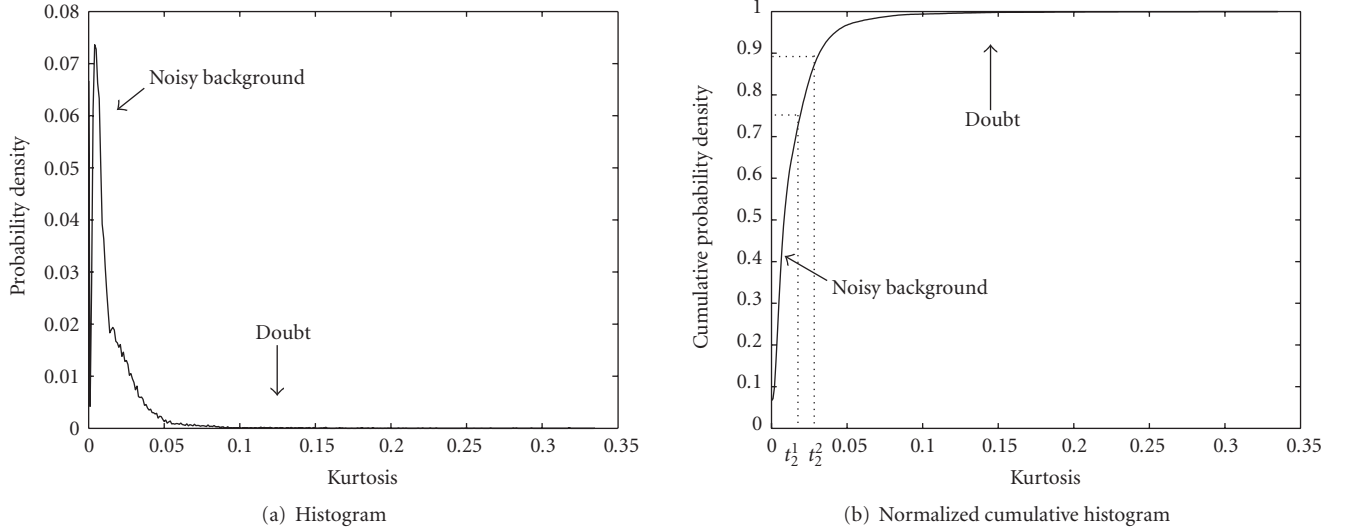


FIGURE 9: Example of histogram and cumulative histogram of a kurtosis image (after focusing and rebuilding, as presented on Figure 4(c)).

3.4. Decision

The results of the fusion step can be used in different ways, producing different end user products:

- (i) “binary” representations can be generated, providing segmented images and giving a clear division of the image into regions likely to contain objects or not;
- (ii) “enhanced” representations of the original SAS image can also be constructed from the results of the fusion. These representations should somehow underline the regions of interest while smoothing the noise, but leave the decision to the human expert.

The “binary” representations only use the results of the fusion process in order to classify each pixel according to the belief, the plausibility, or the pignistic value. A simple solution consists in thresholding the belief or plausibility for the proposition “object,” for instance, all the pixels with a belief above 0.5 are assigned to the class “echo.” A binary image is obtained, separating the echoes from the background. However, this method requires the setting of the threshold by the user. In order to overcome this shortcoming, another strategy consists in associating each pixel to the hypothesis (“object” or “nonobject,” resp.) with the highest belief. This unsupervised method also provides a binary image. The same methods can be used with the plausibility. However, since the space of discernment only contains two elements, plausibility and belief actually provide the same results. The corresponding results are presented on different datasets on Figures 10(d) and 11(d).

Beyond the binary result, a more precise classification can be constructed by assigning every pixel to the class with the highest mass of evidence (including the conflict). The resulting image is divided into four classes: “object,” “nonobject,” “doubt,” and “conflict.” Corresponding results are presented on Figures 10(c) and 11(c). This nonbinary

representation leaves more flexibility to the expert for the final interpretation. A similar strategy has been used in the frame of medical imaging in [21].

These representations are well suited for a “robotics-oriented” detection: regions of interest are defined, and an automatic system, such as an autonomous underwater vehicle (AUV), can be sent to identify the objects. However, such representations lose a lot of potentially valuable information (environment, relief, intensity, etc.). Such information may be useful for a human expert to actually identify the objects and solve some ambiguities. Therefore, other representations can be considered. For instance, we propose to combine the results of the fusion with the original image in order to enhance information. This is achieved by weighting the pixels of the original image by a factor linearly derived from the belief (or the plausibility) of the class “object.” The intensity of pixels likely NOT being echoes is decreased (low belief), thus enhancing the contrast with the pixels most likely being echoes. For instance, Figures 10(b) and 11(b) feature the resulting images with the weighting factor ranging from 0.3 for a null plausibility to 1 for a plausibility of 1. On these images, the background tends to disappear, but all potential objects of interest are preserved. Finally, another solution consists in performing an adaptive filtering of the sonar image in function of the belief. This is described in [26].

4. TOOLS FOR EVALUATION OF THE PERFORMANCES

The decision coming from the results of the fusion process is valid only if the algorithm generating these results is sufficiently efficient. That is why, assessing the performances of the proposed algorithm is a crucial problem. In this section, we propose and discuss different approaches. A manually labeled ground truth can be taken into account or

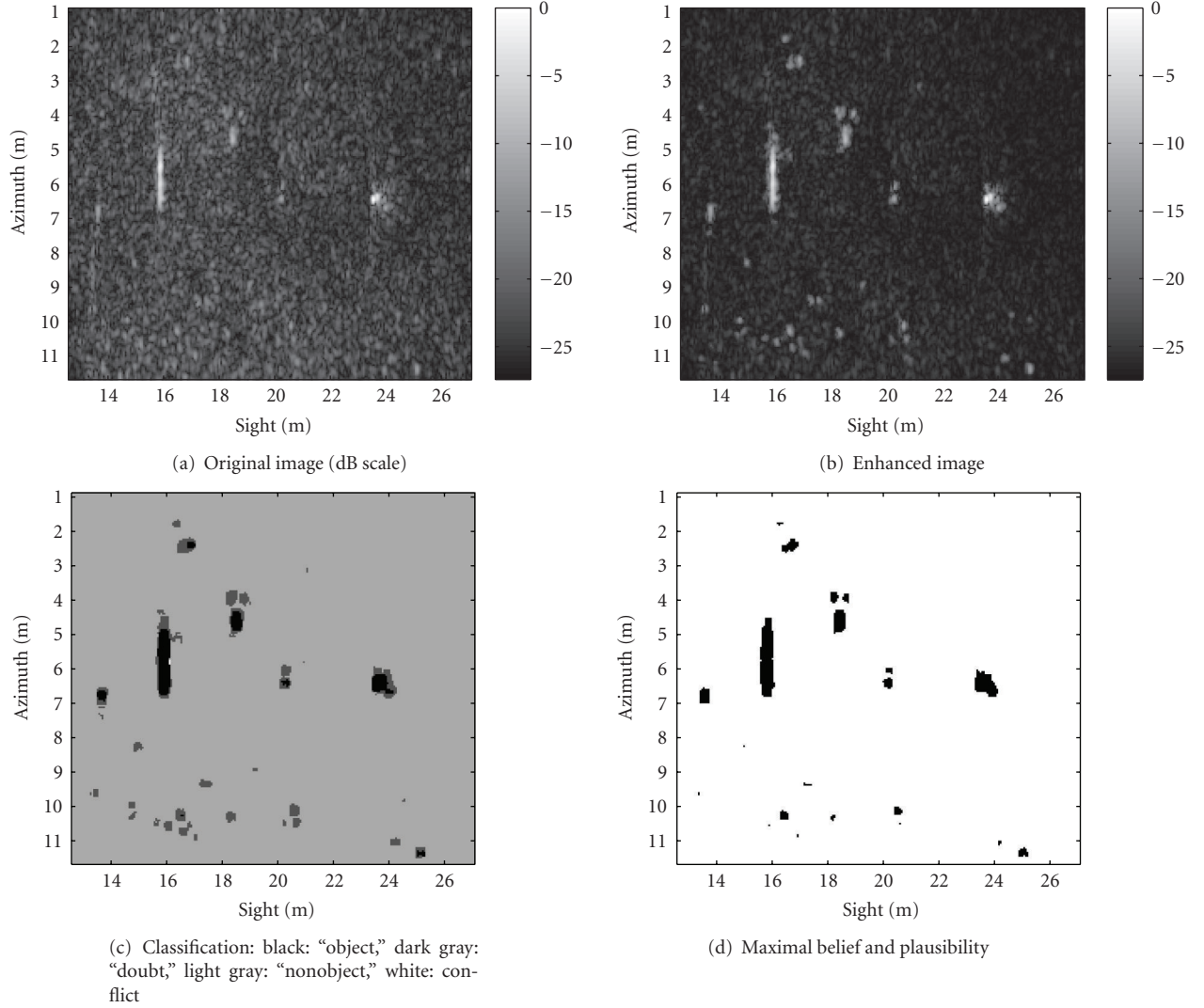


FIGURE 10: Presentation of the fusion results—image 1.

not, the evaluation can work on a direct analysis of the mass functions or on the classification results.

Note that in this section $m_i(A)$ denotes the mass value associated to the proposition A for the pixel i after fusion.

4.1. Intrinsic qualities of the mass functions

The evaluation of the detection performances can first be addressed by directly considering the quality of the resulting mass distribution. The first criterion is the *nonspecificity* [27]. This value estimates the ambiguity remaining in the mass distribution: it is low if the largest part of the mass of evidence is on a singleton or a single hypothesis (certain response); it is high if the mass is on a proposition of higher cardinal (doubt on several hypotheses). The nonspecificity is defined on a mass m by the following expression:

$$N(m) = \sum_{A \in 2^\Omega} m(A) \cdot \log_2 |A|, \quad (7)$$

where $|A|$ is the cardinal of the subset A .

The nonspecificity can take values in the following interval:

$$0 \leq N(m) \leq \log_2 |\Omega|. \quad (8)$$

The bottom limit (zero) is reached in the case of $a_i \in \Omega$ with $m(\{a_i\}) = 1$ (total certainty). It reaches the upper limit with $m(\Omega) = 1$ (total ignorance). The lower the is nonspecificity, the better and more accurate is the detection. A nonspecific mass function gives few false responses (limited risk), but brings limited information (all the hypotheses can be true). On the contrary, a specific response is accurate, but has a higher risk of error.

For the addressed application, the space of discernment is composed of two hypotheses. The nonspecificity is only computed for the mass associated with the proposition “doubt.” Moreover, this value is bound with each pixel of an image. We choose then to define the density of nonspecificity

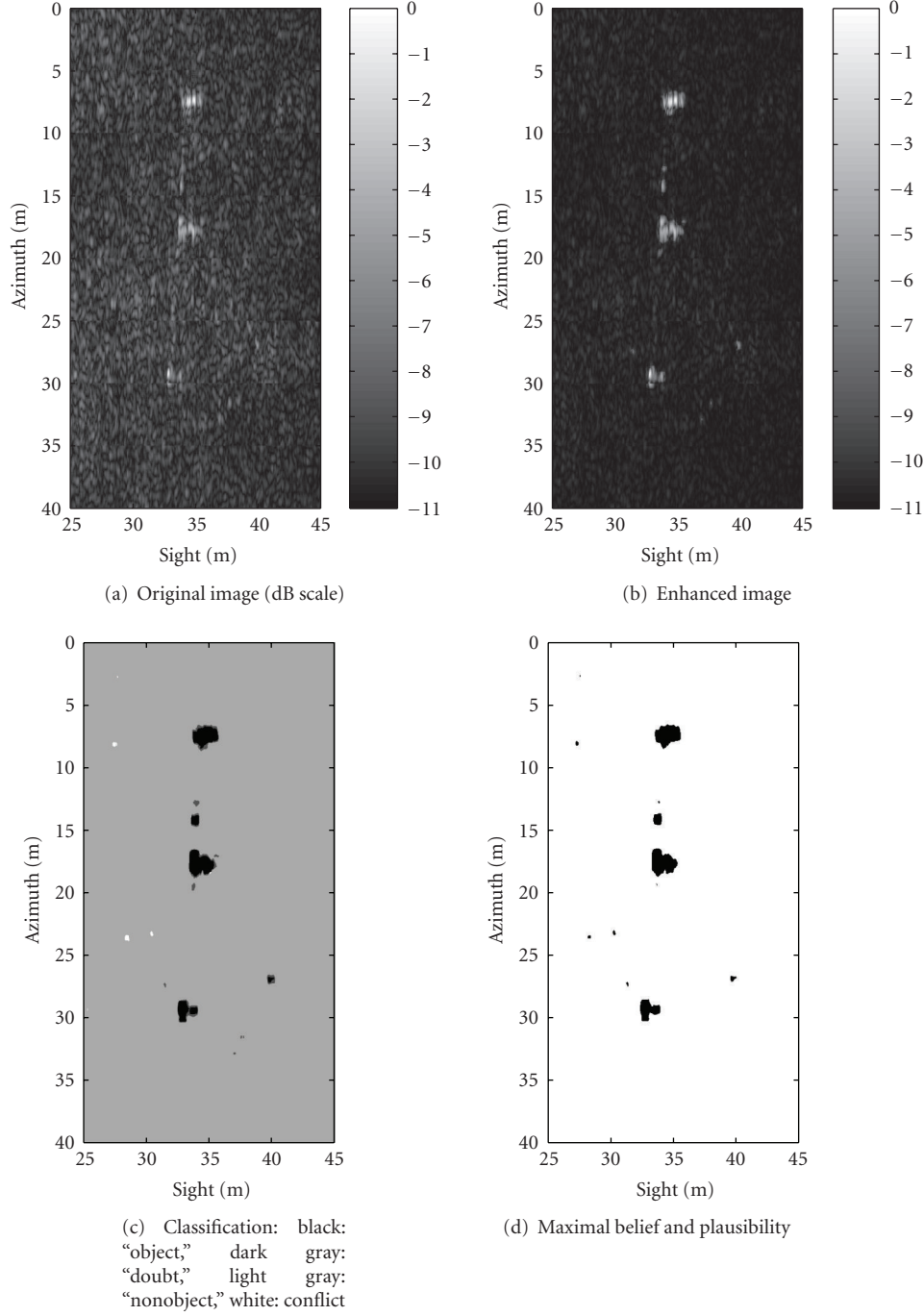


FIGURE 11: Presentation of the fusion results—image 2.

estimating the quality of the fusion result on the whole image. It is defined by the following expression:

$$d_{N(m)} = \frac{1}{n} \sum_{i=1}^n m_i(\text{O} \cup \text{NO}) \quad (9)$$

with n the size of the image (in pixels) and $m_i(\text{O} \cup \text{NO})$ the mass of "doubt" for the pixel i after the fusion. The values of

this density are between 0 and 1. The lower is this density, the more certain is the response of the fusion.

On the other hand, the higher is the specificity, the higher is the risk of conflict. Consequently, the conflict between sources must be analyzed. As previously, we define a density of conflict with the following expression:

$$d_{C(m)} = \frac{1}{n} \sum_{i=1}^n m_i(\emptyset). \quad (10)$$

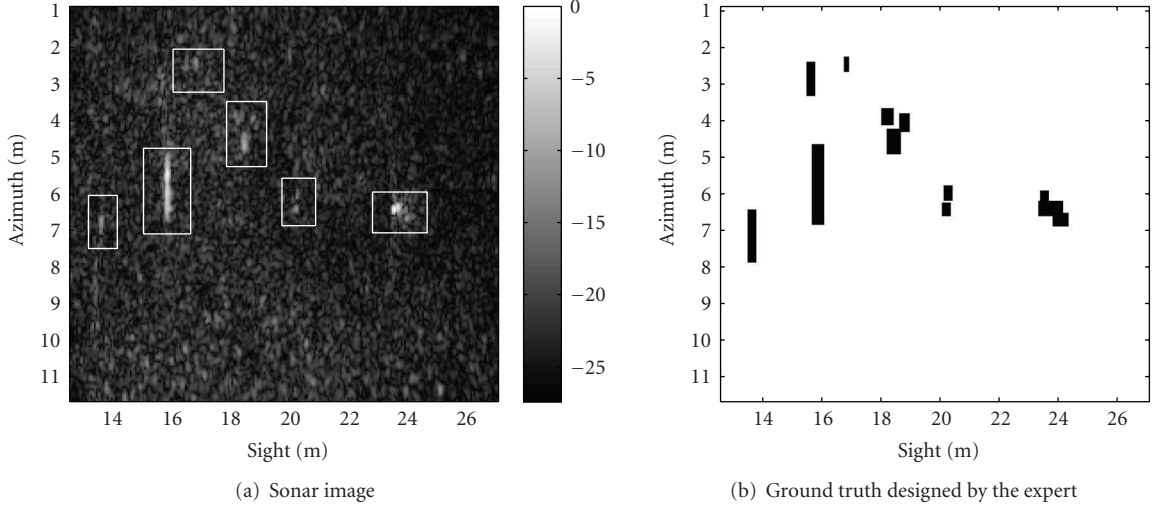


FIGURE 12: Example of image used for the environment truth.

This density is between 0 and 1. Obviously, the lower is this density, the more coherent are the sources of information and the more reliable is the result.

4.2. Assessing the quality of the mass functions when a ground truth is available

In order to validate the results of the fusion process, additional information can be used. For instance a ground truth can be designed by the expert. Figure 12(b) features such a segmented image where the expert roughly isolated the pixels likely to correspond to actual echoes (class “objects” (O), in black) from the background (class “nonobjects” (NO), in white).

If B denotes the environment truth ($B \in \Omega$, e.g., see Figure 12: $B = O$ if the pixel is in black, $B = NO$ if the pixel is in white), we define the rate of nonspecificity knowing the environment truth B :

$$N\left(\frac{m}{B}\right) = \sum_{A/B \subset A} m(A) \cdot \log_2 |A|, \quad (11)$$

$N(m/B)$ corresponds to the sum of the elements including B , weighted by their cardinal. For instance, for one given pixel, if $B = O$, the rate of nonspecificity is estimated using the masses of $A = O$ and $A = O \cup NO$, respectively.

This expression is applied to SAS images and a rate of nonspecificity density associated with the hypothesis B can be defined by

$$d_{N(m/B)} = \frac{1}{n} \sum_{i=1}^n m_i(O \cup NO) \cdot \delta_i(B) \quad (12)$$

with $\delta_i(B) = 1$ if the pixel i has label B (“object” or “nonobject”) in the environment truth, $\delta_i(B) = 0$, otherwise. In this way, only the pixels with the correct assignment B are taken into account in the density estimation. This density consequently allows to characterize the nonspecificity previously estimated (see (9)): it can either come from doubt

on object detection (the most dangerous situation) or on the background.

It is obvious that addition of the density $B = O$ and the density $B = NO$ is equal to the density of nonspecificity of (9):

$$d_{N(m)} = d_{N(m/O)} + d_{N(m/NO)}. \quad (13)$$

In a similar way, we define a rate of error knowing the environment truth by the following expression:

$$\text{Er}\left(\frac{m}{B}\right) = \sum_{A \cap B = \emptyset} m(A) \log_2 (|A| + 1). \quad (14)$$

Considering our application, the rate of density of error, associated with B , is defined by

$$d_{\text{Er}(m/B)} = \frac{1}{n} \sum_{i=1}^n m_i(\bar{B}) \cdot \delta_i(B) \quad (15)$$

with $B \in \{O, NO\}$ and $\bar{B}$ the complementary set of B . As a matter of fact, for a given B , only the mass associated to $\bar{B}$ is taken into account: B and $O \cup NO$ have at least one common element with B . The total error density can also be calculated by adding the two errors corresponding to $B = O$ and $B = NO$:

$$d_{\text{Er}(m)} = d_{\text{Er}(m/O)} + d_{\text{Er}(m/NO)}, \quad (16)$$

where $d_{\text{Er}(m)}$ is an estimation of the detection quality, considering potential mistakes on the pixel nature (“object” or “nonobject”). It should be as low as possible.

In this part, it is assumed that the designed ground truth actually corresponds to the truth. However, in real cases, this might be different as the expert might hesitate on the actual nature of some pixels (fuzzy boundaries of the objects of interest, false alarms, etc.). This results in errors that appear in the error parameters.

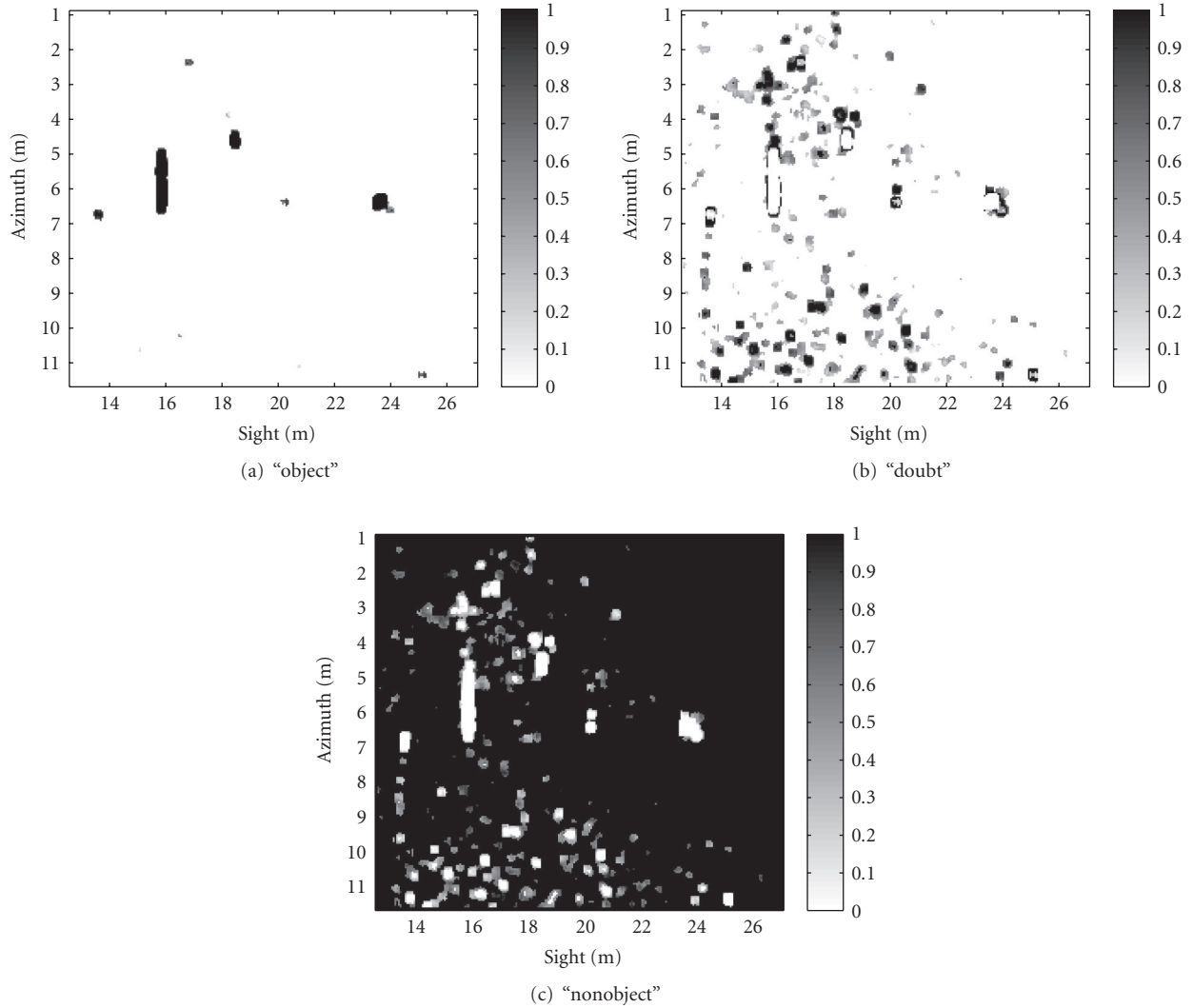


FIGURE 13: Mass images obtained for each proposition after the fusion of the mean standard deviation parameter (segmentation) in image 1.

A last criterion measuring the performance comes from the assumption that a decision is taken for each pixel, considering the corresponding mass functions. The images of belief and plausibility associated with the hypothesis “object” are segmented by applying a threshold. 50 different threshold values, between 0 and 1, are applied. For each threshold, the detection and false-alarm probabilities are computed on the resulting binary image. This is achieved by comparing the segmentation with the environment truth. The plot of these 50 points in the false-alarm rate versus detection probability plane features the ROC curve that is classically used to assess the performances of detection systems in sonar imagery.

5. RESULTS ON SONAR IMAGES

The fusion process presented in this paper is applied on various sonar images provided by the Délégation Générale pour l’Armement (DGA, France). These images are formed by an SAS system [13].

The first image (image 1, Figure 3(a)) features several buried or partially buried objects. This image represents a seabed region of about $10\text{ m} \times 10\text{ m}$, with a resolution of about 6 cm in both dimensions. In this image, the echoes are hardly visible apart from a partially buried cylindrical mine on the left (at 16 m in sight). For each pixel and for each parameter, we firstly estimate the mass associated to each proposition (“object,” “doubt,” and “nonobject,” resp.) by using the mass functions previously defined (Figures 13, 14, and 15, resp.). These images are combined using the orthogonal rule in order to obtain the mass images associated to each proposition (Figure 16). This results in an image of belief (corresponding to the mass of the class “object”) and an image of plausibility (corresponding to the sum of the masses of the classes “object” and “doubt”) associated to the proposition “object” (Figure 17).

One should underline that all the objects in the image are efficiently detected: belief and plausibility are close to 1 in the regions likely to contain echoes. The plausibility highlights some spurious regions at the bottom of the image. These

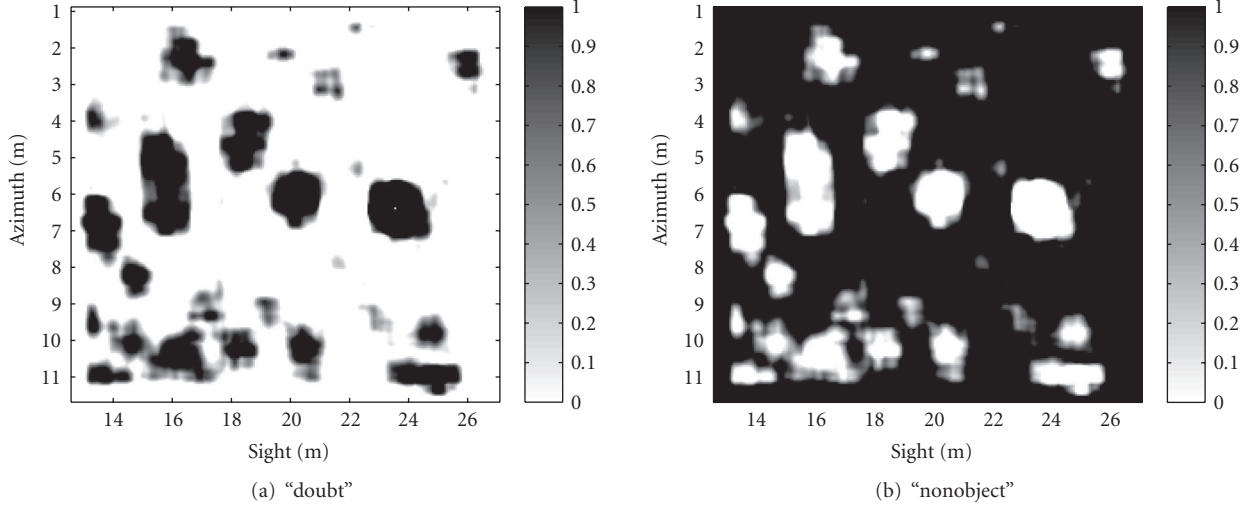


FIGURE 14: Mass images obtained for each proposition with the skewness parameter in image 1.

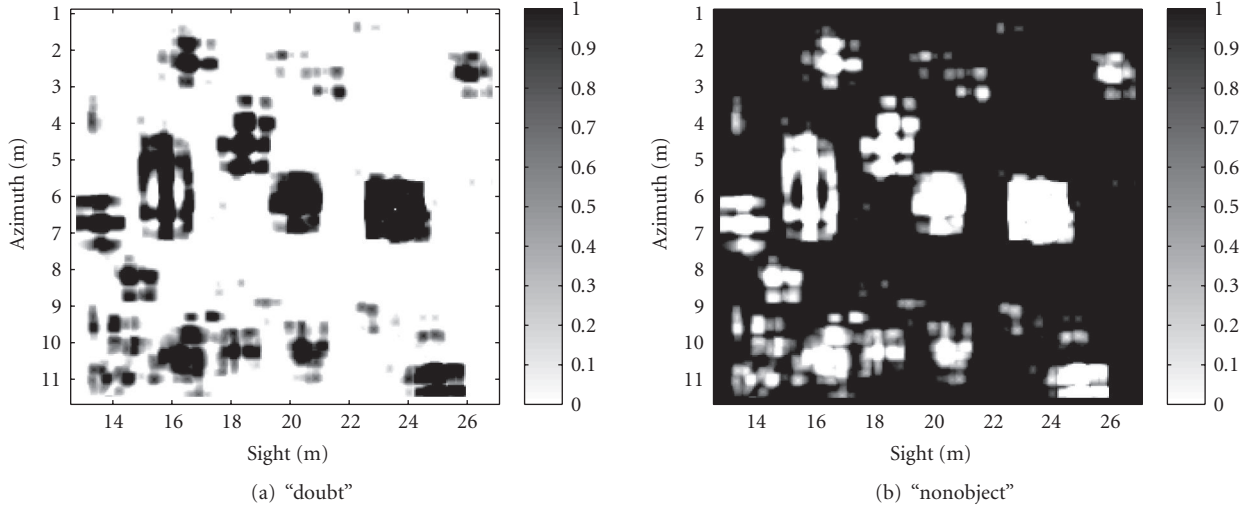


FIGURE 15: Mass images obtained for each proposition with the kurtosis parameter in image 1.

regions have a small area and could be easily removed, for instance, by a morphological filter.

The first- and second-order parameters are complementary to the third- and fourth-order ones. Actually, the doubt on Figure 13 (1st and 2nd order) is decreased by the mass "nonobject" brought by the higher-order statistical parameters (Figures 14 and 15). On the other hand, the doubt coming from HOS is limited by the mass "object" and "nonobject" provided by the first orders. The first-order parameters provide precise information, but with some false alarms (Figure 13), whereas higher orders provide a few false alarms (consider the "doubt" image), but imprecise information (Figures 14 and 15). It illustrates the usual duality between certainty and accuracy, and how a fusion process can take advantage of multiple complementarity sources.

Some conflict appears in the result of the fusion (Figure 16(d)). However, it remains low (the sum of the

masses of the focal elements is strictly inferior but close to 1), and isolated. This result shows the good concordance of the parameters.

5.1. Evaluation of the performances on the sonar image

A first evaluation of the fusion process consists in analyzing the contribution of each parameter to the final result. This is achieved by combining the parameters two by two. As previously observed, the addition of one HOS parameter decreases the mass "doubt" (compare Figure 18(b) with Figure 13(b)). The fusion of three parameters further decreases this mass (Figure 16(b)). The more parameters are added to the fusion process, the more accurate is the response. Note that the addition of one parameter to the fusion process "selects" more accurately the masses: the "object" mass that differs from the values 0 or 1 are fewer.

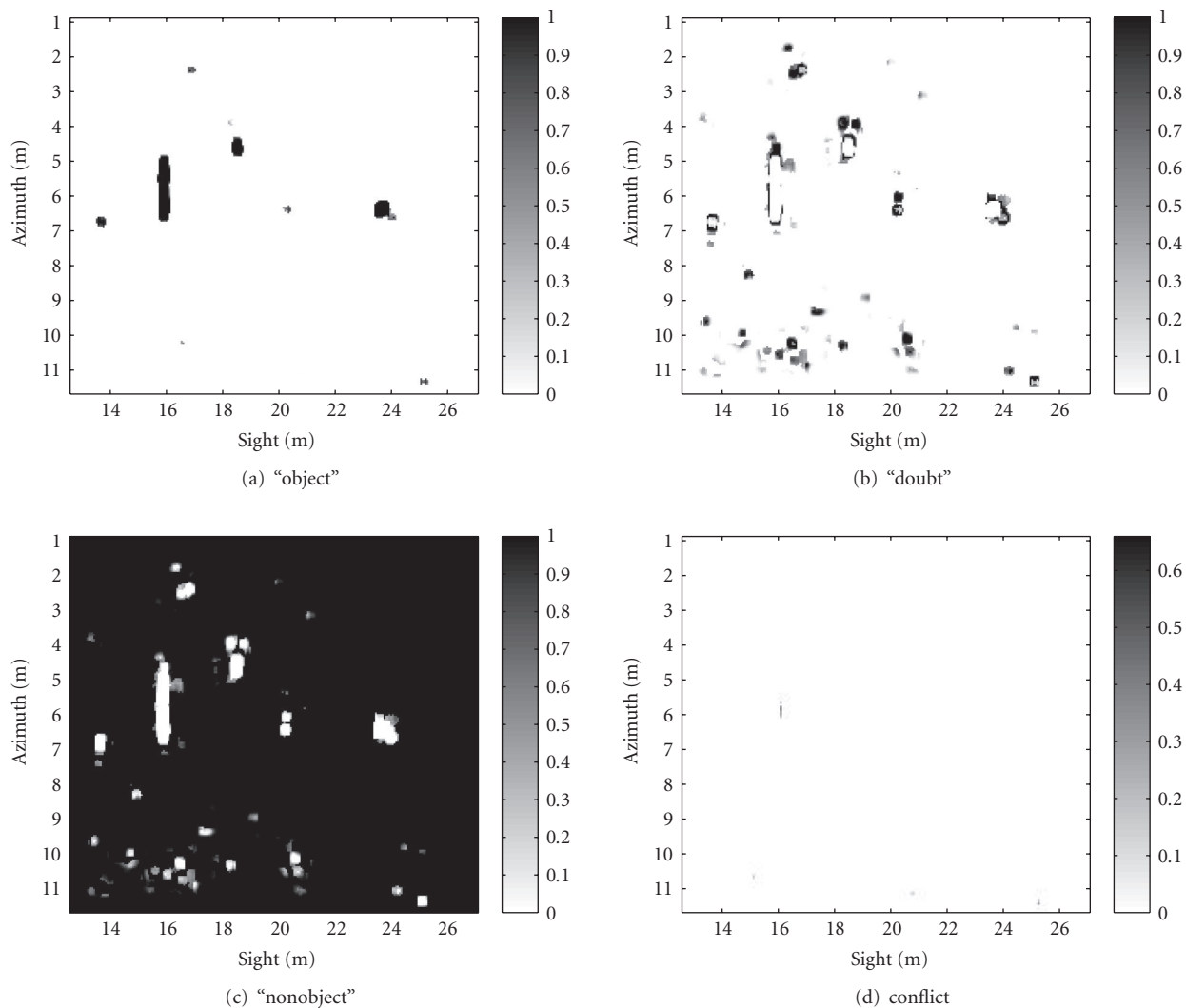


FIGURE 16: Mass images obtained after fusion of the three parameters in image 1.

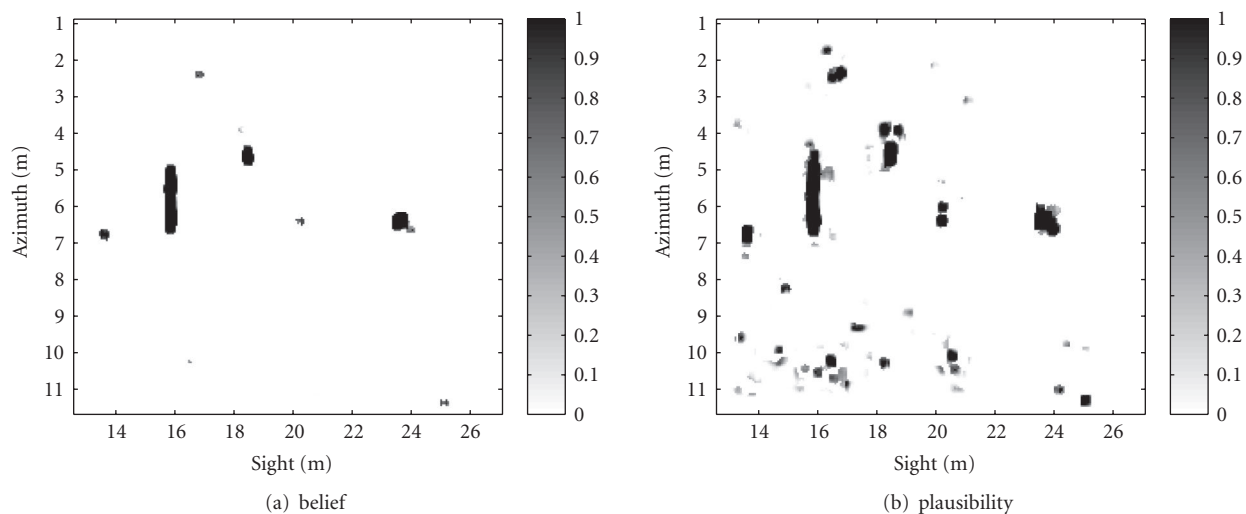


FIGURE 17: Belief and plausibility object images obtained after fusion of the three parameters in image 1.

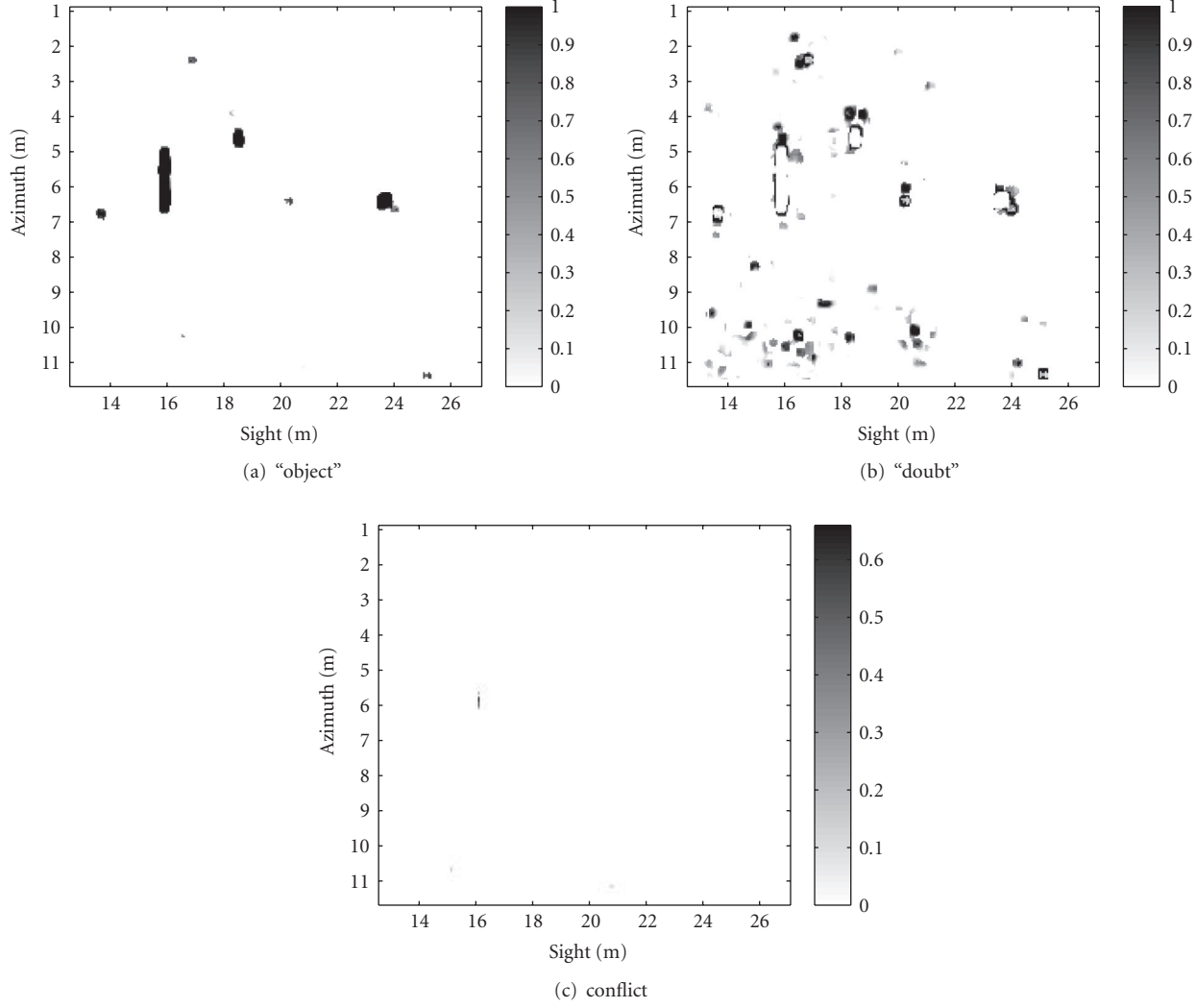


FIGURE 18: Mass images obtained for each proposition after the fusion of the mean standard deviation (segmentation) and kurtosis parameters in image 1.

TABLE 1: Performances of the fusion in image 2, (1-2: mean standard deviation (segmentation), 3: skewness, 4: kurtosis).

densities ($\times 10^{-3}$)	1-2	3	4	1-2 + 3	1-2 + 4	3 + 4	1-2 + 3 + 4
conflict	0	0	0	0.406	0.527	0	0.528
nonspecificity	8.1	159.0	162.8	3.9	3.5	122.0	3.4
/O	1.8	13.5	12.0	1.8	1.5	12.0	1.5
/NO	6.3	145.4	150.8	2.2	1.9	110.0	1.9
error	6.5	0.0446	1.6	6.0	6.2	1.6	6.2
/O	4.4	0.0446	1.6	4.4	4.6	1.6	4.6
/NO	2.0	0	0	1.6	1.6	0	1.6

A quantitative evaluation can also be completed by estimating conflict and nonspecificity densities, independently from the environment truth, or a combination of these values (rate of density of nonspecificities and error). The results are listed in Table 2.

The results confirm the previous qualitative remarks as follows:

- (i) nonspecificity decreases when new parameters are added. Note that this density is high for the two HOS parameters and their fusion;
- (ii) conflict can increase with the addition of one parameter, but this is not obvious in this application. That proves the good reliability of the chosen parameters.

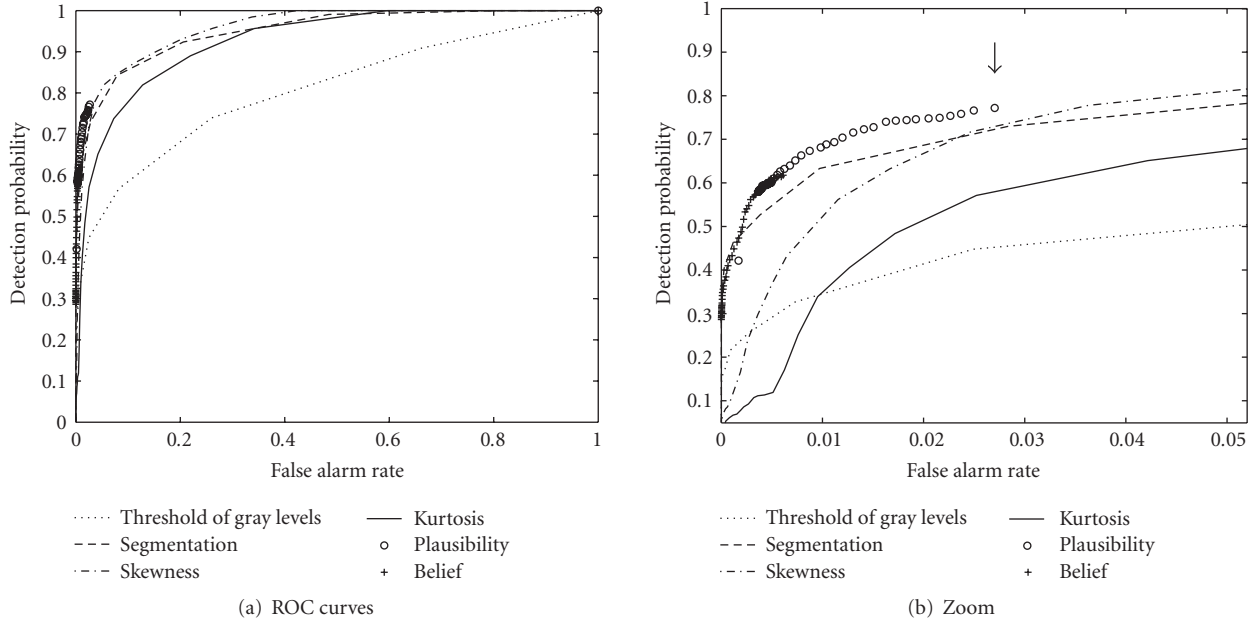


FIGURE 19: ROC curves of each of the three parameters compared with the results of the fusion process (belief and plausibility) in image 1.

TABLE 2: Performances of the fusion in image 1, (1-2: mean standard deviation (segmentation), 3: skewness, 4: kurtosis).

densities ($\times 10^{-3}$)	1-2	3	4	1-2 + 3	1-2 + 4	3 + 4	1-2 + 3 + 4
conflict	0	0	0	0.0299	0.0885	0	0.105
nonspecificity	51.0	166.1	166.1	23.8	20.9	121.3	19.0
/O	7.9	18.8	17.5	6.7	6.1	17.4	6.1
/NO	43.0	147.3	148.6	17.1	14.8	103.9	12.9
error	5.0	2.2	3.5	6.2	6.8	3.5	6.8
/O	4.5	2.2	3.5	5.7	6.3	3.5	6.3
/NO	0.557	0	0	0.520	0.535	0	0.518

These values also estimate the amount of information brought by each parameter: if adding one parameter does not significantly decrease the density of nonspecificity, the corresponding parameter can be considered as bringing very little information. Moreover, if the density of conflict increases, this parameter is contradictory with the others and the reliability of this parameter (or one of the other) should be questioned.

The environment truth is a source of information that can be used to assess the performances of the system. The addition of one HOS parameter slightly decreases the error, which remains low for the HOS. As a matter of fact, the fuzzy definition of the mass functions keeps the error bounded (if the mass “doubt” is 1, the error is null). On the contrary, the relatively high value of error on the areas selected as “object” can be explained by the large size of the regions selected by the expert. This rough selection actually includes a part of the region selected as “background” by the fusion process; but this should not be considered as a bad detection: the echoes are

well detected, but are only smaller than the masks of the original reference image. This will be confirmed by the ROC curves (the maximum detection probability is smaller than one).

The nonspecificity is greater for the “nonobject” pixels on the reference image than for the “object” pixels. This is a promising conclusion for the fusion process: the result is more accurate if a potentially dangerous object is present.

Finally, ROC curves of the fusion results are built and compared with the curves obtained with each parameter alone (segmentation with the 1st and 2nd order, the skewness, or the kurtosis). They are also compared with the ROC curves obtained with the standard detector consisting in directly thresholding the original data.

The first comment on the results presented in Figure 19 concerns the lack of points between low values of false alarms (until 0.03) and the point of probability equal to 1. This is a consequence of the pixels declared as “echo” by the expert, but classified as “nonobject” by the system. In order to include these pixels as “object” by the system, all

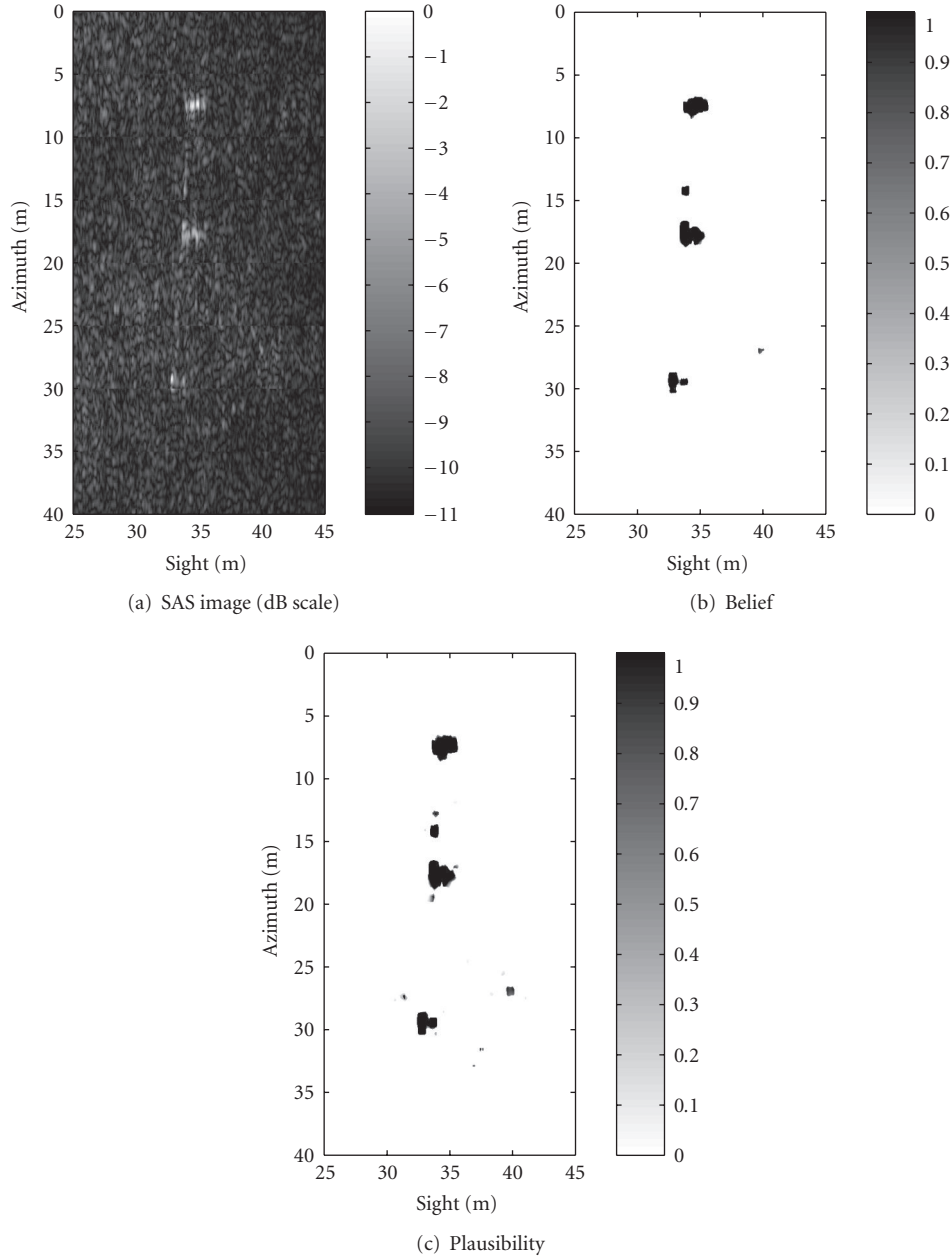


FIGURE 20: Belief and plausibility images obtained after fusion of the three parameters in image 2.

the pixels of the image must be selected (this is, achieved with a threshold of zero). These pixels are not significant at all and come only from the rough design of the regions containing echoes. This results in the maximum false-alarm and detection probabilities being far from the point (1, 1) (see the arrow on Figure 19(b)). In the same way, minimum detection and false-alarm probabilities exist for belief and plausibility obtained with a threshold of 1.

The second comment is that the false-alarm rates and detection probabilities are lower for belief than plausibility. This is linked to the certainty/accuracy duality previously mentioned. Moreover, note that the plausibility and the belief curves are both above all the other curves: this assesses

the improvement of the detection performances obtained thanks to the fusion process.

5.2. Results on other data

In this section, the proposed fusion process is tested on two more SAS images. *Image 2* (Figure 20) represents a region of $40 \text{ m} \times 20 \text{ m}$ of the seabed with a pixel size of about 4 cm in both directions [28, 29]. It contains three cylindrical mines: one mine is lying on the sea floor (top of image), another one is partially buried (approximately in the middle of the image), and the last one is completely buried under the sea floor (lower part of image).

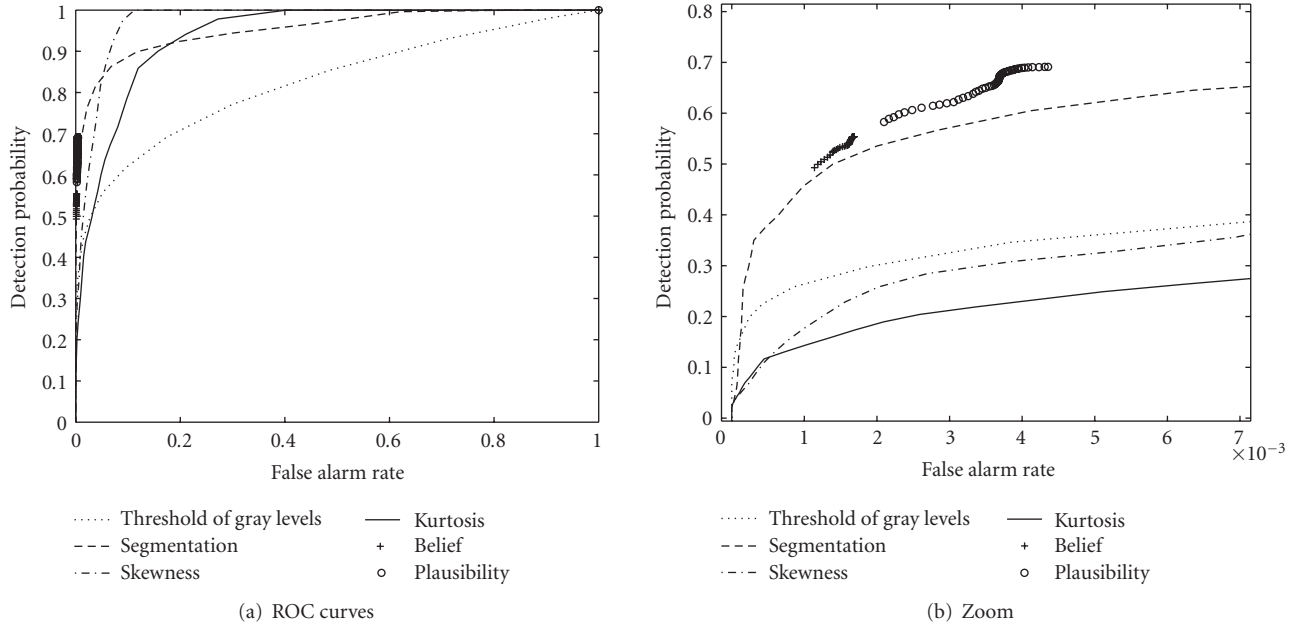


FIGURE 21: ROC curves of each of the three parameters compared with the results of the fusion process (belief and plausibility) in image 2.

Figure 20 represents the belief and plausibility after fusion, and Figure 21 presents the corresponding ROC curves. Moreover, quantitative criteria estimated for *image 2* are presented in Table 1 and can be compared with the results of the first image. The fusion process have been performed with mass functions defined previously, in function of the corresponding standard deviation thresholds and higher-order statistics histogram.

The same comments and conclusions hold for this new image. The detection performances are improved (in particular, see the belief image). However, the fusion with the skewness parameter does not significantly affect the result in *image 2*: the nonspecificity, error, and conflict densities are similar whether two or three parameters are aggregated.

6. CONCLUSION AND PERSPECTIVES

The proposed fusion architecture aims at taking advantage of the complementary properties of sources, based on statistical properties, in order to improve the detection performances. Being able to handle conflicts between sources and doubt between different hypotheses, the belief theory is well suited to represent and characterize the information provided by the different sources. It also provides a fusion rule. The fused data can be used either to take a decision or to enhance the data adaptively, leaving the final decision to an expert.

The design of the mass functions is fairly simple and flexible. A general knowledge about the acquisition system and the induced statistical properties on the SAS image enables the setting of the few parameters (trapeze-shaped functions). Confronted to different datasets, these settings were not modified, thus assessing the robustness of the whole procedure.

The evaluation of the proposed architecture is based on new parameters, some of them classically taking a manually labeled ground truth into account, some others being independent from this ground truth and aiming at directly assessing the quality of the available information. These last criteria determine intrinsic properties of the mass functions, such as nonspecificity and conflicts densities. The first set of criteria concerns the properties conditioned by the ground truth: rates of nonspecificity and error densities, probabilities of detection and false alarm.

The fusion architecture has been tested on two real SAS images and convincing results have been obtained: the fusion actually improves the detection performances of the different sources taken separately.

The proposed process may be improved by incorporating new parameters (statistical, morphological, criteria characterizing the spatial distribution of the features, etc.) coming either from a deeper knowledge of the data or from new sonar images (multiple acquisitions). The interest of such a fusion structure lies in its flexibility: the addition of new parameters is easy to work out and does not need any change of structure or parameterization. Moreover, it is possible to estimate the quantity of information brought by each of the new parameter. This allows to reach the next levels in the detection and classification process, as described in the introduction, by deciding if the regions previously segmented actually contain a sought object and by identifying this object (mine, kind of mine, etc.).

ACKNOWLEDGMENTS

The authors wish to thank Groupe d'Etudes Sous-Marines de l'Atlantique (DGA/DET/GESMA, France) and TNO,

Security and Safety (The Netherlands) for providing SAS data in this work supported by GESMA.

REFERENCES

- [1] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy, "Unsuper-vised Markovian segmentation of sonar images," in *Proceedings of the 22nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, vol. 4, pp. 2781–2784, Munich, Germany, April 1997.
- [2] F. Maussang, J. Chanussot, and A. Hétet, "Automated segmen-tation of SAS images using the mean-standard deviation plane for the detection of underwater mines," in *Proceedings of the Oceans Conference (OCEANS '03)*, vol. 4, pp. 2155–2160, San Diego, Calif, USA, September 2003.
- [3] T. Aridgides, M. Fernández, and G. Dobeck, "Adaptive three-dimensional range-crossrange-frequency filter process-ing string for sea mine classification in side scan sonar imagery," in *Detection and Remediation Technologies for Mines and Minelike Targets II*, vol. 3079 of *Proceedings of SPIE*, pp. 111–122, Orlando, Fla, USA, April 1997.
- [4] R. T. Kessel, "Using sonar speckle to identify regions of interest and for mine detection," in *Detection and Remediation Technologies for Mines and Minelike Targets VII*, vol. 4742 of *Proceedings of SPIE*, pp. 440–451, Orlando, Fla, USA, April 2002.
- [5] S. W. Perry and L. Guan, "Pulse-length-tolerant features and detectors for sector-scan sonar imagery," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 1, pp. 138–156, 2004.
- [6] G. L. Foresti, V. Murino, C. S. Regazzoni, and A. Trucco, "A voting-based approach for fast object recognition in under-water acoustic images," *IEEE Journal of Oceanic Engineering*, vol. 22, no. 1, pp. 57–65, 1997.
- [7] G. Ginolhac, J. Chanussot, and C. Hory, "Morphological and statistical approaches to improve detection in the presence of reverberation," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 4, pp. 881–899, 2005.
- [8] T. Aridgides, M. Fernández, and G. Dobeck, "Side-scan sonar imagery fusion for sea mine detection and classification in very shallow water," in *Detection and Remediation Technologies for Mines and Minelike Targets VI*, vol. 4394 of *Proceedings of SPIE*, pp. 1123–1134, Orlando, Fla, USA, April 2001.
- [9] T. Aridgides, M. Fernández, and G. J. Dobeck, "Recent processing string and fusion algorithm improvements for automated sea mine classification in shallow water," in *Detection and Remediation Technologies for Mines and Minelike Targets VIII*, vol. 5089 of *Proceedings of SPIE*, pp. 65–76, Orlando, Fla, USA, April 2003.
- [10] T. Aridgides, M. Fernández, and G. Dobeck, "Processing string fusion approach investigation for automated sea mine classification in shallow water," in *Proceedings of the Oceans Conference (OCEANS '03)*, vol. 2, pp. 1111–1118, San Diego, Calif, USA, September 2003.
- [11] N. Milisavljević, I. Bloch, S. van den Broek, and M. Acheroy, "Improving mine recognition through processing and Dempster-Shafer fusion of ground-penetrating radar data," *Pattern Recognition*, vol. 36, no. 5, pp. 1233–1250, 2003.
- [12] V. Kaftandjian, Y. M. Zhu, O. Dupuis, and D. Babot, "The combined use of the evidence theory and fuzzy logic for improving multimodal nondestructive testing systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 5, pp. 1968–1977, 2005.
- [13] J. Chatillon, M.-E. Bouhier, and M. E. Zakharia, "Synthetic aperture sonar for seabed imaging: relative merits of narrow-band and wide-band approaches," *IEEE Journal of Oceanic Engineering*, vol. 17, no. 1, pp. 95–105, 1992.
- [14] J. W. Goodman, "Some fundamental properties of speckle," *Journal of Optical Society of America*, vol. 66, no. 11, pp. 1145–1150, 1976.
- [15] F. Maussang, J. Chanussot, A. Hétet, and M. Amate, "Mean-standard deviation representation of sonar images for echo detection: application to SAS images," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 4, pp. 956–970, 2007.
- [16] F. Maussang, J. Chanussot, and A. Hétet, "On the use of higher order statistics in SAS imagery," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 5, pp. 269–272, Montreal, Quebec, Canada, May 2004.
- [17] F. Maussang, J. Chanussot, A. Hétet, and M. Amate, "Higher-order statistics for the detection of small objects in a noisy background application on sonar imaging," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, Article ID 47039, 17 pages, 2007.
- [18] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, NY, USA, 1973.
- [19] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [20] P. Smets, "Data fusion in the transferable belief model," in *Proceedings of the 3rd International Conference on Information Fusion (FUSION '00)*, vol. 1, pp. PS21–PS33, Paris, France, July 2000.
- [21] 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*, vol. 17, no. 8, pp. 905–919, 1996.
- [22] F. Tupin, H. Maître, J.-F. Mangin, J.-M. Nicolas, and E. Pechersky, "Linear feature detection on SAR images: application to the road network," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 36, no. 2, pp. 434–453, 1998.
- [23] J. Chanussot, G. Mauris, and P. Lambert, "Fuzzy fusion techniques for linear features detection in multi-temporal SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1292–1305, 1999.
- [24] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2828–2838, 2006.
- [25] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 1, Charles Griffin, London, UK, 2nd edition, 1963.
- [26] F. Maussang, J. Chanussot, S. C. Visan, and M. Amate, "Adap-tive anisotropic diffusion for speckle filtering in SAS imagery," in *Proceedings of the Oceans Conference (OCEANS '05)*, vol. 1, pp. 305–309, Brest, France, June 2005.
- [27] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information*, Physica, Heidelberg, Germany, 1999.
- [28] A. Hétet, M. Amate, B. Zerr, et al., "SAS processing results for the detection of buried objects with a ship-mounted sonar," in *Proceedings of the 7th European Conference on Underwater Acoustics (ECUA '04)*, pp. 1127–1132, Delft, The Netherlands, July 2004.
- [29] J. C. Sabel, J. Groen, M. E. G. D. Colin, et al., "Experiments with a ship-mounted low frequency SAS for the detection of buried objects," in *Proceedings of the 7th European Conference on Underwater Acoustics (ECUA '04)*, pp. 1133–1138, Delft, The Netherlands, July 2004.