3-D Underwater Object Recognition

David Boulinguez, *Member, IEEE*, and André Quinquis, *Member, IEEE*

***Abstract*—In this paper, we propose an automatic supervised classification of objects lying on the sea floor or buried in sediment layers. This pattern recognition provides a way to distinguish natural and manufactured objects and then should be helpful to detect mine, pipe-line, or wreckage. Proposed methods combine different techniques: pattern information extraction, relevant parameter search, and supervised classifier. Parameters are automatically selected using a principal component analysis to reduce misclassification rate and to simplify classifier structure. Performances of different parameters (two-dimensional and three-dimensional) are compared and discussed from testing on synthetic and real data bases.**

***Index Terms*—Buried objects, feature selection, parametric sonar, supervised classification, two-dimensional (2-D) and three-dimensional (3-D) pattern recognition.**

I. Introduction

One OF the most difficult challenges in the underwater acoustic field is active classification for differentiation of various types of man-made objects. For instance, underwater object identification has been of great interest for a few years to marines (detection of mines), to acousticians (boulders), or to archaeologists (wreckages). Image and signal processing succeed in identifying objects lying on the sea bed. Usual techniques are based on target shadow analysis. In contrast, the recognition of buried targets remains complex. Indeed we need to take several factors into account: attenuation in sediment, resolution, transducer size etc .

In this paper, we deal with the design of a complete identification system from data acquisition to target feature computation, and classification rates.

Our approach is a real innovation in the field of buried target detection and classification. Recent studies are dealing with the same kind of problem. Many of them use acoustic scattering to recognize objects by analyzing acoustic resonance in the time-frequency domain, but these processes are usually limited to simple shaped objects. Recently, some studies tried to analyze the performance of this approach with more complex shapes [1], but the generalization to every kind of object seems to be really difficult. In the same way, [2] is probably one of the more original methods, where the investigation is concerned with imaging a buried object by inversion of measured scattered acoustical waves. Relevant parameters could be computed from these images to classify objects but frequencies used are really higher than those of the parametric sonar and the application in a real environment will probably be more difficult.

For the same reason, acoustic imagery technique is not suitable: it uses high frequencies that are too strongly attenuated inside the sediment.

Another method using a synthetic aperture sonar (SAS) has recently appeared. It provides two-dimensional (2-D) shape information. However, the major difficulty with SAS relates to micronavigation, the platform trajectory estimation during the SAS integration time. Moreover, 2-D shape information may not be sufficient to identify buried targets [3].

Finally, [4] and [5] propose another original approach using a parametric sonar: many parameters are computed from bispectrum, time-frequency space (…) and used to detect and to classify buried objects. However, only a two-class separation is realized and, relevance and utility of all these parameters still need to be analyzed for a more precise classification.

The originality of our work is based on the possibility to analyze precisely, the shape of buried targets, in order to achieve an efficient separation into many classes. The purpose is the development of a complete identification of objects embedded in marine sediment, thanks to an adapted technology using low frequencies and giving access to relevant target features. The proposed system is based on a parametric sonar associated with an original experimental approach. This approach is the first originality of this work. Indeed, it offers the possibility to realize a classification from three-dimensional (3-D) shape of buried targets. Moreover, we propose an automatic way of classifying objects. This leads naturally to find parameters able to discriminate between different shapes. Two original sets of parameters are compared in order to analyze the 3-D shape. On one hand, Fourier descriptors associated with a fusion method, and on the other hand, 3-D invariant parameters computed directly from the 3-D representation.

Thus, after a detailed presentation about the acoustic system in the first section, this paper describes the computation of pertinent parameters in Section III. Then, Section IV compares classification results using a supervised classifier.

II. Acoustic System and Data

The parametric array has many advantages as an acoustic source: high relative bandwidth, narrow beam with no sidelobes, and small transducer surface. All these features imply high spatial and temporal resolutions which are very useful in applications such as object detection and classification [6].

Briefly, in our application, a carrier frequency is generated and mixed in a double balance mixer with a raised cosine bell

Manuscript received October 15, 2000; revised June 15, 2001. This work was supported by the Direction des Systèmes de Forces et de la Prospective (DSP), Brest, France.

D. Boulinguez is with the Signal and System Department, Institute Supérieur d'Electronique du Nord (ISEN), 59046 Lille Cedex, France (e-mail: David.Boulinguez@isen.fr).

A. Quinquis is with the Electronics Engineering Department, Etudes et des Techniques d'Armement (ENSIETA), 29806 Brest Cedex, France (e-mail: quinquis@ensieta.fr).

Digital Object Identifier 10.1109/JOE.2002.805097

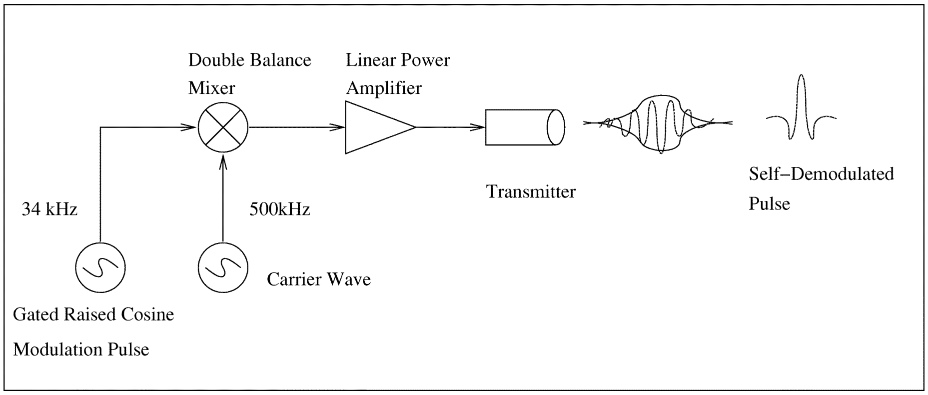


Fig. 1. Schematic of the transmission.

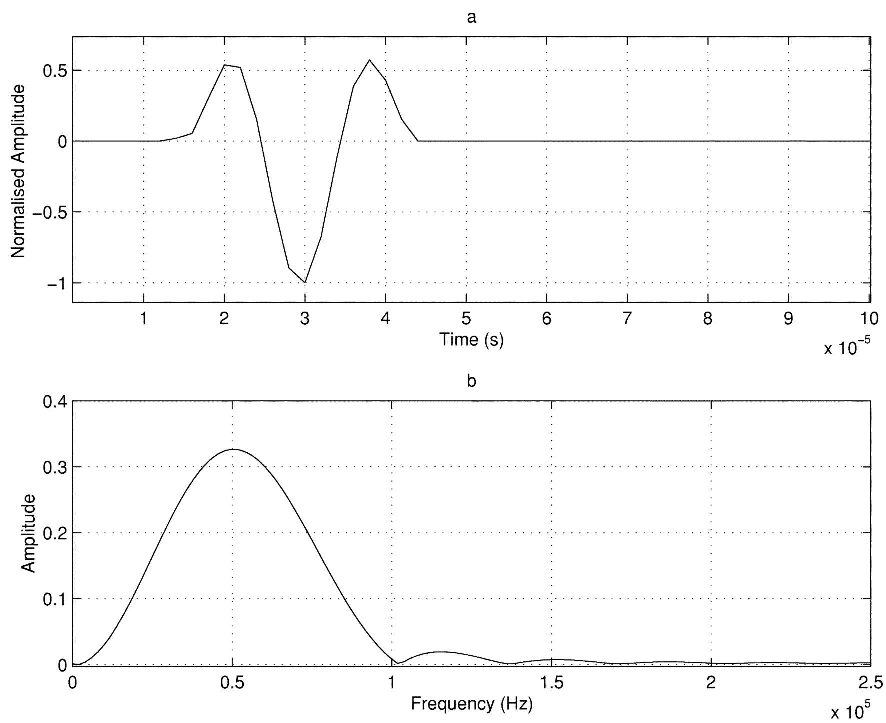


Fig. 2. Ricker (a) in the time domain and (b) in the frequency domain.

envelope created from a function generator (Fig. 1). The modulation is sent to a power amplifier and then to a transducer. Due to water nonlinear propagation characteristics, a low-frequency self-demodulated pulse is generated, which corresponds to the second derivative of the square envelope of the created signal [7].

In our experiment, the parametric sonar has a primary frequency around 500 kHz, whereas the secondary frequencies are centered at 50 kHz. The generated pulse is a Ricker pulse whose temporal and spectral features are described in Fig. 2. It is interesting to mention that the sonar beamwidth is really narrow (about 3°).

During the experiment, the sonar is moved, step by step, in both horizontal directions (x, y), near over the water/sediment interface (see Fig. 3). Signal frequencies are low enough to penetrate the sediment and the object without important attenuation. Thanks to the narrow beamwidth, a small surface of the object is reached by the beam. We can, therefore, realize precise multiline vertical scans over the object as shown in Fig. 4 and obtain cross-sectional images using adjacent reflected signals. For each target, many cross sections are available since the sonar is moved in both directions. So, with successive cross sections, we have access to 3-D shape information.

Real data have been collected from several targets during trials made in a pool: cylinders, spheres and truncated cones (these shapes are currently used for mines) and our purpose is therefore to separate these three classes. Fig. 5 shows a cross section of a real water-filled target buried in sandy sediment. One can clearly notice both interfaces due respectively to the upper surface and the lower surface of the object. However, the real data base was not large enough to develop and to test our classification algorithms. So, a simulated target data base has been created and our following processing are based on both data bases.

The simulated shapes have been generated **in three dimensions** with binary data with the following rule: target surface

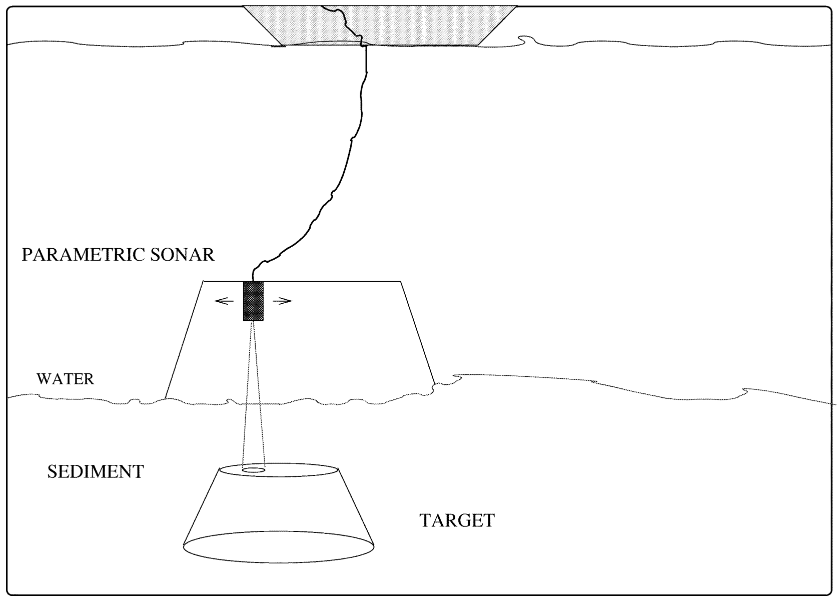


Fig. 3. Experimental environment: sonar moving.

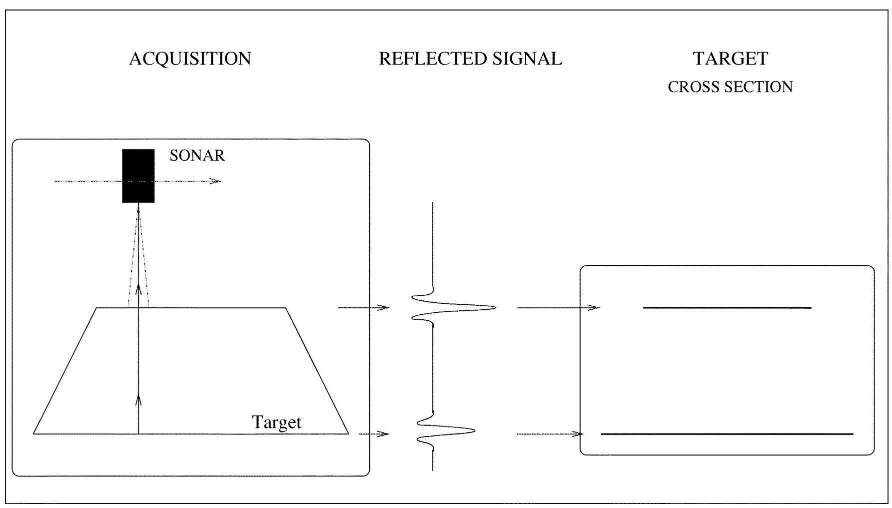


Fig. 4. Experimental approach: acoustic signal acquisition.

reflections are represented with voxels with value "1" (black pixels on Fig. 6) and the value of voxels that do not belong to the target surface is "0" (white pixels on Fig. 6). In order to have many different objects in the same class, we have created targets with various dimensions and we have introduced perturbations on 3-D shapes by disturbing target surface voxels. The position of each target voxel is defined as a Gaussian random number: its mean is the voxel position on undisturbed shape and its standard deviation is equal to σ_P voxels (for our simulations, the maximum deviation is around $3 \times \sigma_P = 6$ voxels). With this rule, we create a noise level labeled with two parameters:

- $P_{\%}$ which corresponds to the percentage of disturbed voxels;
- σ_P which corresponds to the importance of this perturbation.

Fig. 7 shows the result of a 3-D simulated truncated cone (after Delaunay triangulation).

III. Feature Extraction: From 2-D to 3-D Space

This section deals with the search for key parameters: these features must be characteristic of the target shape and must be invariant to the target position, orientation and size. So we need parameters which are invariant of 3-D translation, 3-D rotation and dilation.

Before computing relevant target parameters, we need to process raw data to extract the 2-D target shape (and then, the 3-D shape). So, the first step of the proposed procedure consists in visualizing properly the object inside the three-dimensional data.

A. Automatic "Tracking" of Target Echoes

In the following algorithm, the objective is to localize precisely target reflections inside each cross section so as to build 2-D edges and then, a 3-D surface. Consequently we can materialize the target shape.

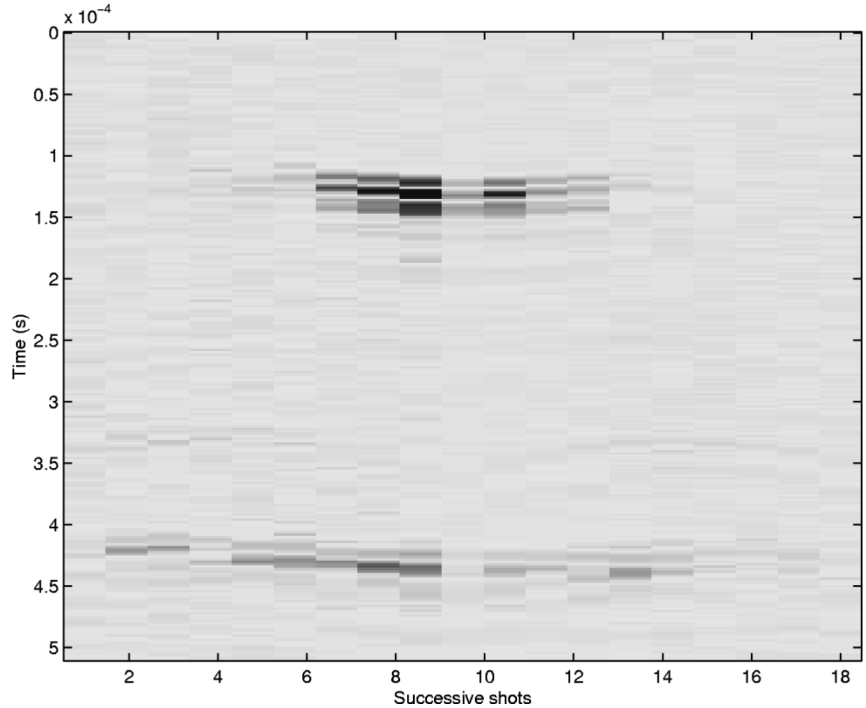


Fig. 5. Experimental result: cross section of a real water-filled truncated cone.

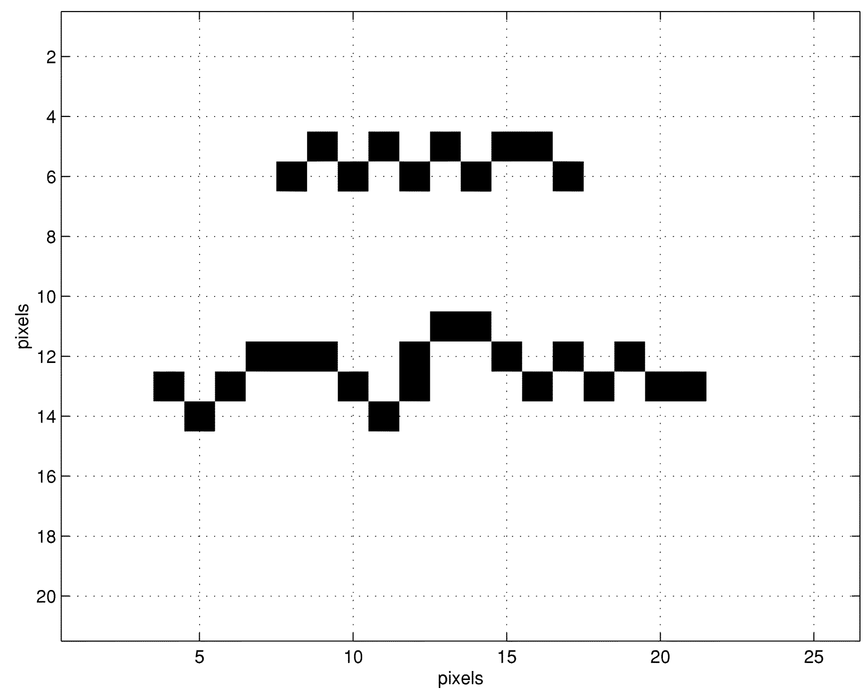


Fig. 6. Example of cross section from a simulated truncated cone with noise: $P_{\%} = 60\%$ $\sigma_P = 0.7$.

Previous work allowed us to propose an algorithm based on a time-scale analysis [8], [9]. Since cross sections are created step by step, the purpose is to use this sonar moving to detect and to follow both target reflections. With reference to this temporal evolution, the processing can be seen as a "tracking" algorithm of target echoes.

This processing uses a wavelet packet decomposition to analyze adjacent shots (i.e., successive columns of cross-sectional images) and to keep only correlated information associated with target reflections. This "tracking" algorithm has many advantages as follows.

- It reduces noise on image.
- It enhances useful information.

Followed by an edge linking algorithm, this method provides a realistic shape visualization.

Fig. 8 gives an example of a processed image where the truncated cone can easily be recognized (original image is shown in Fig. 5). Then, so as to close the shape, either we join the extremities of both pixel chains (which correspond to target echoes) according to a distance minimization criteria (to close the 2-D edge), or we apply a Delaunay triangulation [10] (to close the 3-D shape). A result of the complete "tracking" algorithm has been given in Fig. 7

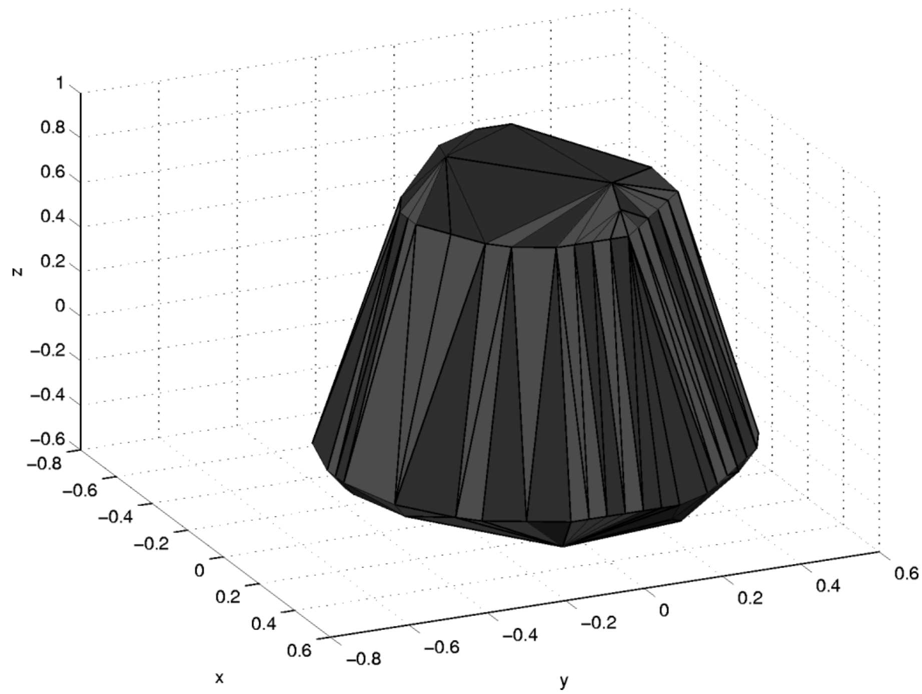
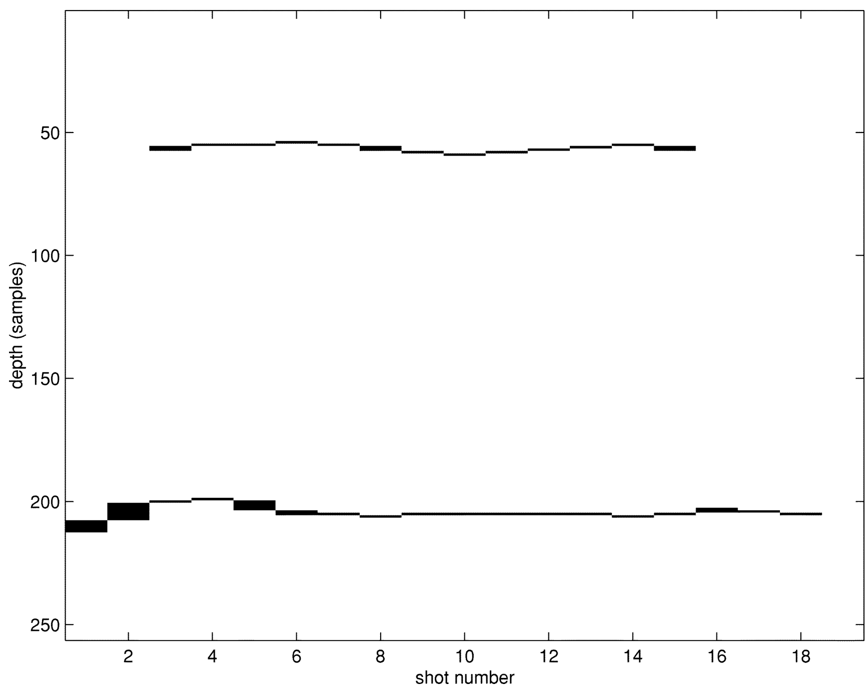


Fig. 7. Simulated truncated cone with noise: $P_{\%} = 60\%$ $\sigma_P = 0.7$.



8. Results of the tracking processing.

The next stage consists now in computing relevant parameters to characterize this shape.

B. Relevant Parameters

Object recognition is an important topic in computer vision and automatic scene analysis but 3-D object identification remains a difficult task because of the data amount to process. So, the commonly used idea is to reduce it into relevant parameters that can characterize the 3-D target shape. And naturally, selecting key parameters constitutes a critical element in designing a robust active classification.

We notice two kinds of methods for this problem.

- Methods using 2-D projections of 3-D objects (typically their silhouette).
- Methods using the full 3-D information to compare it with created models.

Since they are computationally less intensive, the first type of methods have been widely preferred [11], [12]. Indeed, a great deal of work has been dedicated to 2-D pattern recognition, either using structural or statistical approaches.

In the structural approach, features such as holes and corners are detected on the 2-D image of the object and linked to provide its structural description [13]. This structural description is generally a graph whose nodes represent the spatial organization of the features. Recognition is then achieved by graph matching techniques using a database models.

In the statistical approach, a feature vector of fixed length is computed from the image. The components of this feature vector can be, for example, Fourier descriptors [14], moments of Hu [15], or Zernicke moments [16]. Recognition is then performed by a classifier (neural networks, ...).

Few studies have dealt with the 3-D pattern recognition by using full 3-D information: they are specially devoted to the computation of 3-D moments [17], [18] but most of these approaches do not provide useful feature vectors for discriminating 3-D shapes (in fact these moments are often difficult to derive). More recently, an original approach was developed by G. Burel and H. Henocq [19], for deriving 3-D invariants in a simple way. This method was proved to be a generalization of the well-known 3-D moment-invariant technique.

In the following sections, we adapt this technique to the 3-D underwater target recognition. In the same way, we develop another approach based on an improvement of Fourier descriptors. This method combines several 2-D sections of the same object by using a fusion technique. These feature sets will be compared in the last section of this paper.

1) Fourier Descriptors:

a) Mathematical Review: Fourier Descriptors [20] have been already used so as to describe the 2-D silhouette of an object: it consists in a decomposition of an edge into a Fourier series. A **closed** edge can be considered as a list of **uniformly spaced** pixels with complex coordinates or as a discrete complex signal U_m (the real part corresponds to pixel row and the imaginary part corresponds to pixel column). The Fourier coefficients of this signal are given by (1) where N is the number of samples of this signal. These terms are called Fourier descriptors. The edge can be rebuilt from Fourier descriptors using inverse Fourier transform

$$C_n = \frac{1}{N} \sum_{m=0}^{N-1} U_m e^{-j2\pi\,(nm/N)}. \qquad (1)$$

The main advantage of this approach is the possibility of making these descriptors rotational, translational, and dilational invariant. Therefore, the same kind of object will be nearly described by the same descriptors whatever its size, its orientation or its position in the image are. Table I explains how each transformation influences the edge and modifies the Fourier descriptors.

U_m and U'_m correspond respectively to the original edge and its transformation (translation, rotation, dilation, and shift in starting point). C_n and C'_n correspond respectively to Fourier descriptors of U_m and U'_m.

These invariances are verified if the following **normalizations** are realized (derived properties are detailed in [14] or in [21]).

- The first coefficient of the series is ignored so that translation invariance could be checked.
- All coefficients are divided by C'_1 to realize rotation and dilation invariance.
- All coefficients are multiplied by $e^{j(1-n)\phi'_2}$, where ϕ'_2 is the second coefficient phase: a change of the starting point will no more influence the descriptors.

TABLE I
EDGE GEOMETRIC TRANSFORMATION CONSEQUENCES

Transformation	Influence on the edge	Influence on Fourier descriptors
Translation	$U'_m = U_m + T$	$C'_0 + T$
Rotation (ϕ)	$U'_m = U_m e^{j\phi}$	$C'_n = C_n e^{j\phi}$
Dilation	$U'_m = \rho U_m$	$C'_n = \rho C_n$
Shift in starting point	$U'_m = U_{(m+k)\ mod\ N}$	$C'_n = C_n e^{2\pi j \frac{kn}{N}}$

To achieve a robust classification of 2-D shapes, feature vectors will be composed by these **normalized Fourier descriptors**.

b) Information Compression: The main problem to achieve an optimal classification is to keep only descriptors with relevant information. We propose a 2-D pattern recognition based on the following principle: Fourier descriptors associated with high frequencies correspond to edge details. These details are not useful to recognize the shape and could even disturb their classification: as we can see on Fig. 9, only the first 20 descriptors are enough for a correct recognition.

Using simulated targets, the quantity of erroneous pixels on reconstructed edges provide us an error rate. We can verify on Fig. 10 that this rate decreases quickly when the number of descriptors used for the reconstruction increases. It becomes really low when this number is higher than 30. The idea is therefore to propose a feature vector composed only by Fourier descriptors associated with low frequencies: so as to keep only this useful information, we will select parameters thanks to a principal component analysis (see Section IV).

However, this 2-D pattern recognition has some limits. For example, using only one cross section may be not enough to distinguish cylinder and sphere. In order to introduce 3-D information, we propose to use a pair of orthogonal sections from the same object to characterize its shape (Fig. 11). In that way, a shape will be represented by two series of Fourier descriptors. Classification results improvement will be shown in Section IV.

2) Decision Fusion: To avoid overdimensional feature vectors, we can not apply the previous idea on many cross sections from the same object. Indeed, classifier complexity and performances are directly influenced by the feature vector size. However, results are expected to be improved if we use more information about the 3-D shape. In order to exploit as many 3-D information as available, we propose to combine individual results obtained from several section pairs so as to provide a more accurate classification decision.

This fusion system must be able to integrate knowledge from various information sources. A variety of methods for combining evidence produced by multiple information sources has been studied by different researchers [22]–[25]. Techniques such as voting methods, Dempster–Shafer theory, fuzzy integral

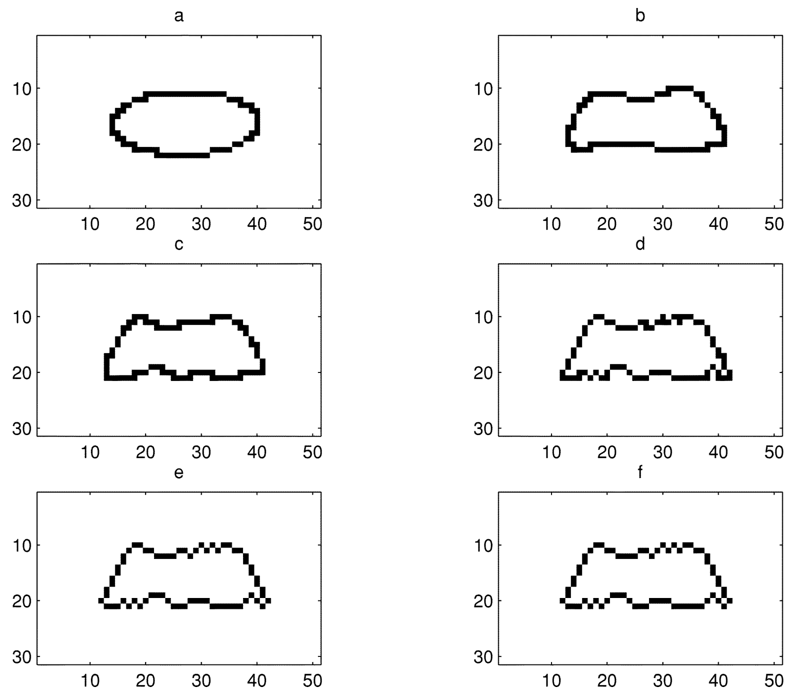


Fig. 9. Reconstruction of a simulated truncated cone with the first 2, 10, 20, 64, 128 coefficients, and original (f).

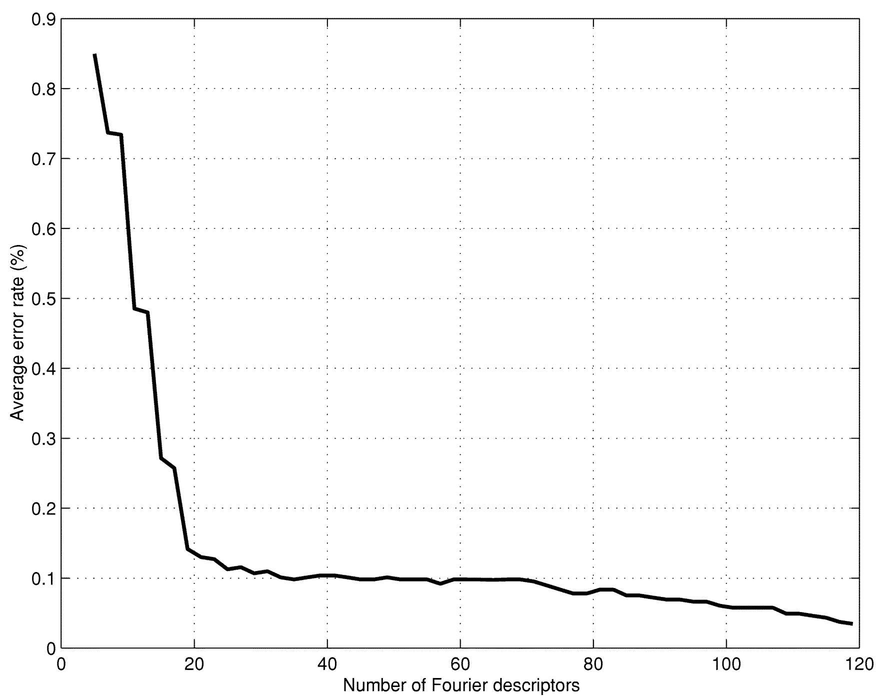


Fig. 10. Evolution of the average error rate defined on reconstructed 2-D edges (computed on 1000 cross sections per class) versus the number of Fourier descriptors used for the reconstruction.

were successfully employed in various recognition systems. All these techniques can deal with the conflicts existing in evidence and can generate better consistent results. However, the fuzzy logic method, based on a fuzzy integral, has been chosen because it takes into account not only the importance of each individual source, but also the importance of their combinations.

In this section, the Sugeno fuzzy integral [22] is employed to implement evidence combination of multiple pairs of Fourier descriptors. The reliability of each pair is treated as fuzzy measure in fuzzy integral (further details about fuzzy integral and its applications can be found in [26] and [27]).

The basic idea of the proposed algorithm is to develop n independently classifiers trained with Fourier descriptors from different pairs of cross-sectional images. Each classifier provides a decision about the target class and the fusion method takes into account all these decisions and their combination to decide finally about the more realistic target class.

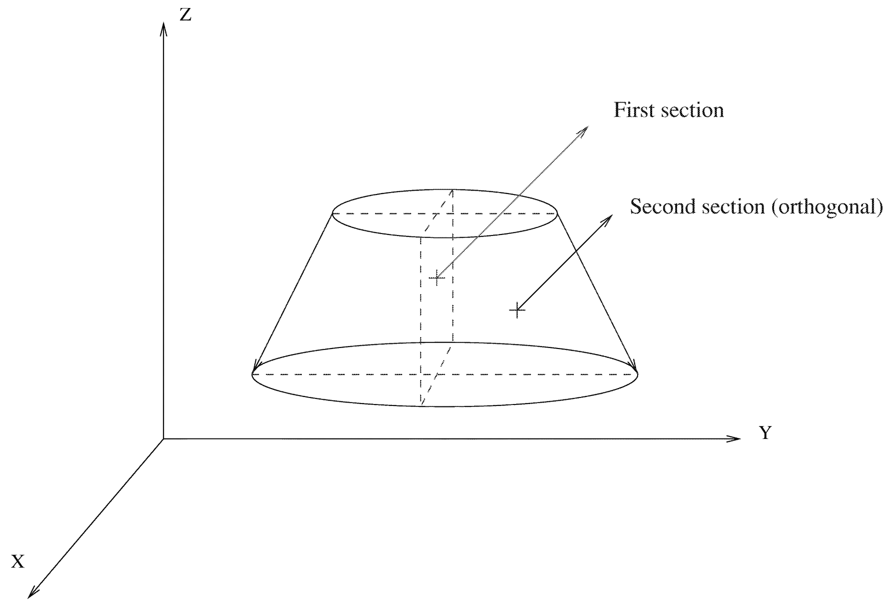


Fig. 11. Example of orthogonal cross sections used to characterize 3-D objects.

Fig. 12 describes the fusion method: χ_k is the fuzzy integral associated with the kth class. It is defined by

$$\chi_k = \int h_k(q) \circ g_k(.) = \max_{i=1,\, n} \left[\min\left\{h_k(q_i),\, g_k(A_i)\right\}\right] \quad (2)$$

where q_i denotes the ith classifier. $h_k(q_i)$ denotes the real ith classifier decision about the object belonging to the kth class (it measures the belonging probability toward the kth class according to the ith classifier).

g_k denotes a fuzzy measure associated with the kth and A_i is a set of classifiers $\{q_1,\, q_2,\, \ldots,\, q_i\}$. According to fuzzy measure properties (3), $g_k(A_i)$ denotes the significance of the classifier set A_i for the final decision. In practice, each $g_k(q_i)$ is defined as the expected correct classification rate of the classifier q_i for the kth class

$$g(A_i) = g(q_i) + g(A_{i-1}) + \lambda.g(q_i).q(A_{i-1}), \qquad \lambda > -1. \quad (3)$$

Finally, each fuzzy integral χ_k can be interpreted as finding the maximal level of agreement between objective evidence and expectation. Then, the highest χ_k provides the class of the analyzed target.

Fig. 12 describes the complete algorithm when only three decisions are merged. The classifier is detailed in Section IV.

This approach provides an improved classification taking into account many cross-section pairs from the same target. It is an innovative way of classifying 3-D objects by separating 3-D data into many 2-D views. Better classification rates are expected but there are some limits: Fourier descriptors can only be applied when there are two reflections (upper and lower surfaces) inside the cross-sectional image. Sometimes, only one reflection can be detected for targets like the truncated cone. So the full 3-D information is not necessarily used and performances of this method must be compared with other key parameter sets taking into account the full 3-D target shape.

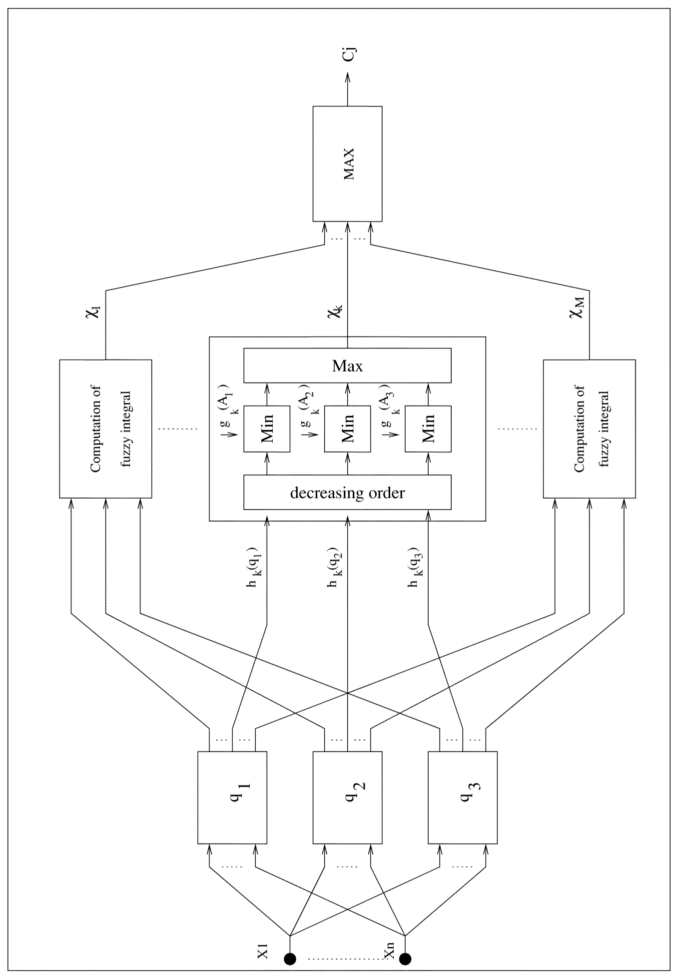


Fig. 12. Fusion algorithm: $g_k(A_i)$ is the fuzzy measure of the group of classifiers $A_i = \{q_1,\, \ldots,\, q_i\}$ toward the class k (it quantifies the importance of this group for the final decision. In fact, these are associated with the classification performances of A_i).

3) 3-D Invariant: In this section, we use no more cross sections individually but all the available cross sections to have access to a **3-D** (binary) representation of the object.

a) Spherical Harmonic Basis: We propose now to create feature vectors with a new set of relevant parameters developed in 1995 [19]. As explained before, a good feature vector should be invariant. Invariances toward dilation and translation are quite easy to obtain from binary 3-D data. The origin of coordinates is placed at the gravity center of the object and we realized a volume normalization by choosing as unit of length the distance between this origin and the farthest point of the target.

However, rotational invariance remains much more difficult to achieve. Of course, for some objects, the rotation may be fixed by diagonalizing the inertia matrix. However, many real objects have similar inertia values on at least two principal axes. Hence, the determination of orientation using this method does not provide results accurate enough to discriminate between similar shapes.

So, we prefer another way of computing rotational invariant parameters. The main objective is to find a suitable representation space where the rotational operator can be expressed easily: a parallel with quantum mechanics is realized [28].

Firsly, it is interesting to notice that the 3-D object can be described by a function $\Psi(x, y, z)$ of $\mathbb{R}^3$. This function is binary (1 at the surface of the object, 0 elsewhere). This function is transformed into $\Psi'(x', y', z')$ using the rotational operator R. A strong link exists between R and the angular momentum (L_x, L_y, L_z) (we use here the same notation as usually found in quantum physics books and in [19] where the 3-D invariant principle is more detailed)

$$R = e^{-i\alpha(u_x L_x + u_y L_y + u_z L_z)} \tag{4}$$

where $[u_x, u_y, u_z]$ is the axis of rotation and α the angle of rotation.

In quantum mechanics, observations concerning angular momentum are defined by the action of two operators: $\mathbf{L^2}$ and L_z. They are commuting Hermitian operators. Hence, there exists an orthonormal basis composed by their common eigenvectors.

These eigenvectors called *Spherical Harmonics* Y_{lm}, are computed using recurrent equations [29] and are defined for $l = 0, \ldots, L_{\max}$ and $m = -l, \ldots, +l$. $L_{\max}$ corresponds to the description level of the shape (increasing the value of $L_{\max}$ provides a more precise description, but also increases computation time). The decomposition of Ψ onto the basis $\{Y_{lm}\}$ is obtained by scalar products of the shape with the vectors of the basis

$$\Psi = \sum_{l,m} C_l^m Y_{lm}. \tag{5}$$

In this base, the rotational operator can be expressed easily using a Taylor series decomposition

$$\begin{aligned} R &= e^{-i\alpha(u_- L_+ + u_+ L_- + u_z L_z)} \\ &= 1 - i\alpha(u_- L_+ + u_+ L_- + u_z L_z) \\ &\quad - \frac{\alpha^2}{2!}(u_- L_+ + u_+ L_- + u_z L_z)^2 - \cdots \end{aligned} \tag{6}$$

where
- $L_+ = L_x + iL_y$;
- $L_- = L_x - iL_y$;
- $u_+ = (u_x + iu_y)/2$;
- $u_- = (u_x - iu_y)/2$.

If the rotational operator is applied on a spherical harmonic, we can deduce that

$$\begin{aligned} R(Y_{lm}) &= Y_{lm} - i\alpha(u_- L_+ + u_+ L_- + u_z L_z)(Y_{lm}) \\ &\quad - \frac{\alpha^2}{2!}(u_- L_+ + u_+ L_- + u_z L_z)^2 (Y_{lm}) - \cdots \end{aligned} \tag{7}$$

We see in (7) that each Y_{lm} is transformed by rotation into a linear combination of $Y_{l,-l}, Y_{-l,-l+1}, \ldots, Y_{l,l}$. Therefore, the space of differentiable functions can be decomposed as a direct sum of globally rotational invariant orthogonal subspaces: the rotational operator can take the simple form of a block matrix [19].

b) Computation of Parameters: 3-D Invariant computation is based on tensor theory whose complete description may be found in [30], [31]. Let us recall the definition of a tensor.

Let us consider a vector space $\mathcal{V}$ and a basis $\{e_i\}$. A new basis $\{\tilde{e}_i\}$ is related to the original basis by (the Einstein summation convention is assumed)

$$\tilde{e}_i = \alpha_i^j e_j \tag{8}$$

or

$$e_i = \beta_i^j \tilde{e}_j \tag{9}$$

where α_i^j denotes a linear transformation and β_i^j its inverse.

Then, any form $T_{l_1,\ldots,l_p}^{m_1,\ldots,m_q}$ is said to be a tensor of covariant rank p and contravariant q if it transforms itself according to the following equation:

$$\tilde{T}_{k_1,\ldots,k_p}^{n_1,\ldots,n_q} = \alpha_{k_1}^{l_1} \cdots \alpha_{k_p}^{l_p} \beta_{m_1}^{n_1} \cdots \beta_{m_q}^{n_q} \qquad T_{l_1,\ldots,l_p}^{m_1,\ldots,m_q} \tag{10}$$

where $m_1, \ldots, m_q$ and $n_1, \ldots, n_q$ denote, respectively, the contravariant components of the original tensor and of its transform; $l_1, \ldots, l_p$ and $k_1, \ldots, k_p$ denote, respectively, the covariant components of the original tensor and its transform.

Thanks to the tensor contraction theorem, we know a tensor of order 0 (i.e., an invariant) can be easily computed by multiplying a contravariant tensor of order 1 with a covariant tensor of order 1.

The shape Ψ can be precisely considered as a tensor of order 1 which contravariant and covariant components are respectively C_l^m and $(C_l^m)^*$. So, with this theorem, we can derive a first set of invariant parameters called N

$$N(l) = \sum_{m=-l}^{l} C_l^m (C_l^m)^*. \tag{11}$$

With the second theorem below, an infinity of invariants can be build. We propose two other series of parameters called P

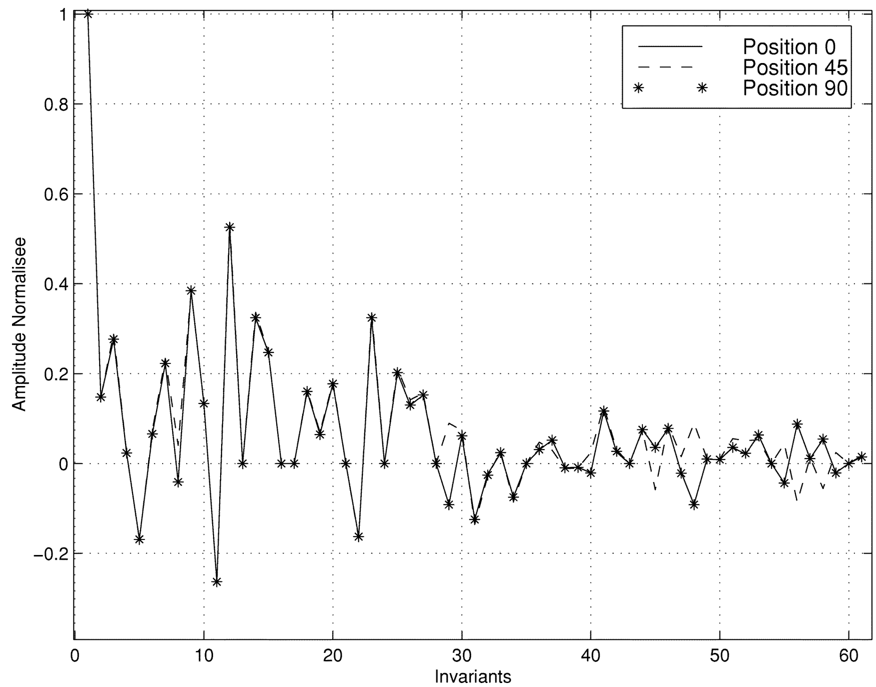


Fig. 13. Example of 3-D invariants computed from a simulated truncated cone ($L_{\max} = 3$).

and Q in (12) and (13)

> **Theorem :** If T_l^m is a tensor of the subspace ϵ_l, then $\Pi(l_1, l_2)_l^m = \sum_{m_1, m_2} \langle l_1\, m_1\, l_2\, m_2 | l\, m\rangle$ $T_{l_1}^{m_1} T_{l_2}^{m_2}$ is also a tensor of $\epsilon_l \langle l_1\, m_1\, l_2\, m_2 | l\, m\rangle$ are Clebsh–Gordan coefficients [29]

$$P(l, l_1, l_2) = \sum_{m=-l}^{l} \Pi(l_1, l_2)_l^m (c_l^m)^* \tag{12}$$

$$Q(l, l_1, l_2, l_3, l_4) = \sum_{m=-l}^{l} \Pi(l_1, l_2)_l^m (\Pi(l_3, l_4)_l^m)^*. \tag{13}$$

Many symmetry relations can be established and used to restrict index values. For instance, $N(l)$ is computed for $l > 0$, $P(l, l_1, l_2)$ is computed for the following indexes:

$$\begin{aligned} &1 \le l_2 \le l_1 \le l \le L_{\max} \\ &l_1 - l_2 \le l \le l_1 + l_2 \\ &l_2 \ne l_1 \qquad \text{or } l \text{ even} \\ &l_1 \ne l \qquad \text{or } l_2 \text{ even} \end{aligned}$$

and $Q(l, l_1, l_2, l_3, l_4)$ is computed for the following indexes:

$$\begin{aligned} &0 \le l_2 \le l_1 \le L_{\max} \\ &0 \le l_3 \le l_4 \le l_1 \\ &1 \le l \le L_{\max} \\ &l_1 - l_2 \le l \le l_1 + l_2 \\ &l_3 - l_4 \le l \le l_4 + l_4 \\ &l_2 \ne l_1 \qquad \text{or } l \text{ even} \\ &l_4 \ne l_3 \qquad \text{or } l \text{ even.} \end{aligned}$$

It is worth noting that the higher the value of $L_{\max}$ is, the more 3-D parameters we compute. Indeed, when the 3-D shape is very complex, details can not be characterized with only few parameters and we need to use a high level $L_{\max}$ to represent the shape precisely.

Fig. 13 presents these invariants for three different positions (0°, 45°, 90°) of a truncated cone. We can verify here the invariance of these parameters. The residual difference between the curves can be explained by the resampling of the 3-D object.

The next section will compare the performances of these sets of parameters using simulated and real data collected on three kinds of targets.

TABLE II
AVERAGE CLASSIFICATION RATES OBTAINED ON SIMULATED DATA ($\boldsymbol{\sigma_P} = (\mathbf{1}, \mathbf{P_{\%}} = \mathbf{90\%})$) WITH TWO CROSS SECTIONS PER TARGET

Real class \ *Recognized class*	*Truncated cone*	*Cylinder*	*Sphere*
Truncated cone	**90%**	5%	5%
Cylinder	5%	**90%**	5%
Sphere	3%	5%	**92%**

TABLE III
AVERAGE CLASSIFICATION RATES OBTAINED ON SIMULATED DATA ($\boldsymbol{\sigma_P} = \mathbf{1}, \mathbf{P_{\%}} = \mathbf{90\%}$) WITH ONLY ONE CROSS SECTION PER TARGET

Real class \ *Recognized class*	*Truncated cone*	*Cylinder*	*Sphere*
Truncated cone	**80%**	15%	5%
Cylinder	5	**75%**	20
Sphere	5%	25%	**70%**

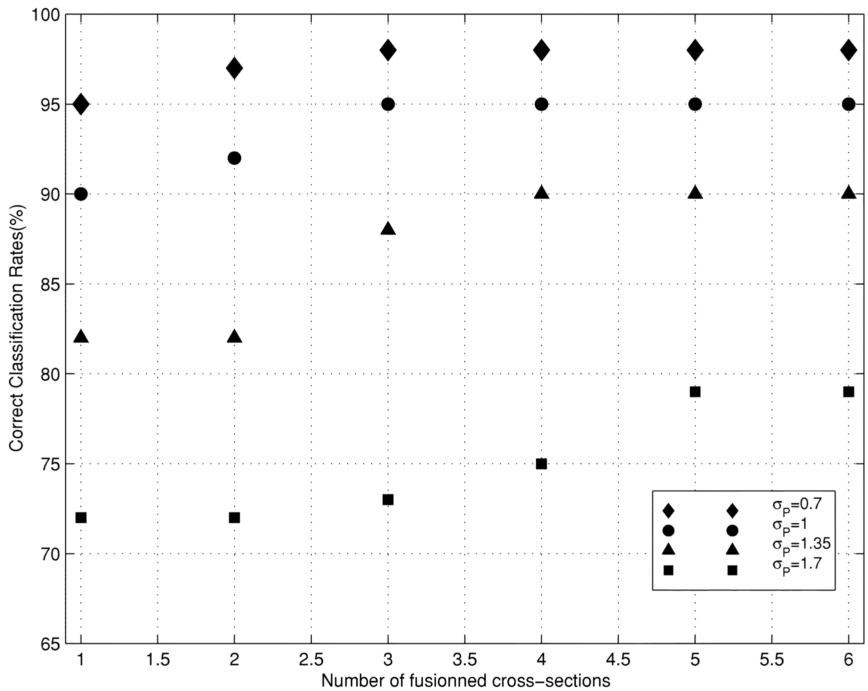


Fig. 14. Evolution of truncated cone correct classification rates versus number of fused sections.

TABLE IV
3-D INVARIANT APPROACH: CORRECT CLASSIFICATION RATES VERSUS. RESOLUTION LEVEL $L_{\max}$

Classes \ Resolution Level L_{max}	$L_{max}=2$	$\mathbf{L_{max}=3}$	$L_{max}=4$	$L_{max}=5$	$L_{max}=6$
Truncated cone	94%	**95%**	95%	95%	92%
Cylinder	93%	**93%**	93%	92%	91%
Sphere	93%	**94%**	94%	93%	92%

IV. Classification Results

First series of tests are realized on simulated data to compare proposed algorithms, then results are given from real data.

A. The Classifier

Several classifiers could be used in our application. However, we have chosen to realize a supervised classification: that means we know *a priori* the belonging of each object of data bases (real and simulated) to the corresponding class. So we prefer a classification based on neural networks or nonparametric methods.

As we do not know whether our three classes (sphere, cylinder and truncated cone) are properly separated, we decided to use the classification algorithm called Q^* [32]. This algorithm realizes a supervised self-organizing learning of representative prototypes from a data set. It has the same functionality as the learning vector quantization, a set of multidimensional samples of valued features will be compressed to a reduced set of prototypes; these are dynamically created and modified during the iterative training process. The data set is then auto-classified by this set of prototypes: in case of a classification error, an additional prototype is added to the actual set of prototypes.

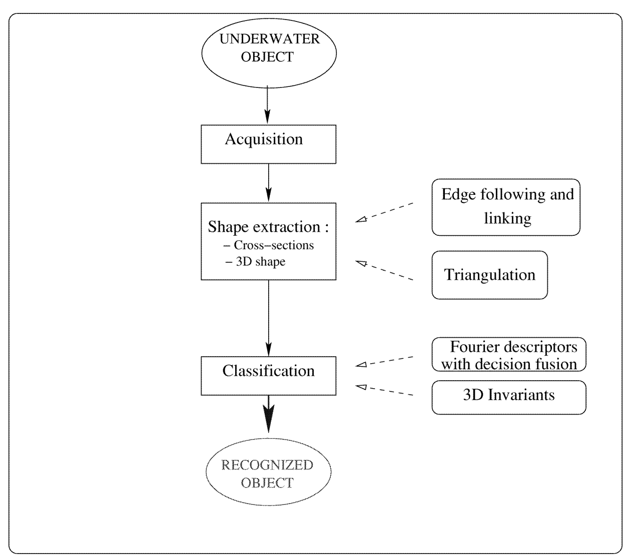


Fig. 15. Complete classification algorithm for underwater targets.

Moreover, Q^* has several properties.

- It is nonparametric which is advantageous if the number of available training examples is small.

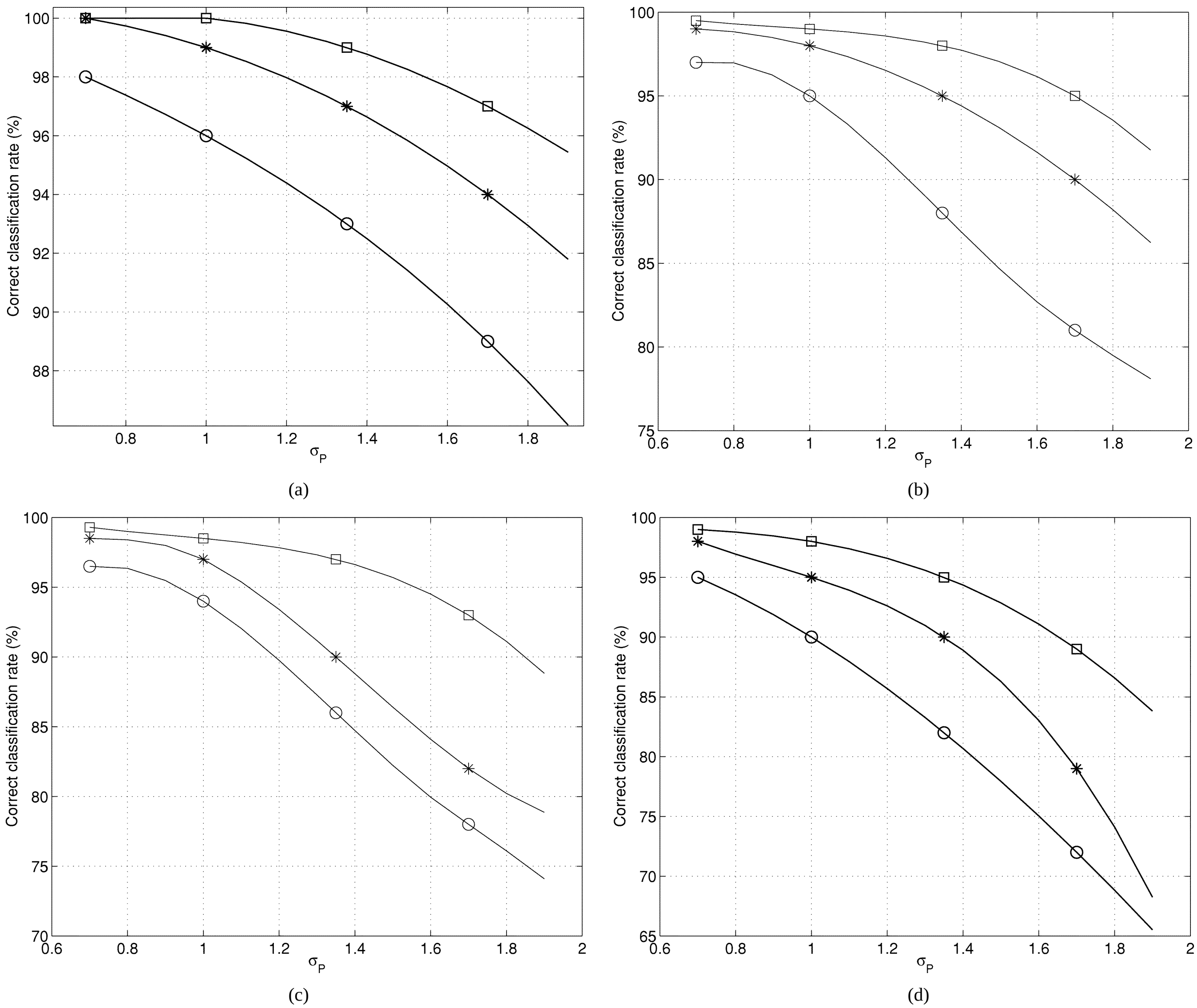


Fig. 16. **Truncated cone class**: Correct classification rate evolution versus noise level ($\boldsymbol{\sigma}_{\mathbf{P}}$, $\mathbf{P}_{\%}$). (a) $\mathbf{P}_{\%} = 60\%$. (b) $\mathbf{P}_{\%} = 70\%$. (c)$\mathbf{P}_{\%} = 80\%$. (d) $\mathbf{P}_{\%} = 90\%$.

- It handles multimodal class distributions.
- It is self-organizing: no parameters have to be specified by the user unlike many methods.

Finally, during the classification test, we compute distances between each prototype and the test vector, and this vector is classified into the nearest prototype class. In conclusion, this is an efficient classification method especially when the feature vector class distributions are unknown.

B. Feature Selection

As we have seen in previous section, each feature vector is composed by numerous relevant parameters. For instance, with $L_{\max} = 3$, we can build 61 3-D invariants. We propose to apply a principal component analysis (PCA) to reduce this dimension and simplify the classifier structure [33]. The purpose is to keep only useful information and another consequence of this feature selection may be a suppression of disturbing information and the improvement of the classification results.

The dimension of the resulting vectors is chosen from a proportion of the covariance matrix global inertia.

This selection will be applied to Fourier descriptors as well as 3-D invariant feature vector. The following paragraph discusses this.

C. Discussion

First tests have been made to prove the second cross section introduction utility to classify the objects using Fourier descriptors. Results obtained with simulated data ($\boldsymbol{\sigma}_{\mathbf{P}} = \mathbf{1}$, $\mathbf{P}_{\%} = \mathbf{90}\%$) are summarized in the following tables Tables II and III. They can be read as following: on the first line of Table II, we can see that **90%** of truncated cones are really recognized as targets of the right class, whereas 5% are recognized as cylinders and 5% as spheres. We can verify that classification rates are improved using two orthogonal cross sections by object; the same kind of results have been noticed with all the other noise levels.

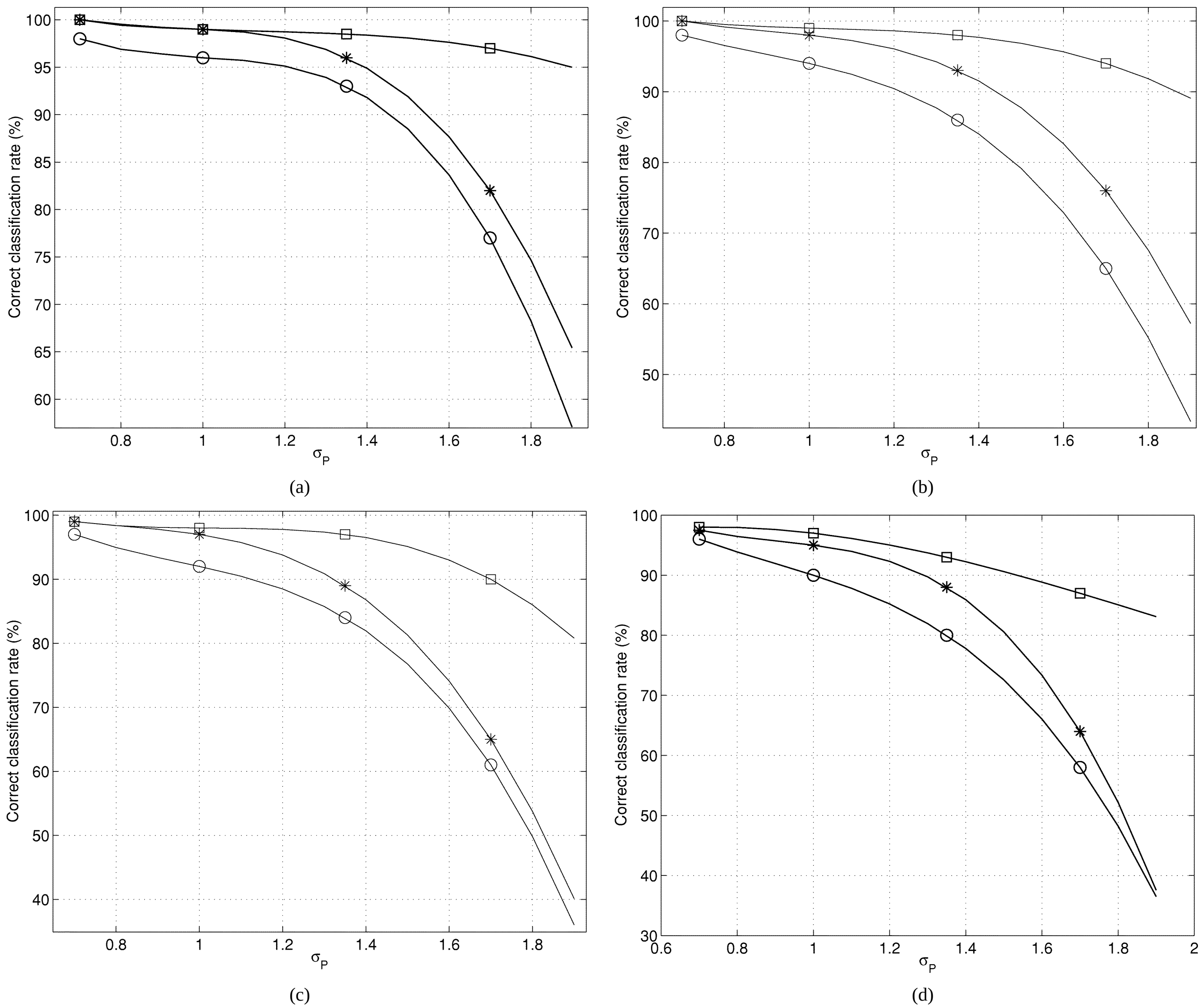


Fig. 17. **Cylinder class**: Correct classification rate evolution versus noise level ($\boldsymbol{\sigma_P}$, $\mathbf{P}_{\%}$). (a) $\mathbf{P}_{\%} = 60\%$. (b) $\mathbf{P}_{\%} = 70\%$. (c)$\mathbf{P}_{\%} = 80\%$. (d) $\mathbf{P}_{\%} = 90\%$.

In the same way, we introduce decision fusion to use more information about the 3-D shape. Fig. 14 presents classification rates versus the number of fused cross sections for different values of $\boldsymbol{\sigma_P}$ in the case of truncated cones.

The first remark is about the classification improvement in comparison with the method without fusion (only one pair of sections is analyzed by object). The fusion succeeds in compensating the false decisions obtained using noisy sections. As we can see on this example, the improvement can reach nearly 10%, and the same conclusions have been made for the other target classes. Finally, we can conclude that we should merge as many sections from the same object as possible to achieve the best classification rates. However the classification rate does not increase anymore after fusion 4 or 5 views. This limit proves that new views do not bring more useful information about the target shape. 2-D views do not succeed in describing the whole 3-D shape, so it really justifies the use of 3-D invariants.

The last important point concerns the resolution level choice for the 3-D invariant computation. This level $L_{\max}$ is related to the 3-D shape complexity degree and the shape noise level. Table IV sums up the classification results obtained for different values of $L_{\max}$. In order to take the noise influence into account, tests have been realized on severely noisy data ($\sigma_P = 1.35$ and $P_{\%} = 90\%$). We can notice that performances are nearly constant for $L_{\max}$ varying from 2 to 6, and best classification rates are obtained for $L_{\max} = 3$. Our shape simplicity can explain this remark: we do not need a high representation level $L_{\max}$ for these objects. And too many details (that means high value of $L_{\max}$) can even degrade the classification rates.

We can therefore deduce another consequence. Even if this low value is enough to describe our targets in the spherical harmonic space, a higher value of $L_{\max}$ should be probably required to characterize objects with a more complicated shape.

D. Comparative Results

In this last paragraph, we propose to compare the different algorithms with simulated data; then we present the classification rates obtained on real data using Fourier descriptor approach.

1) Simulated Data: The data base is composed by 1000 feature vectors for each class. The complet identification algorithm

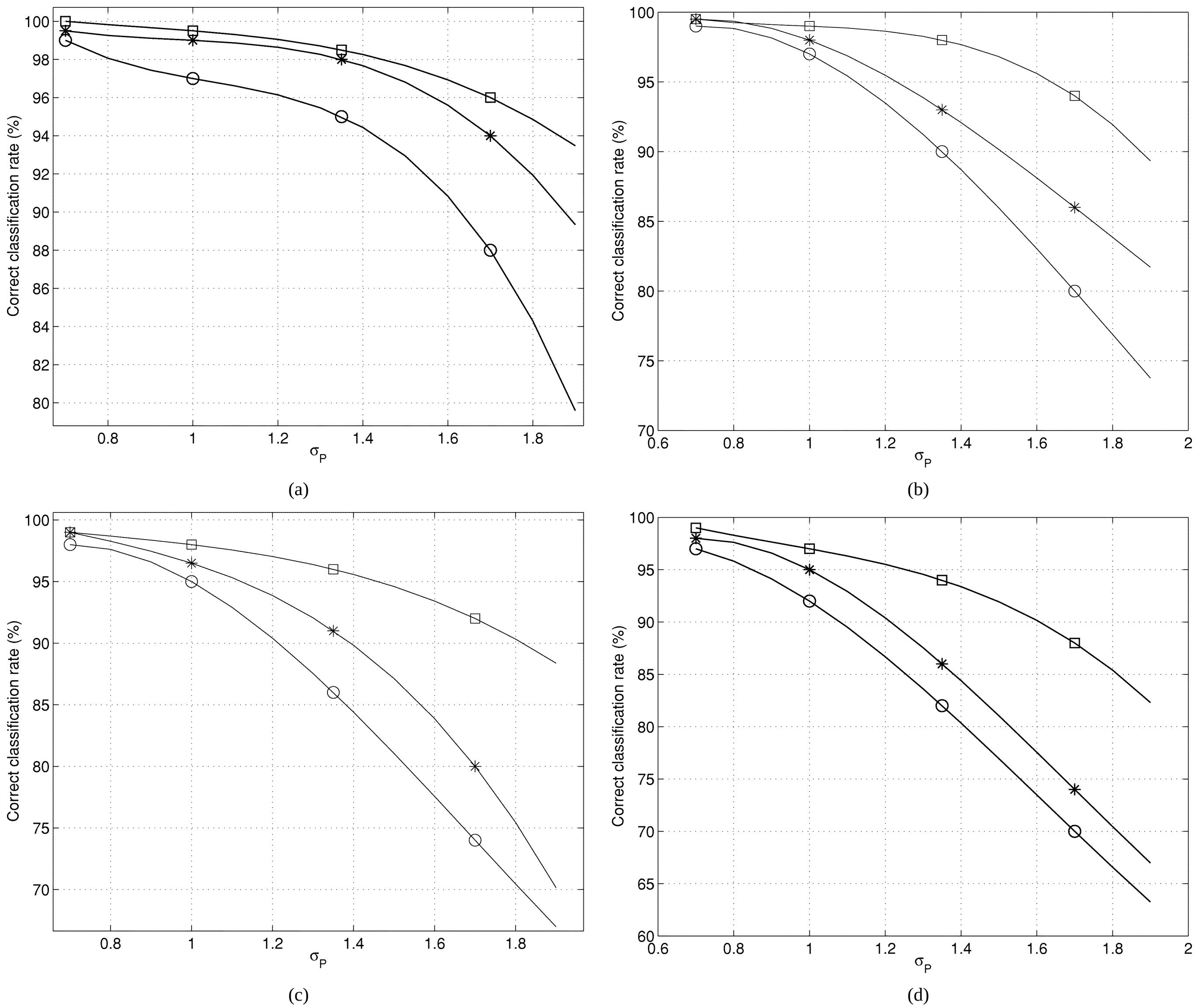


Fig. 18. **Sphere class**: Correct classification rate evolution versus noise level ($\boldsymbol{\sigma_P}$, $\mathbf{P_{\%}}$). (a) $\mathbf{P_{\%}} = 60\%$. (b) $\mathbf{P_{\%}} = 70\%$. (c)$\mathbf{P_{\%}} = 80\%$. (d) $\mathbf{P_{\%}} = 90\%$.

(Fig. 15) is applied and the correct classification rates computed for each target class are presented in Figs. 16–18.

We can mainly remark that these good performances prove the relevance of Fourier descriptors and 3-D invariants to characterize these underwater objects: indeed, at low noise level, classification rates are over 90%. Since simulations take into account objects with various sizes and orientations, the quality of these results proves the parameter robustness toward geometrical transformation invariance.

The next discussion concerns the performance evolution when the noise increases. Of course, the three methods are equivalent when the noise is low but main differences appear when σ_P and $P_{\%}$ increase. Indeed, these figures reveal that Fourier descriptors are less robust toward noise than 3-D invariants. Even if the fusion succeeds in improving the classification rates and noise robustness, 3-D invariants remain the best parameters to describe the targets. In this way, we prove that the more important the information quantity about the shape is, the better the classification is. 3-D Invariant results are over 80% whatever the noise level is. However, this 3-D approach needs a computational time two or three times more important than the 2-D approach.

2) Real Data: Because of the poor quantity of real data, the Fourier descriptor approach has only been able to be tested. The data base is composed by 200 feature vectors for each object class. Classification results are detailed in Table V. With this method, we reach again good performances since rates are over 92% for all the classes and remarks of the previous paragraph remain valid. It proves not only the relevance of Fourier descriptors in a real environment but the validity of our simulations too. To convince us, the last table (Table VI) presents the classification results obtained when the classifier is trained with simulated data ($\sigma_P = 1$, $P_{\%} = 90\%$) and tested on real data. Of course, the rates are lower than those of the previous table but they are still good. Moreover, with a more precise simulation model, we can plan a classifier trained only with simulated data: it could be a useful approach to classify targets if we lacked real data.

Finally, we can propose a complete recognition system from data acquisition to target classification (Fig. 15). In this system,

TABLE V
AVERAGE CLASSIFICATION RATES OBTAINED ON REAL DATA

Real class \ *Recognized class*	*Truncated cone*	*Cylinder*	*Sphere*
Truncated cone	**95%**	3%	2%
Cylinder	2	**97%**	1%
Sphere	5%	3%	**92%**

TABLE VI
AVERAGE CLASSIFICATION RATES OBTAINED ON REAL DATA (TRAINING REALIZED ON SIMULATED DATA)

Real class \ *Recognized class*	*Truncated cone*	*Cylinder*	*Sphere*
Truncated cone	**80%**	10%	10%
Cylinder	10	**90%**	0%
Sphere	5%	20%	**75%**

users can choose the kind of parameters: the 3-D invariants approach should be preferred when users need high classification performances but if computation time is really a key parameter, the

V. CONCLUSION

In this paper, we have proposed an original way to classify buried objects by using acoustic data collected from a parametric sonar. The main innovation appears in the experimental approach that gives us access to 3-D view of underwater targets. Another key point is about the system generality: this experimental approach is intended to targets lying on the seafloor as well as **buried targets**.

Another originality of our work concerns the robust 3-D pattern recognition using either 2-D views or the full 3-D shape. In that way, we introduced two kinds of invariant parameters to describe 3-D target shape: Fourier descriptors and 3-D invariants. Tests realized on simulated data proved these parameter relevances. Moreover, we have shown that taking into account more information about the target is a way of improving classification task.

A real innovative 3-D pattern recognition using decision fusion is proposed. Indeed, the combination of Fourier descriptors and decision fusion improves classification rates. Finally, we describe another kind of key parameters (3-D invariants) that use the full 3-D information and give more precise decision about the target class. Moreover, these 3-D parameters are proved to be more robust toward noise.

So we propose a complete classification system where users can choose between two kinds of parameters depending on computation time and allowed error rate.

Problems for further research include new trials to collect data that could help us to compare methods in a real environment. In this work, we have tried to propose an automatic classification approach but the choice of some parameters (for instance the resolution level $L_{\max}$) needs to be validated. Moreover, in a near future, this parametric sonar based acquisition system should be compared with other acoustic systems such as synthetic aperture sonar.

ACKNOWLEDGMENT

The authors would like to thank the Groupe d'Etudes Sous-Marines de l'Atlantique (GESMA) and Guigne Int'l for their participation during acoustic acquisitions.

REFERENCES

[1] G. Nicq and M. Brussieux, "A time-frequency method for classifying objects at low frequencies," in *Proc. OCEANS'98* , vol. 1, 1998, pp. 148–152.
[2] R. Guillermin, P. Lasaygues, J. Sessarego, and A. Wirgin, "Characterization of buried objects by a discretized domain integral equation inversion method using born approximation (ddieba)," in *Proc. 5th Eur. Conf. Underwater Acoustics*, vol. 2, July 2000, pp. 863–868.
[3] M. Pinto, A. Bellettini, S. Fioravanti, S. Chapman, D. Bugler, Y. Perrot, and A. Hetet, "Experimental investigations into high resolution sonar systems," in *Proc. MTS/IEEE OCEANS'99*, vol. 2, 1999, pp. 916–922.
[4] A. Trucco, "A pattern recognition strategy for the detection of buried objects," in *Proc. 5th Eur. Conf. Underwater Acoustics*, vol. 2, July 2000, pp. 1097–1102.
[5] M. Granara, A. Pescetto, F. Repetto, G. Tacconi, and A. Trucco, "Statistical and neural techniques to buried target detection and classification," in *Proc. IEEE/OES OCEANS'98 Conf.*, vol. 3, 1998, pp. 1269–1273.
[6] O. Bergem and N. G. Pace, "Installation and calibration of a parametric array for shallow water backscatter measurements," in *Proc. OCEANS'96*, vol. 2, 1996, pp. 773–777.
[7] P. Cervenka, "Contribution théorique et expérimentale à l'étude de l'émission et de la réception paramétriques," Ph.D. dissertation, Université Pierre et Marie Curie, Paris VI, France, 1988.
[8] A. Quinquis and D. Boulinguez, "Multipath channel identification with wavelet packets," *IEEE J. Oceanic Eng.*, vol. 22, pp. 342–346, Apr. 1997.
[9] D. Boulinguez and A. Quinquis, "Underwater object classification using wavelet packets and fourier descriptors," in *Proc. ICIAP'99 10th Int. Conf. Image Analysis and Processing*, 1999.
[10] J. Barry, "Construction of three-dimensional Delaunay triangulations using local transformations," *Comput. Aided Geometric Des.*, vol. 8, pp. 123–142, 1991.
[11] C. S. Lin and C. L. Hwang, "New forms of shape invariants from elliptic Fourier descriptors.," *Pattern Recogn.*, vol. 20, no. 5, pp. 535–545, 1987.
[12] J. Wood, "Invariant pattern recognition: A review," *Pattern Recog.*, vol. 29, no. 1, pp. 1–17, 1996.
[13] R. C. Bolles and R. A. Cain, "Recognizing and locating partially visible objects: The local-feature-focus method," *Int. J. Robot.*, vol. 1, no. 3, pp. 57–82, 1982.
[14] O. Mitchell and T. Grogen, "Evaluation of Fourier descriptors for target recognition in digital imagery," Rome Air Development Center, Rome, Italy, Tech. Rep., 1983.
[15] M. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. IT-8, pp. 179–187, Feb. 1962.
[16] A. Khotanzad and Y. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, May 1990.
[17] Y. Sadjadi and E. L. Hall, "Three-dimensional moment invariants," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 2, pp. 127–136, Mar. 1980.
[18] C. H. Lo and H. S. Don, "3-d moment forms: their construction and application to object identification and positioning," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 1053–1064, Oct. 1989.
[19] G. Burel and H. Henocq, "Three-dimensional invariants and their application to object recognition," *Signal Processing*, vol. 45, p. 22, 1995.
[20] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 269–281, Mar. 1972.
[21] G. Burel, D. Carel, and J. Y. Catros, "A connectionist system for recognition of 2-D workpieces," Thomson-CSF, Paris, France, Tech. Rep., vol. 22, Dec. 1990.
[22] M. Sugeno, "Fuzzy measures and fuzzy integrals—A survey," in *Fuzzy Automata and Decision Processes*, M. Gupta, Ed. New York: North Holland, 1977, ch. 6, pp. 89–102.
[23] G. Rogova, "Combining the results of several neural networks," *IEEE Trans. Neural Networks*, vol. 7, pp. 777–781, Sept. 1994.

[24] S. B. Cho and J. H. Kim, "Multiple network fusion using fuzzy logic," *IEEE Trans. Neural Networks*, vol. 6, pp. 497–501, Mar. 1995.
[25] X. Zhang, L. Lin, and J. Wang, "A decision fusion approach for target classification," *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, vol. 1, pp. 667–671, Oct. 1996.
[26] K. Leszczynski, P. Penczek, and W. Grochulski, "Sugeno's fuzzy measure and fuzzy clustering," *Fuzzy Sets Syst.*, vol. 15, pp. 147–158, 1995.
[27] H. Tanahi and J. M. Keller, "Information fusion in computer vision using fuzzy integral," *IEEE Trans. Syst., Man, Cybern.*, vol. 2, pp. 733–741, May 1990.
[28] D. Cyganski and J. Orr, "Applications of tensor theory to object recognition and orientation determination," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, pp. 662–673, Nov. 1985.
[29] A. R. Edmonds, *Angular Momentum in Quantum Mechanics*. Princeton, NJ: Princeton Univ. Press, 1974.
[30] A. Messiah, *Mécanique Quantique*. Paris, France: Dunod, 1964.
[31] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Mécanique Quantique*. Paris, France: Hermann, 1986.
[32] T. W. Rauber, "Inductive pattern classification: Methods, features, sensors," Ph.D. dissertation, Universidade Nova de Lisboa, Lisbon, Portugal, 1994.
[33] G. Saporta, *Probabilités: Analyze des Données et Statistique*. Paris, France: Technip, 1996.

David Boulinguez (M'02) was born in France in 1973. He received the graduate degree in electronic engineering from Ecole Nationale Suprieure des Ingnieurs des Etudes et des Techniques d'Armement (ENSIETA), Brest, France, in 1996, and the Ph.D. degree in electronics from Brest University, Brest, France, in 2000.

He worked at ENSIETA for 4 years, where he jointly taught and conducted research in the areas of detection and classification in underwater acoustics and radar. He is now pursuing his teaching and research activities in signal processing and telecommunications at Institute Supérieur d'Electronique du Nord (ISEN), Lille Cedex, France.

André Quinquis (M'95) was born in Brest, France, in 1961. He received the Ph.D. degree in electronics from Brest University, Brest, France, in 1989.

Since 1992, he has been with the Ecole Nationale Suprieure des Ingnieurs des Etudes et des Techniques d'Armement (ENSIETA), Brest, France, where he has been jointly teaching and working in the areas of detection, adaptive signal processing, wavelets, and parameter estimation of underwater or radar observations.