



# Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results

Isabelle Leonard, Andreas Arnold Bos, Ayman Alfalou

## ► To cite this version:

Isabelle Leonard, Andreas Arnold Bos, Ayman Alfalou. Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results. SPIE Defense, Security, and Sensing, Apr 2010, Orlando, United States. pp.76780O, 10.1117/12.849688. hal-00781621

**HAL Id: hal-00781621**

**<https://hal.science/hal-00781621>**

Submitted on 28 Jan 2013

**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.

# Interest of correlation-based automatic target recognition in underwater optical images: theoretical justification and first results

I. Leonard<sup>a</sup>, A. Arnold-Bos<sup>b</sup> and A. Alfalou<sup>a</sup>

<sup>a</sup>ISEN, 20 rue Cuirasse Bretagne, CS 42807, 29228 BREST CEDEX 2, FRANCE;

<sup>b</sup>THALES UNDERWATER SYSTEMS, Route de Sainte Anne du Portzic, CS 43814, 29238 BREST CEDEX 3, FRANCE

## ABSTRACT

In this paper, we explore the use of optical correlation-based recognition to identify and position underwater man-made objects (*e.g.* mines). Correlation techniques can be defined as a simple comparison between an observed image (image to recognize) and a reference image; they can be achieved extremely fast. The result of this comparison is a more or less intense correlation peak, depending on the resemblance degree between the observed image and a reference image coming from a database. However, to perform a good correlation decision, we should compare our observed image with a huge database of references, covering all the appearances of objects we search. Introducing all the appearances of objects can influence speed and/or recognition quality. To overcome this limitation, we propose to use composite filter techniques, which allow the fusion of several references and drastically reduce the number of needed comparisons to identify observed images. These recent techniques have not yet been exploited in the underwater context. In addition, they allow for integrating some preprocessing directly in the correlation filter manufacturing step to enhance the visibility of objects. Applying all the preprocessing in one step reduces the processing by avoiding unnecessary Fourier transforms and their inverse operation. We want to obtain filters that are independent from all noises and contrast problems found in underwater videos. To achieve this and to create a database containing all scales and viewpoints, we use as references 3D computer-generated images.

**Keywords:** Automatic target recognition, underwater optical images, fast correlation-based methods, composite filter.

## 1. INTRODUCTION

In underwater mine warfare, unmanned underwater vehicles (UUVs) are used as a complement to divers, to detect mines and eventually destroy them. Currently, Remotely Operated Vehicles (ROVs) are used. They typically have a sonar and a camera. They require the constant attention of an operator and they have a short range since they are tethered to a mothership. Thus the planned evolution to ROVs are Autonomous Underwater Vehicles (AUVs). In a typical mine hunting scenario, the mothership uses its sonars to study an area in order to detect and classify mines. Once mines have been detected, ROVs are launched to confirm the classification and eventually destroy the mine. During the transit to the target, the vehicles are guided by their sonar. When close to the mine, video can be used. Because of the reduced visibility in the underwater medium, the range at which video is usable is limited to a few meters only. Video is used to detect, localize and identify the underwater object. The specificities of the underwater medium such as light absorption and scattering lead to images with noise, no contrast and unusable color information. Depending on the location and the recording environment (like water turbidity), these effects can be more or less accentuated in the images. For these reasons, and because correlation is sensitive to these problems, images need preprocessing. Our studies concern a system that is able to perform detection, localization and identification at same time while preprocessing images. Moreover, we add a constraint on the processing time: we want a real-time process.

---

Further author information: Send correspondence to I. Leonard  
E-mail: isabelle.leonard@isen.fr

Research on pattern recognition has been carried out for various purposes like handwritten, fingerprint or face recognition, to name only a few well-known ones. Recognition is a difficult problem since there are many objects to identify and lots of points of view of them and imaging conditions can modify their appearance. In this article, we compare a few methods, selected according to criteria like the processing time and the number of required operations, the idea being to provide elements justifying the choice behind our own approach.

Our paper is organized as follows. First, we present the setting, the studied targets and the context of underwater optical imaging. Then we expose the state of the art and discuss our choice of using optical correlation-based methods to recognize underwater objects. We explain our integrated preprocessing features to improve target images without increasing processing time, our filter manufacturing and our decision of trying the use of computer-generated images instead of underwater images. Finally we present first results obtained with data acquired at sea that seem to validate our approach.

## 2. CONTEXT OF STUDY

### 2.1 Use of video in the underwater environment

Video is used in several applications. Some are briefly developed here.

A first category is visual servoing. For some biological surveys applications, it is required to follow fishes. The fish must be kept in the center of the image recorded by the camera.<sup>1</sup> This application needs to have a robot that is able to adapt to the speed of the fishes and to their orientation changes. Cables and pipes following is another example.<sup>2</sup>

A last example is station keeping:<sup>1</sup> the robot modifies its position depending on the chosen feature displacement, which is measured in comparison with its initial position.

Target recognition is the second application of video, as in the case with mine warfare, but also the retrieval of lost objects, such as flight recorders.

### 2.2 Underwater image

In underwater imaging, visibility is a major problem. It varies from about 30 m in very clear water to about 0.75 m in very turbid water like a harbor water.

The received image  $I_R$  is the addition of the attenuated image  $I_A$ , the forward scattered image  $I_{FS}$  and the backscattered component  $I_{BS}$ :<sup>3</sup>

$$I_R = I_A + I_{FS} + I_{BS} \quad (1)$$

The light attenuation increases exponentially with distance (according to the Beer-Lambert law) and limits visibility. The backscattered component corresponds to the light that has been reflected by particles towards the video camera (see Figure 1). This hides objects present in the scene. The forward scattered component is the light reflected by the object that has been diffused by particles on its way back to the camera, which produces blurred images (see Figure 1).

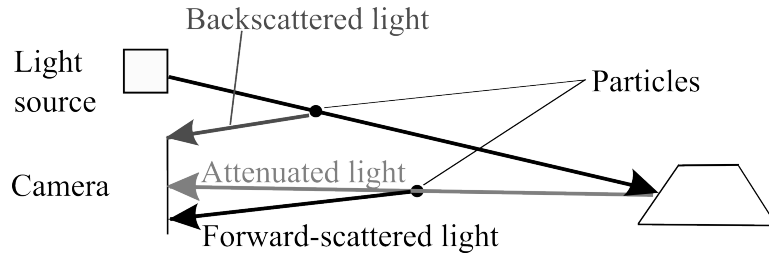


Figure 1. Diagram of the light propagation in an underwater medium

It is possible to add artificial lighting in the scene; this increases the visibility range but it produces a non-uniform lighting. Video detects also particles on which light is reflected, called marine snow.<sup>1</sup>

These consequences of the underwater medium are annoying, especially backscattering that is the most important effect. Thus underwater images need preprocessing.

### 2.3 Studied targets

The studied targets are underwater mines (Figure 2). There are many kinds of mines. In this work we focus on three of them: spherical mines (Figure 2-a), cylindrical mines (Figure 2-b) and Manta mines (Figure 2-c) (the Manta mine is a truncated cone).

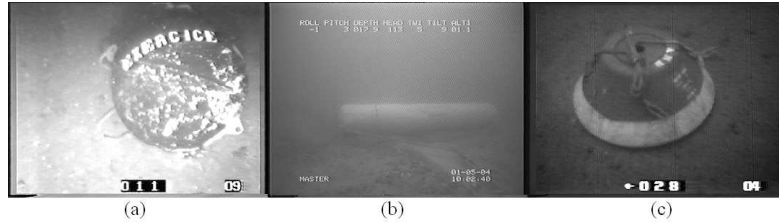


Figure 2. Grayscale images representing three studied targets that come from underwater videos: a spherical mine (a), a cylindrical mine (b) and a Manta mine(c)

### 3. IMAGE RECOGNITION METHODS: A BRIEF STATE OF ART

Pattern recognition is a problem which has received a great deal of attention in the literature. The purpose of this paper is not to give an exhaustive review of all existing methods. In this paragraph, we expose a few methods selected according to criteria like short time processing and the ability to provide the location and/or the orientation of the mine. First we expose methods that have already been used for underwater target recognition purposes. Then we expose some other methods used in other fields, which could have the potential of being adapted for the underwater environment.

Some methods applied on underwater images are based on contours analysis. Thales's module for underwater mine detection preprocesses images before applying segmentation to extract contours.<sup>4</sup> These contours are compared with lines and ellipses, to keep only man-made object contours. The center of contours determines the mine location. To identify a mine, they compare the mine shape with reference shapes and they correlate this information with navigational information, in order to obtain a measure of match quality. We do not use this method because it needs contours extraction and processing of these contours: in the underwater environment, the contours may be too faintly visible for the match to be good.

In their article,<sup>5</sup> Olmos *et al.* suggest and validate a method that consists in distinguishing man-made objects from natural objects in underwater videos. Therefore, they extract edges with a Canny detector and a measure of uniformity which consists in comparing the uniformity of the intensities of each side of the contour. Assuming that man-made objects have straight edges, they use lengths, regularity and the number of found contours to classify natural and man-made objects. As the previous method, this one needs a Canny edge detector. This detector may not be able to find all the interesting contours in very noisy images.

Bazeille *et al.* identify known objects with color analysis.<sup>6</sup> They base their studies on the Beer-Lambert law and showed that a color has a set of compatible colors in the underwater medium. Their algorithm has a training step to find the compatible color range of a known object (through a logarithmic transform and principal analysis component). Then it analyzes image pixels to search objects. Since we have color and grayscale images, we cannot use exclusively color-based methods.

Underwater imaging is also used for tracking. Trucco et Plakas made a review of some recent papers.<sup>2</sup> We can note that developed methods are essentially applied to mosaics, cables and pipes following or positioning. These methods are based on Hough transforms, on corner detection, and on correlation-based algorithms. Hough transform-based methods need another processing to analyze results that may complicate a system. Hough transforms are essentially based on contours, which, as we said above, may be too faintly visible in the image. Corner detectors "lock" on any feature of the image where the gradient is strong in any direction. Any feature such as text painted on the object will be caught by the corner detector. Since the presence of such features is unpredictable, it makes corner detectors hard to use for image recognition.

All existing pattern recognition methods have not yet been applied to underwater target recognition. Therefore, we look after some other methods from terrestrial applications.

In a recognition application for vehicles in terrestrial images, Olson *et al.* proposed an approach based on the use of both edge maps of images and references.<sup>7</sup> They compare the position and the gradient of pixels from the input image to the position and the gradient of pixels from references. The decision is based on a threshold on the number of corresponding pixels. This method has two drawbacks: it needs a good edge detection method and it does comparison on a per-pixel basis.

In another approach, Cole *et al.* proposed a method adapted to Lego pieces recognition.<sup>8</sup> Therefore, they have a database divided in classes. Each class represents one Lego piece and contains an average image and a cluster image. Each image to identify is compared to average images with a normalized cross correlation and a threshold. Then the image is compared to clusters with which there might be a match to obtain the name of the Lego piece. This method needs two comparisons: the first one with a huge database and the second one with a reduce database that leads to the decision.

In the Ref. 9, Schiele and Crowley use color histogram matching to recognize objects. The first step is to calculate the color image histograms. Then they compare them to reference histograms with a  $\chi^2$  test. Different tests<sup>9</sup> prove that this method is robust to orientation change, scale change, partial occlusion and viewing point change. This method is a simple, robust, and has a short processing time. But it depends on color and light intensity, that are underwater image problems.

SpikeNet is a real-time method, developed by Thorpe *et al.*<sup>10,11</sup> They have developed a very fast program that simulates neuron network with very few parameters, to recognize faces. The first version of their algorithm has four steps: a step to know at which state the image pixel correspond to (by analogy with the retina which have two states of its cells: on and off), two steps to recognize contour orientations and then specific features that the program had learnt and a final step to analyze results. Following this version, they proposed an optimized version that is able to recognize all kinds of objects if learnt specific features contain oriented structures. This method needs a good contour visibility and objects from which we can extract specific features. In our case, this means specific features for each mine, each orientation and each scale and as a consequence a huge database of reference features.

In the same category, Viola and Jones used the AdaBoost classifier to rapidly detect faces in grayscale images.<sup>12</sup> They select some classifiers that they assemble in a cascade. Thus they achieve a low error rate on their training database. The drawback to this approach is that AdaBoost, like neural-network-based methods, is of the black box kind: it is hard to debug.

## 4. THEORETICAL JUSTIFICATION: WHY CORRELATION?

### 4.1 Choice

As shown in part “context of study” (part 2, p. 2), in the water medium, color is attenuated and objects’ color changes on each video. Moreover, in our studies, we have color images as well as grayscale images. Thus color methods<sup>6</sup> cannot be used. Schiele and Crowley’s method<sup>9</sup> is not able to provide orientation information. We do not use methods based on edge maps<sup>7</sup> and/or binary images.<sup>5</sup> Indeed, noise present in images affects object contours. These are not clear in all images, which can perturb recognition results.

For these reasons, we chose the correlation method. It compares an image to a reference image. The presence of a correlation peak indicates the resemblance and therefore the mine’s orientation with respect to the video camera. The correlation peak location also gives the location of the mine in the image. This method mainly recognizes contours that can be obtained from grayscale images. Recognition can be obtained even if entire contours are not visible in the studied image.

### 4.2 Correlation

Correlation is a signal and image processing method that compares a target image with a reference image. The result of this comparison is a more or less intense correlation peak, depending on the resemblance degree between these two images. Mathematically, correlation can be written:

$$c(x_0, y_0) = h(x_0, y_0) *^c s(x_0, y_0) \quad (2)$$

$$c(x_0, y_0) = \int \int_{-\infty}^{+\infty} h^*(x, y) s(x + x_0, y + y_0) dx dy \quad (3)$$

where “c” is the result of the correlation operation, “\*” is the correlation product, “\*” is the complex conjugate operator, “h” is a filter or reference, “s” is the image to analyze, “(x<sub>0</sub>, y<sub>0</sub>)” are spatial coordinates and “(x, y)” are integration variables. The correlation operation can be expressed with a Fourier transform:

$$C(\mu, \nu) = H^*(\mu, \nu) \cdot S(\mu, \nu) \quad (4)$$

where  $\mu$  and  $\nu$  are the coordinates in the frequency plane,  $C$ ,  $S$  and  $H^*$  are the respective Fourier transforms of functions  $c$ ,  $s$  and  $h^*$ .

The classical matched filter<sup>13</sup> has been modified by introducing information in order to obtain some robustness to noise or orientation change for instance.<sup>14</sup> Here we develop only three filters: classical matched filter,<sup>13</sup> phase only filter (POF)<sup>15</sup> and optimal trade-off filter (OT filter).<sup>16</sup> Other filters have been developed, in the literature, like the binary phase only filter<sup>17</sup> and the inverse filter.<sup>18</sup> These filters, expressed in the Fourier plane, may then replace “ $H^*$ ” in equation 4.

The most known filter is the classical matched filter ( $F_{CMF}$ ) defined in the Fourier plane as:<sup>13</sup>

$$F_{CMF}(\mu, \nu) = \frac{\alpha S^*(\mu, \nu)}{B(\mu, \nu)} \quad (5)$$

where “ $S^*(\mu, \nu)$ ” denotes the complex conjugate of the reference, “B” the spectral density of the background and “ $\alpha$ ” is a constant. This filter is robust but has a low discriminating power.<sup>14</sup>

The phase only filter ( $F_{POF}$ ) is defined in the Fourier plane as:<sup>15</sup>

$$F_{POF}(\mu, \nu) = \frac{S^*(\mu, \nu)}{|S^*(\mu, \nu)|} \quad (6)$$

where “ $|S^*(\mu, \nu)|$ ” is the module of the reference spectrum. This filter gives a sharp correlation peak, it is very discriminative<sup>15</sup> but also noise sensitive.

The optimal trade-off filter ( $F_{OT}$ ) is defined in the Fourier plane as:<sup>16</sup>

$$F_{OT}(\mu, \nu) = \frac{S^*(\mu, \nu)}{\alpha B(\mu, \nu) + (1 - \alpha)|S^*(\mu, \nu)|^2} \quad (7)$$

From now on, in this article, we will use only the POF filter. Since we need a discriminative filter. We don’t use the OT filter because in order to obtain good performances we have to know the spectral density of the background which must be very close to the one in the actual images.<sup>19</sup> In underwater applications, it is very difficult to know the spectral density of the background because of the ocean floor, turbidity and all noises that perturb an image and vary throughout the video.

Correlation is extremely fast. The fast Fourier transform takes  $O(n \log(n))$  arithmetical operations<sup>20</sup> and each image multiplication takes  $O(n^2)$  arithmetical operations, where  $n$  is one of the two size of the image to analyze (here, we suppose that image is a square image).

To perform a good correlation decision, we should compare our image with a huge database of references, covering all the appearances of mines. Using the filter defined previously, this operation takes a lot of time (proportional to the number of references). To overcome this problem, we use at first the classical composite filter techniques<sup>14</sup> and the POF technique. In future works, we plan to use an optimized version of the composite filter called the segmented composite filter.<sup>21</sup> The classical composite filter technique used in this work consists in linearly adding several references (Figure 3): first each reference image is Fourier transformed. Then we add all the reference spectra to obtain a composite filter in the Fourier domain.

One problem we can meet when using this technique, is a saturation problem. The values stored in the filter are quantized in a fixed amount of values, *e.g.* 255 for 8-bits images. If there are too many references, this

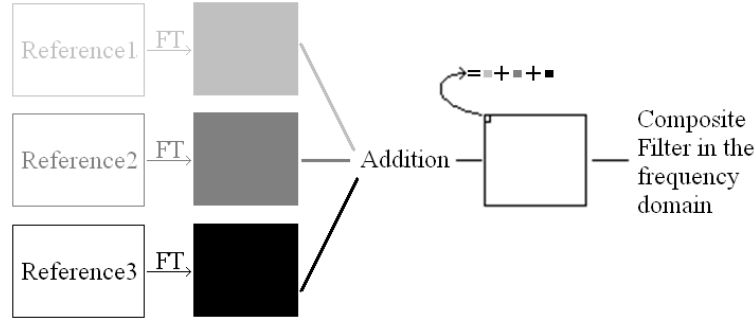


Figure 3. Composite filter technique

maximum value will be reached and adding another reference to the filter will not allow to recognize it in an image. To overcome this problem, we can use the segmented filter technique. In this latest method, the fusion of different references is performed according to a segmentation criterion. This criterion consists in dividing the Fourier plane into separate areas: a first area is common to the different spectra, where the different images have similar spectra (these areas being more or less important depending on the degree of resemblance between the reference images), and other areas are specific to the reference image spectra (for each area, specific information to the reference image spectrum is introduced).

The correlation method has some drawbacks, which can be annoying like the non-robustness to noise, rotation and scale modifications or the fact that it considers only one point of view. Objects can be seen with any point of view and they can have any scale, depending on distance between the object and the video camera. therefore, we have a huge database of references. Composite filter techniques allow reducing the database size and therefore the number of needed operations. The result of the correlation indicates the resemblance degree between the object and the reference image thus we obtain orientation information. The location of the correlation peak gives the location of the object in the image. The method also allows for inserting information in filters. Indeed, in the algorithm, we multiply the target image spectrum and the filter spectrum in the Fourier domain. The “context of study” section of this paper (part 2, p. 2) shows that underwater images cannot be used in the raw state and need preprocessing. Many of preprocessing we need have to be performed in the Fourier domain. It is possible to integrate them directly into filters and reduce the whole processing time by avoiding several Fourier transforms. Moreover, the correlation method has some advantages like its small computation time and robustness to translation.

The optical correlation is an efficient technique, thanks to the decision reliability (it is the aim of all developed filters), cheap but efficient. In addition to digital implementation, it is also possible to use opto-electronic interfaces and for applications that need a rapid identification (for instance person identification). However, for general public applications, the optical implementation will be expensive and may complicate the whole system. Thus, we chose to implement the Vander Lugt setup<sup>13</sup> numerically, which consists in the multiplication of the image spectrum and the complex conjugate filter spectrum. We could have digitally implemented the Weaver and Goodman setup<sup>22</sup> but this setup need the image and the reference in the same input plane. Therefore, the system has to work with large matrix and this takes more processing time. Moreover, this configuration needs some additional processing in the correlation plane. In the correlation plane, there are two cross-correlation peaks (in case of resemblance) and an autocorrelation peak that could perturb the decision.

## 5. CORRELATION APPLIED TO UNDERWATER MINE RECOGNITION

Our algorithm is divided in several steps (Figure 4).

We have an offline part that consists in preprocessing reference images and compositing them to create filters. The second part of our algorithm consists in preprocessing images. Then we perform correlation by multiplying filters and the preprocessed image. We apply the PCE criterion to decide if searched mines are in the studied image.

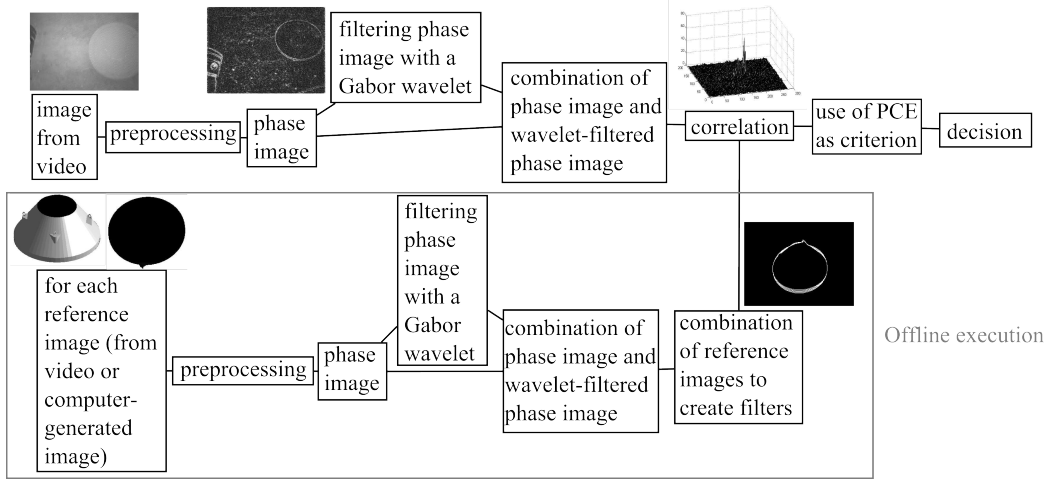


Figure 4. Our algorithm diagram

### 5.1 Preprocessing

To save processing time, images are resized to suppress navigational information. To suppress the moiré effect, we suppress the high frequencies corresponding specifically to the moiré pattern in the spectrum as proposed by Sidorov and Kokaram.<sup>23</sup> The moiré effect is due to the use of analog cameras; it can be avoided by the use of digital cameras.

To enhance the contrast of the images, we implement the preprocessing algorithm described in Ref. 24. From the initial image, we apply a low pass filter to create a low pass image. This low pass image will be used twice. Firstly, it will be subtracted to the initial image to suppress backscattered light in the additive model. Secondly, it will divide the initial image in order to suppress the illumination component in the illumination-reflectance model. After clipping values, final image is the result of the division or the result of the subtraction function of the gradient of each image.

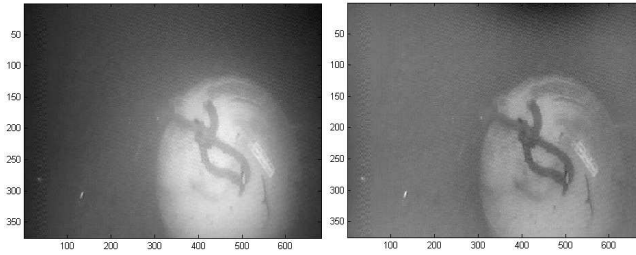


Figure 5. before (left) and after (right) contrast enhancement

### 5.2 Phase images and filtering with a Gabor wavelet

To efficiently recognize objects, we work on grayscale images where the information is related to contours. However, contrarily to typical contour-based methods, we do not threshold a gradient, but use the phase of the image. Edges correspond to sharp phase change while information contained by the amplitude are not interesting in our case.<sup>25</sup> Using the phase is a way to avoid any thresholding: indeed, thresholding is a problem when the signal to noise ratio for the edge strength is low, as is the case for underwater images. To achieve this, we work on the Fourier domain. It is the processed image, once processed in the Fourier plane, that can be expressed as:  $I = r.e^{i\theta}$ . The amplitude information is contained by  $r$  and the phase information by the exponential function. To keep only the phase information, all values are divided by their amplitude:  $I = e^{i\theta}$ .

Since there is still some noise, we use a band pass created with a Gabor wavelet:  $f(x) = \cos(\frac{2\pi x}{\lambda})e^{-\frac{x^2}{\sigma^2}}$ , where “ $\lambda$ ” is the wavelength and “ $\sigma$ ” is the standard deviation of the Gaussian exponential function. This operation



remove some noise but the edges are not as straight as in the phase image. To solve this problem we combine these two images by addition.

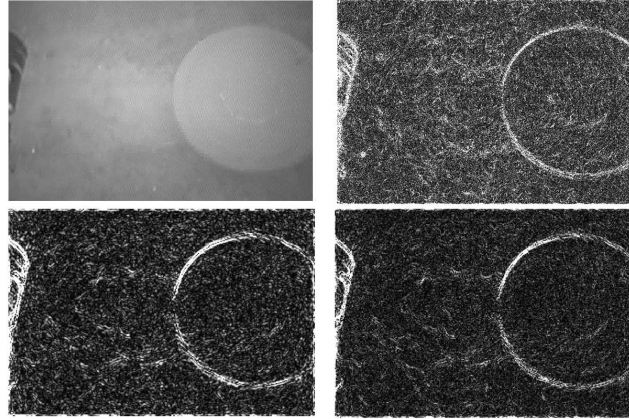


Figure 6. initial image (top left), phase image (top right), phase and wavelet image (bottom left), combination of phase image and wavelet image (bottom right)

On Figure 6, there are images of different steps of our preprocessing. We can notice the apparition of contours in the phase image. The wavelet image filters some noise but edges are not as sharp as in the phase image. The combined image has advantages of the phase image and the wavelet image: we can see quite sharp edges and less important noise.

### 5.3 Introduction of preprocessing in the filter step

Suppressing the moiré effect, normalizing the amplitude, filtering by a wavelet, is all done in the Fourier domain. All these operations could be done one after the other before computing the Fourier transforms of the images. But, since the Fourier transform is linear, it is much more efficient to do them in one step directly after computing the Fourier transform of the image, which is already needed to do the correlation.

### 5.4 Sector filter

Composite filter techniques allow for reducing the database size by grouping several reference images in a single filter. However, as we want to know the mine orientation, we use sector filters. This approach means that we divide the reference database into so-called sectors: one sector contains several reference images that represent the mine under orientations that are close to each other, and with one scale. There is one filter per sector. Thus, the presence of the correlation peak will indicate the presence of the mine, its orientation and its approximate location.

To create filters, we automatically assemble close reference images obtained by computing edges of reference images. The criteria used to assemble two references are the size and the orientation of the mine shape. For each image in the reference database, we compare the mine size and its orientation to reference mines that have already been studied. If the values of the criteria are close to the criteria of an already processed image, we add this image to the corresponding filter. If it does not correspond, we create a new filter. This way, it is possible to use a huge reference database. These filters are created offline and once for all.

### 5.5 Decision with the PCE criterion

To measure the performance of our correlation algorithm, we need a method to evaluate the resemblance. We have chosen the peak to correlation energy (PCE) criterion:

$$\text{PCE} = \frac{\text{correlation peak energy}}{\text{correlation plane energy}} \quad (8)$$

This criterion evaluates energy contained in the correlation peak, in comparison to energy contained in the correlation plane.

There are other existing criteria like the signal to noise ratio or the Horner efficiency.<sup>18</sup>

PCE criterion varies depending on each correlation plane. It needs some modifications. Indeed, the PCE criterion is not a normalized criterion. Thus PCE values can strongly vary from one video to another and according to the implemented algorithm. To compare different methods, some normalization is needed. This could be done with the normalized cross correlation that use as normalization tools the local mean and the standard deviation of the image and of each reference image.

## 6. IMAGES FROM VIDEOS AND COMPUTER GENERATED IMAGES

It is not advisable to use reference images extracted from videos, since those images will be corrupted by noise and not necessarily visible from all orientations and angles. Also, the background is also present, which is a problem when trying to correlate the reference with an observed image. As a result, we decided to use computer-generated images as references. This choice is guided by the fact that computer-generated images are independent from noise and provide us with all scales and all viewpoints we want. Thus, we obtain an extensible database: for each mine we are able to create images for each scale and each point of view.

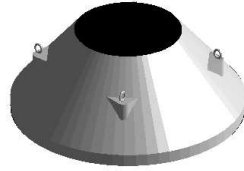


Figure 7. A computer-generated image of a Manta mine

The use of several colors and light gives more details when we perform the contour image. With more details, it is easier to discriminate objects when these details are to be detected in videos. We first create a computer-generated images based database representing the Manta mine. For future works, we will generate database for the spherical and the cylindrical mines.

## 7. FIRST RESULTS

### 7.1 Results obtained with filters created with images from videos

As a first, quick experiment, we created filters with images from a video sequence to recognize the Manta mine from image 650 to image 850. On Figure 8, we can notice that filters do indeed recognize these images. Images around image 1700 have PCE values indicating some degree of recognition as well. This result could be expected since the appearance of the mine in this place is close to the appearance of the mine contained in filters. The black line represents the comparison of Manta based filters and a video of a cylinder mine. Obtained values of PCE means that the mine is not recognized as a Manta mine.

### 7.2 Results obtained with filters based on computer-generated images

We created five filters with computer-generated images that represent the main mine appearances we can see in the studied video. Each filter contains five reference images. Between two reference images there is an angle difference of  $2^\circ$ . Figure 9 shows that the mine is recognized almost everywhere it is present in the video. However, in images located from 1200 to 1400, there is no mine but we see the video camera and it is recognized as a mine by the system. The cylinder mine is not recognized as a Manta mine. This is explained by the fact that the Manta mine is mainly view as a circle or a cone and there is no view of the circular side of the cylinder in the video used to obtain these results.

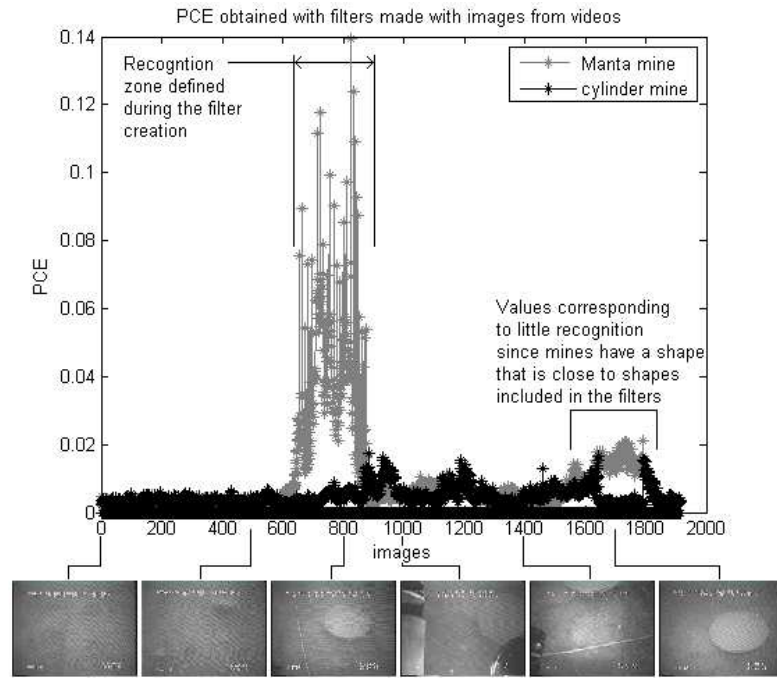


Figure 8. Results obtained with filters created with images from videos

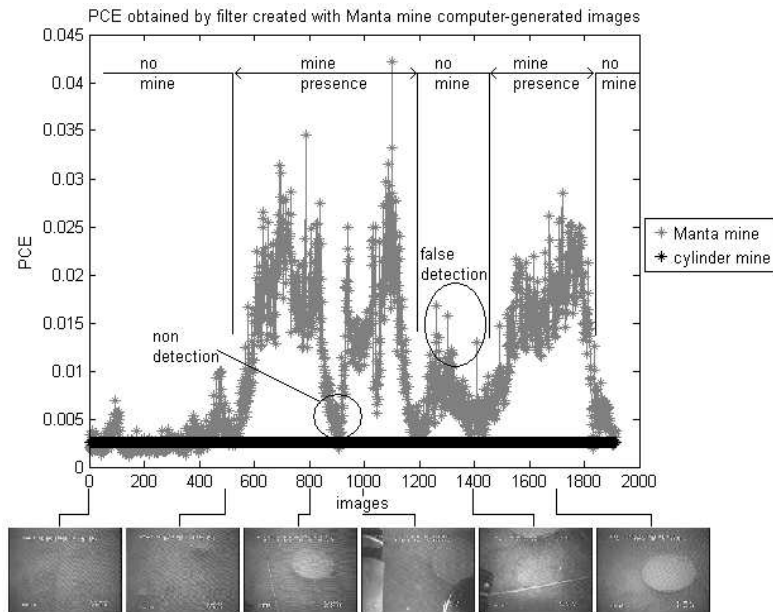


Figure 9. Results obtained with filters created with computer-generated images

### 7.3 Results

Our algorithm has been run using Matlab, version R2007a. The computer we use is a Intel Core 2 Quad CPU cadenced at 2.66GHz. The recognition rate and the no detection rate correspond to the number of images containing a mine that have been correctly or not identified in comparison to the total number of images containing a mine. The error rate corresponds to images containing another type of mine identified as a manta mine in comparison to the number of studied images. The false detection rate corresponds to images without mines that have been identified as an image containing a mine, in comparison to the number of images that does not contain a mine. In Table 1, the error rate has been obtained using a video where the circular side of the

Table 1. Results

Method	Video images based sector filters	Computer-generated images based sector filters
Number of filters	5	5
Execution time	0.7s per image	0.4s per image
Recognition rate	50.4%	79.5%
No detection rate	49.6%	20.5%
Error rate	6%	15%
False detection rate	0%	2.9%

cylindrical mine is presented. That is not the case of the video used to record the Figures 8 and 9. These results are encouraging: they validate the use of the correlation method and composite filter techniques to automatic target recognition in the underwater medium. Moreover, the use of computer-generated images to create filters provides good recognition rates. Error rates are low since there are few images representing the spherical side of the cylinder mine. Seen from some point of views, spherical mines, Manta mines and cylindrical mines have almost identical appearances and this can produce errors.

## 8. CONCLUSION

Our aim is to realize an algorithm that is able to detect, identify and locate underwater mines with a short processing time. Orientation information is important too. With these criteria, we chose the correlation method, in comparison with other methods selecting on these criterion.

To obtain a short processing time we have integrated preprocessing that need a Fourier transform in the filter step. This integration avoids several Fourier transforms and their inverse. To correlate the images with a huge database in a short execution time, we use composite filter techniques which drastically reduce the size of the database. First results seem to validate the use of this technique. Moreover, the use of computer-generated images to create filters is validated by recognition of the mine in videos.

To optimize our filters and recognition, we have several opportunities like the use of a segmentation criterion to composite our filters instead of adding reference images. We will look after a decision criterion since the PCE criterion provides us a decision but without normalization, we cannot compare different methods.

For the guidance, we have two options. The first one is to use the correlation peak location in our algorithm. The second option is to use Weaver and Goodman's setup. In this case, we compare an image with the previous image. The correlation peak locations will change according to the displacement of the mine between both images.

## ACKNOWLEDGMENTS

The raw mine images contained in this publication are derived from data property of the French State that have been provided by the GESMA (Groupe d'Etudes Sous-Marines de l'Atlantique) within TOPVISION project coordinated by Thales Underwater Systems SAS. This project is related to Techno-Vision Programme launched by French Ministry of Research and French Ministry of Defense. More detailed information can be found on <http://topvision.gesma.fr> site.

This work is supported in part by the Regional Council of Brittany.

## REFERENCES

1. R. Marks, M. Lee, and S. Rock, "Using visual sensing for control of an underwater robotic vehicle," in *Proceedings of IARP second workshop on mobile robots for subsea environments*, 1994.
2. E. Trucco and K. Plakas, "Video tracking: a concise survey," *IEEE Journal of Oceanic Engineering* **31**(2), pp. 520–529, 2006.
3. B. McGlamery, "A computer model for underwater camera systems," in *SPIE Ocean Optics VI*, **28**, 1979.
4. A. Arnold-Bos and N. Mandelert, "Sonar and video perception for an autonomous mine disposal vehicle," in *UDT Europe*, 2008.
5. A. Olmos and E. Trucco, "Detecting man-made objects in unconstrained subsea videos," in *BMVC*, 2002.
6. S. Bazeille, I. Quidu, and L. Jaulin, "Identification of underwater man made object using color," in *Proceedings of the Institute of Acoustics 29*, 2007.
7. C. Olson and D. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Transactions on image processing* **6**(1), pp. 103–113, 1997.
8. L. Cole, D. Austin, and L. Cole, "Visual object recognition using template matching," in *Australian conference on robotics and automation*, 2004.
9. B. Schiele and J. Crowley, "Object recognition using multidimensional receptive field histograms and its robustness to view point changes," in *1995 Rosenon workshop on computational vision*, 1995.
10. S. Thorpe, "Ultra-rapid scene categorization with a wave of spikes," in *BMVC*, 2002.
11. A. Delorme, J. Gautrais, R. Van Rullen, and S. Thorpe, "Spikenet: a simulator for modeling large networks of integrated and fire neurons," *Neurocomputing* **26-27**, pp. 989–996, 1999.
12. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings IEEE Conference on computer vision and pattern recognition*, 2001.
13. A. Vander Lugt, "Signal detection by complex filtering," *IEEE Transactions on information theory*, pp. 139–145, 1964.
14. B. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlator," *Applied Optics* **31**, pp. 4773–4801, 1992.
15. J. Horner and P. Gianino, "Phase-only matched filter," *Applied Optics* **23**(6), pp. 812–816, 1984.
16. P. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak and horner efficiency," *Optics Letters* **16**(11), pp. 829–831, 1991.
17. J. Horner, B. Javidi, and J. Wang, "Analysis of the binary phase-only filter," *Optics Communications* **91**, pp. 189–192, 1992.
18. B. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," *Applied Optics* **29**(20), pp. 2997–3006, 1990.
19. A. Grunnet Jepsen, S. Tonda, and V. Laude, "Convolution-kernel-based optimal trade-off filters for optical pattern recognition," *Applied Optics* **35**(20), pp. 3874–3879, 1996.
20. J. Cooley and J. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of computation* **19**(90), pp. 297–301, 1965.
21. A. Alfalou, G. Keryer, and J. de Bougrenet, "Optical implementation of segmented composite filtering," *Applied Optics* **38**(29), pp. 6129–6135, 1999.
22. C. Weaver and J. Goodman, "A technique for optically conciolving two functions," *Applied Optics* **5**(7), pp. 1248–1249, 1966.
23. D. Sidorov and A. Kokaram, "Suppression of moiré patterns via spectral analysis," in *Proceedings of SPIE in visual communications and image processing*, 2002.
24. A. Arnold-Bos, J. Malkasse, and G. Kervern, "Towards a model-free denoising of underwater optical images," in *Proceedings of the IEEE conference on Ocean (Europe)*, 2005.
25. A. Oppenheim and J. Lim, "The importance of phase in signals," in *Proceedings of the IEEE*, **69**(5), pp. 529–541, 1981.