

## Research Article

# A Fully Automated Method to Detect and Segment a Manufactured Object in an Underwater Color Image

Christian Barat<sup>1</sup> and Ronald Phlypo<sup>2</sup>

<sup>1</sup>Laboratoire I3S, Université de Nice Sophia-Antipolis, 06903 Sophia-Antipolis, France

<sup>2</sup>MEDISIP-IBBT-IbiTech, Ghent University, 9000 Ghent, Belgium

Correspondence should be addressed to Christian Barat, barat@i3s.unice.fr

Received 1 July 2009; Revised 30 November 2009; Accepted 2 February 2010

Academic Editor: Warren Fox

Copyright © 2010 C. Barat and R. Phlypo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a fully automated active contours-based method for the detection and the segmentation of a moored manufactured object in an underwater image. Detection of objects in underwater images is difficult due to the variable lighting conditions and shadows on the object. The proposed technique is based on the information contained in the color maps and uses the visual attention method, combined with a statistical approach for the detection and an active contour for the segmentation of the object to overcome the above problems. In the classical active contour method the region descriptor is fixed and the convergence of the method depends on the initialization. With our approach, this dependence is overcome with an initialization using the visual attention results and a criterion to select the best region descriptor. This approach improves the convergence and the processing time while providing the advantages of a fully automated method.

## 1. Introduction

The objective of this work is to present a method which detects and segments manufactured objects (particularly underwater mines) in underwater video images. Actually, the underwater video is increasingly used as a complementary sensor to the sonar especially for detection of objects or animals; see [1–4]. However, the underwater images present some particular difficulties including natural and artificial illumination, color alteration, light attenuation [5], and marine snow. The method must tackle these problems and is developed under two constraints: no human intervention, low processing time. For segmenting an object in an image, we first need to detect the presence of the object. The method is composed of two main steps:

- (1) object detection,
- (2) object segmentation if an object is detected.

For detecting or tracking objects in underwater images, generally, a preprocessing step is applied to enhance the images. For example, in [6], a homomorphic filtering is used to correct for the illumination, wavelet denoising,

anisotropic filtering to improve the segmentation, adjustment of image intensity, and some other processing. A self-tuning image restoration filter is applied in [7], but the illumination is considered to be uniform, which is a restrictive hypothesis. In [4] the constant background features are estimated for each frame by computing the sliding average over the ten preceding frames, and the average is subtracted from the frame. This preprocessing cannot be applied in our images due to the large size of our objects and the slow movement of the object in the image, because the average will contain the object. Also, we want to detect an object in only one image. The main problem encountered in our images is the shadow, a problem which is not discussed in previous works. We propose to use the visual attention approach [2, 8] with some modifications. Once the saliency map is extracted (using the visual attention method), we apply a Linear Discriminant Analysis (LDA) [9] to estimate the probability of object presence. For the second step, we use active contour. Since publication of the work of Kass et al. [10], extensive research on snakes (active contours) has been developed to segment images. The early approaches [10, 11] minimized an energy function

to move the active contour toward the object's edges. In [12], region information has been used to overcome some problems inherent to the edge approach. In this paper we put forward a region based snake method in the Minimum Description Length (MDL) framework. This approach is adapted to our problem of segmentation of uniform color objects. Since we require low processing times, we use explicit snake (polygonal). Implicit approach (level set) needs more computational time even with the fast marching method [13]. The problem encountered in the classical snake method is the dependence on the contour initialization. We propose to use the visual-attention-based method to find the region of interest to segment in the image (there is generally only one manufactured object to be found in underwater images). Moreover, in the classical region-based snake approach, the region descriptor is fixed for the whole video sequence, while in this paper we propose to select the best region descriptor adapted to each image. The idea is to use the information extracted with the visual attention to select, based on a kurtosis criterion, the region descriptor for each image to segment. This approach reduces the shadow effect during the segmentation step. The methods are presented in Section 2, our approach is developed in Sections 2.4 and 2.5. Experimental results on real images are reported in Section 3.2.

## 2. Methods

The general method for image segmentation is composed of six steps, the first three steps corresponding to the object detection and the last three corresponding to the object segmentation:

- (1) extraction of a saliency map using visual attention,
- (2) detection of the most salient part of the map,
- (3) classification in class "object" or "no object",
- (4) selection of a region descriptor,
- (5) initialization of the snake on the most salient part of the map,
- (6) segmentation of the image by the snake.

In what follows we supply the details for each of the above subparts.

**2.1. Visual Attention.** The method we adapt is based on the work of Itti et al. [8]. Low-level vision features (orientation, brightness, color channels tuned to red, green, blue and yellow hues) are extracted from the original color image at several spatial scales (depending on the image size). The different spatial scales are obtained through the use of a dyadic Gaussian pyramid. Then we combine the different maps, after a normalization step, to obtain a saliency map. The final step is the determination of the most salient part of the image using a Winner-Take-All (WTA) neural network; see Figure 1.

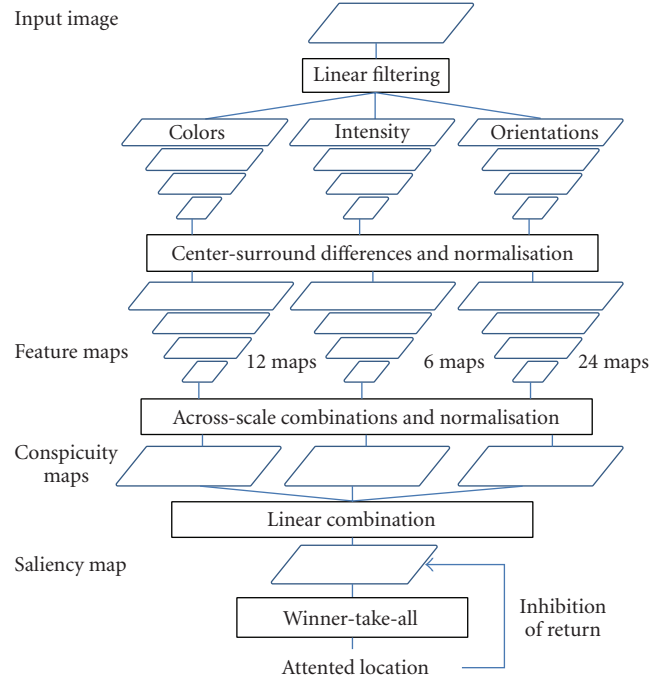


FIGURE 1: Visual attention method.

The initial image is a color image ( $r$  = red,  $b$  = blue,  $g$  = green channels), from which we can extract the intensity image using the following equation:

$$I = \frac{(r + g + b)}{3}, \quad (1)$$

and four color channels are created:

$$\begin{aligned} R &= r - \frac{(g + b)}{2}, \\ G &= g - \frac{(r + b)}{2}, \\ B &= b - \frac{(g + r)}{2}, \end{aligned} \quad (2)$$

$$Y = \frac{(r + g)}{2} - \frac{|r - g|}{2} - b.$$

One Gaussian pyramid  $I(q)$  is estimated from intensity image  $I$ , where  $q \in [0 \dots 8]$  is the scale. Four Gaussian pyramids  $R(q)$ ,  $G(q)$ ,  $B(q)$ , and  $Y(q)$  are created from these color channels. From  $I$ , four orientation-selective pyramids  $O(q, \theta)$  are also created using Gabor filtering at  $\theta = 0, 45, 90,$  and  $135$  deg. Feature maps for each pyramid are calculated using center-surround operation as difference (pixel by pixel) between fine scale  $q_f = \{2, 3, 4\}$  and coarse scale  $q_c = \{q_f + \{3, 4\}\}$

$$\begin{aligned} \mathcal{I}(q_c, q_f) &= |I(q_c) \ominus I(q_f)|, \\ \mathcal{O}(q_c, q_f, \theta) &= |O(q_c, \theta) \ominus O(q_f, \theta)|, \end{aligned} \quad (3)$$

where  $\Theta$  is the center-surround operation. For the center-surround operation the difference between maps at different scale is obtained by interpolation at finer scale followed by a pixel by pixel subtraction. For color channels, the feature maps are calculated for Green/Red and Blue/Yellow opponency

$$\begin{aligned}\mathcal{R}\mathcal{G}(q_c, q_f) &= \left| (R(q_c) - G(q_c)) \Theta (G(q_f) - R(q_f)) \right|, \\ \mathcal{B}\mathcal{Y}(q_c, q_f) &= \left| (B(q_c) - Y(q_c)) \Theta (Y(q_f) - B(q_f)) \right|.\end{aligned}\quad (4)$$

Forty four feature maps are finally obtained, 6 for the intensity, 12 for the color, and 24 for orientation. Then, we calculate the conspicuity maps  $(\overline{\mathcal{O}}(q), \overline{\mathcal{I}}(q), \overline{\mathcal{R}\mathcal{G}}(q), \overline{\mathcal{B}\mathcal{Y}}(q))$  by linear combination of the feature maps at scale  $q = 4$ , followed by a normalization between 0 and 1. Compared to the classical approach, the normalization related to the maximum of each feature map is not applied here; see [8]. This normalization is not needed since our subsequent processing step is invariant to scale; see Section 2.5. Finally, we calculate the conspicuity color map  $\overline{\mathcal{C}}$  and the saliency map  $\mathcal{S}$ :

$$\begin{aligned}\overline{\mathcal{C}}(q) &= \frac{1}{2}(\overline{\mathcal{R}\mathcal{G}}(q) + \overline{\mathcal{B}\mathcal{Y}}(q)), \\ \mathcal{S}(q) &= \frac{1}{3}(\overline{\mathcal{O}}(q) + \overline{\mathcal{I}}(q) + \overline{\mathcal{C}}(q)).\end{aligned}\quad (5)$$

**2.2. Detection of the Most Salient Part.** In this step we select only the most salient pixel of the saliency map. The WTA neural network is not used since we have only one object to detect:

$$P_* = \arg \max_{(x,y)} \{ \mathcal{S}_{(x,y)}(q) \}.\quad (6)$$

**2.3. Classification.** A classic Linear Discriminant Analysis (LDA) is used to classify the area around the maximum detected in the previous step (Section 2.2), as “object” (1) or “not object” (0). As this is a supervised approach, the parameters of the model are estimated using reference videos. An object is considered detected if

$$(M_0 - M_1)^T X + \frac{1}{2}(M_1^T M_1 - M_0^T M_0) > \ln \frac{P_1}{P_0},\quad (7)$$

where  $P_i$ ,  $i = 0, 1$  is the prior probability for classes 0, and 1,  $M_i$  the vector of expected features values and  $X$  the feature vector.

**2.4. Initialization by Visual Attention.** The idea is to use the saliency map to initialize the active contour. Generally, in our application, the underwater image is composed of only one object in a dark and noisy background. The visual attention scheme is well-adapted to these particular images where a single region of interest is contrasted with a background. We propose to use only the most salient part of the image to initialize the position of the active contour.

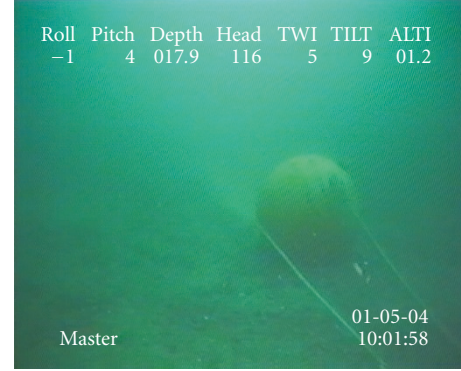


FIGURE 2: Initial image.

**2.5. Adaptive Region Descriptor for Active Contour.** Usually the choice of the region descriptor depends on the application and is fixed a priori. In this work, we introduce a data-driven region descriptor. The information to choose the descriptor will be deduced from the saliency map. Once we have detected the most salient part, we can select the most informative conspicuity map. Since we are dealing with illuminated objects superposed on a dark background, an informative map would mean that the pixels are easy classifiable into either a class object or background (nonobject). In other words, the probability density function of the pixel values would ideally be unimodal with a mode at the background value ( $I \approx 0$ ) and the object constituting the distribution tail ( $I \approx 1$ ), whereas a noninformative pixel map would result in a more uniform distribution of its pixel values. A well-adapted criterion to differentiate an informative from a noninformative pixel map is the entropy of the pixel map, which is under some general assumptions related to the more easily calculable normalized kurtosis of a map  $C$  [14, 15]. The pixel map of choice  $C^*$  is

$$C^* = \arg \max_C \frac{E\{(C(x,y) - \mu_C)^4\}}{(E\{(C(x,y) - \mu_C)^2\})^2},\quad (8)$$

where  $\mu_C = E\{C(x,y)\}$ , and  $C(x,y)$  is the pixel intensity for a pixel belonging to the conspicuity map  $C \in \{\overline{\mathcal{I}}, \overline{\mathcal{O}}, \overline{\mathcal{R}\mathcal{G}}, \overline{\mathcal{B}\mathcal{Y}}\}$ . In [16] the one map is found that contributes most to the activity at the most salient location  $(x_w, y_w)$  looking back at the conspicuity maps. Examining the feature maps that gave rise to the conspicuity map  $C_{k_w}$  with  $k \in \{I, C, O\}$  leads to the one that contributes most to its activity at the winning location. The *winning* feature map is segmented using region growing and adaptive thresholding. In Section 3.4, we compare this approach with our approach and show the improvement.

**2.6. Active Contour.** A parametric snake is a curve  $\Gamma(s) = [x(s), y(s)]$ ,  $s \in [0, 1]$ , which minimizes an energy functional

$$E[\Gamma(s)] = E_{\text{int}}[\Gamma(s)] + E_{\text{ext}}[\Gamma(s)],\quad (9)$$

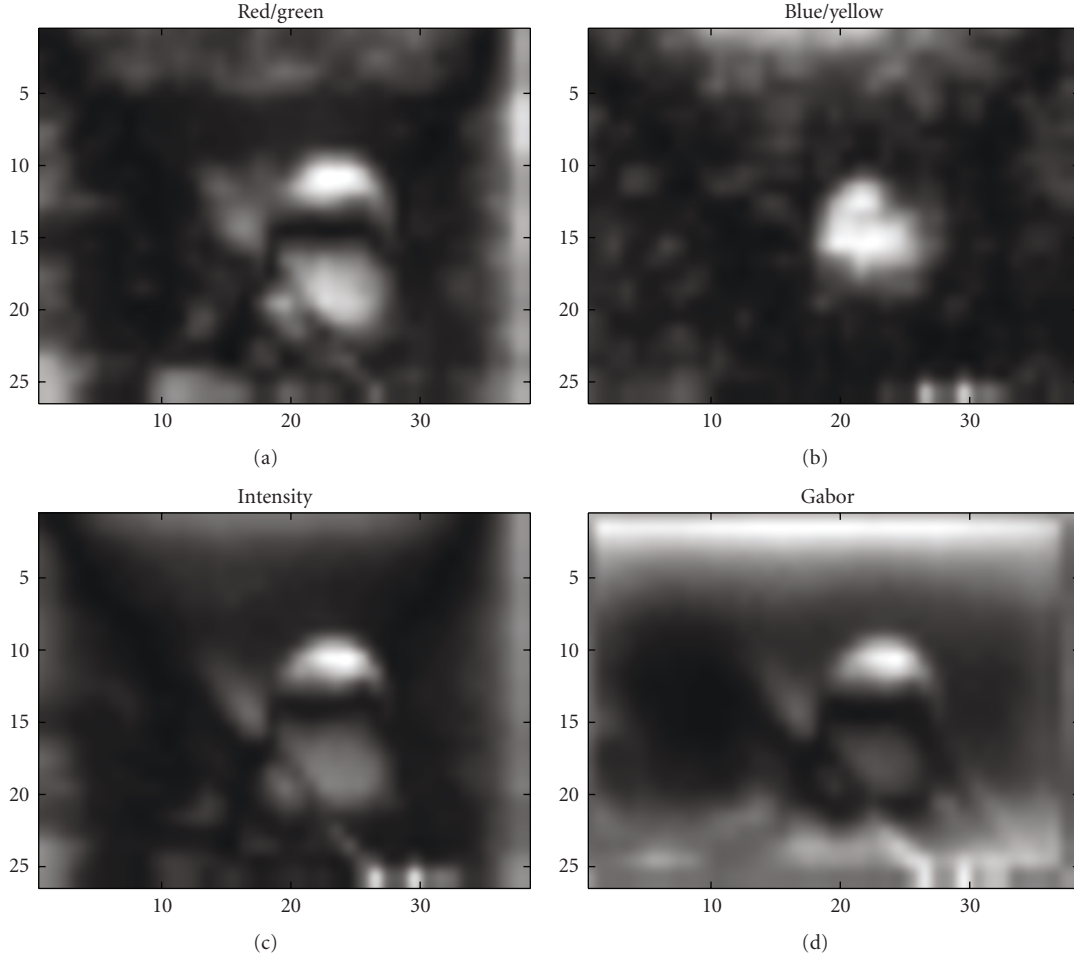


FIGURE 3: Conspicuity maps.

by moving through the spatial domain of an image. The total energy consists of two energies: internal  $E_{\text{int}}$  and external  $E_{\text{ext}}$ . The internal energy preserves the smoothness and continuity. The external energy is derived from the information in the image  $I$ . A typical energy function for a snake using image gradients is given as [11]

$$E[\Gamma(s)] = \oint_{\Gamma(s)} \left\{ \frac{1}{2} (\alpha |\Gamma_s|^2 + \beta |\Gamma_{ss}|^2) - \lambda |\vec{\nabla} I \cdot \vec{\nabla} I| \right\} ds, \quad (10)$$

where  $\Gamma_s = (\dot{x}(s), \dot{y}(s))$ ,  $\Gamma_{ss} = (\ddot{x}(s), \ddot{y}(s))$ , and  $\alpha, \beta, \lambda$  are positive coefficients. For a snake using the region information in the image based on Minimum Description Length (MDL) we have [12]

$$E[\Gamma, \{\alpha_i\}] = \sum_{i=1}^N \left\{ \frac{\mu}{2} \int_{\delta R_i} ds - \log P(\{I_{(x,y)} : (x,y) \in R_i\} | \alpha_i) + \lambda \right\}, \quad (11)$$

where  $R_i$  is the segmented region  $i$ ,  $\mu$  is the code length for unit arc length,  $N$  is the number of regions, and  $\alpha_i$  the

parameters of the distribution describing the region  $i$ .  $\lambda$  is the code length needed to describe the distribution and code system for region  $R_i$ . The external force is processed using a window  $W_{(x,y)}$  of  $m$  pixels around the control point. Then the equation is

$$E[\Gamma, \{\alpha_i\}] = \sum_{i=1}^N \left\{ \frac{\mu}{2} \int_{\delta R_i} ds - \int \int_{R_i} \frac{1}{m} \times \int \int_{W_{(x,y)}} \log P(I_{(u,v)} | \alpha_i) du dv dx dy + \lambda \right\}. \quad (12)$$

If we assume that each pixel  $(x, y)$  has a multiplicative mixture distribution, we obtain

$$E[\Gamma, \{\alpha_i\}] = \mu \oint_{\Gamma} ds - \int \int_R \sum_{i=1}^N \pi_i(u,v) \log P(I_{(u,v)} | \alpha_i) du dv + \lambda N, \quad (13)$$

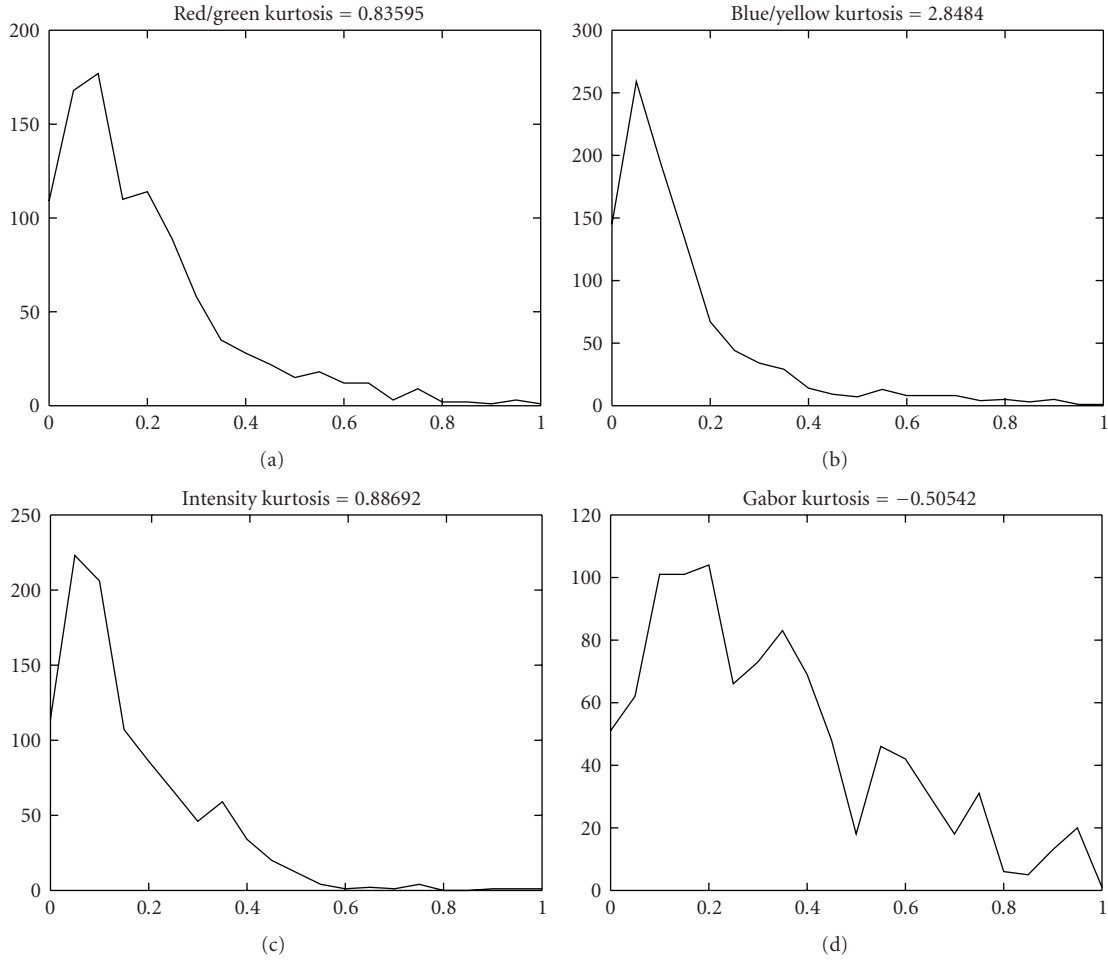


FIGURE 4: The histograms and the estimated kurtosis for the conspicuity maps.

where

$$\Gamma = \bigcup_{i=1}^N \partial R_i,$$

$$\pi_{i(u,v)} = \frac{||R_i \cap W(u,v)||}{m}, \quad (14)$$

$$\sum_{i=1}^N \pi_{i(u,v)} = 1.$$

The motion equation for point  $\vec{v} = (x, y)$  is

$$\frac{d\vec{v}}{dt} = -\frac{\partial E[\Gamma, \{\alpha_i\}]}{\partial \vec{v}}. \quad (15)$$

Using (11)

$$\frac{d\vec{v}}{dt} = \sum_{k \in Q(\vec{v})} \left\{ -\frac{\mu}{2} \kappa_{k(\vec{v})} \vec{n}_{k(\vec{v})} + \log P(I_{(\vec{v})} | \alpha_k) \vec{n}_{k(\vec{v})} \right\}, \quad (16)$$

where  $Q(\vec{v}) = \{k | \vec{v} \text{ lies on } \Gamma_k\}$ ,  $\kappa_{k(\vec{v})}$  is the curvature,  $\vec{n}_{k(\vec{v})}$  is the unit normal to  $\Gamma_k$  at point  $\vec{v}$ .

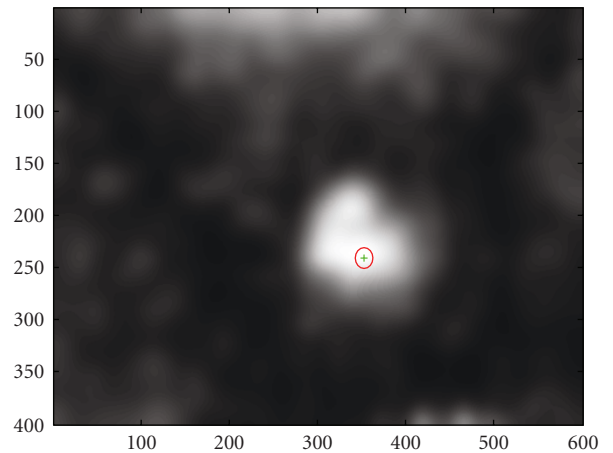


FIGURE 5: Maximum detected: green cross and the initial snake in red. The image is presented at initial scale ( $s = 0$ ) after up sampling operation on Blue/Yellow conspicuity map.

As the objects of interest are simple manufactured objects, we propose to add a constraint on the snake form.

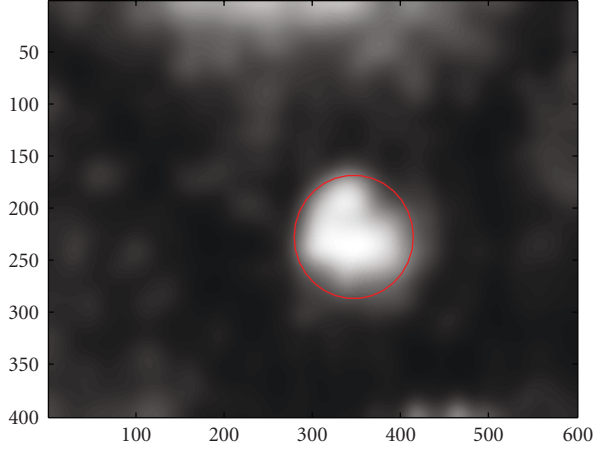


FIGURE 6: Segmentation using the conspicuity map.

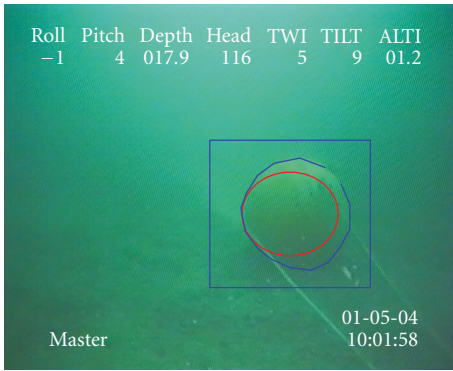


FIGURE 7: Segmentation presented on the real image. In red the obtained segmentation, in blue the ground true. The blue square is the area of interest which can be used for the recognition step.

This allows a smooth segmentation of the image. The constrained form is ellipsoidal. Related works in literature already propose to add to the energy a penalty function increasing with the distance of the curve to an ellipse [17, 18]. We propose to estimate directly the ellipse parameters and not the control points of the curve as in [19], but allowing for a rotation of the ellipse as in [18]. Using the parametric expression of ellipses as function of  $\theta$ , we have

$$\begin{aligned} \Gamma(s, t) &= \begin{pmatrix} x(\theta, t) \\ y(\theta, t) \end{pmatrix} \\ &= \begin{pmatrix} x_c(t) + a(t) \cos(\theta) \cos(\phi(t)) - b(t) \sin(\theta) \sin(\phi(t)) \\ y_c(t) + a(t) \cos(\theta) \sin(\phi(t)) + b(t) \sin(\theta) \cos(\phi(t)) \end{pmatrix}, \end{aligned} \quad (17)$$

where  $(x_c(t), y_c(t))$  is the ellipse center,  $a$  and  $b$  the half-length of the ellipse axes, and  $\phi(t)$  the angle between the  $x$ -axis and the major axis of the ellipse. As we add a constraint

FIGURE 8: Segmentation of a mine in basin without water.  $C_{\text{good}} = 0.96$  and  $C_{\text{false}} = 0.02$ .FIGURE 9: Segmentation of a mine in basin.  $C_{\text{good}} = 0.86$  and  $C_{\text{false}} = 10^{-3}$ .

of form, we can eliminate, respectively, the internal energy and the internal force in (9) and (15). Then (15) becomes

$$\frac{d\vec{v}}{dt} = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} -\frac{\partial E_{\text{ext}}[\Gamma, \{\alpha_i\}]}{\partial x} \\ -\frac{\partial E_{\text{ext}}[\Gamma, \{\alpha_i\}]}{\partial y} \end{pmatrix}. \quad (18)$$

Then, we express the evolution of the ellipse parameters if we consider a discretized curve  $\Gamma$  to be an  $M$ -tuple  $(\vec{v}_1, \vec{v}_2, \dots, \vec{v}_N)$  of points:

$$F_{x_{\vec{v}_i}} = \log P(I_{(\vec{v}_i)} | \alpha_k) \langle \vec{n}_{\vec{v}_i}, \vec{x} \rangle,$$

$$F_{y_{\vec{v}_i}} = \log P(I_{(\vec{v}_i)} | \alpha_k) \langle \vec{n}_{\vec{v}_i}, \vec{y} \rangle,$$

$$\frac{dx_c(t)}{dt} = \frac{k_{x_c}}{N} \sum_{i=1}^N F_{x_{\vec{v}_i}},$$

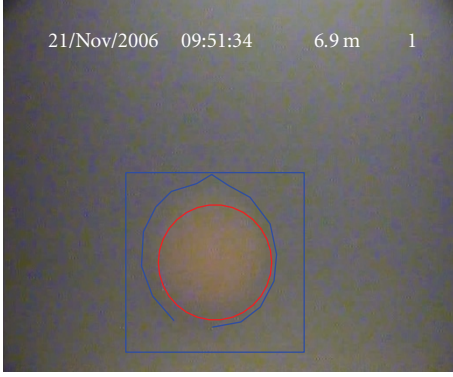


FIGURE 10: Segmentation of a circular mine in basin.  $C_{\text{good}} = 0.64$  and  $C_{\text{false}} = 0$ .



FIGURE 12: Segmentation of a mine at sea.  $C_{\text{good}} = 0.66$  and  $C_{\text{false}} = 0.005$ .

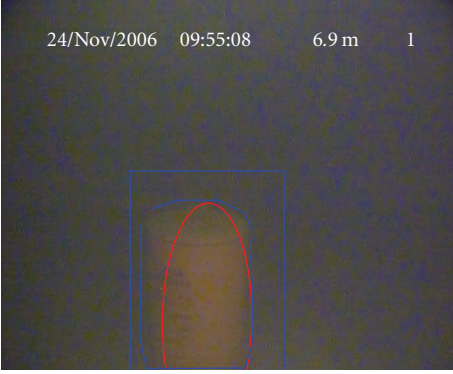


FIGURE 11: Segmentation of a mine in basin.  $C_{\text{good}} = 0.65$  and  $C_{\text{false}} = 0$ .



FIGURE 13: Segmentation of a mine at sea.  $C_{\text{good}} = 0.57$  and  $C_{\text{false}} = 7.10^{-3}$ .

$$\begin{aligned} \frac{dy_c(t)}{dt} &= \frac{k_{y_c}}{N} \sum_{i=1}^N F_{y_{\vec{v}_i}}, \\ \frac{da(t)}{dt} &= \frac{k_a}{N} \sum_{i=1}^N \cos(\theta_{\vec{v}_i}) (\cos(\phi(t)) F_{x_{\vec{v}_i}} + \sin(\phi(t)) F_{y_{\vec{v}_i}}), \\ \frac{db(t)}{dt} &= \frac{k_b}{N} \sum_{i=1}^N \sin(\theta_{\vec{v}_i}) (-\sin(\phi(t)) F_{x_{\vec{v}_i}} + \cos(\phi(t)) F_{y_{\vec{v}_i}}), \end{aligned} \quad (19)$$

$k_{x_c}$ ,  $k_{y_c}$ ,  $k_a$ , and  $k_b$  are coefficients of ponderation controlling the speed of the active contour. The rotation around the center of the ellipse can be found by calculating the angular momentum for a solid object as with Newton's second law. For an object with a moment of inertia  $J$  on which a torque  $\tau$  is exercised, we have

$$\begin{aligned} J \frac{d^2\phi}{dt^2} &= \tau, \\ \frac{d\phi}{dt} &= \frac{\tau}{J} \cdot \Delta t. \end{aligned} \quad (20)$$

Since for an ellipse with uniform density and mass  $M$  the moment of inertia is

$$J = \frac{M(a^2 + b^2)}{4}, \quad (21)$$

and the torque in any point  $\vec{v}$  can be calculated as

$$\tau = R_{\vec{v}} \cdot F_{\perp_{\vec{v}}}, \quad (22)$$

we obtain

$$\frac{d\phi}{dt} = \frac{R_{\vec{v}} \cdot F_{\perp_{\vec{v}}}}{(1/4)M(a^2 + b^2)} \Delta t. \quad (23)$$

with

$$\begin{aligned} F_{\perp_{\vec{v}}} &= F_{y_{\vec{v}}} \cos(\theta_{\vec{v}} + \phi) - F_{x_{\vec{v}}} \sin(\theta_{\vec{v}} + \phi), \\ R_{\vec{v}} &= \sqrt{(a(t) \cos(\theta_{\vec{v}}))^2 + (b(t) \sin(\theta_{\vec{v}}))^2}, \end{aligned} \quad (24)$$

where  $F_{\perp_{\vec{v}}}$  is the force at the point  $\vec{v}$  normal to  $R$ . We can generalise for forces applied along the contour

$$\frac{d\phi(t)}{dt} = \frac{k_{\phi}}{N} \sum_{i=1}^N \frac{R_{\vec{v}_i} \cdot F_{\perp_{\vec{v}_i}}}{a(t)^2 + b(t)^2}, \quad (25)$$

where  $k_{\phi}$  is a coefficient controlling the speed of evolution of the contour including the constants  $\Delta t$  and  $4/M$ .

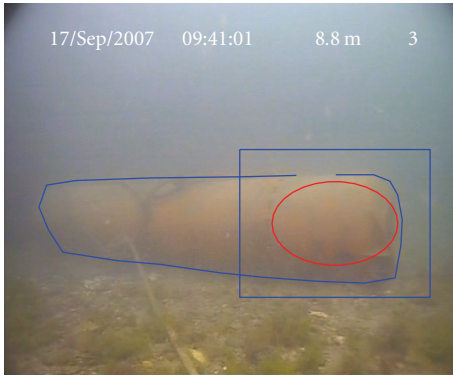


FIGURE 14: Segmentation of a mine at sea.  $C_{\text{good}} = 0.25$  and  $C_{\text{false}} = 0$ .



FIGURE 15: Segmentation of a mine at sea with pre-processing and segmentation using RGB information.  $C_{\text{good}} = 0.53$  and  $C_{\text{false}} = 0$ .



FIGURE 16: Segmentation of a mine at sea with pre-processing and segmentation using RGB information.  $C_{\text{good}} = 0.2$  and  $C_{\text{false}} = 0$ .

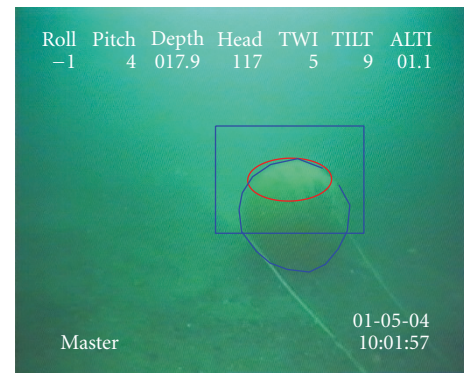


FIGURE 17: Segmentation of a mine using the feature map selected by maximum activity.  $C_{\text{good}} = 0.26$  and  $C_{\text{false}} = 0.003$ .

### 3. Results

In this section, we present first some criterion of performance and then we show the results obtained on real images.

**3.1. Criterion.** The criterion expresses the segmentation quality:

$$\begin{aligned} C_{\text{good}} &= \frac{\text{card}\{\Omega_{\text{in}} \cap \Omega_o\}}{\text{card}\{\Omega_o\}}, \\ C_{\text{false}} &= \frac{\text{card}\{\Omega_{\text{in}} \cap \Omega_b\}}{\text{card}\{\Omega_b\}}, \end{aligned} \quad (26)$$

where  $\Omega_{\text{in}}$  is the internal region of the snake,  $\Omega_o$  the region of the object to detect,  $\Omega_b$  the background region and  $\text{card}(\Omega)$  defines the cardinality (the number of elements in set  $\Omega$ ).

**3.2. Illustration of the Method on Real Image.** The method has been tested on real images of underwater mines (the information contained in this publication are using ac-rov data recorded at Lanvéoc (France) by the GESMA (Groupe d'Etudes Sous-Marines de l'Atlantique)). Since we have only a partial knowledge on the recording conditions, we cannot use an illumination model or any a priori information to restore images. The first result presented is a mine in seawater environment with low visibility, some acquisition noise and compression noise, see Figure 2. We select the best

conspicuity map, Figure 3, using the kurtosis criterion (8). Figure 4 shows the histograms for the conspicuity maps and the calculated kurtosis. The maximum kurtosis for this image is obtained for the Blue/Yellow map. Once the maximum on the saliency map is detected, we initialize an active contour on this position; see Figure 5. Then after convergence of the snake, we obtain the segmentation of the object displayed in Figures 6 and 7. For this image we obtain  $C_{\text{good}} = 0.67$  and  $C_{\text{false}} = 10^{-4}$ .

**3.3. Results on Different Images.** In this section, we present the segmentation results on a set of underwater images. The images have been recorded with different conditions of depth, illumination, noise, acquisition system, camera, and so forth.

The method is robust to the acquisition conditions. Even under bad lighting conditions, the mine is detected, for an example without enough light; see Figure 11. An example with artificial light is shown in Figure 9 and another with natural light Figure 12. The misclassification results are very low, from 0% to 2%; see Figures 8 and 10; and we detect a minimum of 25% of the object; see Figure 14.

**3.4. Comparison with Other Methods.** In this section we compare our method with two other alternative approaches: the approach proposed by [6], also developed for the same images, and the approach proposed by [16] using



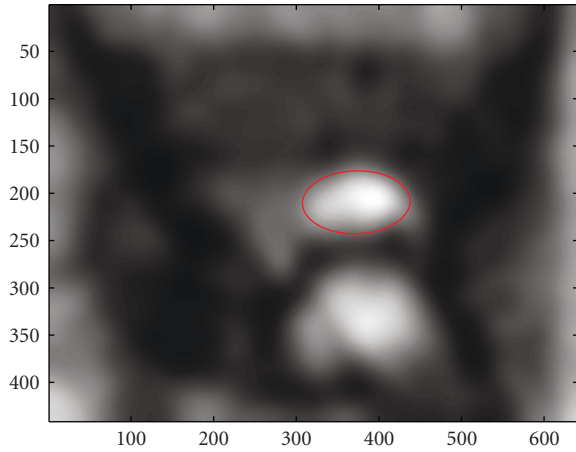


FIGURE 18: Selected feature map ( $\mathcal{R}\mathcal{G}(3,7)$ ) by maximum activity and segmentation.

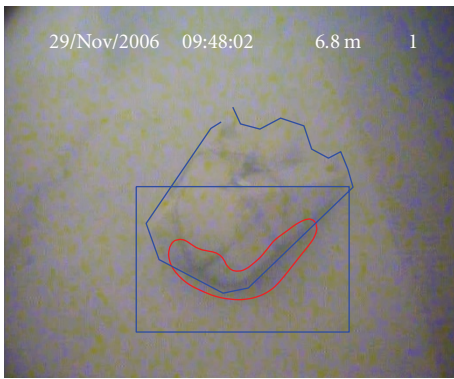


FIGURE 19: Segmentation of a gray mine at sea.  $C_{\text{good}} = 0.25$  and  $C_{\text{false}} = 0.002$ .

the maximum of activity to select the feature map; see Section 2.5.

**3.4.1. First Method.** The image is corrected using the following pre-processing and the segmentation is applied using RGB features:

- (i) removing aliasing effect due to digital conversion of the images,
- (ii) converting color space from RGB to YCbCr,
- (iii) correction of nonuniform illumination using homomorphic filtering,
- (iv) wavelet denoising,
- (v) anisotropic filtering to improve image segmentation,
- (vi) adjusting image intensity,
- (vii) converting from YCbCr to RGB,
- (viii) equalizing color mean.

We obtain similar results with the method proposed in [6] for objects without shadows; see Figures 8 and 15. However, our method is more robust to shadow; see, for example,

Figure 16, and compare with Figure 7. For some images the segmentation using active contour cannot converge as the contours are blurred and the contrast is low between the object and the background; see Figures 12 and 13. Using the conspicuity map, we can reduce this effect.

**3.4.2. Second Method.** We implement the method to select the feature map based on the maximum activity and use active contour segmentation. As in the previous approach, we obtain very similar results for objects with a uniform illumination and no shadows. But, as illustrated in Figures 17 and 18 compared to the results presented in Figures 6 and 7, this method focuses on the illuminated part of the object. The criterion to select the map is local compared to our criterion (8) and the selection will thus focus on detail.

**3.5. Results Obtained during the Evaluation of the Project TOPVISION.** The results next were obtained by our method applied on test videos for the project TOPVISION (Operational Trials of UnderWater Videos for Identification of Harmfull Article) [20]. Remark that these videos were never used during the development of the method. The project is composed of 4 steps.

- (i) object detection,
- (ii) position finding,
- (iii) characterisation,
- (iv) identification.

For the detection step, we only try to detect if an object is present in the image. We use the output of the LDA classifier for the detection step; see Section 2.3. For the position finding, the position is correct if we are inside the object. We are using the maximum of the saliency map as position; Section 2.4. Only the two first steps have been evaluated on 11 videos ( $\approx 20000$  images), and we have obtained 64.77% of good object detection (and 2.82% of false alarms) and 85.28% of position finding. The method presented in [6] has been tested on the same images for the detection step and obtains 56.55% of good object detection and 0.79% of false alarms. The results are more robust in terms of false alarm but the percentage of detection is inferior. The method uses the detection of geometric forms to determine the object presence. This needs more computation than our method because, after the preprocessing step (see Section 3.4), it extracts lines and circles.

**3.6. Limits of the Method.** The method provides satisfactory results in numerous cases, but when the object is of a color indistinguishable from that of the background it presents some limitation illustrated in Figure 19. It is difficult for the method to find the object. To obtain an acceptable result we had to remove the ellipsoidal constraint.

## 4. Conclusion

In this paper we have presented a fully automated method to detect and segment manufactured objects in underwater

images. The method uses two well-known approaches with some original adaptations: the visual attention scheme, and the active contours. We have described the successive steps of the method and presented some results on real images. The method presents good performance even with noisy images and is robust to different acquisition conditions (illumination, camera settings, shadow, etc). The limitations of the approach have also been evoked. Essentially to ensure good results the object to segment must be uniform and sufficiently contrasted with respect to the background. For the future, we are developing a method to track these objects in videos based on the conspicuity map and particle filtering. Currently, we can track the object using the estimated snake in the previous frame as initialization in the actual frame.

## Acknowledgment

The authors would like to thank V. Zarzoso for the interesting discussions on this work.

## References

- [1] S. Bazeille, L. Jaulin, and I. Quidu, "Identification of underwater manmade object using a colour criterion," in *Proceedings of the 4th International Conference on Bio-Acoustics*, Proceedings of the Institute of Acoustics, Loughborough, Uk, April 2007.
- [2] D. Edgington, D. Walther, D. Cline, R. Sherlock, and C. Koch, "Detecting and tracking animals in underwater video using a neuromorphic saliency-based attention system," in *Proceedings of the American Society of Limnology and Oceanography (ASLO) Summer Meeting*, Savannah, Ga, USA, June 2004.
- [3] A. Olmos and E. Trucco, "Detecting man-made objects in unconstrained subsea videos," in *Proceedings of the British Machine Vision Conference*, pp. 517–526, 2002.
- [4] D. Walther, D. R. Edgington, and C. Koch, "Detection and tracking of objects in underwater video," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 544–549, 2004.
- [5] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Proceedings of the IEEE Computer Society Conference on Vision and Pattern Recognition*, vol. 1, pp. 536–543, 2004.
- [6] S. Bazeille, I. Quidu, L. Jaulin, and J.-P. Malkasse, "Automatic underwater image pre-preprocessing," in *Proceedings of the SEA TECH WEEK Characterisation du Milieu Marin*, Brest, France, October 2006.
- [7] E. Trucco and A. T. Olmos-Antillon, "Self-tuning underwater image restoration," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 2, pp. 511–519, 2006.
- [8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, Calif, USA, 1990.
- [10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [11] C. Xu and J. Prince, "Gradient vector flow: a new external force for snakes," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 66–71, San Juan, Puerto Rico, USA, June 1997.
- [12] S. C. Zhu and A. Yuille, "Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.
- [13] J. A. Sethian, "Evolution, implementation, and application of level set and fast marching methods for advancing fronts," *Journal of Computational Physics*, vol. 169, no. 2, pp. 503–555, 2001.
- [14] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalised ica algorithms," in *Advances in Neural Information Processing Systems*, 1996.
- [15] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. 23, no. 9, pp. 881–890, 1974.
- [16] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Computer Vision and Image Understanding*, vol. 100, no. 1-2, pp. 41–63, 2005.
- [17] N. Ray, S. T. Acton, and K. Ley, "Tracking leukocytes in vivo with shape and size constrained active contours," *IEEE Transactions on Medical Imaging*, vol. 21, no. 10, pp. 1222–1235, 2002.
- [18] C. Pluempitwiriwawej, J. M. F. Moura, Y.-J. L. Wu, S. Kanno, and C. Ho, "Stochastic active contour for cardiac mr image segmentation," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, vol. 2, pp. 1097–1100, September 2003.
- [19] D. Zugaj and V. Lattuati, "Contours actifs: suivi de structures linéiques déformables. Extension à la résolution de problèmes d'optimisation sous contraintes," in *Proceedings of the Congrès GRETSI 97: Seizième Colloque sur le Traitement du Signal et des Images*, vol. 16, pp. 407–410, Grenoble, France, 1997.
- [20] GESMA, <http://topvision.gesma.fr/articles.php?lng=en&pg=6>.