



**HAL**  
open science

## Detection of floating mines in infrared sequences by multiscale geometric filtering

Dominique Florins, Antoine Manzanera

► **To cite this version:**

Dominique Florins, Antoine Manzanera. Detection of floating mines in infrared sequences by multiscale geometric filtering. SPIE Conference in Defence Security and Sensing: Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XVII, Apr 2012, Baltimore, United States. 10.1117/12.918420 . hal-01119659

**HAL Id: hal-01119659**

**<https://hal.science/hal-01119659v1>**

Submitted on 23 Feb 2015

**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.

# Detection of floating mines in infrared sequences by multiscale geometric filtering

Dominique Florins<sup>a</sup> and Antoine Manzanera<sup>a</sup>

<sup>a</sup>ENSTA ParisTech, 32 Boulevard Victor, Paris, France

## ABSTRACT

Automatic detection of floating mines by passive sensing is of major interest, yet remains a hard problem. In this paper, we propose an algorithm to detect them in infrared sequences, based on their geometry, provided by spatial derivatives. In infrared images, floating mines contrast with the sea due to the difference of emissivity at low incidence angles: they form bright elliptical areas. Using the available data and the geometry of our camera, we first determine the scales of interest, which represent the possible size of mines in number of pixels. Then, we use a temporal and a morphological filter to perform smoothing in the time dimension and contrast enhancement in the space dimensions, at the selected scales, and calculate for every pixel the Hessian matrix, composed of the second order derivatives, which are estimated in the classical scale-space framework, by convolving the image with derivatives of Gaussian. Based on the eigenvalues of the Hessian matrix, representing the curvatures along the principal directions of the image, we define two parameters describing the eccentricity of an elliptical area and the contrast with sea, and propose a measure of “mine-likeness” that will be high for bright elliptical regions with selected eccentricity. At the end, we only retain pixels with high mine-likeness, stable in time, as potential mines. Using a dataset of 10 sequences with ground truth, we evaluated the performance and stability of our algorithm, and obtained a precision between 80% and 100%, and a per-frame recall between 30% and 100%, depending on the difficulty of the scenarios.

**Keywords:** infrared images, passive mine detection, multiscale derivatives, ellipticity measure

## 1. INTRODUCTION

The automatic detection of floating mines is a problem of crucial importance for the security of naval ships. Doing it by passive sensing using camera is an objective of major interest, for security and cost reasons, but it remains a difficult problem, because of the limited visibility of the mines, and its variability of appearance induced by the waves and the changing weather conditions.

Different video based techniques have been proposed, see the review of Borghgraef *et al*<sup>1</sup> for a detailed study. Those methods generally aim to distinguish the mine from its environment using motion criteria. We can classify them into two categories: the pixel-based background subtraction algorithms, e.g. ViBe,<sup>2</sup> and the behaviour subtraction methods.<sup>3</sup> The background subtraction is a change detection algorithms in which a per-pixel statistical model is estimated for the environment (the background) of the mine. This background model is then used to classify every pixel as foreground or background. Only the foreground pixels are retained, providing a binary mask which can then be tracked from one frame to the other and classified on geometric or cinematic criteria. On the other hand, the behaviour subtraction principle is to monitor the level of scene activity in a sliding time window, and to compare this activity to a background behaviour model corresponding to the normal behaviour of the sea, and derived from training sequences.

Our approach differs significantly in the sense that the core of the detection is geometric. The temporal aspects (motion part) only appears in pre- or post-processing complements. In infrared images, an emerging floating mine contrasts with the sea due to the difference of emissivity at low incidence angles: it forms a bright elliptical area. In this paper, we present an algorithm to detect automatically the presence of floating mines in infrared images based on geometrical characterisation and spatiotemporal filtering. We evaluate the detection

---

Further author information:

Dominique Florins: E-mail: dominique.florins@ensta-paristech.fr

Antoine Manzanera: E-mail: antoine.manzanera@ensta-paristech.fr

performance of our algorithm using a dataset of 10 sequences of variable difficulties, showing that it can provide significant assistance in an operative detection system.

Our algorithm is designed to work in a system installed on board, with a camera in the mast pointing forward the ship to detect floating mines. In Section 2, we present the geometrical and physical constraints of the scene that are used to justify the approach and define its main parameters. Section 3 gives the mathematical background used in the algorithm, which is presented in details in Section 4. Section 5 presents the results and the evaluation of the method.

## 2. GEOMETRY AND PHYSICS OF THE SCENE

In ship-mounted detection systems, the camera is usually placed 10 meters above the sea surface, in the mast of a ship, pointing forward to detect floating mines at a distance between 500 and 1000 meters. This leads to very sharp observation angles of  $1^\circ$  or less. Floating mines have different shapes and sizes, and are mainly submerged (60% to 80%), resulting in a small observable angle. As the detection algorithm is based on the geometry, the size of the mine in number of pixels is needed, in particular the radii of the outer circle  $\sigma_{out}$  and inner circle  $\sigma_{in}$  of the mines. This is estimated using available data of mines and the parameters of the camera. At a distance between 500 and 1000 meters, most mines are a few pixels wide.

At low incidence angles, the contrast between water and mines is important in infrared images. When the sea is flat, the emissivity of the water is close to zero, and the reflectance is almost one, then what will be seen in the image is in fact the thermal emissions of the sky near the horizon, whose temperature is generally much lower than the sea surface.<sup>4</sup> The mines are usually at sea temperature or higher, and thus contrast highly with the sea surface. This is not the case anymore when the sea moves much: some waves can induce observation angles close to  $45^\circ$ , allowing the sea thermal emissions to become observable. This fills the image with large portions of the same intensity as the mine.

## 3. MULTISCALE DIFFERENTIAL GEOMETRY

Our algorithm is mainly based on geometrical characterisation of bright elliptical areas in the image. The differential geometry is particularly well suited to the detection of such features. In the differential framework, an image is seen as a differentiable function  $f$  from  $\mathbb{R}^2$  to  $\mathbb{R}$ . In such framework, the local behaviour of the function is determined by its partial derivatives, as stated by the Taylor expansion:

$$f(x + c_x, y + c_y) = \sum_{k=0}^r \sum_{i=0}^k \binom{k}{i} c_x^{k-i} c_y^i \frac{\partial^k f}{\partial x^{k-i} \partial y^i}(x, y) + o(\|(c_x, c_y)\|^r) \quad (1)$$

where  $(x, y)$  are the coordinates of a pixel in the image and  $(c_x, c_y)$  a small displacement. We will use the following notation:

$$f_{ij} = \frac{\partial^{i+j} f}{\partial x^i \partial y^j}. \quad (2)$$

In digital images, using the classical scale space framework,<sup>5</sup> the partial derivatives are estimated at a specific scale  $\sigma$ , by the mean of a convolution of  $f$  with the derivative of a two-dimensional Gaussian  $G_\sigma$  of standard deviation  $\sigma$ :

$$f_{ij}^\sigma = f \star \frac{\partial^{i+j} G_\sigma}{\partial x^i \partial y^j}. \quad (3)$$

The detection of elliptical regions is a typical use of the second order derivatives. We then estimate the Hessian matrix  $\mathcal{H}^\sigma$  for different scales  $\sigma$ :

$$\mathcal{H}^\sigma = \begin{bmatrix} f_{20}^\sigma & f_{11}^\sigma \\ f_{11}^\sigma & f_{02}^\sigma \end{bmatrix} \quad (4)$$

The Hessian matrix defines the second order (curvature) behaviour of  $f$ , since the second derivative of  $f$  along any couple of directions represented by two unit vectors  $\mathbf{u}$  and  $\mathbf{v}$  can be computed as  $\frac{\partial^2 f}{\partial \mathbf{u} \partial \mathbf{v}} = \mathbf{u}^T \mathcal{H} \mathbf{v}$ . The eigenvalues

of the Hessian matrix, denoted  $\lambda^\sigma$  and  $\Lambda^\sigma$ , with  $\lambda^\sigma < \Lambda^\sigma$ , represent the curvatures along the principal directions of  $f$ , and are given by:

$$\lambda^\sigma = \beta^\sigma - \alpha^\sigma \quad (5)$$

$$\Lambda^\sigma = \beta^\sigma + \alpha^\sigma \quad (6)$$

with :

$$\beta^\sigma = \frac{1}{2}(f_{20}^\sigma + f_{02}^\sigma) \quad (7)$$

$$\alpha^\sigma = \frac{1}{2}\sqrt{(f_{20}^\sigma - f_{02}^\sigma)^2 + 4(f_{11}^\sigma)^2} \quad (8)$$

The second order local geometry of  $f$  is determined by the eigenvalues of the Hessian matrix: if their sign is different, it is a hyperbolic (tubular) region, whereas if their sign is the same, it is an elliptical region. Bright elliptical regions of scale  $\sigma$  in the image are then characterised by  $\Lambda^\sigma < 0$  and  $\lambda^\sigma < 0$ . Those regions are potential mines.

## 4. ALGORITHM

### 4.1 Spatiotemporal filtering

The purpose of the pre-processing step is to eliminate small fluctuations due to the waves, while enhancing the appearance of the mines. To this end, we perform smoothing in the time dimension and contrast enhancement in the space dimensions. The time smoothing is done recursively, using an exponential moving average:

$$M_\alpha(t) = \alpha M_\alpha(t-1) + (1-\alpha)I_t \quad (9)$$

where  $I_t$  is the image at time  $t$  and  $\alpha \in [0,1]$  a time depth parameter, giving more or less importance to new values. The moving average eliminates small fluctuations in the image.

The spatial part is a non linear contrast enhancement of  $M_\alpha(t)$  using the morphological contrast filter<sup>6</sup> :

$$M_\alpha^\sigma = \begin{cases} \delta_{D(\sigma)}(M_\alpha) & \text{if } (\delta_{D(\sigma)}(M_\alpha) - M_\alpha) < (M_\alpha - \varepsilon_{D(\sigma)}(M_\alpha)) \\ \varepsilon_{D(\sigma)}(M_\alpha) & \text{if } (\delta_{D(\sigma)}(M_\alpha) - M_\alpha) \geq (M_\alpha - \varepsilon_{D(\sigma)}(M_\alpha)) \end{cases} \quad (10)$$

where  $\delta_{D(\sigma)}(M_\alpha)$  is the dilation of  $M_\alpha$  by a disk of radius  $\sigma$  and  $\varepsilon_{D(\sigma)}(M_\alpha)$  its erosion by the same structuring element. Regions of size  $\sigma$  ( $\sigma$  is determined by the geometry of the scene and camera) will have their contrast enhanced.

### 4.2 Geometrical detection

The goal of the geometrical detection is to detect all bright elliptical areas in a range of scales between  $\sigma_{in}$  and  $\sigma_{out}$ . It is an adaptation of the vessel detection technique of Frangi *et al*<sup>7</sup> to the elliptical case:

A measure of ellipticity is defined based on the eigenvalues of the Hessian matrix. To do this, 2 parameters are defined. The first one is a measure of the geometrical aspect ratio or eccentricity of the elliptical bright areas:

$$\mathcal{R}^\sigma = \frac{\lambda^\sigma}{\Lambda^\sigma} \quad (11)$$

This parameter is close to 1 for circular areas and nearly zero for flat ellipses.

The second parameter is a measure of the contrast in the image. For the background pixels, the spatial derivatives should be small, whereas bright areas have large values of derivatives and then large eigenvalues. The parameter is thus set to the Frobenius norm of the Hessian matrix, defined as:

$$\mathcal{S}^\sigma = \|\mathcal{H}^\sigma\|_F = \sqrt{(\lambda^\sigma)^2 + (\Lambda^\sigma)^2} \quad (12)$$

Based on these 2 parameters, the following measure of ellipticity can be defined:

$$\nu^\sigma = \begin{cases} 0 & \text{if } \lambda^\sigma \geq 0 \text{ or } \Lambda^\sigma \geq 0 \\ (\exp(-\frac{(\mathcal{R}^\sigma - \rho)^2}{2\xi^2})) (1 - \exp(-\frac{(\mathcal{S}^\sigma)^2}{2\gamma^2})) & \text{otherwise} \end{cases} \quad (13)$$

where  $\xi$  and  $\gamma$  are thresholds to be chosen to highlight potential mines.  $\rho$  is a parameter corresponding to the expected aspect ratio of the mines, according to the geometry of the scene. It was set to  $\frac{4}{5}$  in our experiments. The measure  $\nu^\sigma$  ranges between 0 and 1, being high for elliptical areas of scale  $\sigma$ .

By thresholding these measures, a multiscale detection image can be obtained:

$$det = \bigcup_{\sigma_{in} \leq \sigma \leq \sigma_{out}} \{ \mathbf{x}; \nu^\sigma(\mathbf{x}) > T_{geometry} \} \quad (14)$$

The different steps of the geometrical detection are illustrated in Figure 1.

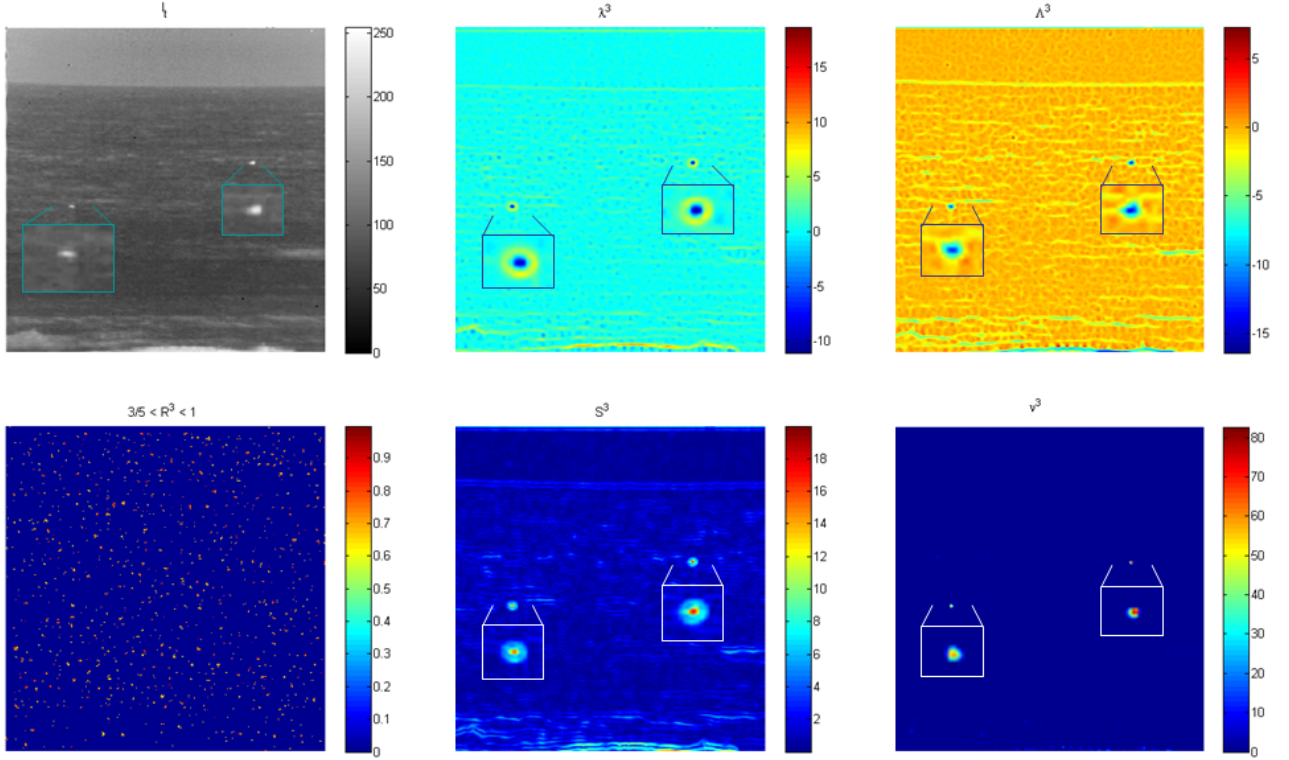


Figure 1. Geometrical detection: [tl] infrared image of 2 floating mines, [tc] small eigenvalue  $\lambda^3$ , [tr] large eigenvalue  $\Lambda^3$ , [bl] geometrical aspect ratio  $\mathcal{R}^3$ , [bc] contrast  $\mathcal{S}^3$ , [br] ellipticity measure  $\nu^3$

### 4.3 Temporal stability

The previous detection image is instantaneous and can be cluttered by many false alarms. We choose to retain only detections which are often present in a spatiotemporal cone of length  $\tau$ . To obtain them, we consider the set  $F = \{det_t, det_{t-1}, \dots, det_{t-\tau+1}\}$ , where  $det_i$  is the detection at time  $i$ .

To allow a certain displacement of the mine from one frame to the other, at every time step  $t$ , we dilate the  $\tau - 1$  past valid detection images preceding the current detection by a disk of radius  $r$ . The dilation of the past detections being repeated from frame to frame, the  $det_{t-\tau+1}$  image is finally dilated by a disk of radius  $r(\tau - 1)$ . We then keep only the pixels which appear a sufficient number of times in the  $\tau$  past dilated detections, as follows:

$$det_t^C = \{\mathbf{x}; \sum_{0 \leq k \leq \tau-1} \delta_{D(kr)}(det_{t-k}(\mathbf{x})) > T_{valid}\} \quad (15)$$

In summary, and as illustrated in Figure 2, we create a cone of length  $\tau$ , and count the number of times that the intersections of this cone with the sequence  $F$  contain a detection. If it is more than  $T_{valid}$  (obviously  $1 \leq T_{valid} \leq \tau$ ), we consider the detection as valid.

Finally to get the complete mine at instant  $t$  if it as been validated, we reconstruct geodesically  $det_t^C$  in  $det_t$ , i.e. we keep the all the complete connected components of  $det_t$  as soon as they contain one valid pixel. The final image of the valid detection is called  $det_{val}(t)$ :

$$det_{val}(t) = \mathcal{R}^{det_t}(det_t^C) \quad (16)$$

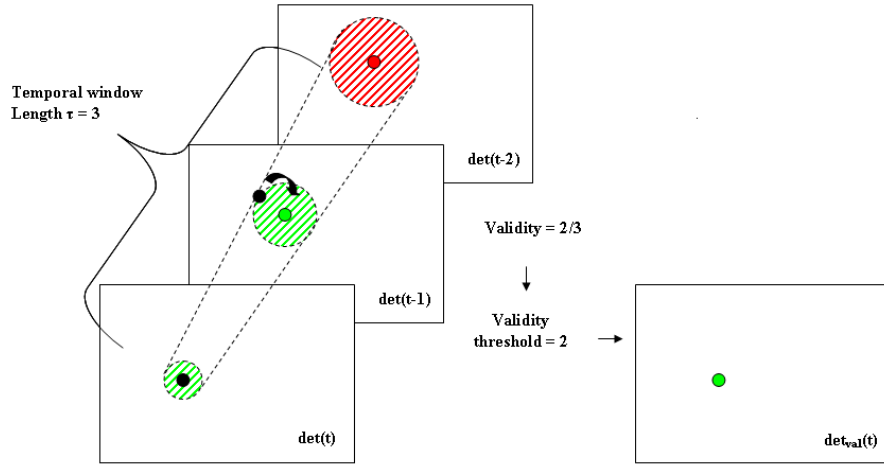


Figure 2. Temporal stability: a spatiotemporal cone is created along current detection at time  $t$ . The image at time  $t - 1$  has a detection in the cone (black). The number of detection in the cone is larger than the threshold of validity. Detection at  $t$  and pixel  $\mathbf{x}$  is then validated.

#### 4.4 Tracking

The final step consists in tracking the valid detection and is illustrated in Figure 3. It is indeed important to keep detecting the mine when it gets occluded, typically by a high wave. To do this, we calculate a moving average of  $det_{val}(t)$  then compute the boundary of the corresponding set:

$$M_{det}(t) = 0.9M_{det}(t - 1) + 0.1det_{val}(t) \quad (17)$$

$$Track(t) = \delta_{D(3)}(M_{det} > T_{track}) - \varepsilon_{D(3)}(M_{det} > T_{track}) \quad (18)$$

## 5. VALIDATION AND RESULTS

### 5.1 Validation protocol

To evaluate the algorithm described above, we used 10 infrared video sequences of floating test targets taken in various environmental conditions, and for which we have a ground truth, i.e. a binary image showing the location of mines. These sequences have been captured in coastal waters using stationary thermal cameras, except the sequence “Ship” which was taken from a frigate board.

For the evaluation, we compared the stable detection and tracking results with the ground truth. So we have two sequences of binary images to compare:  $Track$  which includes all the images of valid detections with their track, and  $Det_{gt}$  which contains all images of the ground truth.

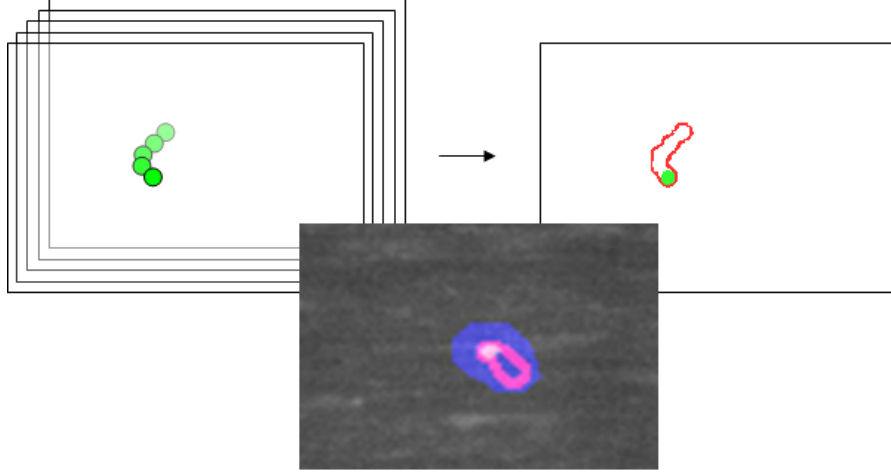


Figure 3. Tracking of the valid detections (principle and exemple).

Figure 4 shows the different steps needed to understand the evaluation protocol. We start by dilating the ground truth mask  $Det_{gt}(t)$  by a radius  $r$  in order to induce a certain tolerance in the localisation of the mine. We compute  $int(t)$  the intersection of the dilated ground truth with the tracks (See Fig. 4(3) and (4)), and reconstruct it in  $Det_{gt}(t)$  (See Fig. 4(5)). The resulting set corresponds to the true positives. We also reconstruct  $int(t)$  in  $Track(t)$  (See Fig. 4(6)), which provides the pixels of  $Track(t)$  considered as true positives. These two sets are denoted by  $TP_{track}(t)$  and  $TP_{gt}(t)$ .

We determine the set of pixels considered as false negatives by  $FN(t) = Det_{gt}(t) - TP_{gt}(t)$  (See Fig. 4(7)), and the set of pixels considered as false positives by  $FP(t) = Track(t) - TP_{track}(t)$  (See Fig. 4(8)).

We then calculate the following quantities:

$$nb\_TP(t) = \#(TP_{track}(t)) + nb\_TP(t-1) \quad (19)$$

$$nb\_FP(t) = \#(FP(t)) + nb\_FP(t-1) \quad (20)$$

$$nb\_FN(t) = \#(FN(t)) + nb\_FN(t-1) \quad (21)$$

where  $\#(X)$  represents the number of connected components of set  $X$ . These measures are cumulated over the whole sequence.

Finally, we calculate the following two validation measures:

$$precision(t) = \frac{nb\_TP(t)}{nb\_TP(t) + nb\_FP(t)}, \quad (22)$$

which indicates the average percentage of mine tracks  $Track(t)$  that are actually mines, and:

$$recall(t) = \frac{nb\_TP(t)}{nb\_TP(t) + nb\_FN(t)}, \quad (23)$$

which indicates the average percentage of mines detected in  $t$ .

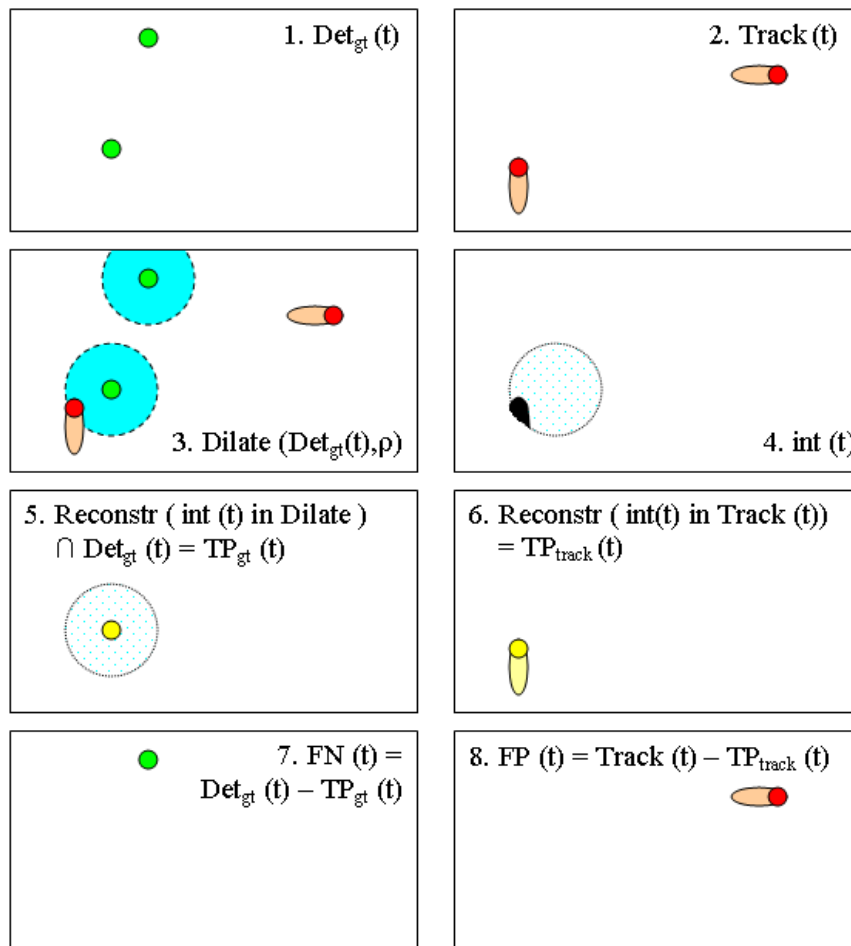


Figure 4. The evaluation protocol.



Table 1. Maximum and minimum precision and recall obtained with the algorithm on all sequences parameterised by the validity threshold.

Sequence	$p_{min}$	$p_{max}$	$r_{min}$	$r_{max}$
Ship	51.93	89.88	21.71	41.32
MW42	28.05	100.00	22.45	30.39
MW15	79.31	100.00	95.00	95.83
LW15	25.95	100.00	100.00	100.00
LW19	40.38	100.00	100.00	100.00
LW16	64.00	100.00	100.00	100.00
LW22	89.09	100.00	100.00	100.00
LW37	71.70	80.54	94.74	96.20
LW05	100.00	100.00	97.16	100.00
MW04	100.00	100.00	97.46	99.17

## 5.2 Results

We worked on a data set of 10 sequences, referred to as their acronyms as shown in the first column of Table 1. *MW* means that the image was taken in the mid infrared band [ $3 - 5\mu m$ ] and *LW* means that the image was taken in the long infrared band [ $8 - 12\mu m$ ]. These sequences were taken from a fixed station on the shore. The sequence *Ship* is the most interesting and more realistic one to us because it was captured by a long infrared band camera installed on a frigate. Anyway the complete set of sequences show varying difficulties, some sequences contain several mines, and for some of them the mines are significantly hidden by the swell.

During the construction of the ground truth masks, we chose to make them always visible, even when the mine is hidden, because the mine is still present, and it allows us to evaluate the performance of the tracking step. However, this has a strong impact on the calculation of the recall, which generally cannot reach 100% when the mine disappears significantly during the sequence. In this case, the recall can reach at most a few percent in addition to the percentage of visibility of the mine in the sequence. For example, in the sequence *MW42*, the mine is visible on 66 images among the 205 frames, that is a visibility of 32%. In the best cases, the tracking allows the recall to slightly exceed this percentage.

Figure 5 shows the ROC curves of the different sequences and Table 1 gives the extreme values obtained with the algorithm on all sequences parameterised by the validity threshold  $T_{valid}$ .

## 6. CONCLUSION

In this paper, we have described the geometry and physics involved in the automatic detection of floating mines in infrared sequences. We have used it to justify our geometrical approach to distinguish potential mines from the background, taking into account the shape of the mines and the contrast with the sea. This approach is combined with a spatiotemporal filter to retain time stable areas. We evaluated our algorithm on different sequences: the high precision and recall allow to envisage the use of this technique as a useful assistance for a human operator, providing reliable warnings to focus his attention.

To increase the level of confidence and use it in fully automated detection systems, this method can be used as a first step to provide regions of interest with sufficiently high confidence to allow higher-level, object-based classification methods. Further improvements involve real time embedded implementation and evaluation of the on-board system.

## ACKNOWLEDGMENTS

This research takes part of the MRN06 project sponsored by the Belgian Ministry of Defence and was conducted at the French Ecole Nationale Supérieure de Techniques Avancées (ENSTA-ParisTech).

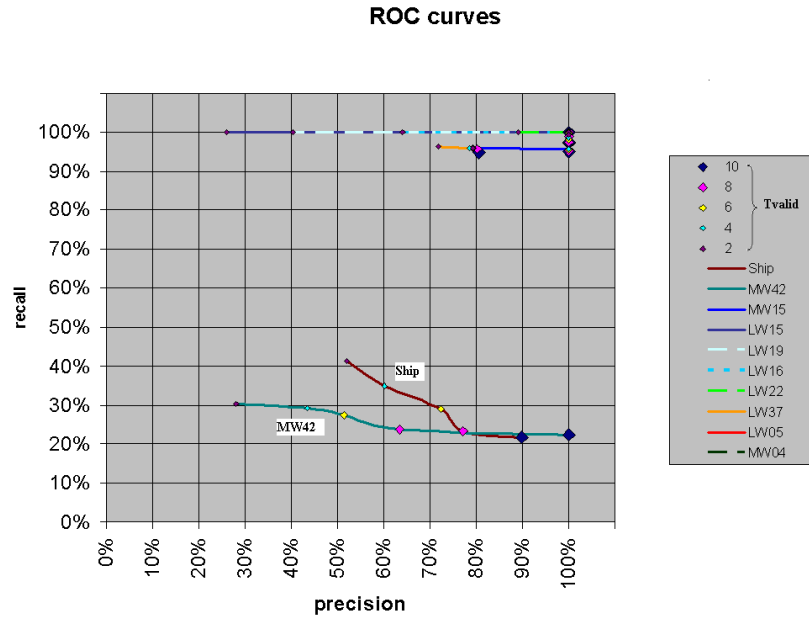


Figure 5. ROC curves of the detection algorithm on all sequences

## REFERENCES

- [1] Borghgraef, A., Barnich, O., Lapierre, F., Van Droogenbroeck, M., Philips, W., and Acheroy, M., “An evaluation of pixel-based methods for the detection of floating objects on the sea surface,” *EURASIP Journal on Advances in Signal Processing* , 1–12 (2010).
- [2] Barnich, O. and Van Droogenbroeck, M., “Vibe: A universal background subtraction algorithm for video sequences,” *IEEE Transactions on Image Processing* **20**, 1709–1724 (June 2011).
- [3] Jodoin, P., Konrad, J., and Saligrama, V., “Modeling background activity for behavior subtraction,” in [*Second ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC'08)*], 1–10 (2008).
- [4] Schweicher, E., “Optoelectronics.” Royal Military Academy Course, Brussels (1992).
- [5] Lindeberg, T., “Feature detection with automatic scale selection,” *Int. J. of Computer Vision* **30**(2), 77–116 (1998).
- [6] Meyer, F. and Serra, J., “Contrasts and activity lattice,” *Signal Processing* **16**(4), 303–317 (1989).
- [7] Frangi, A., Niessen, W., Vincken, K., and Viergever, M., “Multiscale vessel enhancement filtering,” *Medical Image Computing and Computer-Assisted Intervention* **1496**(2), 130–137 (1998).