Uncertainty Fusion based Object Recognition and Tracking in Maritime Scenes using Spatiotemporal Active Contours

Ikhlef Bechar[1], Frederic Bouchara[1], Thibault Lelore[1], Vincente Guis[1] and Michel Grimaldi[2]
[1]*LSIS Laboratory, Toulon University, Toulon, France*
[2]*PROTEE Laboratory, Toulon University, Toulon, France*

Keywords: Airborne Video System, Maritime Surveillance, Vessel Recognition, Dynamic Background, Chromatic Uncertainty, Dynamic Texture Uncertainty, MAP Estimation, Energy Minimization, Spatiotemporal Active Contours.

Abstract: This article addresses the problem of near real time video analysis of a maritime scene using a (moving) airborne RGB video camera in the goal of detecting and eventually recognizing a target maritime vessel. This is a very challenging problem mainly due to the high level of uncertainty of a maritime scene including a dynamic and noisy background, camera's and target's motions, and broad variability of background's *versus* target's appearances. We propose an approach which attempts to combine several types of spatiotemporal uncertainty in a single probabilistic framework. This allows to achieve a likelihood ratio with respect to any possible spatiotemporal configuration of the $2D+T$ video volume. Using the MAP estimation criterion, such a problem can be recast as as an energy minimization problem that we solve efficiently using a spatiotemporal active contour approach. We demonstrate the feasibility of the proposed approach using real maritime videos.

1 INTRODUCTION

Maritime surveillance is an important customs application aiming at an efficient monitoring of maritime traffic, and securing sea coasts and harbors from fraudulent activities such as smuggling, thefts, piracies, intrusions, and human trafficking (Bloisi and Iocchi, 2009; Pires et al., 2010). Traditionally, it consists of a workflow of laborious tasks performed by human operators (e.g., coast guards). Recently, semi-automated and automated airborne video-surveillance systems have gained increasing popularity in maritime surveillance. The latter generate huge volumes of video data that thus need be analyzed automatically and in near real time in the goal of recognizing maritime targets and ranking their activity (e.g., usual versus fraudulent activity) (see Fig.1).

In this paper, we describe the video processing system that we have developed for automatic maritime object (e.g., vessel) recognition using an airborne visible light (i.e., RGB) video camera. The hardware architecture chosen for the project allows to continuously acquire video-streams of a maritime scene in the visible light spectrum (400-700 nm) involving a single target at once. Each movie of a target is stored on a local (airborne) computer and it is analyzed in quasi real time on board for recognition purposes (see (Bechar et al., 2013) for more details).

1.1 Motivations and Related Work

Most currently existing video based maritime surveillance systems are based on static (i.e., grounded) RGB cameras. From an algorithmic point of view, they have borrowed existing video-surveillance techniques originally developed for rather "gentle" environments, and thus they might not be well suited to highly dynamic scenes such as maritime environments, such as background subtraction techniques (Stauffer and Grimson, 1999), optical flow (Lucas and Kanade, 1981), statistical learning (Bloisi et al., 2012), and son on. Furthermore, there exists other surveillance systems which are based on different (though more expensive) data acquisition modalities such as infrared imagery (Smith and Teal, 1999), multiple sensor information fusion such as radar/AIS(B. J. Rhodes, 2007), and RGB/thermal infrared imagery (Bechar et al., 2013) in order either to cover larger areas of the sea or to account for RGB system's unreliability.

In this work, we consider a non-static RGB video system and our goal is to devise an automatic tech-

Bechar I., Bouchara F., Lelore T., Guis V. and Grimaldi M..
Uncertainty Fusion based Object Recognition and Tracking in Maritime Scenes using Spatiotemporal Active Contours.
DOI: 10.5220/0004755406820689
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 682-689
ISBN: 978-989-758-003-1

(a) (b) (c) (d)

Figure 1: Four maritime RGB video shots showing: (a) a cargo; (b) a medium-sized vessel; (c) a yacht; (d) a sailboat.

nique which combines several types of RGB video information in order to yield a likelihood ratio with respect to any spatiotemporal trajectory of a video that it corresponds to a maritime target's. Eventually, we would like to exploit such a result in order to detect as most accurately as possible (ideally delineate) the target in a video with an overwhelming probability. Let us emphasize that in contrast to static video systems in which objects are only filmed when they enter the field of view of the camera (in which case, a prior modeling via learning of the background without objects, combined with a proper background subtraction technique may be envisaged for object detection), our airborne (and thus non-static) video system acquires a specific maritime scene of a target directly, based on a radar based notification. Therefore, video processing should determine alone the spatiotemporal location of a target in the whole video.

To this end, we advocate the use of an approach which attempts to fuse several types of uncertainty regarding main chromaticity (main color) and dynamic texture (i.e., stochastic deviations around principal color) in a common probabilistic approach. This allows to yield a probability ratio with respect to all possible spatiotemporal configurations (i.e., both the spatial location and the temporal trajectory) of a target in a video. After that, using the maximum a posteriori (MAP) estimation criterion (in the log-likelihood form), we are able to recast the original problem as an energy minimization problem which is solved using an active contour approach (Mumford and Shah, 1989; Vese and Chan, 2002).

2 MATHEMATICAL MODEL

Given video data $\mathcal{X}_{\overline{1,T}} := \{X_t, t = \overline{1,T}\}$, where for all $t = \overline{1,T}$, X_t stands for the t-th video frame, the goal is to recognize and to track a maritime object throughout the whole video. Let us then denote by Ω the image domain, which is identical for all video frames. We adopt a probabilistic approach, and our goal is to assign to any spatiotemporal configuration $\mathcal{O}_{\overline{1,T}} \subseteq \Omega \times \overline{1,T}$ a probability that it corresponds to the spatiotemporal trajectory of the target object, given the observed video data. Eventually, we extract the spatiotemporal trajectory $\mathcal{O}^*_{\overline{1,T}} \subseteq \Omega \times \overline{1,T}$ that is most likely to correspond to the actual target's one.

Such a problem can indeed be formulated using the maximum a posteriori (MAP) principle as it will be described in the sequel. It should be mentioned however that our MAP approach 1 differs from another well-known energy based approach which is based on naive Bayes classification, in the sense the latter requires an explicit modeling simultaneously of the foreground $p(x_t/target)$ and background $p(x_t/backrgound)$ probabilities, whereas our approach solely focuses on foreground model $p(x_t/target)$. This has thus the advantage to alleviate the burden of having to estimate a background model (which does not seem to be a trivial task when dynamic backgrounds are considered) via a MAP estimation framework (which can also be seen as a hypothesis testing against hypothesis H_0 which is foreground model).

Having said this, by using the classical Bayes rule, one can write

$$\mathbb{P}\left(\mathcal{O}_{\overline{1,T}}/\mathcal{X}_{\overline{1,T}}\right) = \frac{\mathbb{P}\left(\mathcal{O}_{\overline{1,T}}\right)\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/\mathcal{O}_{\overline{1,T}}\right)}{\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}\right)} \quad (1)$$

where: $\mathbb{P}\left(\mathcal{O}_{\overline{1,T}}/\mathcal{X}_{\overline{1,T}}\right)$ stands for the a posteriori likelihood of spatiotemporal region $\mathcal{O}_{\overline{1,T}} \subseteq \Omega \times \overline{1,T}$, $\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/\mathcal{O}_{\overline{1,T}}\right)$ stands for the likelihood of video data given $\mathcal{O}_{\overline{1,T}}$, $\mathbb{P}\left(\mathcal{O}_{\overline{1,T}}\right)$ stands for a priori probability, and finally $\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}\right)$ stands for the likelihood of video data. The goal is to find the spatiotemporal configuration which maximizes formula 1. Note that the latter formula is often rewritten using the log-likelihood (or energy) form (which turns out to be more intuitive and easier to optimize in practice) as follows

$$-\log\left[\mathbb{P}\left(\mathcal{O}_{\overline{1,T}}/\mathcal{X}_{\overline{1,T}}\right)\right] = -\log\left[\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/\mathcal{O}_{\overline{1,T}}\right)\right] - \log\left[\mathbb{P}\left(\mathcal{O}_{\overline{1,T}}\right)\right] - \log\left[\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}\right)\right] \quad (2)$$

where now:

- $E\left(O_{\overline{1,T}}\right) := -\log\left[\mathbb{P}\left(O_{\overline{1,T}}/\mathcal{X}_{\overline{1,T}}\right)\right]$ stands for the total energy of $O_{\overline{1,T}}$ given that object is located at $O_{\overline{1,T}}$;
- $E_{fid}\left(O_{\overline{1,T}}\right) := -\log\left[\mathbb{P}\left(O_{\overline{1,T}}\right)\right]$ stands for a data fidelity term of the total energy of $O_{\overline{1,T}}$;
- $E_{reg}\left(O_{\overline{1,T}}\right) := -\log\left[\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/O_{\overline{1,T}}\right)\right]$ stands for a spatiotemporal regularization term;

Our main task in the remainder consists in the modeling of the a priori probability $\mathbb{P}\left(O_{\overline{1,T}}\right)$ of the a posteriori likelihood $\mathbb{P}\left(O_{\overline{1,T}}/\mathcal{X}_{\overline{1,T}}\right)$ and respectively, since $\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}\right)$ stands for a constant that we may ignore in the remainder.

2.1 Modeling $\mathbb{P}\left(O_{\overline{1,T}}\right)$

One firstly notes that $\mathbb{P}\left(O_{\overline{1,T}}\right)$ models purely spatiotemporal geometric information (or the spatiotemporal trajectory) of a target in a video (of course, up to camera motion). For instance, if one knows in advance that the target corresponds to some rigid object moving according to a similarity motion (rotation, translation and scale), thus it could be very useful to incorporate this information in $\mathbb{P}\left(O_{\overline{1,T}}\right)$ in a way which penalizes spatiotemporal trajectories $O_{\overline{1,T}} \subseteq \Omega \times \overline{1,T}$ that do not fit to the aforementioned prior geometric knowledge. Nevertheless, this suggests a proper parametrization of the total energy (using for instance a similarity matrix) and this is known to be computationally untractable. Therefore, it is often replaced with a computationally efficient approximate model (e.g., using a random Markov field model) but which nonetheless may yield quite satisfactory practical performances.

Therefore in this paper we propose to model directly $E_{reg}\left(O_{\overline{1,T}}\right) := -\log\left[\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/O_{\overline{1,T}}\right)\right]$ as the sum of a multiplicative factor of the total $2D$ surface of the spatiotemporal volume $O_{\overline{1,T}}$ and of a multiplicative factor of its total $3D$ volume. This gives rise to a regularization term which is composed of a traditional total variation (TV) term (Cremers et al., 2011) and a linear term in a continuous (relaxed) form of the total energy 2 (see section 3 for more details). While the former finds a pretty interpretation as a discontinuity preserving smoothing term of a spatiotemporal trajectory of target, the latter penalizes the total volume of the spatiotemporal trajectory of a target. Let us note that because of some (mainly computational) considerations that will be motivated in subsection 2.3, we assume that all video frames are aligned with each other prior to model's optimization. Therefore, such a regularization term operates directly on the aligned video.

2.2 Modeling $\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/O_{\overline{1,T}}\right)$

In this subsection, we describe our approach for modeling $\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/O_{\overline{1,T}}\right)$ in equation 1. As mentioned earlier in this section, the latter models the a posteriori likelihood of video data $\mathcal{X}_{\overline{1,T}}$ if a target is located at spatiotemporal location $O_{\overline{1,T}}$ of the video volume $\Omega \times \overline{1,T}$. It is clear that in the absence of a closed-form expression of a theoretical data observation model, one may only resort to an approximate formula of it. This is a difficult task because of the big variability of both appearance (i.e., intensity) models of the dynamic maritime background (i.e., the sea) and of a target (i.e., a vessel).

Indeed, depending on the weather conditions and on a target's speed and size, the background's dynamics may exhibit drastic differences from one video to another one (i.e., varying from a quasi-static blue sea to a rough, wavy and dark sea background). On the other hand, a vessel's appearance (in terms of main colors and their spatiotemporal variations) cannot be anticipated, because of color variability and important illumination changes inherent to a maritime scene. Therefore it makes sense to seek an expert system which can trade-off (merge) different types of RGB video uncertainty in order to achieve generally a good detection of a vessel. Having said this, the starting idea for achieving such a goal is that both chromaticity (i.e., as the principal pixel's color) and spatiotemporal (dynamic) texture turn out generally to be good features of both sea background and target. Before going into more details about this idea in the goal of achieving a good approximate model of $\mathbb{P}\left(\mathcal{X}_{\overline{1,T}}/O_{\overline{1,T}}\right)$, let us first consider the following (additive) video data model:

$$\mathcal{X}_{\overline{1,T}} = \mathcal{C}_{\overline{1,T}} + \mathcal{K}_{\overline{1,T}} + n_{\overline{1,T}} \tag{3}$$

where $\mathcal{C}_{\overline{1,T}}$ stands for the principal color (or the chromaticity) of the video pixels, $\mathcal{K}_{\overline{1,T}}$ stands for a (statistical) spatiotemporal (dynamic) texture (Derpanis and Wildes, 2012) which superimposes onto the principal color of video pixels, and $n_{\overline{1,T}}$ models system's plus environment's noise. Furthermore, we view $\left(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}\right)$ as a couple of random vector variables having some probability distribution $\mathbb{P}\left(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}/\text{target}\right)$ with respect to a tar-

get object, and some other probability distribution $\mathbb{P}\Big(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}/\text{background}\Big)$ with respect to background. An elaborated method would attempt to estimate in a common (parametric) framework simultaneously for $\big(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}\big)$ and the spatiotemporal location of a target. However this might be too costly computationally for near real time video processing. Therefore, if we assume that one may figure out deterministically, using video preprocessing (e.g., color clustering and texture filtering) the couple $\big(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}\big)$ from $\mathcal{X}_{\overline{1,T}}$ according to decomposition 3, then one may be able to exploit both principal color and spatiotemporal texture information in order to characterize target's likelihood in a video. This can be seen (after ignoring the noisy component $n_{\overline{1,T}}$ in formula 3) by replacing $\mathbb{P}\Big(\mathcal{X}_{\overline{1,T}}/O_{\overline{1,T}}\Big)$ with

$$\mathbb{P}\Big(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}/O_{\overline{1,T}}\Big) := f\big(\mathcal{C}_{\overline{1,T}}, \mathcal{K}_{\overline{1,T}}\big) \qquad (4)$$

where $f(\cdot)$ is some function which models an expert's knowledge regarding both chromaticity and (dynamic) texture at target's spatiotemporal locations.

In this work, we model such an expert function $f(\cdot)$ based on the following remarks:

- The brighter (whiter) a pixel is, the less likely it is to belong to the foreground as brightness is generally (but not always) characteristic of the foam;
- The more blue a pixel is, the more likely it is to belong to the background (i.e., to the foam);
- The more scattered in the image plan a pixel's color is, the less likely it is to belong to a target, as generally the sea occupies a spatially scattered region of a video frame;
- Target's dynamic texture is normally zero, up to imaging artifacts (such as illumination changes).

Therefore, we propose to take $f(c,k) := g\big(w(c), b(c), s(c), k\big)$ where g stands for some positive real valued function trading off $W(c), B(c)$, $S(c)$ and k which stand respectively for some functions of how much a color c is bright (white) ($W(c)$), blue ($B(c)$), scattered in the image plane ($S(c)$), and (dynamic) texture (k). As aforementioned, we have assumed that one may be able to extract each of the components c and k from any spatiotemporal video location using video preprocessing. Obviously, there is no single method for choosing the functions $W(c), B(c)$, and $S(c)$, therefore we propose to model them based on RGB video information as follows. For computational convenience, we firstly assume statistical independence between pixels, and we propose to take

$$W(c) \propto \exp\Big(-\frac{\big\|c-\overline{c}_w^*\big\|^2}{2\sigma_{w^*}^2}\Big)\mathbf{1}(W)$$

where $\overline{c}_w^*$ stands for the mean intensity of the brightest color class in a video frame (if any, and hence the use of $\mathbf{1}(W)$), and σ_{w^*} stands for its standard deviation.

$$B(c) \propto \exp\Big(-\frac{\big\|c-\overline{c_b}^*\big\|^2}{2\sigma_{b^*}^2}\Big)\mathbf{1}(B)$$

where $\overline{c}_b^*$ stands for the mean intensity of the blue color class in a video frame (if any) and $\sigma_{b^*}^2$ its standard deviation. Such brightest color and blue color classes are identified with respect to each video frame using color recognition techniques based respectively on the sum of three RGB components $c_r + c_g + c_b$ and (roughly) on the ratio $\frac{c_b+c_g}{2c_r}$ and by using a clustering technique (such as the Otsu algorithm). $S(c)$ is modeled as some decreasing function of the standard deviation $s(c)$ of the position in a video frame of color c. We take

$$S(c) \propto \exp\Big(-\frac{s^2(c)}{2}\Big)\mathbf{1}(B)$$

where c_S^* stands for the principal color with biggest standard deviation. Finally, we estimate the dynamic texture k at each video pixel by analyzing the video signals in its neighborhood of size N using the white noise assumption of dynamic texture at target. We take k as some increasing function of the minimum absolute value of the correlation ρ between any pair of neighboring signals (assuming prior alignment of video frames). Obviously, the expected value of k is the expected value of a normalized random variable with mean 0, and therefore one expects ρ to be generally much smaller for target pixels than for dynamic sea pixels (whose dynamic texture instead does not satisfy the white noise assumption). Thus we take

$$k \propto \exp\Big(-\frac{\rho^2}{2}\Big)$$

Finally, we take

$$f(z) \propto 1 - \alpha W(c)B(c)S(c)k \qquad (5)$$

where α stands for some positive constant in $]0,1]$. It should be mentioned however that we don't claim that such a choice of $f(z)$ is the most pertinent one, nevertheless such a made simplification may be seen as the price to pay for speeding up computations for achieving quasi real time system's performance.

2.3 Video Frame Alignment

As above-mentioned, in order to be able to figure out dynamic texture of a spatiotemporal configuration and to reduce problem's combinatoric complexity, therefore prior to video processing, we align all video frames in such a way that imaged points of a maritime scene correspond to a same pixel location across the video. This is achieved using pixel tracking

via optical flow (Lucas and Kanade, 1981). However, it is important to account for illumination changes between video frames in the optical flow model. Therefore, we first proceed by transforming linearly each RGB video frame X_t to a positive gray image I_t in a way which maximizes frame's contrast. Then, by assuming a linear illumination model, we perform optical flow (u,v) at current log-transformed gray video frame $J_t := \log[I_t]$ by assuming the following optical equation:

$$\frac{\partial J_t}{\partial x}u + \frac{\partial J_t}{\partial y}v + \frac{\partial J_t}{\partial t} + \nu = n \tag{6}$$

with n standing for white noise, and ν models illumination variation between consecutive frames. Such a model is estimated as (Lucas and Kanade, 1981) using least square criterion in the neighborhood of a pixel of size N. One notes that a bicubic interpolation and image sampling to obtain a lower resolution video frame is used prior to optical flow estimation. This allows to reduce overall video noise and to flatten highly textured sea regions (such as foam) in order to yield good estimation of spatiotemporal video gradient, and thereby to obtain a reliable optical flow. The latter is recomputed for the original video and later used to align video frames with each other.

3 ACTIVE CONTOUR IMPLEMENTATION

The goal is to minimize the following energy model with respect to all possible spatiotemporal confirguations O:

$$\left\{ \int_O F(z) + \lambda \text{Perim}\big(O\big) + \beta A\big(O\big) \right\} \tag{7}$$

where $F(z) := -log[f(z)]$ with $f(z)$ given by formula 5, λ and stand for some positive constant, $\text{Perim}\big(O\big)$ and $A\big(O\big)$ stand respectively for the total surface and the total volume of spatiotemporal configuration O.However while model's constant λ (which enforces target's spatiotemporal smoothness) is generally easy to tune as a wide range of values may be suitable for the task, the choice of β is not easy though crucial for obtaining a reasonable segmentation of a target. Furthermore, ideally we want to let the system choose adaptively the best value of μ to use, provided that we can inform it accordingly about what good values of β correspond to. Nevertheless, as we are dealing with grey videos (as $-log$ of pixel-wise probabilities), therefore we may detect a target as the most contrasted spatiotemporal configuration in the video, in a sense which we will specify hereafter. We show indeed in this paper that model 7 is equivalent to a traditional Mumford-Chah model, namely the Chan-Vese model 11 (Vese and Chan, 2002). The proof our claim along with the transformation of model 7 into an equivalent Chan-Vese model are detailed in the appendix section. Therefore the energy that we minimize is written as:

$$\left\{ \int_O (f(z)-c_1)^2 + \int_{\overline{O}} (f(z)-c_2)^2 + \lambda \text{Perim}\big(O\big) \right\} \tag{8}$$

where c_1 and c_2 correspond to some constants that are estimated adaptively using video data. Such a model 8 is firstly relaxed (M. Nikolova and Chan, 2006; Pock et al., 2008; Cremers et al., 2011; Chambolle et al., 2010) using variables $u(z;t) \in [0,1]$ as

$$\left\{ \int \Big((f(z)-c_1)^2 - (f(z)-c_2)^2 \Big) u(z) + \lambda TV(u) \right\} \tag{9}$$

where $\lambda TV(u)$ stands for the total variation of $u(z;t)$. For known c_1 and c_2, such a model 9 is convex and thus can be solved for exactly using the following iterative scheme, starting from an initial solution u_0: $u_{j+1} = u_j - \lambda div\Big(\frac{\nabla u_j}{|\nabla u_j|}\Big)$ where ∇ stands for the spatiotemporal gradient sign and div stands for the spatiotemporal divergence operator. The constants c_1 and c_2 are updated simultaneously with u_j as the mean intensities inside and outside current target respectively.

4 RESULTS

The current version of the method is implemented in C++ and runs in quasi real time on a standard $1.7Ghz$ PC and for video resolution of 30 frames/sec. and about 1 mégabyte frame resolution. Fig.2 and Fig.3 show results of the proposed method using two real maritime videos with $\lambda := 1000$.

We have validated the proposed method using a dozen of realistic video sequences under different weather conditions, and target appearances. The method has shown to perform generally very well for detecting a vessel target except in some situations where a target is mainly characterized with a whitish color and the airborne camera lies so far away from the target that the sea foam does not exhibit enough texture that may distinguish it from the target.

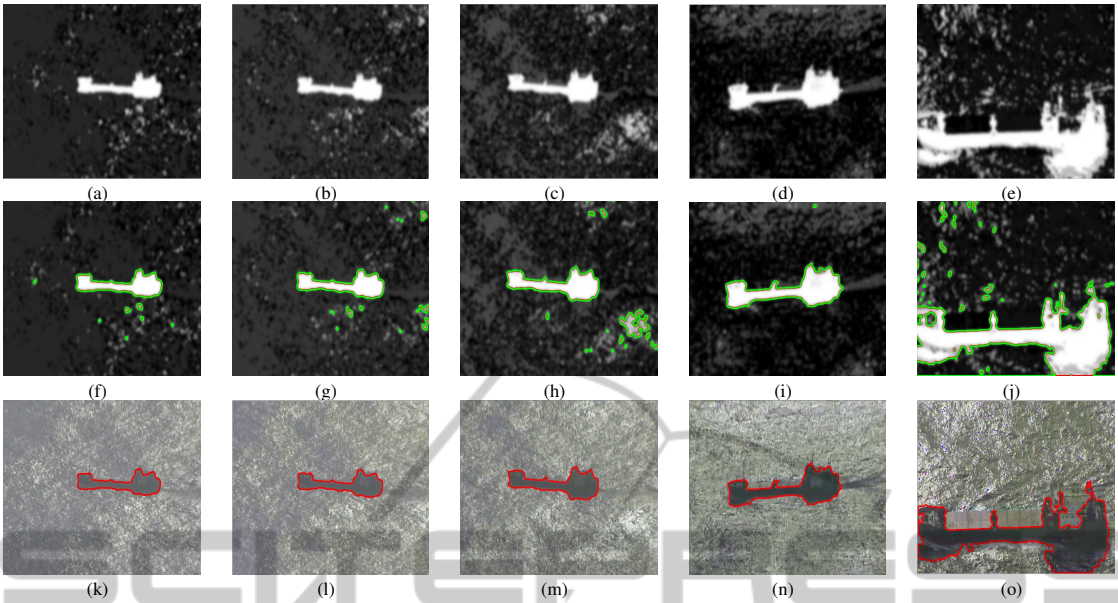

Figure 2: Uncertainty based recognition and tracking of a maritime target (a cargo) using a spatiotemporal active contour. approach. Upper row: A posteriori log-likelihood of the target (normalized between 0 and 255) at video frames number 1, 100, 200, 300, and 400 (resp). Middle row: The converged spatiotemporal active contour; Lower row: Target recognition.

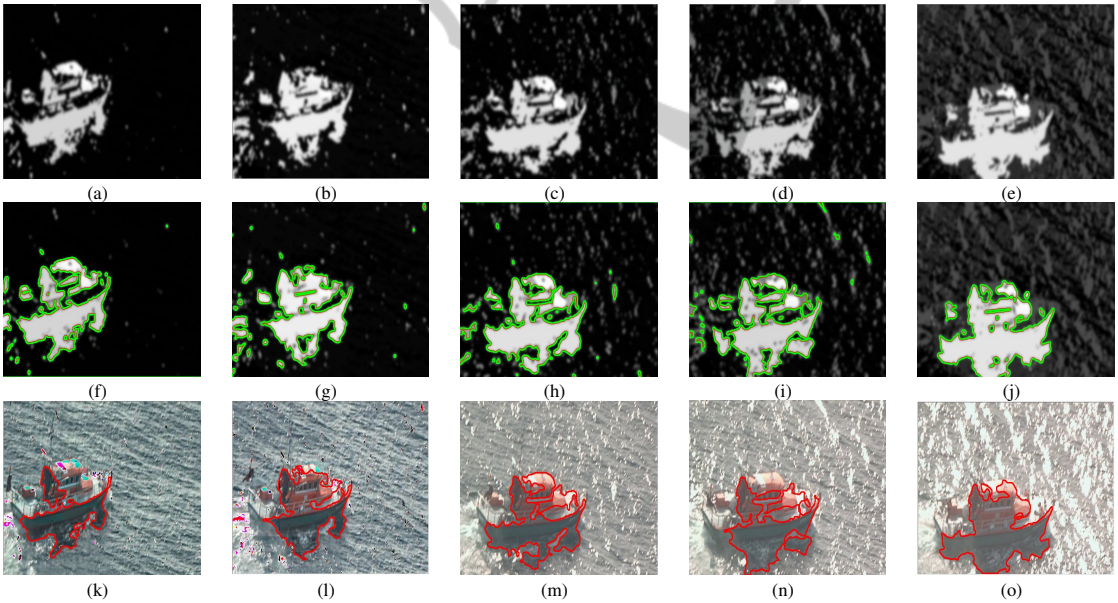

Figure 3: Uncertainty based recognition and tracking of a maritime target (a yacht) using a spatiotemporal active contour. Upper row: A posteriori log-likelihood of the target (normalized between 0 and 255) at video frames number 1, 100, 200, 300, and 400 (resp). Middle row: The converged spatiotemporal active contour; Lower row: Target recognition.

5 CONCLUSIONS

We have described a novel method for detecting vessels in dynamic maritime background using a fusion of uncertainty based approach, and we have solved the problem efficiently using a spatiotemporal active contour approach. Our tests using a dozen of real video sequences of several minutes duration have shown that our methods outperforms some state of the art object tracking techniques such as meanshift/camshift techniques. In the current version of our system, all method's parameters have been hardly coded based

on experience. Nevertheless it makes sense to devise an automatic parameter selection technique in a future version of the system. Also as future work, we plan to consider other types of video information such as more sophisticated texture models and geometric prior knowledge (such as object's rigidity, height and shape) in order to yield as most reliable vessel detection algorithm as possible.

ACKNOWLEDGEMENTS

Thanks to the French customs for funding.

REFERENCES

B. J. Rhodes, e. a. (2007). Seecoast: persistent surveillance and automated scene understanding for ports and coastal areas. Ed., vol. 6578, no. 1. SPIE, p. 65781M.

Bechar, I., Lelore, T., Bouchara, F., Guis, V., and Grimaldi, M. (2013). Toward an airborne system for near real-time maritime video-surveillance based on synchronous visible light and thermal infrared video information fusion. an active contour approach. In Proc. Ocoss'2013, Nice, France.

Bloisi, D. and Iocchi, L. (2009). Argos - a video surveillance system for boat traffic monitoring in venice. In *IJPRAI, vol. 23 (7), pp. 1477–1502*.

Bloisi, D., Iocchi, L., Fiorini, M., and Graziano, G. (2012). Camera based recognition for marine awareness, great lakes and st. lawrence seaway border regions. In *Int. Conf. Infor. Fusion, pp. 1982–1987*.

Chambolle, A., Caselles, V., Novaga, M., Cremers, D., and Pock, T. (2010). An introduction to total variation for image analysis. In *Chapter in Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter*.

Cremers, D., Pock, T., Kolev, K., and Chambolle, A. (2011). Convex relaxation techniques for segmentation, stereo and multiview reconstruction. In *Chapter in Markov Random Fields for Vision and Image Processing*. MIT Press.

Derpanis, K. and Wildes, R. (2012). Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. In *PAMI,34(6):1193-205.*

Lucas, B. and Kanade, T. (1981). An iterative image registration tech- nique with an application to stereo vision. In *In Proceedings of the International Joint Conference on Artificial Intelligence, pp. 674–679*.

M. Nikolova, S. E. and Chan, T. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. In *SIAM Journal of Applied Mathematics 66, 1632–1648*.

Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. In *Comm. Pure. Appl. Math. 42:577-685*.

Pires, N., Guinet, J., and Dusch, E. (2010). Asv: an innovative automatic system for maritime surveillance. In *Navigation, vol. 58(232), pp. 1–20*.

Pock, T., Schoenemann, T., Graber, G., Bischof, H., and Cremers, D. (2008). A convex formulation of continuous multi-label problems. In *ECCV'08*.

Smith, A. and Teal, M. (1999). Identification and tracking of marine objects in nearinfrared image sequences for collision avoidance. In *In 7th Int. Conf. Im. Proc. Applic., pp. 250–254*.

Stauffer, C. and Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. in CVPR'99, pp. 2246–2252.

Vese, L. and Chan, T. (2002). A new multiphase level set framework for image segmentation via the mumford and shah model. In *IJCV, Vol. 50, pp. 271–293*.

APPENDIX

Let us prove the claim we made in section 3. For simplicity's sake and without loss of generality, we consider the following MAP based image segmentation problem:

$$\min_O \lambda \text{Perim}(O) \quad + \quad \beta A(O) \quad + \quad \int_O g(z) \quad (10)$$

where $g(z)$ is a positive function as it corresponds $-\log$ of a probability. Now, let us consider the Chan & Vese image segmentation model

$$\lambda \text{Perim}(O) + \int_O \frac{(f(z)-c_1)^2}{\sigma_1^2} + \int_{\overline{O}} \frac{(f(z)-c_2)^2}{\sigma_2^2} \quad (11)$$

One may rewrite model 11 equivalently (after throwing away the constant term) as follows

$$\begin{aligned}
&\lambda \text{Perim}(O) + \int_O \Big(\frac{(f(z)-c_1)^2}{\sigma_1^2} - \frac{(f(z)-c_2)^2}{\sigma_2^2} \Big) \Big\} \\
&= \Big\{ \lambda \text{Perim}(O) + \int_O \Big(f^2(z) \big(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \big) \\
&\quad - 2f(z) \Big(\frac{c_1}{\sigma_1^2} - \frac{c_2}{\sigma_2^2} \Big) \Big) + \int_O \Big(\frac{c_1^2}{\sigma_1^2} + \frac{c_2^2}{\sigma_2^2} \Big) \\
&= \lambda \text{Perim}(O) + \Big(\int_O \Big(f^2(z) \big(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \big) - 2f(z) \big(\frac{c_1}{\sigma_1^2} - \frac{c_2}{\sigma_2^2} \big) + K \Big) \\
&\quad + \Big(\Big(\frac{c_1^2}{\sigma_1^2} + \frac{c_2^2}{\sigma_2^2} - K + R \Big) A(O)
\end{aligned}$$

where K is the smallest positive constant (perhaps 0) which makes the integrand term $\Big(f^2(z) \big(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \big) - 2f(z) \big(\frac{c_1}{\sigma_1^2} - \frac{c_2}{\sigma_2^2} \big) + K \Big)$ always positive, whatever z.

Now putting

$$\begin{aligned} f^2(z)\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right)-2f(z)\left(\frac{c_1}{\sigma_1^2}-\frac{c_2}{\sigma_2^2}\right)+K &= g(z) \\ \frac{c_1^2}{\sigma_1^2}+\frac{c_2^2}{\sigma_2^2}-K &= \beta \end{aligned}$$

which is always possible via an appropriate tuning of c_1 and c_2 such that $\frac{c_1^2}{\sigma_1^2}+\frac{c_2^2}{\sigma_2^2}-K$ is positive and equal to β, and by solving for the second degree equation with respect to $f(z)$ in order to find the formula of $f(z)$ as a function of $g(z)$, and hence the proof.