Distance Estimation for Marine Vehicles Using a Monocular Video Camera

Ran Gladstone, Yair Moshe, Avihai Barel, Elior Shenhav
Signal and Image Processing Laboratory (SIPL)
Department of Electrical Engineering, Technion – Israel Institute of Technology
Technion City, Haifa, Israel, http://sipl.technion.ac.il/

***Abstract*—Distance estimation for marine vessels is of vital importance to unmanned ship vehicles (USVs) for navigation and collision prevention. This can be achieved by means such as radar or laser sighting. However, due to constraints of the USV, it may be desired to estimate distance using a monocular camera. In this paper, we propose a method that, given a video of a marine vehicle in a maritime environment and a tracker, estimates the distance of the tracked vehicle from the camera. The method detects the horizon and uses its distance as a reference. It detects the contact point of the vehicle with the sea surface by finding a maximally stable extremal region (MSER). Then, it relies on geometries of the earth and on optical properties of the camera to compute the distance. The method was tested on video footage of several sea maneuvers with an average error of 7.1%.**

Keywords— unmanned ship vehicles; USV; distance estimation; horizon detection; marine navigation

I. Introduction

Unmanned ship vehicles (USVs) are vehicles that operate on the surface of the water without a crew. These vehicles are witnessing an increasing interest in both academia and industry due to their numerous applications for research, defense missions, and for commercial purposes [1, 2]. USVs are attractive for surveillance applications, like patrolling and maintaining harbors or other "crucial" sites safeguarded against intruders, as they can identify the level of menace of an unknown radar track, exposing no human operators to possible threats [3]. Highly autonomous USVs are desired in order to decrease human operator's involvement and for improved ability to deal with communication failures. An important requirement for an autonomous USV is the ability to navigate and avoid collisions with obstacles in its environment [4]. Typically, static obstacles such as islands and immobile constructs can be mapped in advance to a world model, but dynamic obstacles, such as boats and water scooters, need to be detected and continuously tracked to avoid collision. Once tracking commences, the range and azimuth of the obstacle from the USV is estimated using GPS and its position serves as input to a navigation algorithm.

The task we deal with in this paper is distance estimation of dynamic obstacles on the sea surface, specifically marine vehicles. Distance estimation of marine vehicles can be achieved by several methods, such as stereo vision [3, 5, 6], laser sighting [7, 8], or radar [3, 7]. The maritime environment is challenging for many types of sensors due to the fast rate of change it experiences from waves, weather and sea traffic. Therefore, it is difficult to achieve accurate detection, tracking, and ultimately distance estimation with only one sensor type. Hence, a common approach to tackling these problems is using multiple types of sensors and performing sensor fusion, namely, the merging of the output of several sensor types to achieve an accurate result [3, 7]. However, due to the constraints of the USV's size and the cost of equipping it with additional sensors, especially if 360 degrees coverage is required, it may be desired to estimate the distance of marine vehicles with a monocular video camera. Another advantage of distance estimation using a monocular camera is that it is a passive sensor, requiring no transmission of any signal, which makes it useful for defense applications. As with other sensor types, processing of vision video footage of maritime environments is challenging due to the fast rate of change of the sea surface and other environmental effects. Maritime videos are affected by high noise, and objects are characterized by a low contrast (due to their large distance to the camera, humidity, fog, bad lighting conditions, etc.). Furthermore, the camera is constantly moving, and the background is highly dynamic, thus making it difficult to distinguish between object pixels and background ones [9]. Therefore, traditional approaches in image and video processing often fail to produce accurate results in such scenarios. Examples of previous vision-based works for detection, tracking, and segmentation of objects, in a maritime environment, can be found in [9-13].

In the following section, we present a novel method to estimate the distance of marine vehicles using a monocular video camera. To the best of our knowledge, only [3, 13] suggest to perform such a task using a monocular video camera. However, [3] lacks a full description of the method and presents no results and [13] assumes a geo-stationary camera. The proposed method receives a grayscale video of a marine environment, along with tracking information of an object in the form of (x, y) coordinates for each frame. Tracking coordinates are assumed to be on, or very close to, the tracked object. The contact point of the tracked vehicle with the sea surface is determined, and then, using the horizon as a reference and the specifications of the camera, we infer the distance of that point. The output of the method is the distance of the tracked object from the camera's location.

The paper is organized as follows. In section II we describe the proposed method for distance estimation. In section III we present the experiment conducted to test the method and the results are compared to ground truth distance. Finally, in section IV we conclude our work.

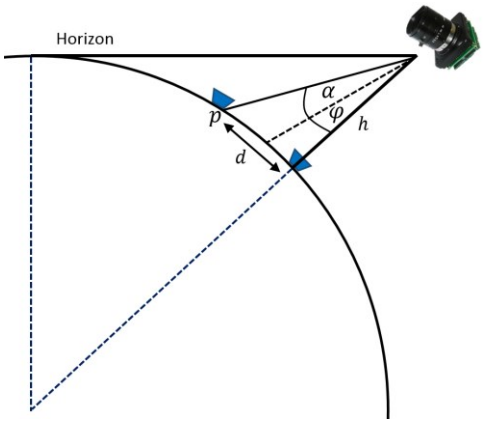


Fig. 1. Problem setup. A camera mounted at height $\boldsymbol{h}$ captures an image of an object at distance $\boldsymbol{d}$. The object is represented by a pixel $\boldsymbol{p}$. φ is the angle from the line connecting the camera to the center of the earth to the beginning of the FOV of the camera. α is the angle between the beginning of the FOV to $\boldsymbol{p}$.

II. Distance Estimation

A. General description

To better explain the basic idea behind the proposed method, we neglect camera movement caused by the movement of the USV. We therefore assume, for the sake of simplicity, that the USV is immobile and that the camera's position remains constant. We will remove this assumption in the following subsections. A general scheme of the problem setup is given in Fig. 1. A camera mounted on a USV at height h captures an image of an object. Consider a pixel p on the sea surface representing the object captured by the camera. This pixel corresponds to a distance d from the USV. To calculate the distance d, we should calculate the angle between the line of sight to a small sea surface area that the pixel p represents and the line connecting the camera to the center of the earth. Thus, if we know the height of the camera h and the angle $\alpha + \varphi$ described above, and under the assumption that each pixel captures approximately the same solid angle in space, we can calculate the distance using simple trigonometry. Notice that not all pixels composing the needed angle are captured by the camera since its field of view (FOV) begins only at an angle φ. However, assuming the camera's position in relation to the USV is constant, φ remains constant and can be measured. Hence, the angle we require is the sum of the angle φ corresponding to the beginning of field of view, and the angle α corresponding to the number of pixels between p and the bottom of the video frame.

Based on the geometrical explanation described above, the major problem that remains is to relate a meaningful pixel p to each tracked object. Since one purpose of our work is to provide support for autonomous navigation, the pixel that should be chosen is the pixel representing the closest point of the tracked object to the camera. A good pixel to choose is a pixel representing the area where the object touches the sea surface. Intuitively, pixels representing the sea surface closer to the horizon correspond to greater distances from the camera than pixels farther from the horizon. Neglecting protruding sharp features the object might have, the best pixel for our application is the pixel farthest from the horizon, thus representing the closest possible distance of the tracked object to the camera.

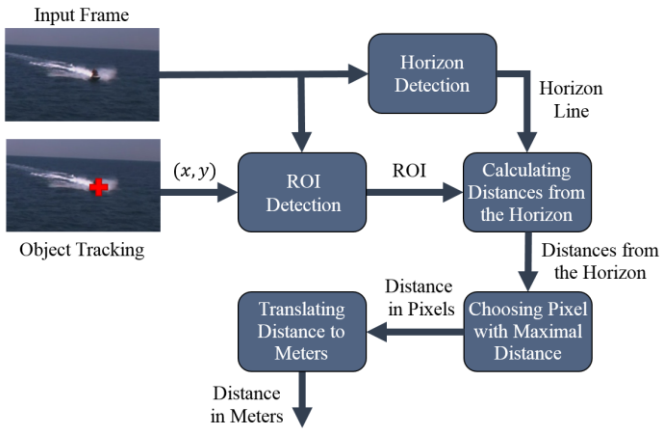


Fig. 2. Distance estimation.

A scheme of the proposed method is depicted in Fig. 2. The inputs are a video frame and a tracker output coordinates in that frame. First, the horizon line is detected in the video frame. Then, we detect a region of interest (ROI) of the tracked object. This ROI is obtained using extraction of stable regions and selection of the region whose centroid is the closest to the tracker output coordinates. We proceed to inspect the pixels on the boundary of the ROI and calculate the distance in pixels of each such pixel from the horizon line. We choose only the pixel the farthest from the horizon line and use the pixel's coordinates, along with the camera specifications and horizon incline angle to calculate the distance of the tracked object in meters. In the following subsections we further elaborate on these steps.

B. Horizon Detection

Correct horizon detection is critical for accurate algorithm performance. The horizon in a video frame may be tilted and raised or lowered due to the USV's movement. To calculate the correct distance, pixels need to be counted along a line perpendicular to the horizon. Furthermore, the USV may have a pitch angle which determines the angle of the beginning of FOV. Several previous works in the literature deal with horizon detection in a maritime environment. A recent comparison of state-of-the-art horizon detection techniques is given in [14, 15]. However, the techniques are tested on easy scenarios and usually perform poorly in more realistic maritime video footage. In this subsection we describe a horizon detection method that we have developed. It is based on [14, 15] with several improvements and modifications to increase its robustness to harsh environmental conditions.

Fig. 3 depicts the proposed method for horizon line detection. Horizon detection begins with morphological erosion of the image using a small circular mask. This is done in order to remove small features that may be located near the horizon line and to reduce false positives from sea waves. Sun glare was not apparent in the input images of [14, 15], so next we use histogram equalization to reduce its effects in the image. We proceed by applying Canny edge detection to get an edge map of the image. Typically at this point the edge map contains large parts of the horizon line, along with many small edges detected due to waves on the sea surface. Unlike [14, 15], we eliminate many of those irrelevant edges by removing all edges whose

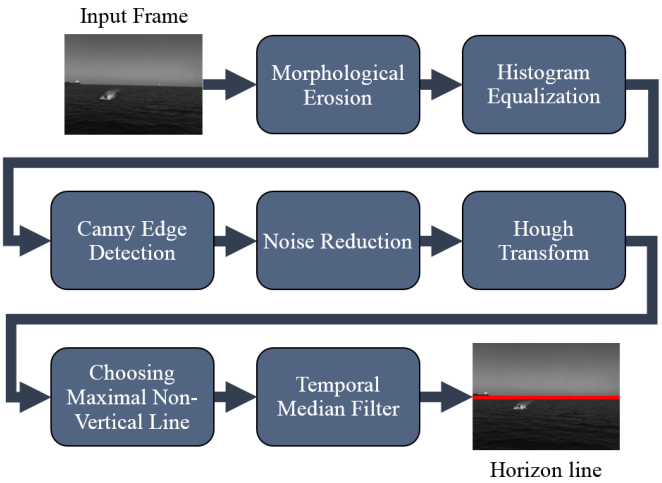


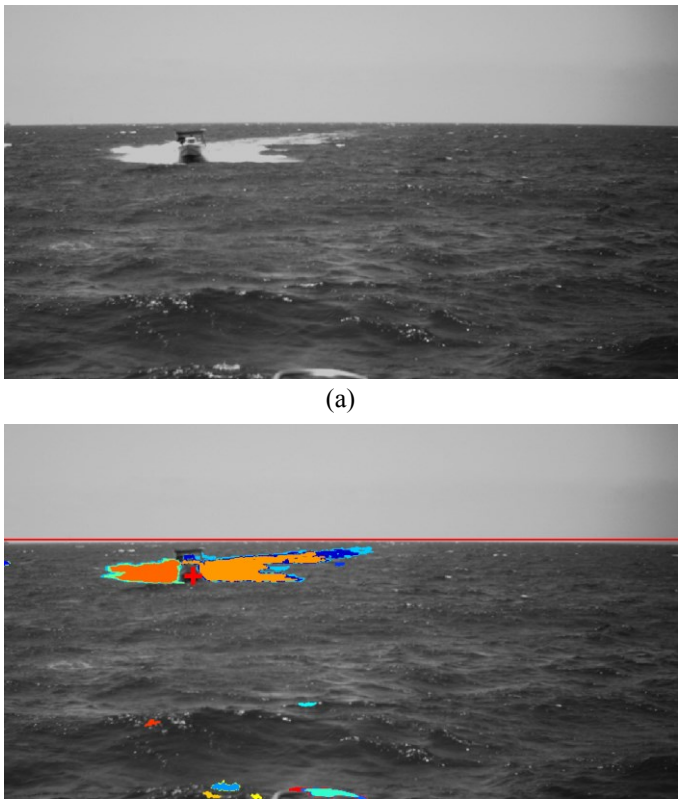
Fig. 3. Horizon line detection.

number of pixels is smaller than a pre-determined threshold, thus reducing the noise in the edge map. Next, the Hough transform of the edge map is computed in order to detect straight edges. The Hough transform detects straight lines using a voting procedure. We ignore lines that are nearly vertical, which appear in case of sun glare in the image. Then, we choose the parameters of angle and height of the line that correspond to the maximal number of votes. Finally, for smoothing the decision and removing transient false detections, we apply a temporal median filter of size 3 on the height and angle of the detected horizon lines.

C. ROI Detection

Ideally, we would like to segment the tracked object from the maritime background. However, performing such segmentation is a difficult problem, as common approaches for foreground segmentation are not effective for maritime environment footage. Moreover, even previous works that deal specifically with segmentation in a maritime environment such as [9, 11] are inappropriate for our problem setup. They assume an immobile camera, are confused by the object wake or may result in erroneous segmentation of the object. A more suitable approach to the problem at hand is to segment either the object or relatively small parts of its wake. Both are acceptable for the next stage of the algorithm, where we choose a pixel representing the closest point of contact of the tracked object with the sea surface.

Since we cannot assume prior knowledge on the shape, size, speed, initial distance and other properties of the tracked object, a multiscale image processing method is required. Hence, the method chosen is the extraction of maximally stable extremal regions (MSER) [16]. This is an efficient multiscale blob detection algorithm usually used for stereo matching and object recognition. It finds regions that remain stable over a certain number of thresholds. A unique contribution of this paper is the use of MSER in a maritime environment. MSER is suitable for our purpose since a typical marine environment tends not to have many MSER regions due to the gradual change of its pixel intensities. However, man-made objects and objects' wakes tend to be detected as MSER regions, as demonstrated in Fig. 4. Once all MSER regions in the image are detected, we choose as



(a)

(b)

Fig. 4. (a) A video frame of a sailboat and (b) MSER regions (in color) detected in it. Note that the regions overlap. The coordinates supplied by the tracker are marked as a red cross. The detected horizon is marked as a red straight line.

ROI the one whose centroid is closest to the coordinates supplied by the tracker. MSER extraction has two parameters: threshold increment and maximal area variation of the extremal areas. We set the values of those parameters according to whether the image experiences sun glare and, once an estimate of the distance of the tracked object is available, according to its estimated distance in previous frames. An image experiencing sun glare has many detections of irrelevant MSER regions, and thus we increase the threshold delta and decrease the maximal area variation. Likewise, a closer object has more small visible features which can be interpreted as MSER regions which may be irrelevant. Therefore, we increase the threshold increment and the maximal area variation as a function of the distance estimate of the tracked object in previous frames. Currently we use a set of conditions to select between a few values of the maximal area variation and the threshold increment, but in future work a continuous transformation will be tested.

D. Distance Calculation for a Representing Pixel

From the ROI we choose only the farthest pixel from the horizon line. We then rotate and translate the image according to the angle of the detected horizon line so it is not tilted and its height in pixels in the frame remains constant through all the video frames. We count the number of pixels from the chosen pixel to the bottom of the frame and transform the number of pixels to an angle according to the camera's FOV. The distance of the tracked object is given by (see Fig. 1):

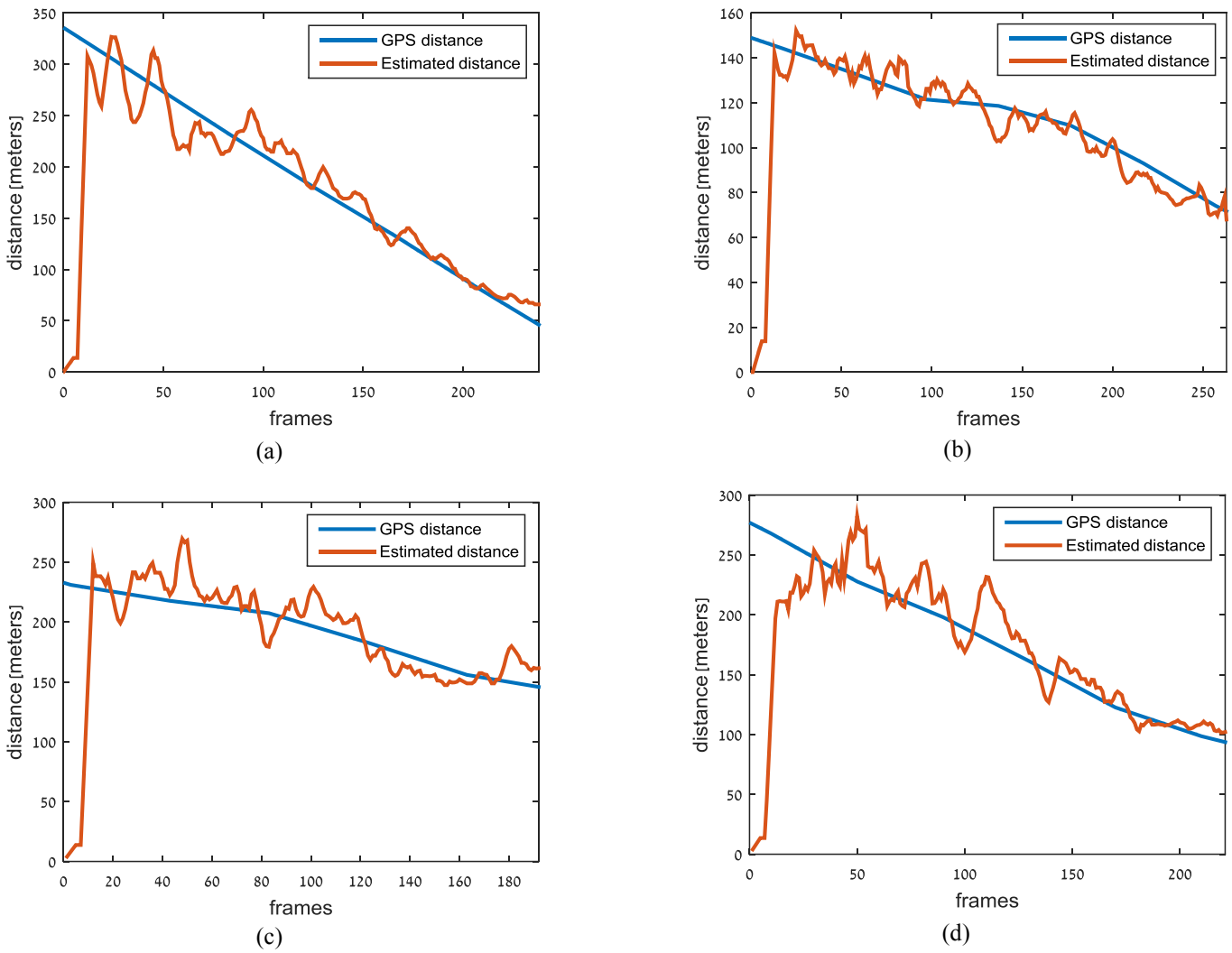


Fig. 5. Distance estimation vs. ground truth measured with a GPS. For each video we measure the mean absolute percentage error (MAPE) and the standard deviation (SD). (a) Approaching rubber boat. MAPE=8.7%, SD=7.1%. (b) Sailboat sailing in parallel to the camera. At the same time, the recording boat is approaching the sailboat. MAPE=4.8%, SD=4%. (c) Sailboat sailing in parallel to the camera. The horizon line in this video is the line differentiating sea from land. MAPE=7%, SD=4.8%. (d) Approaching rubber boat. The horizon line in this video is the line differentiating sea from land. The video contains challenging land background, making horizon detection difficult. Mean absolute relative error is MAPE=9%, SD=7%.

$$d = h \tan(\alpha + \varphi) \tag{1}$$

Since the distance of the tracked object is continuous, we apply a median filter and an averaging filter of a five frame window to smooth the results.

III. Results

A. Experiment Setup

An experiment involving three marine vehicles was conducted on a day with clear weather and relatively calm sea. A CCD camera was installed at a height of about 2.7 meters above sea level on one ship, referred to as the recording vessel, along with GPS sensors. The camera has an FOV of 24 degrees, resolution of 800x600, frame rate of 15Hz and was stationary in relation to the recording vessel. Oscillations due to waves that affect the recording vessel affect the position of the camera. The GPS on the recording vessel was installed very close to the location of the camera. GPS units were also installed on the other two vessels, referred to as the maneuvering vessels. All GPS units are commercial GPS units with 7.8 meters positioning error with 95% confidence interval. The maneuvering vessels were a large sailboat and a small rubber boat that conducted several maneuvers in the field of view of the camera installed on the recording vessel. Ground truth distance between the camera and each of the maneuvering vessels was calculated according to the GPS coordinates of the recording vessel and the maneuvering vessels. The resulting video and the distance of each maneuvering vessel were synchronized, so that each frame in the video could be associated with a ground truth distance. Tracker output coordinates were obtained using a commercial tracker. The proposed distance estimation method was then applied offline to the experiment data, and the output estimated distance was compared to the ground truth. The full algorithm was tested on four videos from the experiment, each with duration of 12-17 seconds. Details of the scenario in these videos are given in the caption of Fig. 5.

B. Analysis of the Results

The results of the proposed method for the four experiment videos show an average absolute error relative to the GPS distance in the range of 4.8% to 9%, giving an overall average of 7.1% and a standard deviation of 5.8%, depending on the scenario shown in the videos, as presented in Fig. 5. Note that the algorithm only provides results after receiving tracker data and processing a few frames. The oscillations on the graphs seem to be caused by the fluctuating height of the camera, due to movement of the USV. These oscillations, and thus the mean error, can be reduced in future work by considering h, the height of the camera, as a variable. h can be measured by using an inertial navigation system (INS).

Horizon detection proved reliable in all scenarios. In the scenarios with land in the background it was somewhat less stable, resulting in an increased relative error in distance estimation. This effect can be mitigated in future work, using a color edge detector which will improve the accuracy of the horizon detection in such situations.

Detected MSER regions on the tracked object sometimes do not cover all the object, especially when the object is close to the camera and small features on it become visible, such as light reflection patches of metallic parts. Another issue with extracting MSER regions is that vessels producing less wake and foam on the water are more difficult to detect as they have less chance of producing MSER regions.

Running time of the proposed method is about 0.5-2 seconds per frame, written in MATLAB, running on a standard quad-core Windows desktop PC. This is a relatively fast running time for video processing applications thus, with an efficient implementation, we expect it to be feasible to run the proposed method on a USV in real-time.

Horizon detection is performed at a time complexity of $O(MNlogMN)$ where $M \times N$ is the number of pixels in each frame. ROI detection is of time complexity of $O(MN)$. The rest of the blocks in Fig. 2 are also of time complexity of $O(MN)$. Therefore, the time complexity of the entire algorithm is $O(MNlogMN)$.

IV. Conclusions

In this paper, we have proposed a novel method for distance estimation of marine vehicles in a challenging maritime environment using a monocular video camera. An important stage of this method is a robust horizon line detection algorithm. Another contribution of this paper is a novel use of MSER regions for image processing in a maritime environment. These regions are used, together with supplied tracker output coordinates, for finding the closest contact point of the vehicle with the sea surface. A simple trigonometric formulation based on the geometries of the earth and on optical properties of the camera, allows us to compute the distance to the tracked object. The method was tested on video footage of several sea maneuvers. It was able to estimate successfully the distances of marine vehicles using no prior knowledge on the vessel's size, shape or velocity. Mean absolute percentage error of 7.1%. and standard deviation of 5.8% were achieved for real experiment data. The accuracy of the results is suitable for autonomic USV navigation.

Acknowledgment

The authors would like to thank Rafael for the collaboration and support. The authors would also like to thank Prof. David Malah, head of SIPL, and Nimrod Peleg, chief engineer of SIPL, for their support, advice and helpful comments. Finally, we would like to thank Shai Rozenberg for productive discussions.

References

[1] J. E. Manley, "Unmanned surface vehicles, 15 years of development," in *OCEANS 2008*, 2008, pp. 1-4.

[2] V. Bertram, "Unmanned surface vehicles–a survey," *Skibsteknisk Selskab, Copenhagen, Denmark,* pp. 1-14, 2008.

[3] J. Larson, M. Bruch, R. Halterman, J. Rogers, and R. Webster, "Advances in autonomous obstacle avoidance for unmanned surface vehicles," DTIC Document2007.

[4] T. Statheros, G. Howells, and K. M. Maier, "Autonomous ship collision avoidance navigation concepts, technologies and techniques," *Journal of navigation,* vol. 61, p. 129, 2008.

[5] T. Huntsberger, H. Aghazarian, A. Howard, and D. C. Trotz, "Stereo vision–based navigation for autonomous surface vessels," *Journal of Field Robotics,* vol. 28, pp. 3-18, 2011.

[6] G. Casalino, A. Turetta, and E. Simetti, "A three-layered architecture for real time path planning and obstacle avoidance for surveillance USVs operating in harbour fields," in *Oceans 2009-Europe*, 2009, pp. 1-8.

[7] L. Elkins, D. Sellers, and W. R. Monach, "The Autonomous Maritime Navigation (AMN) project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles," *Journal of Field Robotics,* vol. 27, pp. 790-818, 2010.

[8] C. R. Sonnenburg and C. A. Woolsey, "Modeling, identification, and control of an unmanned surface vehicle," *Journal of Field Robotics,* vol. 30, pp. 371-398, 2013.

[9] M. Grimaldi, I. Bechar, T. Lelore, V. Guis, and F. Bouchara, "An unsupervised approach to automatic object extraction from a maritime video scene," in *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, 2014, pp. 1-6.

[10] D. Socek, D. Culibrk, O. Marques, H. Kalva, and B. Furht, "A hybrid color-based foreground object detection method for automated marine surveillance," in *Advanced Concepts for Intelligent Vision Systems*, 2005, pp. 340-347.

[11] Z. L. Szpak and J. R. Tapamo, "Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set," *Expert systems with applications,* vol. 38, pp. 6669-6680, 2011.

[12] D. Frost and J.-R. Tapamo, "Detection and tracking of moving objects in a maritime environment using level set with shape priors," *EURASIP Journal on Image and Video Processing,* vol. 2013, pp. 1-16, 2013.

[13] S. Amarasinghe, N. D. Kodikara, and D. Sandaruwan, "Location estimation in a maritime environment using a monocular camera," in *Advances in ICT for Emerging Regions (ICTer), 2014 International Conference on*, 2014, pp. 21-28.

[14] E. Gershikov, T. Libe, and S. Kosolapov, "Horizon Line Detection in Marine Images: Which Method to Choose?," *International Journal on Advances in Intelligent Systems,* vol. 6, 2013.

[15] I. Lipschutz, E. Gershikov, and B. Milgrom, "New methods for horizon line detection in infrared and visible sea images," *Int. J. Comput. Eng. Res,* vol. 3, pp. 1197-1215, 2013.

[16] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing,* vol. 22, pp. 761-767, 2004.