Maritime Vessel Tracking with an Ensemble of WiSARD Classifiers in Video

Rodrigo da Silva Moreira, Nelson Francisco Favilla Ebecken

***Abstract*—**Ships are vulnerable when they are moored in bays and ports or when they sail in high risk areas due to a low capacity to perform maneuvers and the restrictive laws of certain countries to prevent radar sensor usage. The development of a video surveillance system is necessary to curb drug trafficking and ilegal immigration, avoid collisions and support other sensor types. This paper proposes a maritime vessel tracker named EWRN (Ensemble Wizard's Rodrigo Nelson), composed of an ensemble of WiSARD weightless neural network classifiers. A failure detector analyzes vessel movement with a Kalman filter and corrects the tracking, if necessary, using FFT matching. The use of the WiSARD neural network to track objects is uncommon. The additional contributions of the present study include: a performance comparison between EWRN and four state of the art trackers, an experimental study of the features that improve maritime vessel tracking, the first use of an ensemble of classifiers to track maritime vessels and a new quantization algorithm that compares the values of pixel pairs.

***Keywords*—**Ram memory, WiSARD Weightless Neural Network, object tracking, quantization.

I. Introduction

SMALL boats and people on jet skis can attack maritime vessels and warships. Small and agile boats are difficult to detect and track correctly with a radar sensor. The development of an autonomous video surveillance system is essential to improve port security and coastal defense, support other sensor types and curb unwanted event occurrences such as illegal fishing and immigration, pirate and terrorist attacks, ship collisions and drug trafficking [1, 2, 3, 4, 5, 6].

When Tracking is generally a challenging problem and has been widely studied. It is used for surveillance [5], target tracking [7], simulation of human gait motion [8], trajectory tracking for a mobile robot [9] and even to recognize hand gestures [10]. Object tracking is an essential surveillance system component. The functions of a tracker include object localization, object association and movement estimation to define the region of interest or ROI (the frame region wherein the tracker will try to localize the vessel).

According to [11], tracking by detection interprets the problem as a binary classification problem and is one of the most attractive and researched fields in visual computation. The tracker tries to find the function that best separates the features from the CVT class that represent the vessel being tracked from the CNVT class features that represent the other moving vessels as well as the environment [12]. The tracker has one vessel detector (DVT) that is trained to recognize the vessel being tracked with an ensemble of classifiers. The features extracted from N different candidate regions CRi(t) at each frame form the test dataset (1). The vessel position is defined using Bayes' rule [13] as the candidate region that provides the highest probability difference: P(yi=CVT | CRi(t)) - P(yi=CNVT | CRi(t)).

$$CR_i(t) = \{\, CR(t) \mid \|L(CR(t)) - L(CR(t-1))\| < r \,\} \quad (1)$$

L(v) is the centroid position of the vessel v, t is the frame index and r is the ROI size. Machine learning techniques are used to update the trackers when the vessel's appearance or shape changes [1]. The DVT can be trained online, offline or using a combination of both techniques. The classifiers are updated during tracking if they are trained online [11, 12, 13, 14, 15, 16], otherwise, they are updated offline before the tracker initiates [5, 15]. The training dataset must be representative and contain the largest number of features from both classes.

The proposed tracker EWRN uses tracking-by-detection. EWRN has tree components: an object detector composed of an ensemble of WiSARD weightless neural network (first RAM memory), a second RAM memory (RAM2) and a position predictor. At each frame, the object detector returns the target position and writes it into RAM2. The object detector acts as a classifier, recognizing a different class of bit pattern. At the first frame, the ensemble is trained with the quantized pixels (target model) inside a frame region defined manually by the operator, the selection window. At the following frames, the object detector receives as input the quantization result of all pixels inside the ROI. The object detector tries to recognize the target in a different region inside the ROI. Comparing the responses, the object detector defines the target position. The position predictor tries to estimate the object position in the next frame. The center of the ROI in the next frame will be moved to this position (Fig. 1). A failure detector analyzes vessel movement with a

This research was supported by Federal University of Rio de Janeiro UFRJ and Brazilian Naval Research Institute IPqM.

R. S. Moreira is with the Brazilian Naval Research Institute IPqM, Rio de Janeiro, RJ, Brazil (phone: 055-21-21265826; e-mail: rocypa@yahoo.com.br).

N. F. F. Ebecken is with the Department of Civil Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil. (e-mail: nelson@ntt.ufrj.br).

Kalman filter and corrects the tracking, if necessary, using FFT matching.

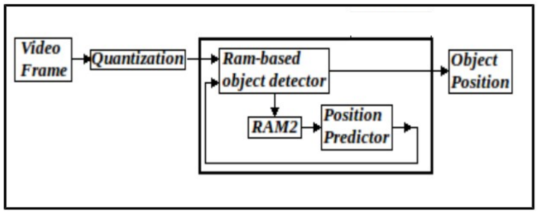


Fig.1 Object localization

II. Surveillance Systems

The main surveillance system functions are detection, processing the frame image, classification, tracking and analysis of vessel behavior (Fig. 2). The initial detector can be composed of a movement detector [1, 4, 6, 17, 18], an object detector composed of an ensemble of classifiers trained offline [5, 15] or be based on dataset models [2]. The frame regions with the highest probabilities of containing maritime vessels are handled by an image processor [1, 4, 6, 17, 18] that eliminates noise, segments the regions and detects connected components. Features are extracted from the connected components and classified as either maritime vessel or environmental features [3, 5]. The tracker locates the vessel at each frame using a model and sends trajectory information to the analyzer [3, 4, 5, 6, 17, 18, 19]. The analyzer then decides whether the vessel is a threat.

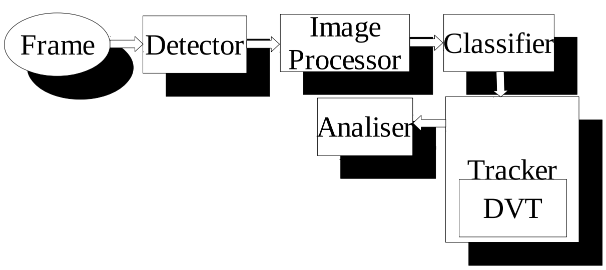


Fig.2 Components of a video surveillance system

III. State of the Art

The sea is a dynamic environment, which makes initial vessel detection and tracking a challenge [5]. Certain approaches detect the horizon line [5, 6] to decrease the search area. Some algorithms use frequency information [2], gradients [6] and histograms [20], however, recent studies have shown that detection by background subtraction and the use of normal distributions to model the environment are efficient [1, 4, 6, 17, 18]. Some authors [1] divide the frame into N regions and extract features such as entropy, energy, homogeneity and/or contrast. Each region that have different features from other regions contain a maritime vessel.

Conventional video detection and tracking algorithms are not suitable for maritime environments, which have specific characteristics that must be considered. Ocean dynamics and features are unpredictable, which makes their mathematical modeling difficult [1, 18]. An image can be corrupted by electronic component noise and by environmental clutter caused by haze, fog, rain or low lighting conditions [17, 19]. A camera can be installed over a vibrating [19] or moving platform [6]. White foam over the sea surface, sunlight reflections, illumination-changing conditions, waves, objects above the sea surface, birds and clouds over the horizon can also cause tracking failures [1, 4, 5, 6, 17, 18].

The existing algorithms to track maritime vessels include Mean-shift [21], successive clustering [6], histogram matching [5], active contour [17, 18], template matching [4, 19] and feature point matching [3]. The Kalman filter [3, 6], which is used to estimate a vessel's future position, produces good results.

IV. The Weightless Neural Network WiSARD

The early artificial neural network (ANN) researches have emerged in the 40s with the precursors Warren McCulloch and Walter Pitts. They sought to develop a mathematical neuron model base on natural human brain neurons. Encouraged by the McCulloch and Pitts neuron model, researchers try to develop, through the union of multiple neurons, a network capable of learning and recognizing the patterns provided as input.

Weightless neural network (WNN) is an important branch of research related to neural networks. In these networks, the neuron input and output are sets of binary numbers and there are no weights between neurons. The activation function of each neuron is stored in look-up tables that can be implemented as RAM memories [22]. The training of an ANN involves the adjustment of weights. Unlike these networks, the training process of the WNN is carried by modifying the words stored in the look-up tables, allowing the construction of flexible and fast training algorithms. With the use of look-up tables, any activation function can be implemented at the nodes, since any binary value can be stored in response to any set of input bits during training. The fast training speed is due to the mutual independence between nodes when a new input data is presented. The training process of an artificial neural network changes the weights values and the network behavior relative to patterns previously trained is modified.

Weightless neural network (WNN) is an important branch of research related to neural networks. In these networks, the neuron input and output are sets of binary numbers and there are no weights between neurons. The activation function of each neuron is stored in look-up tables that can be implemented as RAM memories [22]. The training of an ANN involves the adjustment of weights. Unlike these networks, the training process of the WNN is carried by modifying the words stored in the look-up tables, allowing the construction of flexible and fast training algorithms. With the

use of look-up tables, any activation function can be implemented at the nodes, since any binary value can be stored in response to any set of input bits during training. The fast training speed is due to the mutual independence between nodes when a new input data is presented. The training process of an artificial neural network changes the weights values and the network behavior relative to patterns previously trained is modified.

[24] proposed the creation of a visual system that automatically controls the movement of an offshore platform. To control the platform movements, a WiSARD network identifies and follows some reference points on the quantized images of the ship's deck. Using these reference points, the movement is modeled and the charge and discharge operations can be performed. Counting how often each RAM node address is accessed during the training phase [25], it is possible to associate the most accessed addresses with the patterns used in the training phase. Thus, models more representative for each class are created (DRASiW). To reduce the neuron saturation problem [26] proposed a modification of the original WiSARD training. A measure that combines the concepts of cross-validation and Shannon's information theory is used during the training phase to select the bests connections and consequently achieve a better performance. The surveillance system based on WiSARD [27] quantizes only some relevant regions. The quantized pixels are placed at the WiSARD input module. When an abnormal event is detected, such as people moving on railroad, an alert is generated. Three types of networks were compared to the WiSARD network relative to the neuromuscular disorders diagnosis performance [29]. Although the WiSARD network obtains a similar diagnosis performance, the training time was much smaller. The WiSARD network nodes can be easily implemented as RAM memories, and hence, the EWRN hardware implementation is straightforward.

The WiSARD neural network is built grouping a set of basic elements called discriminators. Each discriminator (Fig. 3) recognizes a different class of bit pattern. Each discriminator is a set of k RAM memory nodes (Fig.2), each addressed by N bits (N-input RAM node). Each RAM stores 2N words of one bit [22]. At an image containing k.N pixels, one quantized pixel represents one bit. k sets of N randomly chosen bits are connected to k RAM bus addresses. The discriminator response is a weightless sum of all k RAM accessed words. Binding of k.N bits at bus addresses is called input mapping. Once the input mapping is set, this remains constant during the training and classification phases.

Suppose that a WiSARD discriminator has k RAM nodes. Before the discriminator training phase, the bit "0" is written at all RAM accessed addresses. A training vector set X representing object examples of class A is prepared. Each example has k.N bits. During the training phase, each example from X is placed at the discriminator input, one by one. The bit "1" is written at all RAM accessed address of the discriminator being trained. Another training vector set Y representing object examples of class B will be used to train another discriminator. In the classification phase, one test vector is placed at the WiSARD input. The vector is classified as a class member represented by the discriminator that returns the greatest response. If a WiSARD is trained to recognize handwritten letters, each discriminator recognizes one different letter. Each discriminator is trained with quantized images containing only the letter it needs to recognize. One discriminator is trained with images containing letter A, the other discriminator is trained with images containing letter B, and so on.

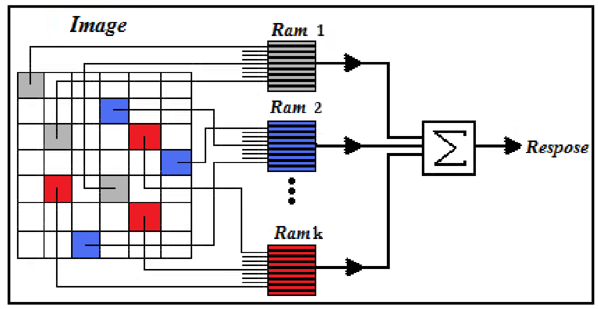


Fig. 3 WiSARD discriminator: a discriminator with k RAMs of 9bit bus address where every bit is connected to a quantized pixel. Being represented only three of nine pixels connected to each RAM.

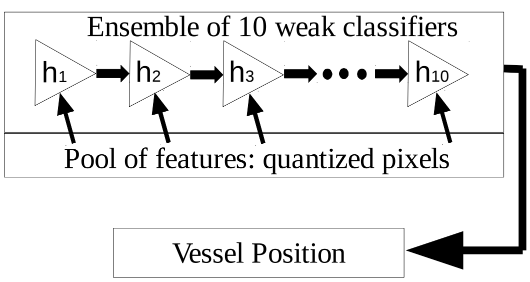
V. The EWRN Tracker

The EWRN tracker determines the vessel position in RGB videos. The DVT is composed of 10 weak classifiers. Each classifier is an independent WiSARD neural network. The features extracted from the candidate regions of each frame are quantized to form 10 binary vectors, the weak classifier's input. To increase their diversity and independence, the training and testing data sets of each weak classifier are unique. The weak classifiers are arranged in sequence (Fig. 4) to eliminate the highest number of unlikely candidate regions with the initial ensemble classifiers and are updated in two cases: when the ensemble response is lower than 90% of the ensemble response obtained one frame after the last update, and when the second highest response (R2) obtained 17x17 pixels from the highest frame response (R1) is greater than 0.9R1.

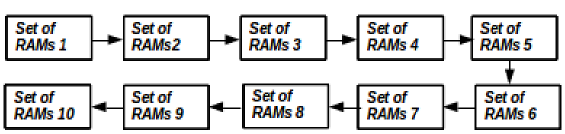
All of the discriminators of a weak classifier are trained with the same bit vector and with the quantization of features extracted from the vessel's bounding box. Each discriminator is responsible for recognizing the vessel at one unique candidate region. All the possible regions inside the ROI are tested.

The DVT detector is composed of 10 weak classifiers; therefore, 10 discriminators try to recognize the vessel at a different frame region inside the ROI (Fig. 5 and Fig. 6). The 10 bit vectors are extracted from each candidate region to be tested by the ensemble. The sum of the 10 discriminators' responses encircled in red is proportional to the probability that the vessel is in the red frame region. The sum of the 10 discriminators' responses encircled in green is proportional to the probability that the vessel is in the green pixel region. DVT determines the vessel position to be the candidate region associated with the line of discriminators that produces the highest response. If one of the 10 discriminators on the same line produces a response that is less than 30% of the highest possible response, or if the DVT analyzes the partial sum and perceives that it will be impossible to exceed the highest response previously calculated, the subsequent discriminators on this line will not test their inputs, and the candidate region is discarded to decrease the average tracking time.

The EWRN tracker stores vessel positions. The failure detector tests if the position indicated by the ensemble is valid by analyzing its movement with a Kalman filter. If the movement is unlikely to occur, the position is corrected using Fast Fourier Transform (FFT) matching. The new position is the candidate region that has the most similar FFT response to the FFT model. The failure detector stores a high confidence FFT model. This FFT model is only updated if the ensemble's response is higher than 95% of the highest possible response.



(a)



(b)

Fig.4 DVT is composed of 10 weak classifiers (a). 10 weak classifiers represented as 10 sets os RAM memories (b).

VI. Experiments and Results

The experiments were implemented with MATLAB 2011, which was installed on a Positivo laptop with Windows XP, and had the necessary computational resources to test the algorithms. The laptop had one 320 GB HD, 4 GB of RAM memory and a 1.6 GHz Intel Atom processor.

Few tracker datasets exist on the Internet, and those that are available are simple and unchallenging. For this reason, I created a dataset with 20 RGB videos with 240x320 pixels. Overall, 9 videos were captured with a digital camera and 11 were downloaded from the Internet. Each video contains one or more features that hinder the tracker. Jet skis produce a large amount of white foam and continuously change their appearance and shape (Fig. 7a). Coast proximity (Fig. 7b), low contrast (Fig. 7c) and partial occlusion from the water and other vessels (Fig. 7d) increase the difficulty of the data.

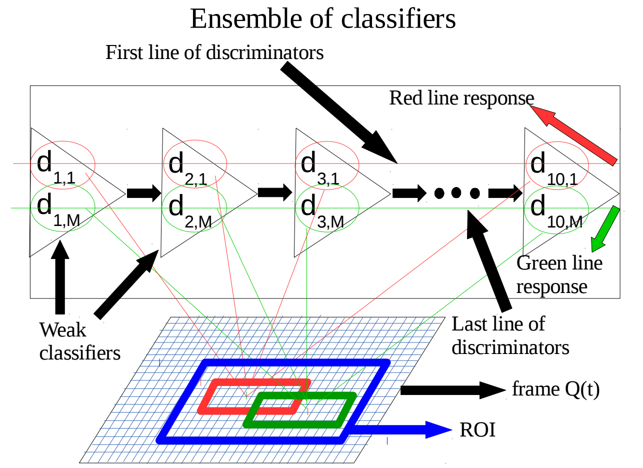


Fig.5 Each discriminator line tests one candidate region inside the ROI

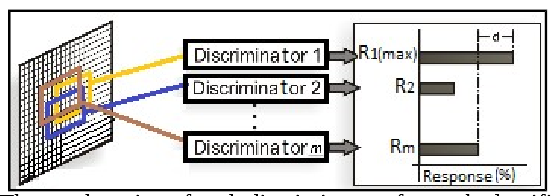
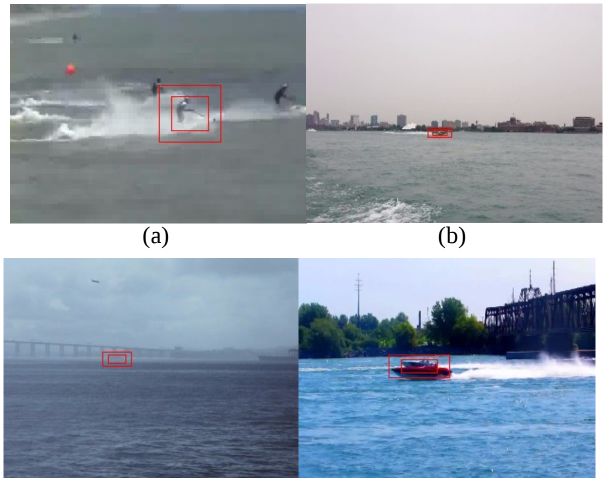


Fig.6 The seach region of each discriminator of a weak classifier has a different color (left). Responses of all discriminators (right).

A. Quality Measures

The tracking quality was evaluated using quantitative measures. The vessel bounding box was manually defined and served as a reference. Three quality measurements were based on the position differences between the reference bounding box, BBR, and the bounding box determined by the tracking algorithms, BBD. The bounding box overlap degree (OD) between the BBR and BBD was defined to be the ratio between their intercession and union [6] (2). The temporal dice proposed by [6] is the frame quantity where the OD is greater than a threshold, T. The threshold used in the present study was T=0.5. This measure was altered to consider all of the video frames (3). [13] and [19] utilized the tracking rate, TR (4), which measures the number of frames for which the tracker precision is high. The other quality measures used in this study were the average tracking and update time, the number of videos whereupon the tracker did not fail, the average number of updates and the number of videos wherein the TDM was greater than 70%.



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

Fig.7 Dataset videos

$$OD=\frac{area(BBD\cap BBR)}{area(BBD\cup BBR)} \quad (2)$$

$$TDM=\frac{\sum_{i=1}^{N} GS(i)}{\text{Quantity of frames}} \quad (3)$$

$$TR=\frac{N(GS>L)}{\text{Quantity of frames}} \quad (4)$$

B. Results

A total of 6 experiments were implemented to determine the factors that improve tracking performance in maritime environments. There are no in-depth studies on this subject in the literature. In total, 4 state of the art trackers were used. The tracker proposed by [5] applies histogram matching using the Bhattacharyya coefficient in HSV color space. [2] proposed a tracker that applies FFT matching to the pixel gradient. [19] determines the vessel position with template matching and uses a particle filter to increase the tracker's robustness to partial occlusions. The tracker proposed by [6] extracts frame clusters and defines the vessel position as the nearest cluster to the last vessel position. The Otsu algorithm was used to calculate a threshold to quantize the pixel gradients.

Experiment 1 aimed to define which color component provides the most efficient tracker. A total of 12 color components were tested: gray and the components R, G, B, Y, Cb, Cr, H, S, V, I, and Q belonging to the color spaces RGB, YCbCr, HSV and YIQ. The component V provides the best tracking quality because this component was found to be superior in 5 of the 8 quality measures.

Experiment 2 was intended to define which feature provides the best tracking. A total of 19 features were compared: the V color component [5], the average [6], the variance [30], the entropy [1, 17], 5 Haar Wavelet feature types [16] and 10 forms of gradient extraction [17]. Entropy, wherein each pixel value is replaced by the entropy of the pixels in a 3x3 window, was found to be the most relevant feature because it was found to be superior in 5 of the 8 quality measures. Each pixel value is replaced by the entropy calculated with the pixels inside a 3x3 window.

The objective of the third experiment was to compare 15 quantization methods: the edge detectors proposed by Sobel, Prewitt and Roberts [7, 31], the Otsu method [6], the ROI average value used as a threshold, a decision tree [32], a Bayes filter [19], an artificial neural network [33], an SVM [30], a classifier based on the Normal distribution [11, 13, 14, 15] and 5 ensembles made up by the last 5 classifier types mentioned above. Each classifier was trained with features from the CVT class, which were extracted from inside the bounding box, and features from the CNVT class, which were extracted far from the vessel bounding box but within the ROI. The ensembles quantized the pixels by voting without weights. It was concluded that the decision tree provides the best quantization quality because it was superior in 4 of the 8 quality measures.

Experiment 4 aimed to discover the WiSARD configuration that provides the most efficient tracking. Two configurations were tested: one WiSARD network and an ensemble of 10 WiSARD networks. Three quantization algorithms were compared: the decision tree (Experiment 3) and two proposed algorithms using 1) pixel pair value comparisons (Fig. 8) and 2) thresholds applied to the absolute differences of pixel pair values. The first algorithm compares the values of the pixel pairs connected to successive RAM addressing pins. If the first pixel value is greater than the second one, the first pixel value is quantized to 1, otherwise, the first pixel value is quantized to 0 (Fig. 8). The second algorithm compares the threshold with the absolute difference of the pixel pair values and modifies the first pixel values exactly as the first quantization method. RAM memories with 3, 7, 11 and 15 addressing pins and RAM generalization were evaluated. 7-RAM provides better tracking. Without generalization, 3-RAM is more efficient because the smaller the RAM size is, the larger the discriminator response is, which allows the discriminator to better compensate for the lack of generalization. The quantization method that performs best is the decision tree; however, the quantization that compares pixel values achieves nearly the same efficiency as the decision tree. Generalizing 1/3 of the RAM addressing pins produces a tracker that makes fewer errors. Without generalization, a vessel model has to be more precise because the tracker fails in the frames wherein the vessel model does not localize it with a high response. If several bits are generalized, the tracker loses its capacity to differentiate the features from candidate regions that are far away from other regions. The RSWC7_2 VA tracker performed best. It is composed of one WiSARD network with 7-RAM memory generalizing 2 bits. The pixels are quantized by comparing values. The ensemble with the identical configuration performs in a similar way. The ensemble is 11 ms faster, but it is less precise (its TDM and TR are 2% lower compared with the single WiSARD network).

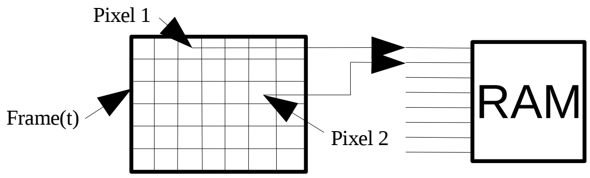


Fig. 8 if pixel 1 > pixel 2 than pixel 1 is quantized to 1. Pixel 1 is quantized to 0 otherwise.

Experiment 5 was intended to select 4 classic trackers for the last experiment. Ensembles composed of 5 classifier types were compared: decision tree [32], Bayes filter [19], artificial neural network [33], SVM [30] and classifiers based on a Normal distribution [11, 13, 14, 15]. The following 5 algorithms were tested for selecting 10 weak classifiers online from a pool with 20 weak classifiers to form the ensemble: Bagging [13, 16], Online Gentle Boosying [13], Online AdaBoosting [14], Online Weighted MIL Boosting [34] and the selection of the 10 best weak classifiers from the pool [13]. The FFT matching [2], template matching [5], histogram matching [19], feature point matching [3], mean-shift [21] and successive clusterization [1, 6] algorithms were also tested. To compare the ensembles, the CSVSME tracker, composed of 10 SVMs, and the CDNSME tracker, composed of 10 normal distributions, were both updated online by selecting the 10 best weak classifiers to form the ensemble, performed best. The CSVSME tracker was more precise, with a TR=0.83 (5% higher than the CDNSME tracker and 2% higher than the ensemble of SVM classifiers updated with Online AdaBoosting) and TDM=0.62 (3% higher than the CDNSME tracker and 2% higher than the ensemble of SVM classifiers updated with Online AdaBoosting). The CDNSME ensemble updated with an average time of only 18 ms, which is nearly 11 ms faster than the ensemble of decision trees updated with Online Gentle Boosting. The RF tracker, which locates the vessel using FFT matching, was the best single classic tracker (not an ensemble). The RF tracker's average tracking time is only 11 ms and it is more precise than the other classic tracking algorithms (TR=81%, which is 2% higher than template matching, and TDM=69%, which is 1% higher than template matching). For the above reasons, the failure detector proposed in this article was implemented with FFT matching.

Experiment 6 aimed to compare the 2 best trackers from experiment 5 with 4 state of the art trackers and 4 trackers based on the WiSARD weightless neural network. The vessel models and ensembles performed best when they were updated only as necessary. Updating at each frame causes drift, but a lack of updating causes tracking failure. The RSWC7_2VA tracker performed best, which suggests that the WiSARD neural network can be used to track maritime vessels. Despite its average tracking time being 15 ms longer than the average RF tracking time, the RSWC7_2VA tracker was more precise (TR=90%, which is 7% higher than CSVSME, and TDM=76%, which is 7% higher than FFT matching). The average execution time (tracking + model update + frame processing times) was 55 ms.

It was necessary to modify the RSWC7_2VA algorithm to cause fewer tracking failures. The tracker failed on 5 of 20 dataset videos. After exhaustive testing, it was concluded that the following modifications were necessary: pixels associated with the same RAM memory cannot be next to each other, the V color component must be used when the frame contrast is low (when the variance sum of the R, G and B pixel color components inside the ROI < 0.225) and the failure detector must be used to (section 5) avoid tracking failures in videos wherein the water nearly occludes the entire vessel. After modifying the 4 WiSARD trackers, the tracker composed of 10 WiSARD neural networks with 7-RAM memory generalizing 2 bits performed best, and represents the final configuration of the EWRN tracker. The pixels were quantized by comparing pair pixel values (an innovation of the present study). The EWRN tracker analyzes 17 frames per second and fails in only 1 video due to a long-lasting total occlusion. The quality measurements related to the model's number of updates and average time decreased because the FFT model used by the failure detector was also being updated. The precision measurements TDM and TR increased nearly 10% (TR=98% and TDM=84%). The average tracking time decreased 5 ms because the ensemble eliminates several candidate regions proposed by the initial weak classifiers.

VII. Conclusion

The EWRN tracker proposed in this paper is able to track maritime vessels with different shapes, appearances and maneuverabilities in dynamic environments. It is robust to partial occlusions with water, other maritime vessels and objects present on the ocean's surface. The tracker reaches a speed of 17 frames per second. EWRN does not fail in 19 of 20 dataset videos and all of its precision quality measurements indicate that it is superior to other trackers.

A long-lasting total occlusion made the EWRN tracker fail (OD < 0.5) because the vessel was occluded in approximately 30 frames and reappeared outside the ROI. In this case, it was necessary to develop an initial detector (Fig. 1) to assist the tracker that did not require any previous information about the environment. The frame image processing applied by the EWRN tracker is simple. The RGB color space is converted to greyscale, bicubic interpolation is applied to the pixels to diminish the influence of noise and the pixels are quantized to create bit vectors used to train and test the WiSARD weak classifiers. The EWRN tracker outperforms 4 state of the art trackers. The main reason for this is that the video dataset chosen was very challenging.

In future work, total long-lasting occlusions must be handled. It is not a trivial task to differentiate the changes in the vessel's state and occlusions, thus, further research is required. Additional RAM memory sizes and quantities of generalized bits can be tested. It is impractical to test all the possible WiSARD configurations if we also consider all the possible features that can be extracted using color components, quantization algorithms and supplementary variables. Faster programming languages such as C/C++ and the use of a computer with superior processing capability as the GPU are required to analyze the improvements that can be found by extracting and applying more than one feature type per frame. Adjusting the size of the bounding box is difficult because the maritime environment is dynamic. Early occlusion detection of maritime vessels with similar colors can be implemented with multiple object tracking techniques. The WiSARD neural network is not widely used and is an additional tool that can be used to develop new trackers.

ACKNOWLEDGMENT

This research was supported by Federal University of Rio de Janeiro UFRJ and Brazilian Naval Research Institute IPqM. The authors wish to thank the professors from the Department of Civil Engineering from UFRJ and our colleagues from IPqM who provided insight and expertise that greatly assisted the research.

REFERENCES

[1] Robert-Inácio, F., Raybaud, A. & Clément, É., "Multispectral Target Detection and Tracking for Seaport Video Surveillance," in *2007 Proceedings of Image and Vision Computting*, pp. 169-174.

[2] Sulivan, M. D. R. & Shah, M., "Visual Surveillance in Maritime Port Facilities," in *2008 Proceedings of SPIE*, vol. 6978, no. 11, pp. 1-8.

[3] Teutsch, M. & Kruger, W., "Classification of Small Boats in Infrared Images for Maritime Surveillance," in *2010 Proceedings of Waterside Security Conference*, pp. 1-7.

[4] Hu, W.-C., Yang, C.-Y. & Huang, D.-Y., "Robust Real-Time Ship Detection and Tracking for Visual Surveillance of Cage Aquaculture," *Journal of Visual Communication & Image Representation*, vol. 22, no. 6, pp. 543-556, 2011.

[5] Bloisi, D., Locchi, L. & Fiorini, M., "Automatic Maritime Surveillance with Visual Target Detection," in *2011 Proceedings of the International Defence and Homeland Security Simulation Workshop*, pp. 141-145.

[6] Fefilatyev, S., Goldgof, D. & Shceve, M., "Detection and Tracking of Ships in Open Sea with Rapidly Moving Buoy-Mounted Camera System," *Ocean Engineering*, vol. 54, no. 1, pp. 1-12, 2012.

[7] Moreira, R. D. S, Ebecken, N. F. F. & França, F. M. G., "Tracking Targets in Sea Surface with the WiSARD Weightless Neural Network," in *2013 Proceedings of the 1st BRICS Countries Countries Congress*.

[8] Tandl, M., Stark, T., Erol, N. E., Loer, F., Kecskeméthy, A., "An Object-Oriented Approach to Simulating Human Gait Motion Based on Motion Tracking," *International Journal of Applied Mathematics and Computer Science*, vol. 19., no. 3, 2009.

[9] Michatek, M., Dutkiewicz, P., Kietczewski, M., Pazderski, D., "Trajectory Tracking for a Mobile Robot with Skid-Slip Compensation in the Vector-Field-Orientation Control System," *International Journal of Applied Mathematics and Computer Science*, vol. 19, no. 4, 2009.

[10] Kasprzak, W., Wilkowski, A., Czapnik, K., "Hand Gesture Recognition Based on Free-Form Contours and Probabilistic Inference," *International Journal of Applied Mathematics and Computer Science*, vol. 22, no. 2, 2012.

[11] Wu, S., Zhu, Y. & Zhang, Q., "A New Robust Visual Tracking Algorithm Based on Transfer Adaptive Boosting," *Mathematical Methods in the Applied Sciences*, vol. 35, no. 17, pp. 2133-2140, 2012.

[12] Liu, R., Cheng, J. & Lu, H., "A Robust Boosting Tracker with Minimum Error Bound in a Co-Training Framework," in *2009 IEEE 12th. International Conference on Computer Vision (ICCV)*, pp. 1459-1466.

[13] Wang, D., Lu, H. & Xiao, Z., "Fast and Effective Color-Based Object Tracking by Boosted Color Distribution," *Pattern Analysis Application*, vol. 16, no. 4, pp. 647-661, 2013.

[14] Grabner, H. & Bischof, H., "On-Line Boosting and Vision," in *2006 Proceedings of Computer Vision and Pattern Recognition*, vol. 1, pp. 260-267.

[15] Grabner, H., Leistner, C. & Bischof, H., "Semi-Supervised On-Line Boosting for Robust Tracking," in *2008 Proceedings of European Conference on Computer Vision*, pp. 234-247.

[16] Babenko, B., Yang, M. & Belongie, S., "Robust Object Tracking with Online Multiple Instance Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619-1632, 2011.

[17] Frost, D., Tapamo & J.-R., "Detection and Tracking of Moving Objects in a Maritime Environment with Level-set with Shape Priors," *EURASIP Journal on Image and Video Processing*, vol. 1, no. 42, pp. 1-16, 2013.

[18] Szpak, Z. L. & Tapamo, J. R., "Maritime Surveillance: Tracking Ships Inside a Dynamic Background Using a Fast Level-Set," *Expert System with Applications*, vol. 38, no. 6, pp. 6669-6680, 2011.

[19] Bacho, A. K., Roux, F. & Nicolls, F., "An Optical Tracker for The Maritime Environment," in *2011 Proceedings of SPIE*, vol. 8050, pp. 1-10.

[20] Smith, A. A. & Teal, M., "Identification and Tracking of Maritime Objects in Near Infrared-Image Sequences for Collision Avoidance," in *2007 7th International Conference on Image Processing and It's Applications*, vol. 1, pp. 250-254.

[21] Bibby, C. & Reid, I. D., "Visual Tracking at Sea," in *2005 Proceedings of the International Conference on Robotics and Automation*, pp. 1841-1846.

[22] Ludermir, T. B., Carvalho, A. P. L., Braga, A. P., De Souto, M. C. P., "Weightless Neural Models: A Review of Current and Past Works," *Neural Computing Surveys 2*, pp. 41-61, 1998.

[23] Aleksander, I., Thomas, W. V. & Bowden, P. A., "WiSARD: a radical step forward in image recognition," Sensor Review, vol. 4, no. 3, pp. 120–124, 1984.

[24] França, H. L., Silva, J. C. & Lengerke, O., "Movement Pursuit Control of an Offshore Automated Platform Via a RAM Based Neural Network," in *2010 11th International Conference Control Automation Robotics & Vision – ICARCV*, pp. 2437-2441.

[25] Grieco, B. P. A., Lima, P. M. V. & De Gregorio, M., "Producing Pattern Examples From Mental Images," *Neurocomputing*, vol. 73, no. 7, pp. 1057–1064, 2010.

[26] Tarling, R. and Rohwer, R., "Efficient Use of Training Data in the N-tuple Recognition Method," *IEEE Electronics Letters*, vol. 29, no. 24, pp. 2093-2094, 1993.

[27] De Gregorio, M., "The Agent WiSARD Approach to Intelligent Active Video Surveillance Systems," in *2007 Proceedings of IAPR Conference on Machine Vision Applications - MVA*, pp. 331-334.

[28] Coraggio, P. & Gregorio, M. D., "WiSARD and NSP for Robot Global Localization," in *2007 International Work-conference on the Interplay between Natural and Artificial Computation IWINAC(part II)*, vol. 4528, pp. 449–458.

[29] Pattichis, C. S., "A Hybrid Neural Network Electromyographic System: Incorporating the WISARD Net," in *1994 IEEE International Conference on Neural Networks*, vol. 6, pp. 3478-3483.

[30] Zhou, J., Xu, T. & Gan, J., "Feature Extraction Based on Local Directional Pattern with SVM Decision-Level Fusion for Facial Expression Recognition," *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 2, pp. 101-110, 2013.

[31] Moreira, R. D. S, Ebecken, N. F. F. & Livernet, F., "A Survey on Video Detection and Tracking of Maritime Vessels," *International Journal of Research and Reviews in Applied Sciences (IJRRAS)*, vol. 20, no. 1, pp. 1-15, 2014.

[32] Breiman, L., "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[33] Leung, H., Dubash, N. & Xie, N., "Detection of Small Objects in Clutter Using a GA-RBF Neural Network," *IEEE Transactions on Aerospace and Electronics Systems*, vol. 38, no. 1, pp. 98-118, 2002.

[34] Zhang, K. & Song, H., "Real-Time Visual Tracking Via Online Weighted Multiple Instance Learning," *Pattern Recognition*, vol. 46, no. 1, pp. 397-411, 2013.