Saliency-based Detection for Maritime Object Tracking

Tom Cane and James Ferryman
Computational Vision Group, School of Systems Engineering
University of Reading, Whiteknights
Reading RG6 6AY, UK
{ t.cane | j.m.ferryman }@reading.ac.uk

Abstract

This paper presents a new method for object detection and tracking based on visual saliency as a way of mitigating against challenges present in maritime environments. Object detection is based on adaptive hysteresis thresholding of a saliency map generated with a modified version of the Boolean Map Saliency (BMS) approach. We show that the modification reduces false positives by suppressing detection of wakes and surface glint. Tracking is performed by matching detections frame to frame and smoothing trajectories with a Kalman filter. The proposed approach is evaluated on the PETS 2016 challenge dataset on detecting and tracking boats around a vessel at sea.

1. Introduction

Maritime piracy continues to place a huge economic and human cost on commercial shipping around the world [1]. The most effective protection for ships is a proper lookout to maximise early warning of a potential attack, allowing time for the crew to prepare accordingly [3]. Radar and crew members with binoculars represent the state of the art technology available to commercial fleets. However, the navigation radar available on ships does not perform well with small, fast-moving objects [24] such as the 'skiffs' used by pirates, and crew members become fatigued after maintaining a lookout for a long period.

Automated visual surveillance offers a new sensing modality for ships which could operate continuously without human intervention and increase the early detection of piracy threats. This is also one of the themes of the PETS 2016 workshop [2] in which one of the challenges is to accurately detect and track mobile objects around a vessel as the first step towards deciding if their activities are normal behaviour, non-dangerous abnormalities, or criminal activity. However, the maritime environment poses a characeristic set of challenges for visual detection and tracking which cause many methods developed for land-based use to perform poorly. The sea presents a highly dynamic background caused by waves, reflections and weather conditions. A wide variety of objects may be encountered, ranging from small buoys and watercraft, to large commercial shipping tankers, so alogrithms must be able to handle a broad range of object profiles. Finally, methods must be robust to camera motion because they are mounted on a mobile waterborne platform.

Many of these difficulties can be addressed by using thermal cameras [19, 22, 23]. Unfortunately, the cost of thermal sensor hardware is prohibitive for many applications. Visible light cameras offer a more affordable alternative which could provide surveillance coverage of a larger region and complement other available sensors, such as radar. This motivates further research into improving their performance for operation in maritime environments. A number of recent studies have used a visual saliency approach [4, 17, 14, 20]. These methods aim to mimic the low-level human visual attention mechanism which is very efficient at locating the most 'interesting' (salient) regions in an image for further high-level processing. In this paper, we propose a new tracker which uses a variant of the saliency method used in [20] to address some of the challenges of maritime scenes. In particular, the proposed approach does not make any assumptions about object size or appearance, is robust to dynamic background (in particular wakes and specular reflections) and generalises to scenes with different viewpoints, backgrounds and conditions. Finally, we evaluate its performance on the PETS 2016 maritime dataset.

2. Related work

Recent approaches for detection and tracking in maritime environments [5, 7, 11, 15, 19, 21, 22, 23] include the use of background modelling and subtraction (which can perform poorly when the background is highly dynamic)

This work was supported by funding from the EU 7th Framework Programme for research, development and demonstration under Grant Agreement No. 607567.

978-1-5090-1437-8/16 $31.00 © 2016 IEEE
DOI 10.1109/CVPRW.2016.159

[7, 11], learning of object profiles (and therefore need substantial training data and are prone to overfitting for particular target classes) [15, 21], are limited to thermal images (and therefore require expensive hardware for an operational surveillance system) [19, 22, 23], or rely on static cameras and are therefore not suitable for deployment at sea [5].

Salient object detection has been widely studied in the literature [6]. Most commonly, the methods are applied to individual images in which there is a single, well-centered salient object to detect. A number of works have implemented saliency methods to detect objects in maritime images and video [4, 17, 14, 20]. However, in order to reliably detect objects, additional methods have been used in conjunction with the saliency step, such as background subtraction using a Mixture of Gaussians [17] or learning a background classifier [4], learning weights for combining features in the saliency map [14], and Robust Principal Components Analysis (RPCA) [20] to identify foreground and background from the separated sparse and low-rank matrices. RPCA is computationally expensive and background subtraction methods can only cope with a limited level of background variation, whilst learning-based approaches involve a substantial training effort and do not generalise well.

3. System description

The proposed tracker (Fig. 1) creates a saliency map for each frame and performs adaptive hysteresis thresholding to locate the salient regions corresponding to potential objects. The list of candidate objects is filtered using some basic constraints and surviving object detections are matched from frame to frame using the Hungarian algorithm. Finally, the tracks are smoothed using a Kalman filter.

3.1. Modified Boolean Map Saliency

The Boolean Map Saliency (BMS) method [25] (used in [20]) exploits the visual property of surroundedness whereby objects in an image are more salient, the more surrounded they are by background regions in a given feature space. In principle, any feature channels can be used (colour, orientation, motion, etc.), but the method in [25] uses the CIELAB colour channels. The colourspace is first rectified using a whitening step, then each of the channels, L, A and B, are normalised to the range $[0, 255]$ and binary thresholded at intervals with a step size of δ. This yields a set of N binary images (Boolean maps), $\{B_i\}_{i=1}^{N}$. An activation map is then created for each Boolean map by identifying the surrounded regions. A black region is surrounded in B_i if it is enclosed by a white region and *vice versa*. The activation map, A_i, is created by setting pixels to 1 if the corresponding pixel is in a surrounded region of B_i, and setting 0 elsewhere. The set of activation maps, $\{A_i\}_{i=1}^{N}$, is then normalised in order to emphasise maps with small activated regions. First, each activation map is split into two sub-activation maps, A_i^+ and A_i^-, according to

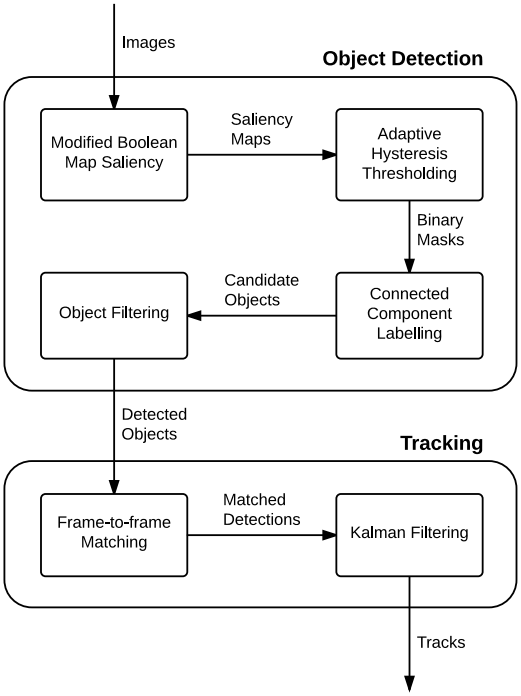


Figure 1. Block diagram presenting the different stages of the proposed algorithm.

$$A_i^+ = A_i \wedge B_i, \tag{1}$$

$$A_i^- = A_i \wedge \neg B_i, \tag{2}$$

where $\wedge$ represents pixel-wise logical AND between two binary maps and $\neg B_i$ is the negation (logical NOT) of B_i. Both sub-activation maps are dilated with a kernel K_{D1} of size $D1$ and divided by their L2-norm. This serves to emphasise clumps of small activated regions whilst reducing the importance of small, scattered regions. The normalised activation map, $\bar{A}_i$, is therefore calculated as

$$\bar{A}_i = \frac{A_i^+ \oplus K_{D1}}{||A_i^+||_2} + \frac{A_i^- \oplus K_{D1}}{||A_i^-||_2}, \tag{3}$$

where $\oplus$ represents the morphological dilation operation. The final saliency map, S, is found by taking the average of all the normalised activation maps and performing a second dilation operation followed by Gaussian smoothing

$$M = \frac{1}{N} \sum_{i=1}^{N} \bar{A}_i, \tag{4}$$

$$S = G_\sigma * (M \oplus K_{D2}), \tag{5}$$

where K_{D2} is a dilation kernel of size $D2$ and G_σ is a Gaussian kernel with standard deviation σ.

One of the weaknesses of the BMS method [25] when applied specifically to maritime scenes is a tendency to

highlight the wakes of the boats and specular reflections (glint) in the water. To counter this, we propose a modification which is designed to suppress these features of the background. Instead of using the CIELAB colourspace, we propose the use of broadly-tuned, intensity-decoupled red, blue and green colour channels used in earlier, biologically-inspired salient object detection approaches [10].

First, hue is decoupled from intensity using the method in [10] by dividing the red, green and blue channels of the image (r, g, and b) by the intensity channel (I). The channels are set to zero for pixels where I is less than $1/10$ of its maximum value, $I_{\max}$, to represent the fact that hue variations are not perceivable at low luminance. The intensity channel is derived from the RGB image as

$$I = \frac{r + g + b}{3} \tag{6}$$

A broadly-tuned colour channel is one that gives maximum response for the pure, fully-saturated hue for which it is tuned, and yields a zero response for black and white [8]. The three broadly-tuned colour channels, $\bar{R}$, $\bar{G}$ and $\bar{B}$, are defined as:

$$\tilde{R} = \begin{cases} \dfrac{r - (g+b)/2}{I} & \text{if } I > \dfrac{I_{\max}}{10}, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

$$\tilde{G} = \begin{cases} \dfrac{g - (r+b)/2}{I} & \text{if } I > \dfrac{I_{\max}}{10} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

$$\tilde{B} = \begin{cases} \dfrac{b - (r+g)/2}{I} & \text{if } I > \dfrac{I_{\max}}{10}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

In the proposed modified version of BMS, $\bar{R}$, $\bar{G}$ and $\bar{B}$ are used instead of the L, A and B channels of the CIELAB colourspace. The colour whitening step is also omitted.

3.2. Adaptive Hysteresis Thresholding

Once the saliency map has been generated, it is binary thresholded to extract candidate object regions. Setting a fixed value for the threshold would not generalise well for different sequences, so the threshold is set to the 99$^{\text{th}}$ percentile of the saliency map. This captures the most salient points in the image, but is likely to miss true object regions which were still highly salient but not in the top 1%. However, a lower threshold is likely to introduce more false detections. Hysteresis thresholding is a common way to address this and is used here for this purpose, as it has been in other recent maritime works [12, 16]. Two thresholds are set; an upper and a lower. The saliency map is binary thresholded at the upper value and the flood-fill algorithm is then used to grow regions to add connected pixels which are above the lower threshold. In the proposed approach, the upper and lower thresholds are set to the 99$^{\text{th}}$ and 98$^{\text{th}}$ percentiles, respectively.

3.3. Object Extraction and Filtering

Candidate objects are extracted from the binary mask by labelling connected components and computing bounding boxes. The set of candidate objects is likely to contain some false detections from the background, so filtering is carried out by applying some simple constraints. False detections from glint tend to have very small bounding boxes. However, objects on the horizon also have small bounding boxes, so setting a global minimum allowable size would not be suitable. Instead, the minimum allowable size is calculated as a function of the distance from the base of the image to the horizon. Bounding boxes with a height less than T_h are removed.

$$T_h = h_{\max} - (h_{\max} - h_{\min}) \left(\frac{H - y_c}{\alpha H} \right)^{\lambda}, \tag{10}$$

where H is the image height, α is the approximate position of the horizon line from the bottom of the image as a proportion of image height, and λ is the fall-off rate. α and λ could be set dynamically using a horizon detection method but in this study, λ is fixed empirically at 1 and α is set for each sequence as per Table 2.

3.4. Tracking

A simple tracking framework is implemented to assess the utility of the saliency approach as a basis for object detection and tracking. In each frame, new detections are assigned to detections and tracks from the previous frame using the Hungarian algorithm [13, 18]. The cost matrix is completed by calculating the Euclidean distance, d, between the centroids of each pair of bounding boxes

$$d = ||\mathbf{p}_c(j) - \mathbf{p}_c(i)||_{L2}\,, \tag{11}$$

where $\mathbf{p}_c(i)$ and $\mathbf{p}_c(j)$ are the centroids of box i and j, respectively. Gating is implemented by introducing a maximum distance threshold for assignment, $d_{\max}$, such that matches are discarded if $d > d_{\max}$.

Matches between new detections in two consecutive frames triggers the creation of a new track which is managed by a standard constant velocity Kalman filter with the following state space and process models:

$$\mathbf{x} = \begin{bmatrix} x_c & y_c & \dot{x}_c & \dot{y}_c & w & h & \dot{w} & \dot{h} \end{bmatrix}^T \tag{12}$$

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_k \tag{13}$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{w}_k \tag{14}$$

$$\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}) \tag{15}$$

$$\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}) \tag{16}$$

where (x_c, y_c) and $(\dot{x}_c, \dot{y}_c)$ are the position and velocity of the bounding box centroid, and w, h, $\dot{w}$ and $\dot{h}$ are the width and height of the bounding box and their respective rates of change. The transition and observation matrices, $\mathbf{F}$ and $\mathbf{H}$, are taken as

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{17}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{18}$$

and the process and observation noise covariances, $\mathbf{Q}$ and $\mathbf{R}$, are initialised with the identity matrix.

When new detections are assigned to existing tracks, the track is updated by estimating the state using the new observation. If a track is not assigned a new detection in the frame, the new bounding box is predicted by the Kalman filter. The filter is allowed to predict up to 5 frames without a new matched detection before the track is terminated.

4. Experimental results

This section presents the experimental results and validation of the proposed tracker on the PETS 2016 maritime dataset. The experimental setup is described, followed by the evaluation and analysis of the results.

4.1. Experimental setup

The visual saliency methods from the literature [4, 17, 14, 25] and the proposed modified method were evaluated on a common maritime dataset to compare their detection performance in maritime scenes. As the aim was to assess the performance of the saliency method only, without additional object detection steps, each algorithm from the literature was implemented up to the saliency map stage (i.e. no background modelling, etc.).

Pixel-level groundtruth was created for each of the IPATCH low-level tracking challenge sequences. A sub-sequence of 500 frames was selected from each sequence for this purpose. Each sub-sequence was chosen so that it included as much variation in object size and appearance as possible and there was at least one object visible in every frame. A mask was also applied to each sequence to remove regions of the host vessel which were visible in the scene. The sub-sequences are described in Table 1.

Table 1. Sub-sequences for saliency map evaluation

PETS IPATCH Sequence	Sub-sequence Frames	Total Target Area Range (pixels)
Sc2a_Tk1-CAM11	994 – 1343	39 – 20573
Sc2a_Tk1-CAM12	571 - 1070	62 – 5583
Sc3_Tk2-CAM14	4838 – 5337	205 – 31081

4.2. Evaluation metrics

For quantitative evaluation, three metrics were selected from the salient object detection literature to evaluate key areas of detection performance: Mean Absolute Error (MAE), Precision-Recall (PR) curve, and Receiver Operating Characteristic (ROC) curve [6]. The PR curve is an important complement to the ROC curve, especially when dealing with highly skewed datasets [9] such as the PETS IPATCH sequences.

Mean Absolute Error (MAE) measures the average deviation between the saliency map and the groundtruth object regions. It is therefore an indication of how well the saliency map models the saliency of the scene. The MAE score for frame n is computed as the average absolute pixel-wise difference between the saliency map, S, and the binary groundtruth mask, G, both scaled to the range $[0, 1]$

$$\text{MAE}_n = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} |S_n(i,j) - G_n(i,j)|, \tag{19}$$

where $S_n(i,j)$ and $G_n(i,j)$ are the saliency and groundtruth values of pixel (i,j) in frame n, and W and H are the image width and height. In addition, the mean MAE score, $\widehat{\text{MAE}}$, is calculated for a sequence by averaging over all frames.

$$\widehat{\text{MAE}} = \frac{1}{N} \sum_{n=1}^{N} \text{MAE}_n, \tag{20}$$

where N is the number of frames in the sequence.

The Precision-Recall curve plots the fraction of the salient pixels that correspond to salient object regions (Precision) against the fraction of the salient object pixels that were correctly identified in the saliency map (Recall). S is scaled to the range $[0, 255]$ and binary thresholded at a range of values, t, to create a set of binary maps, $\{\bar{S}_t\}_{t=1}^{255}$. For each $\bar{S}_t$, Precision and Recall are calculated as

$$\text{Precision}_t = \frac{|\bar{S}_t \cap G|}{|\bar{S}_t|} \tag{21}$$

$$\text{Recall}_t = \frac{|\bar{S}_t \cap G|}{|G|} \tag{22}$$

where $|.|$ is the set cardinality operator which denotes the number of pixels in the map equal to 1. The ROC curve, which plots True Positive Rate (TPR) against False Positive Rate (FPR), is calculated in a similar manner using the same set of threshold values.

$$\text{TPR}_t = \frac{|\bar{S}_t \cap G|}{|G|} \tag{23}$$

$$\text{FPR}_t = \frac{|\bar{S}_t \cap G|}{|\bar{S}_t \cap G| + |\neg\bar{S}_t \cap \neg G|} \tag{24}$$

where $\neg$ denotes the inverse of the binary map. The PR and ROC curves for all frames are averaged to create a single curve for the sequence. The area under the curves is also calculated for numerical comparison.

For qualitative anaylsis of the tracking performance, the proposed tracker was run on sequences from the PETS 2016 [2] challenge. All the IPATCH sequences listed in the low-level video analysis section on detection and tracking have been processed (3 sequences) plus selected sequences from the mid-level category (2 sequences) to provide contrasting detection and tracking challenges. The processed sequences are listed in Table 2.

Table 2. Processed PETS IPATCH sequences. Key. DB: dynamic background; TM: translatory camera motion; ST: single target; MT: multiple targets; OT: occluding targets; SC: scale changes.

PETS IPATCH Seq.	Horizon (α)	No. Frames	Challenges
Sc2a_Tk1-CAM11	0.96	3646	DB, TM, MT
Sc2a_Tk1-CAM12	0.84	3857	DB, TM, MT, OT, SC
Sc3_Tk2-CAM14	0.97	5425	DB, TM, MT, SC
Sc1_Tk3-CAM12	0.79	1659	DB, TM, ST, SC
Sc3b_Tk1-CAM14	0.97	1729	DB, TM, MT, SC

The algorithm parameters were fixed for all sequences at the following values: upper hysteresis threshold $T_1 = 99^{\text{th}}$ percentile, lower hysteresis threshold $T_2 = 98^{\text{th}}$ percentile, minimum allowable bounding box height at horizon $h_{\text{min}} = 0.0$ pixels, minimum allowable bounding box height at base of image $h_{\text{max}} = 60$ pixels, fall-off rate $\lambda = 1$, track time-to-die $TTTD = 5$ frames. The BMS method parameters were kept fixed at the values reported in [25]. The horizon line was set for each sequence individually. The values of α are listed in Table 2. The saliency methods and tracker were implemented in Python and run on a MacBook Pro with 2.6GHz Intel Core i7 processor and 16GB RAM.

4.3. Evaluation and analysis

The numerical results of the saliency method analysis are presented in Fig. 2 and Table 3. The proposed maritime-tailored variant of BMS achieves the best MAE performance on all sequences and better or comparable PR and ROC performance when compared with BMS [25] and other methods [4, 17, 14]. Visual comparison of the saliency maps can be made in Fig. 3. The qualitative analysis confirms that the tracking algorithm is able to handle scale changes during tracking and the broadly-tuned, intensity-decoupled RGB channels help suppress unwanted wake and glint. However, the algorithm struggles to detect dark, small/distant objects and sometimes splits objects into multiple detections. The long-term tracking ability is also limited. Features of the tracking performance can be seen in representative frames in Fig. 4.

5. Conclusions and future work

In this paper, a new maritime object detection and tracking algorithm has been presented which uses visual saliency as the basis for object detection to overcome some of the challenges of maritime scenes. The effectiveness of the approach has been demonstrated on five challenging sequences from the PETS 2016 maritime dataset. The use of broadly-tuned, intensity-decoupled red, blue and green colour channels reduces the number of false detections from wake and reflections, whilst maintaing the ability to detect salient objects (boats). The tracking algorithm also shows robustness in dealing with scale changes, however, distant, dark objects are difficult to detect and objects are sometimes incorrectly split into multiple detections. Future work will focus on improving the long term tracking ability and comparing the proposed algorithm with existing ones on a wider range of maritime data.

References

[1] Oceans Beyond Piracy. The State of Maritime Piracy 2014, http://dx.doi.org/10.18289/OBP.2015.001. 1

[2] PETS16. http://www.cvg.reading.ac.uk/PETS2016. Accessed March 2016. 1, 5

[3] IMO. Armed Robbery Against Ships in Waters off the Coast of Somlia: Best Management Practices to Deter Piracy in the Gulf of Aden and Off the Coast of Somalia Developed by the Industry, IMO MSC.1/Circ.1339, 2011. 1

[4] T. Albrecht, G. A. W. West, and T. Tan. Visual Maritime Attention Using Multiple Low-Level Features and Naïve Bayes Classification. In Proc. International Conference on Digital Image Computing: Techniques and Applications, pages 243–249, 2011. 1, 2, 4, 5, 6

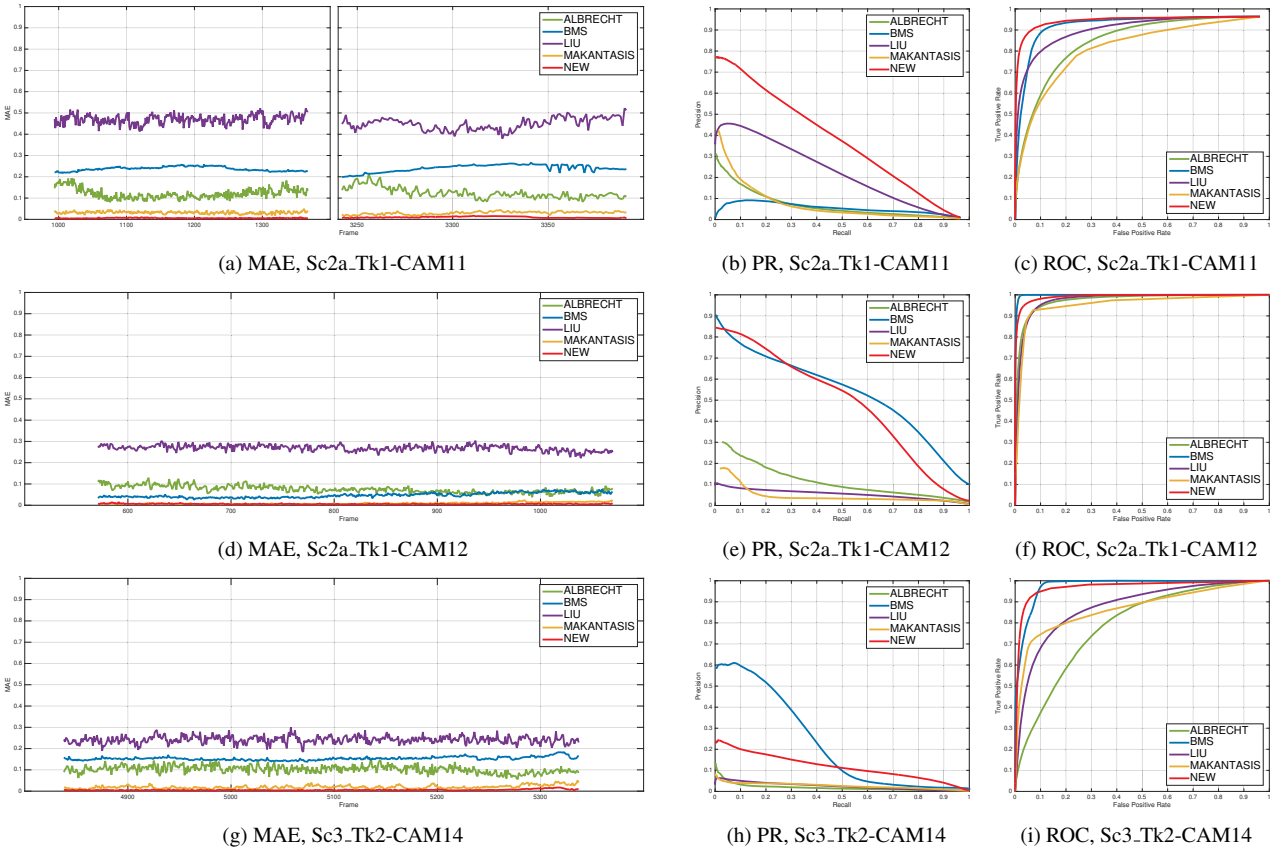
[5] X. Bao, S. Zinger, R. Wijnhoven, et al. Robust moving ship detection using context-based motion analysis and occlusion handling. In Proc. ICMV, pages 90670F–90670F, 2013. 1, 2

[6] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient object detection: A survey. arXiv preprint arXiv:1411.5878, 2014. 2, 4

[7] S. P. V. D. Broek, H. Bouma, R. den Hollander, H. Veerman, K. Benoist, and P. B. W. Schwering. Ship recognition for improved persistent tracking with descriptor localization and compact representations. In Proc. SPIE Electro-Optical and Infrared Systems: Technology and Applications XI, 2014. 1, 2

Table 3. Overall sequence results

PETS IPATCH Sequence	Mean MAE					Area Under PR Curve					Area Under ROC Curve				
	ALB.	BMS	LIU	MAK.	NEW	ALB.	BMS	LIU	MAK.	NEW	ALB.	BMS	LIU	MAK.	NEW
Sc2a_Tk1-CAM11	0.1221	0.2382	0.4639	0.0313	**0.0069**	0.0660	0.0530	0.2217	0.0656	**0.3717**	0.8101	0.8859	0.8683	0.7794	**0.9088**
Sc2a_Tk1-CAM12	0.0751	0.0470	0.2688	0.0091	**0.0058**	0.1062	**0.5334**	0.0545	0.0429	0.4824	0.9750	**0.9984**	0.9747	0.9542	0.9920
Sc3_Tk2-CAM14	0.1027	0.1536	0.2417	0.0196	**0.0057**	0.0180	**0.2228**	0.0254	0.0264	0.1177	0.7854	**0.9764**	0.8769	0.8687	0.9701



(a) MAE, Sc2a_Tk1-CAM11 (b) PR, Sc2a_Tk1-CAM11 (c) ROC, Sc2a_Tk1-CAM11

(d) MAE, Sc2a_Tk1-CAM12 (e) PR, Sc2a_Tk1-CAM12 (f) ROC, Sc2a_Tk1-CAM12

(g) MAE, Sc3_Tk2-CAM14 (h) PR, Sc3_Tk2-CAM14 (i) ROC, Sc3_Tk2-CAM14

Figure 2. MAE (a,d,g), PR Curve (b,e,h) and ROC Curve (c,f,i) results for the saliency methods evaluated on sub-sequences of the PETS 2016 IPATCH dataset: (a-c) IPATCH-Sc2a_Tk1-CAM11; (d-f) IPATCH-Sc2a_Tk1-CAM12; (g-i) IPATCH-Sc3_Tk2-CAM14; Key to methods: ALBRECHT [4], BMS [25], LIU [14], MAKANTASIS [17], NEW - proposed approach.

[8] K. Cheoi and Y. Lee. Detecting perceptually important regions in an image based on human visual attention characteristic. In Structural, Syntactic, and Statistical Pattern Recognition, pages 329–338. 2002. 3

[9] J. Davis and M. Goadrich. The relationship between precision-recall and roc curves. In Proc. ICML, pages 233–240, 2006. 4

[10] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Trans. Pattern Anal. Mach. Intell., 20(11):1254–1259, 1998. 3

[11] P. Kaimakis and N. Tsapatsoulis. Background modeling methods for visual detection of maritime targets. In Proc. ACM/IEEE ARTEMIS, 2013. 1, 2

[12] S. Kim. Analysis of small infrared target features and learning-based false detection removal for infrared search and track. Pattern Anal. and Appl., 17:883–900, 2013. 3

[13] H. W. Kuhn. The hungarian method for the assignment problem. Naval research logistics quarterly, 2(1-2):83–97, 1955. 3

[14] H. Liu, O. Javed, G. Taylor, X. Cao, and N. Haering. Omni-directional surveillance for unmanned water vehicles. In Proc. IWVS, 2008. 1, 2, 4, 5, 6

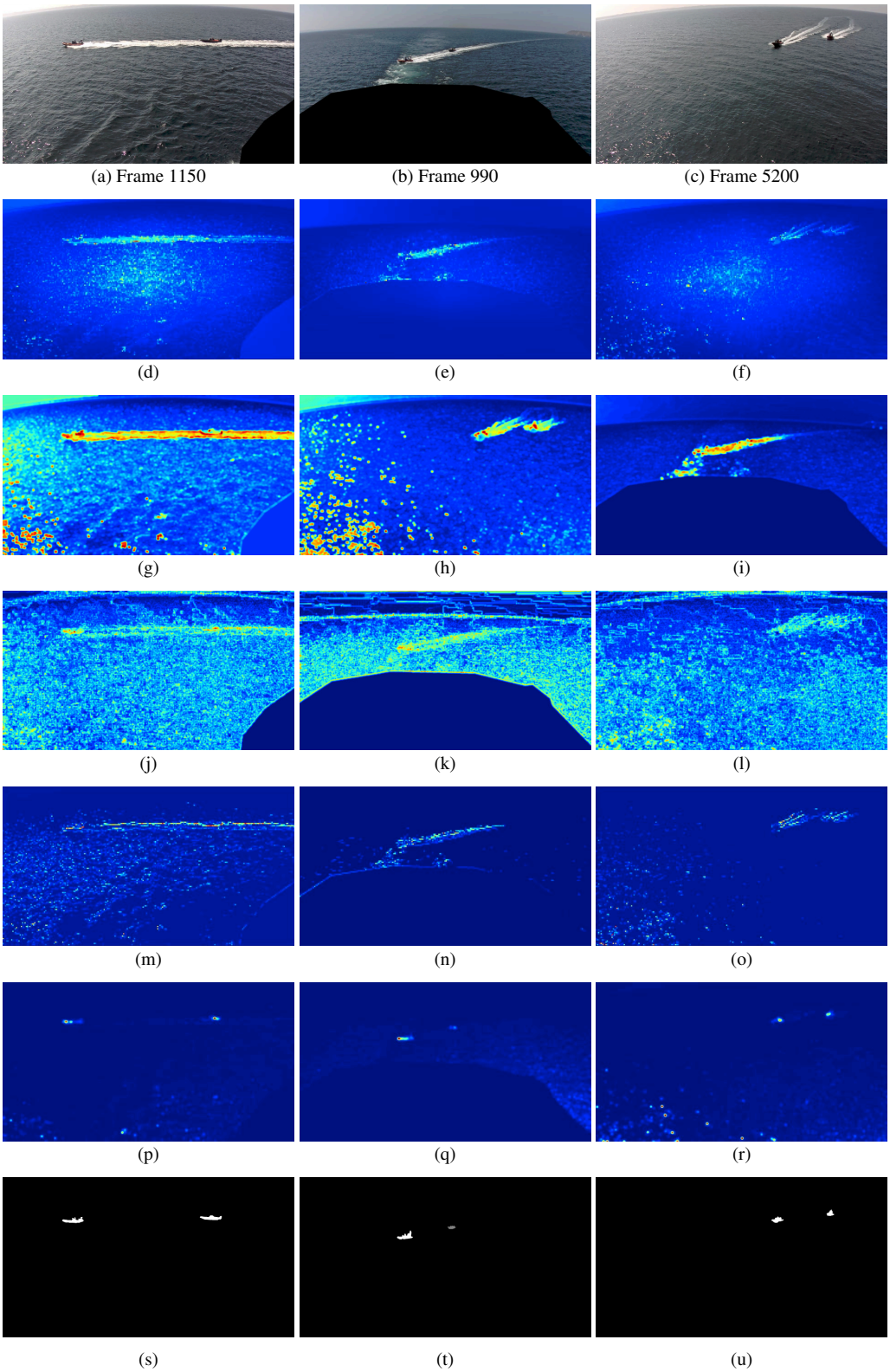
[15] M. J. Loomans, R. G. Wijnhoven, and P. H. de With. Robust automatic ship tracking in harbours using active cameras. In Proc. ICIP, 2013. 1, 2

[16] Q. Luo, T. M. Khoshgoftaar, A. Folleco, and B. Raton. Classification of Ships in Surveillance Video. In Proc. Information Reuse and Integration, pages 432–437, 2006. 3

[17] K. Makantasis, A. Doulamis, and N. Doulamis. Vision-based Maritime Surveillance System using Fused Visual Attention Maps and Online Adaptable Tracker. In Proc. WIAMIS, pages 1–4, 2013. 1, 2, 4, 5, 6

[18] J. Munkres. Algorithms for the assignment and transportation problems. SIAM J. Appl. Math., 5(1):32–38, 1957. 3

[19] B. Qi, T. Wu, B. Dai, and H. He. Fast detection of small infrared objects in maritime scenes using local minimum patterns. In Proc. ICIP, 2011. 1, 2



(a) Frame 1150 (b) Frame 990 (c) Frame 5200

(d) (e) (f)

(g) (h) (i)

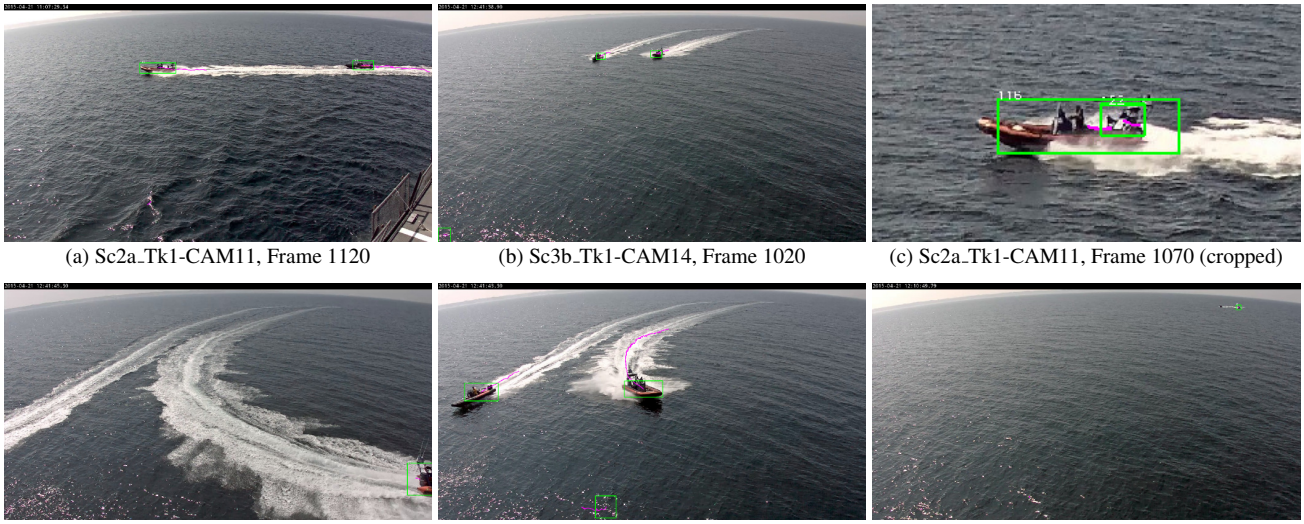
(j) (k) (l)

(m) (n) (o)

(p) (q) (r)

(s) (t) (u)

Figure 3. Representative saliency map examples: Left - IPATCH-Sc2a_Tk1-CAM11; Middle - IPATCH-Sc2a_Tk1-CAM12; Right - IPATCH-Sc3_Tk2-CAM14; (a-c) Original with masked regions shown in black; (d-f) ALBRECHT; (g-i) BMS; (j-l) LIU; (m-o) MAKANTASIS; (p-r) proposed approach; (s-u) groundtruth.



(a) Sc2a_Tk1-CAM11, Frame 1120 (b) Sc3b_Tk1-CAM14, Frame 1020 (c) Sc2a_Tk1-CAM11, Frame 1070 (cropped)

(d) Sc3b_Tk1-CAM14, Frame 1180 (e) CAM14, Frame 1130 (f) Sc3_Tk2-CAM14, Frame 4510

Figure 4. Qualitative tracking results: showing robustness to wake and glint in (a) Sc2a_Tk1-CAM11 and (d) Sc3b_Tk1-CAM14; showing robustness to scale change in Sc3b_Tk1-CAM14 (b,e); examples of failure modes (c) splitting object into multiple detections and (f) failure to detect some dark, small/distant targets. Green bounding boxes mark the estimation of the target in the current frame and magenta lines show the track history of the bounding box centroids.

[20] A. Sobral, T. Bouwmans, and E.-h. ZahZah. Double-constrained rpca based on saliency maps for foreground detection in automated maritime surveillance. In Proc. AVSS, pages 1–6, 2015. 1, 2

[21] M. Sullivan and M. Shah. Visual surveillance in maritime port facilities. In Proc. SPIE Visual Information Processing XVII, 2008. 1, 2

[22] W. Tao, H. Jin, and J. Liu. Unified mean shift segmentation and graph region merging algorithm for infrared ship target segmentation. Opt. Eng., 46(12):127002–1–127002–7, 2007. 1, 2

[23] M. Teutsch, W. Krüger, and B. J. Fusion of region and point-feature detections for measurement reconstruction in multi-target kalman filtering. In Proc. Fusion, 2011. 1, 2

[24] P. Voles, A. Smith, and M. Teal. Nautical Scene Segmentation Using Variable Size Image Windows and Feature Space Reclustering. In Proc. ECCV, pages 324–335, 2000. 1

[25] J. Zhang and S. Sclaroff. Exploiting Surroundedness for Saliency Detection: A Boolean Map Approach. IEEE Trans. Pattern Anal. Mach. Intell., 38(5):889 – 902, 2015. 2, 4, 5, 6