70th Anniversary
The Institute of Electronics and Information Engineers

ICEIC 2016

International Conference on Electronics, Information, and Communication (ICEIC) 2016

Jan 27 (Wed) ▸ Jan 30 (Sat), 2016

Hyatt Regency Resort & Spa, Danang, Vietnam

IEIE IEEE

Vision based boat detection for maritime surveillance

Thanh-Hai Tran
International Research Institute MICA
HUST - CNRS/UMI - 2954 - INP Grenoble
Hanoi University of Science and Technology
Hanoi, Vietnam
thanh-hai.tran@mica.edu.vn

Thi-Lan Le
International Research Institute MICA
HUST – CNRS/UMI – 2945 – INP Grenoble
Hanoi University of Science and Technology
Hanoi, Vietnam
thi-lan.le@mica.edu.vn

Abstract—**Automatic boat detection plays an important role for maritime surveillance. However, maritime environment represents lots of challenges such as wave of water, boat movements, and weather condition. This paper presents a method for detecting moving boats from sequence of images. We reply on background subtraction (BS) techniques to detect main movements in the scene. However, as the BS algorithms detect only motion pixels, they cannot detect when boats stay immobile for a moment or move slightly. In addition, the BS results often suffer strongly from water waves as well as boat wake. To overcome these drawbacks we detect salient regions using saliency detection techniques (SD) on images. We then apply dynamic fusion technique on BS and SD results for a final boat detection. Experiments have been conducted on a very challenging dataset and show promising results (detection rate achieved to 89%).**

Keywords—background subtration; saliency detection; boat detection; maritime surveillance

I. Introduction

Recently, maritime surveillance has been an active research area because of its crucial role in helping authorities to protect interests in Exclusive Economic Zones, territorial waters against resource plundering, smuggling, illegal immigration, piracy, and terrorism. The common goal of a maritime surveillance system is to monitor the activity of a maritime area, to detect abnormal object behaviors and to raise alarms when such behaviors are detected. Boat/ship detection is the first and crucial step in any maritime surveillance system. Results of this step will be used for further processing such as boat/ship classification and tracking, behavior analysis.

There are two main approaches proposed for boat detection. The first approach tries to apply appearance for boat detection while the second one relies on the motion detection.

The work presented in [1] belongs to the first approach. In this work, the authors proposed to combine a spatial partial feature with Adaboost classifier in order to detect the ship. Though a number of features that are proposed for appearance-based object detection can be applied for ship, this approach is not generic because different types of ships have significant differences in their appearance.

In contrary, the second approach bases on an assumption that the boat is normally moving in the video sequences. For this, motion can be robust for boat detection. In [2], the author proposed a method for maritime domain background modeling in order to taking into account dynamic aspect of maritime environment such as water surface, weather issues. The authors in [3] combine Visual Background Extractor as background subtraction and backwash cancellation for ship detection. However, the background subtraction techniques do not work well with very complicated and highly dynamic cluttered background of maritime surveillance video.

In this paper, we propose to extract both temporal motion and spatial cues and incorporate them in a meaningful way to improve detection results. Our main contribution is twofold: i) we perform a comparative evaluation of two well-known BS techniques Mixture of Gaussians (MOG) and Visual Background Extractor (VIBE) for detecting motion objects in the scene; ii) we combine BS with saliency detection techniques Spectral Residual (SR) and Maximum Symmetric Surround (MSS) using a dynamic fusion model.

The remaining of this paper is organized as follows. In section II, we present the overall framework as well as two background subtraction techniques and saliency detection techniques. Then, the experimental results are described in Section III. Section IV gives some conclusions and future works.

II. Proposed method for boat detection

A. General framework for boat detection

We propose a framework for boat detection which exploits both temporal and spatial cues from image sequence. The framework composes of three main components as seen in Fig. 1:

- Background subtraction: As boats are moving objects in the scene, we then apply background subtraction technique to detect moving boats. Two BS algorithms are investigated MOG and VIBE.
- Saliency detection: In several cases, boats could stay immobile for a moment or move slightly, we apply saliency detection technique to detect all salient regions in the scene that could be boat candidates. Two SD algorithms are studied: SR and MSS.
- Fusion: The results of boat detection are finally generated by fusing saliency detection and background subtraction results.

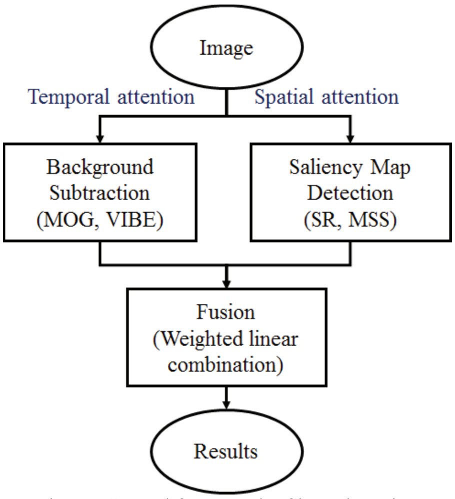

Fig. 1: General framework of boat detection

B. Background subtraction

Background subtraction is a common technique for detecting moving objects in the scene captured by fixed camera. Generally, BS is composed of two main steps: initialization and updating. The initialization step generates the first model of the BS from several frames. The updating step aims to incorporate information of changes of observed scenes into the BS model. Therefore, building accurate model of BS is a key issue.

There exists a numerous methods of BS. In our research, we investigate two methods that have been shown recently to be efficient, mostly in boat detection problem. They are Visual Background Extractor (VIBE), Mixture of Gaussians (MOG). In the following, we summarize the main idea of each method.

1) MOG: Mixture of Gaussians

This method has become a very common approach of BS and applied widely in many surveillance applications [4]. According to this method, each pixel is modelled as a mixture of Gaussians and an online approximation is utilized to update the model. The Gaussian distributions are evaluated to determine which is most likely to result from a background process. The each pixel is classified as background / foreground depending on its Gaussian distribution.

2) VIBE: Visual Background Extractor

The method VIBE was firsty introduced in and explained in more detail in [5]. The main idea of VIBE is to consider the problem of background subtraction as a classification problem. Each background pixel is modelled by a set of samples instead of with an explicit pixel model: $M(x) = \{v_1, v_2, \ldots, v_N\}$. To classify a pixel value $v(x)$ according to its corresponding model $M(x)$, the value of pixel is compared with the closest values within the set of samples defined by a sphere $S_R(v(x))$ of radius R centered at $v(x)$. The pixel value is classified as background if the cardinality of the set intersection of $S_R(v(x))$ with $M(x)$ is larger or equal to a given threshold.

By this way, the method does not need to estimate the pdf of the background pixel but only compare the current value of the pixel to its closest samples within the collection of samples.

C. Saliency detection

As BS techniques can detect only moving objects, they can not detect boats that stay motionless in certain time. We propose to apply a technique of saliency detection to overcome this issue. There are many approaches for saliency detection. In this paper, we chose two methods of SD for studying due to their efficiency and simplicity for implementation.

1) SR: Spectral Redisual

The main idea of spectral residual based saliency detection is to remove redundancies in the sensory input (in our case redundancies are the features which occurs frequently such as water surface) while keep sensitive to features that deviate from the norm.

This method composes of four steps: 1) compute the Fourier transform of the input image that produce two matrices: amplitude and phase of the spectrum; 2) compute log-amplitude of the spectrum; 3) compute spectral residual which is defined as difference of log-spectrum and its smoothed version; 4) saliency map is the inverse Fourier transform composed from spectral residual and phase.

Once saliency map is computed, we follow the same technique presented in [6] to detect proto-objects. Given S(x) the saliency map of the image. The object map O(x) is obtained by thresholding S(x) by a threshold. This threshold is empirically selected based on the average value of the saliency map.

2) MSS: Maximum Symetric Surround

Based on observation that saliency maps generated by several methods suffer from low resolution that limits the range of spatial frequency content retained in the full resolution image. To avoid this, in [7], the authors proposed to treat the entire image and take symmetric surround of each pixel into account. For a given image, a saliency map is obtained as L_2 norm of the difference between the average CIELab vector of the image and the corresponding CIELab image vector in the Gaussian filtered version of the original image.

D. Dynamic fusion

Two previous sections present the extraction of region of interest from spatial and temporal information separately. As we have mentioned, these information must be incorporated to produce final region which is sensitive to little movement of objects without considering non object of interest from background. We follow a dynamic fusion technique as follows:

- If strong motion appears in the video sequence, temporal attention should be more dominant
- Otherwise, if motion is low, the spatial attention should be more important.

We have now two binary maps corresponding to detection results from BS and SD. We combine by performing a simple linear combination weighted by their contribution.

$$Det(I) = k_T BS(I) + k_S SD(I) \qquad (1)$$

Where I is the input image, $BS(I)$ is the result of Background subtraction applied on I, $SD(I)$ is the result of Saliency detection applied on I, k_T and k_S are dynamic weights which are computed as follows.

$$k_T = \frac{Var_T}{Var_T + const} \quad (2) \qquad k_S = \frac{const}{Var_T + const} \quad (3)$$

Where $Var_T = max(BS(I)) - mean(BS(I))$ representing the variance of the temporal attention model, *const* is empirically selected. In our experiment, we set *const* to 0.3.

III. EXPERIMENTS

A. Dataset

To evaluate the proposed methods, we have recorded a dataset of 4 videos, resolution of 572x760 at Baichay beach, Quangninh province in Vietnam. The video frame rate is set at 25fps. 1 video are taken in the morning, 2 videos in the cloudy afternoon and 1 video at the end of the day. The total number of frames are 8975. The camera has top or rear view. There will have one or more boats (2-3, more) moving in the scene. Boats are of categories: Cargo Ship, Fishing Boat, Cruise Ship, Canoe which have very different appearance (size, shape, texture, structure, etc.). This dataset is very challenging because videos are taken in free manner in different conditions of lighting, camera viewpoint, boat categories and movements, waving, reflections, The first column of Fig. 5 illustrates some images extracted from these video.

B. Performance measurements

We evaluate segmentation and detection performance. To evaluate segmentation, we use usual performance measurements of background subtraction that is binary Jaccard Index (JI) which is defined as follows [8]:

$$JI = \frac{FG \cap GT}{FG \cup GT}$$

Where FG is foreground detection result, GT is ground-truth.

To compute this measurement, we have to prepare groundtruth manuallly. For segmentation, we use an opensource tool that semi-automatially select regions of interest and store corresponding binary images. As we can not prepare groundtruth for all 8975 frames of video sequences, we take samples from each sequence and have totally 52 images for evaluation. The second column of the Tab. 1 presents the total numbers of positive pixels belonging to boats that we have annotated manually.

The Jaccard Index is computed on pixels. However, in boat detection problem, it is not necessary to achieve pixel precision. Therefore, we use another measure to evaluate boat detection which is True Positive Rate (TPR). Then we compute rectagular bounding boxes of segmented regions and evaluate the JI on rectangular regions. A detection is considered as true positive if the JI is higher than 0.5.

C. Preliminary results

1) Background subtraction results

The columns 3 and 4 of the Tab. 1 show the results of background subtraction using MOG and VIBE. In almost cases, VIBE gives significantly higher performance than MOG. We observe that at sequence 3, MOG fails absolutely to MOG updates the background based only on the pixel distribution itself without considering the neighbors, as consequent, the detected region is smaller than the ground-truth. VIBE avoids this situation. However, VIBE is not sensitive to small movement. It cannot detect the boats in turning surrounding itself (Seq. 1).

Tab.1 A comparison of MOG, VIBE, VIBE+SR, VIBE+MSS methods on the dataset in term of Jaquard Index.

Seq.	# pos. pixels	MOG	VIBE	VIBE + SR	VIBE+ MSS
1	291565	0.05	0.02	0.20	**0.79**
2	177649	0.26	0.51	**0.68**	0.65
3	89320	0.00	**0.56**	0.34	0.38
4	93027	0.61	**0.70**	0.60	**0.66**
Average		0.23	0.48	0.59	**0.62**

Original image GroundTruth MOG VIBE

Fig. 2: Sequence 1 (first row) and 3 (second row)

In both cases, the JI is not high. For example for the first sequence, as boats move slightly, both MOG and VIBE fail to detect boats. The highest JI is 0.70 for the seq. 4. The main reason is that the BS algorithms detect moving pixels in the scene. When boats move, pixels belonging to boats move also. However, often, these pixels have the same characteristics of color and texture. By appearance, we cannot observe the movement of these pixels on the image.

2) Saliency detection results

Both SD algorithms give interesting results. They give high response on the regions of boats or at its corners. They could be able to detect still objects in the scene that BS cannot. We can see in Fig. 4 that saliency detection techniques applied in the image of the sequence 1 (see original image and ground-truth in the first row of the Fig. 2). In this example, the MSS technique outperforms SR and gives very good result of boats detection. However, both SD technique detect also some other background objects such as house, mountain.

3) Final detection results using dynamic fusion

Fusion helps to improve significantly the performance of segmentation in term of Jaccard index. Sometime it reduces the result (sequence 1) with SR method because as we said that the ground-truth we prepare is only moving boats, not static boats. However, SD algorithm highlight also other non-boat objects such as mountain and house. MMS helps to fill more. The column 5 and 6 show the improvement of using both BS and SD techniques.

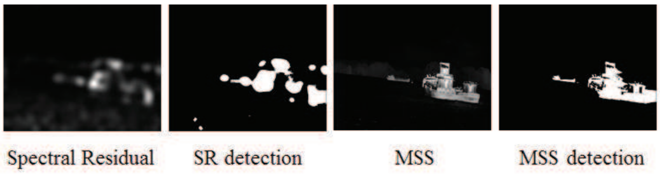

Fig. 3: Saliency detection results on one image from seq. 1 (see the original image in the first row of the Fig. 3)

4) Object detections

As the main objective is not to detect boats at pixel precision. We conduct our evaluatin on object detection. The Tab. 2 shows the result of boat detection in term of TPR. The best result is achieved with VIBE and MSS. The highest TPR is 1 and 0.89 in average. Fig. 5 illuatres some examples of segmentation and detection and effiency of combining VIBE with MSS using dynamic fusion technique.

Tab.2 Result of boat detection using VIBE+SR and VIBE+MSS

Seq.	VIBE + SR	VIBE+ MSS
1	0.00	**1.00**
2	0.73	**0.8**
3	0.76	**0.75**
4	0.83	**1.00**
Average	0.58	**0.89**

IV. CONCLUSIONS

This paper presented a comparative study on two background subtraction methods for moving boats detection. We found that VIBE method is more efficient than GMM method. However, both techniques detect only moving boats with very poor detection rate because of very complicated and highly dynamic cluttered background. The use of saliency map helps to detect boats with lightly movement. We have compared the efficiency of SR and MSS techniques and found that MSS outperforms SR in all cases. Combining BS with SD improves significantly the results from 058 to 0.89 in term of TPR. In the future, we would like to improve the fusion method by analyzing the local importance of motion region instead of global one as in the current work and test with public datasets.

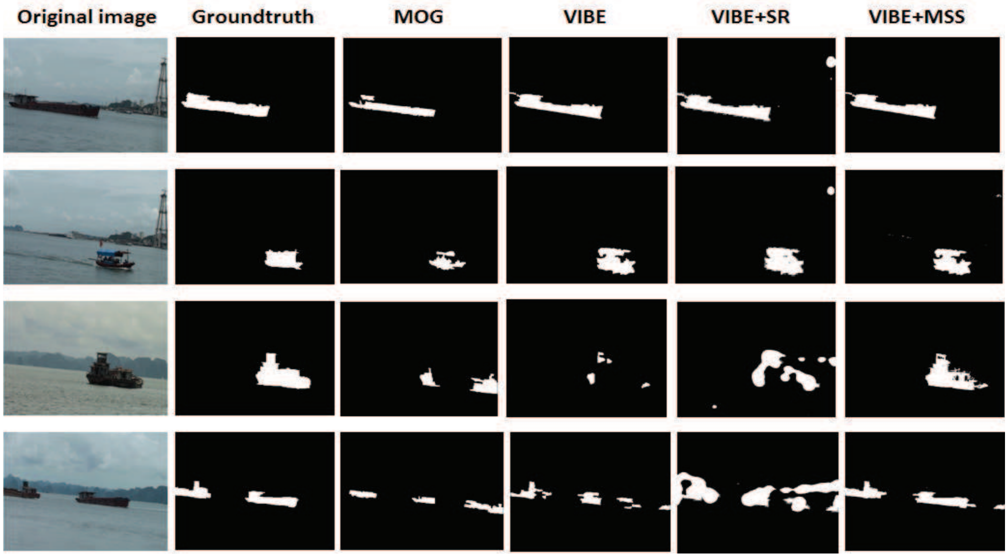

Fig. 4: Results of boat detection using different techniques of background subtraction, saliency detection and fusion

ACKNOWLEDGEMENT

This research is funded by Hanoi University of Science Technology under grant number T2015-095

REFERENCES

[1] H. Wang, F. Zhou, and T. Jin, "Ship detection based on spatial partial features," in *International Symposium on Photoelectronic Detection and Imaging*, 2013.

[2] D. Bloisi, A. Pennisi, and L. Locchi, "Background modeling in the maritime domain," *Machine Vision and Applications*, vol. 25, no. 5, 2014.

[3] S. A. S. J. Mohammad Hassan Saghafi, Seyed Majid Noorhosseini, "Robust Ship Detection and Tracking Using Modified ViBe and Backwash Cancellation Algorithm," in *Second International Conference on Computational Intelligence and Information Technology – CIIT*, 2012.

[4] C. Stauffer and W. E. . Grimson, "Adaptive background mixture models for real-time tracking," in *CVPR*, 1998.

[5] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.

[6] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," in *CVPR*, 2007, pp. 1 – 8.

[7] R. Achanta and S. Susstrunk, "Saliency detection using maximum symmetric surround," in *17th ICIP*, 2010.

[8] K. McGuinness and N. E. O'Connor, "A comparative evaluation of interactive segmentation algorithms," *Pattern Recognition*, vol. 43, no. 2, 2010.