A Hybrid Color-Based Foreground Object Detection Method for Automated Marine Surveillance

Daniel Socek, Dubravko Culibrk, Oge Marques, Hari Kalva, and Borko Furht

Center for Coastline Security Technologies (CCST),
Department of Computer Science and Engineering,
Florida Atlantic University, Boca Raton FL 33431, USA

Abstract. This paper proposes a hybrid foreground object detection method suitable for the marine surveillance applications. Our approach combines an existing foreground object detection method with an image color segmentation technique to improve accuracy. The foreground segmentation method employs a Bayesian decision framework, while the color segmentation part is graph-based and relies on the local variation of edges. We also establish the set of requirements any practical marine surveillance algorithm should fulfill, and show that our method conforms to these requirements. Experiments show good results in the domain of marine surveillance sequences.

1 Introduction

Automatic detection of semantic visual objects within a digital image or video stream still represents one of the great challenges in computer vision. Although the problem of object detection within a video sequence (*foreground object detection*) is often treated separately from the image segmentation problem (*color-texture segmentation*), the two problems exhibit a strong conceptual similarity. In this paper, we present a framework that combines a foreground object detection approach and an image color segmentation technique in order to achieve better detection of semantic objects within a video sequence.

Common solutions to foreground object detection from a digital video are based on some form of background subtraction or background suppression [4,5]. These approaches work well when the camera is in a fixed position, and when there is no background movement (e.g., a footage taken by a stationary camera filming a highway toll plaza on a bright, sunny day). However, if the camera moves, or if the scene contains a complex moving background, the object detection and tracking becomes more difficult. In many real-world applications, such as marine surveillance, a scene can potentially contain both types of background: moving and stationary.

Object detection plays a crucial role in most surveillance applications. Without a good object detection method in place, the subsequent actions such as object classification and tracking would be infeasible. Our main goal was to obtain a robust marine surveillance object detection method that can successfully

J. Blanc-Talon et al. (Eds.): ACIVS 2005, LNCS 3708, pp. 340–347, 2005.

overcome obstacles inferred by the presence of the complex, moving background. In our view, such algorithm should have the following properties in order to be of practical use:

1. *Determine potentially threatening objects within a scene containing a complex, moving background.* In marine surveillance applications, it is essential that the algorithm can deal with moving background such as flickering water surfaces and moving clouds, and still detect potential objects of interest.
2. *Produce no false negatives and a minimal number of false positives.* A surveillance application in general prefers no false negatives so that no potential threat is ever overlooked. On the other hand, having too many false positives would make potential postprocessing activities, such as object classification, highly impractical.
3. *Be fast and highly efficient, operating at a reasonable frame rate.* The object that poses a potential threat must be detected fast so that the appropriate preventive action can be taken in a timely manner. Furthermore, if the algorithm operates at an extremely small frame rate due to its inefficiency, some potential objects of interest could be overlooked.
4. *Use a minimal number of scene-related assumptions.* When designing an object detection method for marine surveillance, making the algorithm dependent upon too many assumptions regarding a scene setting would likely make the algorithm fail as soon as some of the assumptions do not hold.

To our knowledge, the existing literature does not offer an approach that exhibits all of the aforementioned properties. In this paper, we establish a hybrid method that essentially has such properties. We have slightly modified and extended two previously proposed general-purpose approaches, one for color-texture image segmentation and one for a foreground video object detection, and merged them into a hybrid method that is suitable for practical marine surveillance applications.

The paper is organized as follows. Section 2 briefly introduces related work, including the two methods used in our hybrid approach. The framework of our proposed algorithm is described in Section 3, with its experimental results given in Section 4. The last section concludes this paper.

2 Related Work

Some of the early methods for dealing with the instances of non-stationary background were based on smoothing the color of a background pixel over time using different filtering techniques such as Kalman filters [7,9], or Gabor filters [6]. However, these methods are not particularly effective for sequences with high-frequency background changes. Slightly better results were reported for techniques that rely on a Gaussian function whose parameters are recursively updated in order to follow gradual background changes within the video sequence [1]. More recently, this model was significantly improved by employing a Mixture of Gaussians (MoG), where the values of the pixels from background objects are described by multiple Gaussian distributions [2,12,15]. This model was considered promising since it showed good foreground object segmentation results for many outdoor sequences. However, weaker results were reported [8] for video sequences containing non-periodical background changes. This is the case for most of the marine sequences, which exhibit frequent background changes due to waves and water surface illumination, cloud shadows, and similar phenomena.

Voles *et al.* proposed a method suitable for object detection in maritime scenes based on anisotropic diffusion [14]. Unlike Gaussian filtering, anisotropic diffusion preserves well-defined edges and large homogeneous regions over poorly defined edges and small inhomogeneous regions. This approach performs well for horizontal and vertical edges, but it fails for other directions. In addition, unless simplified at the expense of performance, anisotropic diffusion is iterative and time consuming.

In 2003, Li *et al.* proposed a method for foreground object detection employing a Bayes decision framework [8]. The method has shown promising experimental object segmentation results even for the sequences containing complex variations and non-periodical movements in the background. In addition to the generic nature of the algorithm where no *a priori* assumptions about the scene are necessary, the authors claim that their algorithm can handle a throughput of about 15 fps for CIF video resolution, which is a reasonable frame rate for our purposes. Moreover, the algorithm is parallelizable at the pixel level, so that even better frame rates could be achieved if parallelization can be afforded. However, when we applied the algorithm to marine sequences, the object boundaries were not particularly accurate, and the segmented frames contained too many noise-related and scattered pixels. Furthermore, the adaptive threshold mechanism from [10] that was originally used by Li *at al.* performed poorly when fast large objects suddenly entered a scene. As a consequence, the algorithm produced instant flashing frames where most of the pixels were mistakenly classified as a foreground.

For removing these scattered noise pixels, Li *et al.* suggested applying morphological open and close operations [8]. Unfortunately, in doing so, small objects of interest could be lost or the boundaries of larger objects could be degraded and chopped, which could potentially change the outcome of the threat classification postprocess.

In general, the idea of combining the motion-related and texture-related information to improve the segmentation output is not new. In [11], Ross presented a method in which a duality of color segmentation and optical flow motion information was exploited. As a result, a better image segmentation is reported for a variety of natural images (frames) [11]. Ross also presented a comparative study of some of the relevant color-texture segmentation methods suitable for algorithmic synergy with the motion-related information. Among the candidates, the Felzenszwalb-Huttenlocher (F-H) [3] image segmentation algorithm was outstanding for its speed, its clear theoretical formulation, and its performance on natural images. The overview of F-H approach is presented in the following section.

3 Description of the Proposed Algorithm

The proposed hybrid background segmentation method has two distinct phases: (i) primary foreground segmentation based on background modeling; and (ii) post-processing based on color segmentation. A block diagram of the system is shown in Fig. 1.

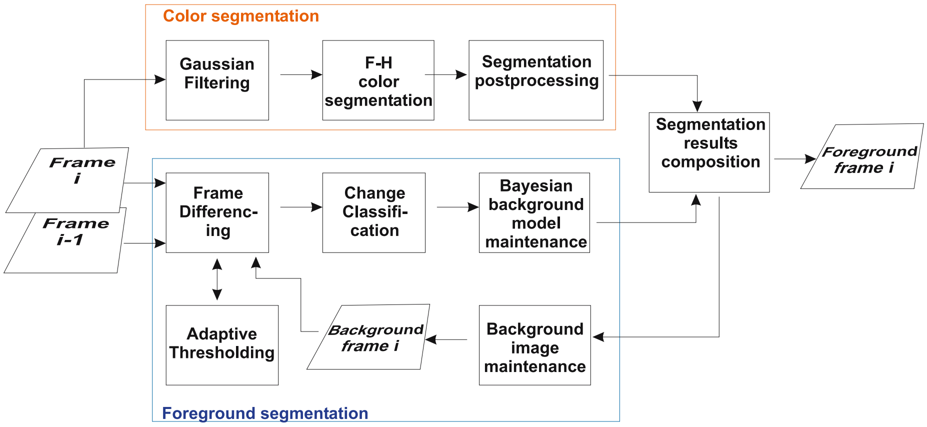

Fig. 1. Block diagram of the proposed foreground segmentation system

Primary foreground segmentation is based on a partial probabilistic model of the background in conjunction with a more classical low-pass filtered background image and a Bayesian decision framework for change classification proposed in [8]. The approach relies on the assumption that for a scene obtained from a static camera, there exist features, which can be used to discern whether a certain pixel belongs to a background or a foreground object and on the idea that the background can be modeled by probabilities of a certain feature value occurring at a specific pixel. Furthermore the background is viewed as consisting of objects that may be moving to an extent but are stationary in general, making it possible to model it by a small number of feature values that occur at a specific pixel with significant probability. This assumption is ground for computational feasibility of the proposed approach.

The used Bayesian classifier is general in terms of allowing for the use of different features to describe the stationary and movement characteristics of the background [8]. The specific features employed in this project are the color descriptors (RGB values) for the stationary background model and the color co-occurrence descriptors (RGB values the pixel takes in two consecutive frames) to describe the background movement.

The foreground segmentation algorithm (Fig. 1) has four main steps: change detection, change classification, foreground segmentation and background model learning and maintenance. The last step is addressed by two modules concerned with two distinct parts of the background model, as indicated in Fig. 1.

The initial stage of the algorithm is concerned with detecting the differences between the current frame and the background reference image kept (to detect the "stationary" differences) as well as the differences between two consecutive frames of the video sequence (to detect the movement). Once the differences are identified, they are used to determine whether the change is something consistent with the background or something that should be deemed foreground, based on the learned probabilistic model of the background. This is followed by a post-processing step used to enhance the effects of foreground segmentation by combining them with the results of color-based segmentation. The final step of the algorithm is the one in which background statistics are learned and the background image updated. In it, the information of how the pixels have been classified is used to gradually change the probabilities of significant feature values encountered to be able to accurately classify changes in the future. In addition to the probability learning process, knowledge of the background is stored by maintaining a reference background image updated through Infinite Impulse Response (IIR) filtering.

The Bayesian decision framework forms the change classification and part of the background model learning and maintenance step. The probabilistic model is used as the sole model for the movement of the background, however, it is only an extension of a more traditional IIR filter model. Therefore, for change detection, the post-processing in the third step and the background image filtering an arbitrary approach could be used. The original approach used automatic thresholding based on noise intensity estimation approach proposed in [10] for change detection while morphological open and close operations were used to enhance the foreground segmentation results. Our hybrid approach uses the original Bayesian classifier and the IIR filter based background image maintenance. However we found that automatic thresholding based on a Poisson distribution model for the spatial distribution of the noise [10] leads to better results in our application domain (the noise appeared more distinguished from the signal). In addition, we choose to enhance the results of the foreground segmentation based on color-based image segmentation algorithm, which, unlike the morphological operations, provides additional information. The authors of the original approach used feature binding to enhance the performance of the algorithm. We have not followed this practice, fearing reduced accuracy of segmentation.

The F-H algorithm [3], indicated by the top box in Fig. 1, uses a simple graph theoretic model that allows for the segmentation in $O(n \log n)$ time, n being the number of image pixels. The F-H algorithm is based on a local variation (the intensity differences between neighboring pixels). Image segmentation was treated as a graph partitioning problem where for each given image, a graph $G = (V, E)$ is defined such that each vertex from V corresponds to an image pixel, and each edge from E corresponds to a connection between two neighboring pixels. Any given edge in E carries a weight given by the intensity difference between pixels it connects. In such setting, image segmentation is equivalent to obtaining a partition of the graph G into disjoint connected components. Given a measure of variation between two disjoint components (called *external variation*) and a measure of the inner variation of a single component (called *internal variation*) it is possible to evaluate a given segmented image, or equivalently a given graph partition. More precisely, a graph partition is over-segmented with too many components if the variation between two disjoint components is small relative to the internal variation of both components. On the other hand, a partition is under-segmented (not enough components) if there is a way to split some of its components and produce a partition which is still not over-segmented. The F-H algorithm essentially generates a partition that is optimal in the sense that it is neither over-segmented nor under-segmented. In this model, *internal variation* of a component is the maximum edge weight in any minimum spanning tree of that component, and the *external variation* between two components is the lowest edge weight connecting them. The threshold function $\tau(C) = k/|C|$ of a component C controls the degree to which the external variation can be larger than the internal variations, and still have the components be considered similar. In our experiments, we have selected the input parameter $k = 100$.

The segment postprocessing for minimizing the number of segments by blindly merging the small segments with the larger neighboring ones, which was proposed by the authors [3], is far too dangerous to apply in marine surveillance applications since the small objects could disappear in the process. For that reason, we modified the postprocessing mechanism to work on a more sophisticated level. Namely, our modified postprocessing was based on the segment features consisting of the first and the second order color moments [13], calculated per each RGB channel. In many instances, better results were obtained when the features also included several bins counting very small vertical, horizontal, and diagonal edge weights within a segment.

4 Experimental Results

To test the approach a number of sequences extracted from a marine surveillance video has been used. The data used is real and pertinent to our problem domain. Frames include water-surface, sky, parts of solid ground and were captured by a stationary camera. The camera was occasionally moved slightly by wind.

Fig. 2 illustrates the enhanced performance of our approach in the case of a frame on which both algorithms perform comparably well. Better results were achieved through the use of color-based segmentation over those achieved by employing morphological operations. Primary motivation for the use of different thresholding approach was the inability of the original approach to adequately select the threshold in a number of specific frames, specifically when an object first enters the scene and when the camera is slightly moved by the wind. Fig. 3 illustrates the performance of the original thresholding approach for these two cases.

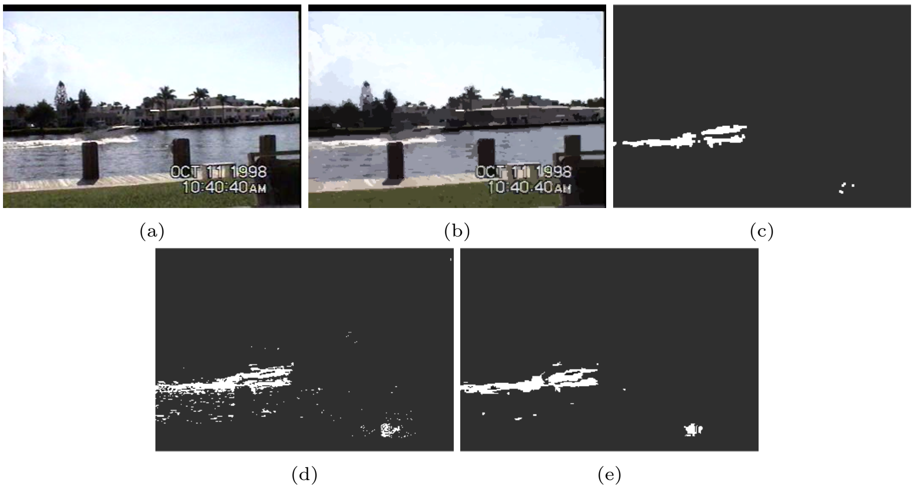

Fig. 2. The experiment performed on a video containing typical marine surveillance footage: (a) the original frame, (b) color segmentation results, (c) foreground obtained using the model described in [8] where morphological operators were used, (d) foreground obtained using new Poisson spatial noise distribution model-based thresholding without enhancement of the background segmentation results, and (e) with color segmentation-based enhancement

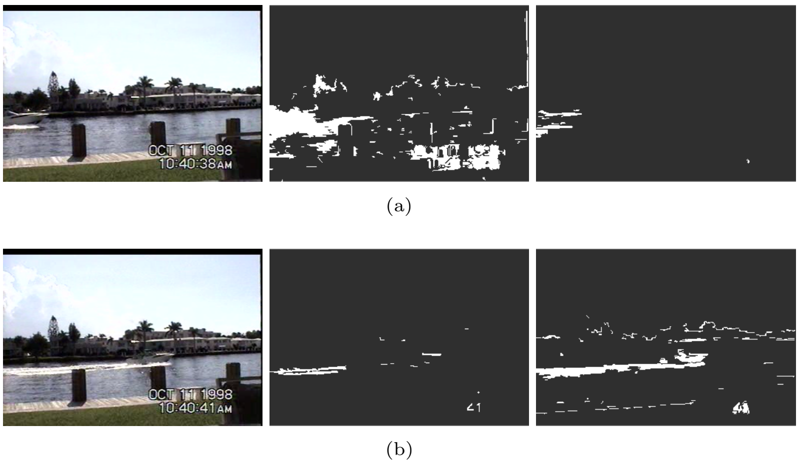

Fig. 3. Results obtained for representative frames: (a) a frame where an object first enters the scene (original frame and foreground obtained with the original and new approach, from left to right), (b) a frame where there a slight movement of the camera occurred (same layout as previous)

5 Conclusions

Object segmentation in the domain of marine surveillance is faced with the task of distinguishing between object of interest and complex moving background. We presented a hybrid method combining color-based single frame segmentation and change detection and classification based foreground segmentation. We evaluated the performance of the proposed method on a set of real marine surveillance sequences and presented a number of representative result frames.

References

1. Boult, T.E., Micheals, R., Gao, X., Lewis, P., Power, C., Yin, W., Erkan, A.: Frame-rate omnidirectional surveillance and tracking of camouflaged and occluded targets. in Proc. of IEEE Workshop on Visual Surveillance (1999) 48-55
2. Ellis, T.J., Xu, M.: Object detection and tracking in an open and dynamic world. in Proc. of the Second IEEE International Workshop on Performance Evaluation on Tracking and Surveillance (PETS'01) (2001)
3. Felzenszwalb, P.F., Huttenlocher, D.P.: Image segmentation using local variation. in Proc. of Computer Vision and Pattern Recognition (CVPR'98) (1998) 98–104
4. Haritaoglu, I., Harwood, D., Davis, L.: W^4: Real-time survaillance of people and their activities. IEEE Trans. Pattern Analysis and Machine Intelligence **22(8)** (2000) 809–830
5. Heikkila, J., Silven, O.: A real-time system for monitoring of cyclists and pedestrians. in Proc. of the Second IEEE Workshop on Visual Surveillance (1999) 74–81
6. Jain, A.K., Ratha, N.K., Lakshmanan, S.: Object detection using Gabor filters. Journal of Pattern Recognition **30(2)** (1997) 295–309
7. Karmann, K.-P., von Brandt, A.: Moving object recognition using an adaptive background memory. in "Timevarying Image Processing and Moving Object Recognition, 2", V. Cappellini (ed.), Elsevier Publishers B.V., Amsterdam, The Netherlands, (1990) 297-307
8. Li, L., Huang, W., Gu, I.Y.H., Tian, Q.: Foreground object detection from videos containing complex background. in Proc. of the Eleventh ACM International Conference on Multimedia (MULTIMEDIA'03) (2003) 2–10
9. Ridder, C., Munkelt, O., Kirchner, H.: Adaptive background estimation and foreground detection using Kalman-filtering. in Proc. of International Conference on Recent Advances in Mechatronics (ICRAM'95) (1995) 193–199
10. Rosin, L.R.: Thresholding for change detection. in Proc. of the Sixth International Conference on Computer Vision (ICCV'98) (1998) 274–279
11. Ross, M.G.: Exploiting texture-motion duality in optical flow and image segmentation. *Master Thesis*, Massachusetts Institute of Technology (2000)
12. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **22(8)** (2000) 747–757
13. Stricker, M.A., Orengo, M.: Similarity of color images. in Proc. of SPIE Storage and Retrieval for Image and Video Databases III **2420** (1995) 381–392
14. Voles, P., Teal, M., Sanderson, J.: Target identification in complex maritime scene. in Proc. of IEEE Colloquium on Motion Analysis and Tracking **15** (1999) 1-4
15. Ya, L., Haizhou, A., Guangyou, X.: Moving object detection and tracking based on background subtraction. in Proc. of SPIE Object Detection, Classification, and Tracking Technologies **4554** (2001) 62–66