Quantitative Performance Analysis of Object Detection Algorithms on Underwater Video Footage

Isaak Kavasidis
Department of Electrical, Electronical and Computer Engineering
University of Catania
Catania, Italy
kavasidis@dieei.unict.it

Simone Palazzo
Department of Electrical, Electronical and Computer Engineering
University of Catania
Catania, Italy
palazzosim@dieei.unict.it

ABSTRACT

Object detection in underwater unconstrained environments is useful in domains like marine biology and geology, where the scientists need to study fish populations, underwater geological events etc. However, in literature, very little can be found regarding fish detection in unconstrained underwater videos. Nevertheless, the unconstrained underwater video domain constitutes a perfect soil for bringing state-of-the-art object detection algorithms to their limits because of the nature of the scenes, which often present with a number of intrinsic difficulties (e.g. multi-modal backgrounds, complex textures and color patterns, ever-changing illumination etc..).

In this paper, we evaluated the performance of six state-of-the-art object detection algorithms in the task of fish detection in unconstrained, underwater video footage, discussing the properties of each of them and giving a detailed report of the achieved performance.

Categories and Subject Descriptors

I.4 [**Image Processing and Computer Vision**]: Segmentation—*Pixel Classification*

General Terms

Algorithms, Performance

Keywords

Underwater, unconstrained, object detection

1. INTRODUCTION

Computer vision gathers huge attention from the worldwide scientific community. In fact, computer vision applications play a fundamental role in every-day activities. Particular attention is given to the object detection subfield of computer vision, which consists in detecting moving objects in video streams.

While object detection algorithms have been applied to diverse domains, little has been done for unconstrained underwater environments. Applying object detection algorithms to underwater videos is a crucial step in works like [12], [14] and [13] in order to study the behavior of underwater species, geologic phenomena (e.g. typhoons) etc...

Unconstrained underwater environments present various difficulties for the detection algorithms. In [7] and [6], the authors studied the visual effects that atmospheric phenomena produce and proposed algorithms in order to alleviate the problem of object detection under bad weather conditions, but no quantitative evaluation has been carried out. In other works, like in [9], the authors evaluated object detection algorithms by using artificial methods to simulate lighting changes (by changing the image brightness) and poor image quality (by introducing Gaussian noise).

While many algorithms performed well under these conditions, it is not certain whether their performance will remain as high when these algorithms are applied on real world, heavily affected by interferences, videos. In fact, it is not rare the case to see state-of-the-art algorithms performing very well in videos containing traffic images, pedestrians etc, suffering a substantial performance degradation when applied to underwater videos containing scenes having one or more of the following properties:

- light changes: real-time video acquisition requires robust object detection under every possible lighting conditions. In fact, the video feeds are captured during the whole day and the detection algorithms should consider the light transition.
- physical phenomena: image contrast is influenced by various physical phenomena occurring during video acquisition. For instance, typhoons, storms or sea currents can easily compromise the contrast and the clearness of the acquired videos;
- grades of freedom: while videos containing traffic images or pedestrians are virtually confined in two dimensions, in underwater videos the moving objects can move in all three dimensions;
- algae formation on camera lens: the contact of sea water with the camera's lens facilitates the rapid formation of algae on top of;
- periodic and multi modal background: arbitrarily moving objects (e.g. stones) and periodically moving ob-

MAED'12, November 2, 2012, Nara, Japan.

jects (e.g. plants subject to flood-tide and drift) are a common finding in the underwater setting.

In this paper, we conduct a performance evaluation of six state-of-the-art object detection algorithms on underwater footage. These algorithms were selected based on their ability to deal with one or more of the aforementioned properties of the underwater environments. The remainder of the paper is as follows: Section 2 briefly describes the object detection algorithms, highlighting their advantages and disadvantages in the unconstrained underwater setting, while in Section 3 a performance evaluation of the aforementioned algorithms applied to a large number of underwater videos is presented. Finally in Section 4 conclusions are drawn and suggestions for improvements are given.

2. ALGORITHM DESCRIPTION

Until now, no object detection algorithm exists that can deal robustly with every single peculiarity in real life videos. That said, the performance of all the devised algorithms depend largely on a specific application domain and they generally deal with one or more characteristic difficulties found in the videos, but not with all.

In this work, six state-of-the-art object detection algorithms were used in order to evaluate their performance in unconstrained underwater videos:

- Gaussian Mixture Model (GMM) [15]
- Adaptive Poisson Mixture Model (APMM) [2]
- Intrinsic Model (IM) [8]
- Wave-back (WB)[10]
- CodeBook (CB) [5]
- Video Background Extraction (ViBe) [1]

The well-known “Gaussian Mixture Model” (GMM) and the “Adaptive Poisson Mixture Model” (APMM) have been implemented, in order to evaluate how mixture based algorithms perform. In fact, multi-modal and periodic backgrounds (i.e. backgrounds that can assume different forms in an aperiodical or periodical manner) can be handled by using a mixture of background models. A background model can be created by using statistical distributions reflecting each pixel’s intensity values. The classification of a pixel as belonging either to the background or to the foreground depends on whether there exists any distribution with enough supporting evidence. The distributions’ weights are modified in order to keep the background model as up-to-date as possible. Using distribution mixtures results in a flexible way to handle sudden and global lighting changes and other casual variations in the scene. Mixture-of-model approaches can potentially converge to any arbitrary distribution providing enough number of observations, but the computational cost grows exponentially as the number of models in the mixture increases. A drawback of these methods is that they ignore the temporal correlation of color values. This does not allow to distinguish a periodic background motion such as swaying plants driven by drift, algae on camera lens, rotating objects and so on, from the foreground object motion.

The GMM algorithm uses Gaussian distributions to model the background and deals very well with videos containing multi-modal backgrounds but it cannot handle frequent or abrupt lighting changes. APMM, which employs Poisson distributions instead, based on the observation that the intensity of pixels is Poisson-distributed, should deal with abrupt illumination variations better.

The Wave-back algorithm applies frequency decomposition on each pixel’s history vector in order to catch periodic background movements. In particular, given a set of previous frames, the Discrete Cosine Transform coefficients of each pixel are calculated and are compared to the respective background coefficients resulting in a distance map and by subsequent thresholding, the foreground objects can be extracted. The Wave-back algorithm should perform well in low-contrast videos with repetitive scenes but should suffer in videos with erratic and fast fish movement and when sudden lighting transitions occur.

In order to handle situations deriving from illumination fluctuations, the Intrinsic Model algorithm was used. In particular, this algorithm consists in dividing the scene in two parts: the reflectance part of the image, which is a static component that remains more or less the same, and the illumination part of the image, which is a dynamic component that represents the current luminosity and varies according to the lighting condition. The background is modeled by calculating the temporal median of these components. In order to model the background the IM algorithm calculates the temporal median of these two components and the number of images involved in the process is defined by a window period.

The Codebook algorithm is an adaptive background subtraction algorithm that maintains a model of the background by keeping the last pixel values in terms of codewords, which represent the RGB coordinates of the pixels and the light intensity range.In order to decide whether a pixel belongs to a moving object or to the background, the Codebook algorithm compares each pixel value with the stored codewords and calculates the color distortion(i.e. the difference between the current pixel’s RGB value and each codewords’s RGB value) and checks if the gray-level value fits inside the light intensity range. In case the color distortion is less than a predefined threshold and the light intensity fits in the intensity range the pixel is classified as a background pixel.

The ViBe algorithm is a pixel-based technique that extracts the background in video sequences. In particular, ViBe maintains a model of the background that it is not only based on the previous pixel values, but also on the neighboring pixels. In order to classify whether a pixel belongs either to the background or the foreground, it calculates the Euclidean distance between the current pixel value and older ones. If these distances result less than a certain threshold, the pixel belongs to the background. In the background updating step, ViBe does not only update the history of the current actual pixel, but also the history of the neighboring pixels, thus exploiting spatial information. Moreover, in order to deal with repeating motion patterns that belong to the background, ViBe does not make any temporal distinction on the background values (i.e. it does not consider recent pixel values more important than older pixel values), so the updates occur randomly.

3. PERFORMANCE EVALUATION

In order to evaluate the performance of the aforementioned algorithms for object detection, we compared the obtained results to a set of hand-labeled videos. The objects of interest were the fish present in the videos, so in order to filter out every non interesting object, the algorithms' outputs were fed to a post processing module [14].

For the evaluation of the detection performance we used eight videos (10 minutes long each) from the Fish4Knowledge[1] project's repository. Four of the videos had resolutions of 320×240 with a 24-bit color depth at a frame rate of 5 fps and the other four had a resolution of 640×480 at a frame rate of 24 fps. The videos were selected based on the presence of specific features to test the performance of each algorithm when extraordinary conditions occurred. In particular, the features considered were: dynamic backgrounds, lighting variations, high water turbidity, low contrast and camouflage phenomena. The ground truth of these videos was drawn by using the tool described in [4].

The performance of the detection algorithms were evaluated, at the blob level, in order to test their capabilities in detecting effectively objects, in terms of detection rate (DR) and false alarm rate (FAR) which are defined as:

$$DR = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{1}$$

$$FAR = \frac{N_{FP}}{N_{TP} + N_{FP}} \tag{2}$$

where N_{TP}, N_{FP} and N_{FN} are the number of true positives, false positives and false negatives.

Algorithm	PDR	$PFAR$	DR	FAR
GMM	90.9%	19.6%	66.6%	8.3%
APMM	83.2%	17.2%	59.4%	21.3%
IM	86.1%	22.2%	44.7%	30.8%
WB	85.9%	27.1%	39.4%	29.7%
VB	93.4%	15.5%	82.5%	9.5%
CB	85.2%	15.6%	62.4%	18.9%

Table 1: Performance of the algorithms on all videos, at their best operating points.

The performance of the algorithms were also evaluated at pixel level, in order to test their potential in preserving the objects' shapes, in terms of pixel detection rate and pixel false alarm rate. We assessed the number of true positives (pixels correctly classified as belonging to the foreground), false positives (background pixels classified as foreground) and false negatives (undetected foreground pixels) for each fish correctly identified by a detection algorithm. According to these values, we then computed the pixel detection rate and the pixel false alarm rate.

For both the pixel level and blob level, we assessed the performance with different threshold values, and the performance achieved were represented by Receiver Operating Characteristic (ROC) curves. The ROC curves at the blob level and at the pixel level are shown in Figure 1 and Figure 2, respectively, while in Table 3 the performance of the considered algorithms at their best operating points are shown.

At the blob level, the performance of the algorithms is generally good in the first four videos which contained scenes under normal weather and lighting conditions and more or less static backgrounds, except from the Wave-back algorithm. In fact, Wave-back could not reach a DR value of 50%, with an acceptable FAR value, in any video. Moreover Wave-Back, together with the Intrinsic Model algorithm, achieved the highest FAR rates in all the settings, but the Intrinsic Model algorithm generally achieved better DR rates. On the other hand, the ViBe algorithm excelled in nearly all the videos, both in terms of DR and FAR. The mixture model based algorithms performed somewhere in the middle, with the Gaussian-based approach resulting slightly better than the Poisson-based algorithm. Last, the Codebook based algorithm gave the best results, comparing it to the other algorithms, in the high resolution videos but the average values were influenced negatively by its performance in the low-resolution videos.

At the pixel level, while all the algorithms show a good pixel detection rate, i.e. they are able to correctly identify pixels belonging to an object, with values in the range between 83.2% (APMM) and 93.4% (ViBe), they suffer from a relatively high pixel false alarm rate (background pixels included into the objects' blobs), especially the Intrinsic Model and Wave-back algorithms, mostly, when the contrast of the video was low, a condition encountered during low light scenes and when violent wheather phenomena were present (typhoons and storms).

4. CONCLUSIONS

In this paper we shown how six state-of-the-art object detection algorithms performed when applied on unconstrained underwater videos. The algorithms, generally, performed well when the videos contained clear-water scenes and uniform backgrounds. When temporal phenomena, like tropical storms and hurricanes, were present, the performance of all the algorithms, in detecting objects, degraded at such point that we could safely say that they become unusable. Given that all the algorithms achieved a PDR of at least 50% as compared to the ground truth and since the algorithms deal with different aspects in the processed videos, it would be interesting to try to combine all the algorithms together by using Adaboost [3], in order to produce a more reliable classifier based on the best characteristics of each single detection algorithm. If such approach produces solid results, an MAP-MRF filter, like the one proposed in [11], could be applied afterwards in order to diminsh the number of false negatives and to better preserve the shapes of the detected objects.

5. REFERENCES

[1] O. Barnich and M. Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 20(6):1709–1724, June 2011.

[2] A. Faro, D. Giordano, and C. Spampinato. Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection. *Appearing on IEEE Transactions on Intelligent Transportation Systems*, 2011.

[3] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting, 1995.

[1] http://fish4knowledge.eu

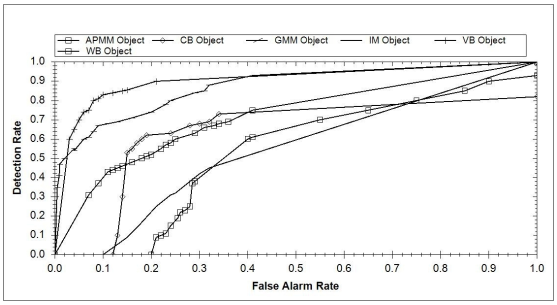

Figure 1: ROC Curves of the performance of the algorithms in object detection.

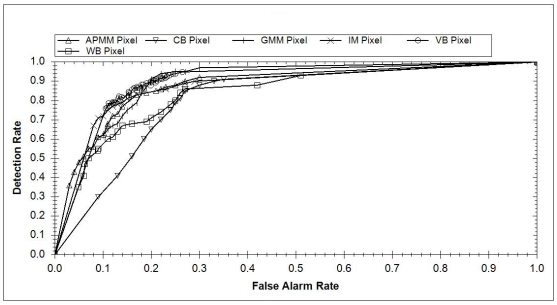

Figure 2: The performance of the algorithms in pixel detection.

[4] I. Kavasidis, S. Palazzo, R. Di Salvo, D. Giordano, and C. Spampinato. A semi-automatic tool for detection and tracking ground truth generation in videos. In *VIGTA '12: Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, pages 1–5, New York, NY, USA, 2012. ACM.

[5] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis. Background modeling and subtraction by codebook construction. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 5, pages 3061 – 3064 Vol. 5, oct. 2004.

[6] S. G. Narasimhan and S. K. Nayar. Vision and the atmosphere. *Int. J. Comput. Vision*, 48(3):233–254, July 2002.

[7] S. K. Nayar and S. G. Narasimhan. Vision in bad weather. In *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, ICCV '99, pages 820–, Washington, DC, USA, 1999. IEEE Computer Society.

[8] F. Porikli. Multiplicative background-foreground estimation under uncontrolled illumination using intrinsic images. In *in Proc. of IEEE Motion Multi-Workshop*, 2005.

[9] F. Porikli. Achieving real-time object detection and tracking under extreme conditions. *Journal of Real-Time Image Processing*, 1:33–40, 2006. 10.1007/s11554-006-0011-z.

[10] F. Porikli and C. Wren. Change detection by frequency decomposition: Wave-back. In *Proc. of Workshop on Image Analysis for Multimedia Interactive Services*, 2005.

[11] Y. Sheikh and M. Shah. Bayesian object detection in dynamic scenes. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 74 – 79 vol. 1, june 2005.

[12] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. Fisher. *Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos*, volume 2, pages 514–519. 2008.

[13] C. Spampinato, D. Giordano, R. Di Salvo, Y.-H. J. Chen-Burger, R. B. Fisher, and G. Nadarajan. Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, ARTEMIS '10, pages 45–50, New York, NY, USA, 2010. ACM.

[14] C. Spampinato, S. Palazzo, B. Boom, J. van Ossenbruggen, I. Kavasidis, R. Di Salvo, F.-P. Lin, D. Giordano, L. Hardman, and R. Fisher. Understanding fish behavior during typhoon events in real-life underwater environments. *Multimedia Tools and Applications*, pages 1–38. 10.1007/s11042-012-1101-5.

[15] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Cat No PR00149*, 2(c):246–252, 1999.