Evaluation of an object detection system in the submarine environment

Farah Rekik
Computer and embedded system laboratory, National Engineering School of Sfax
3052 Sfax,Tunisia
farah.rekik@enis.tn

Walid Ayedi
Computer and embedded system laboratory, National Engineering School of Sfax
3052 Sfax,Tunisia
ayadiwalid@yahoo.fr

Mohamed Jallouli
Computer and embedded system laboratory, National Engineering School of Sfax
3052 Sfax,Tunisia
mohjallouli@gmail.com

ABSTRACT

The object detection in underwater environment requires a perfect description of the image with appropriate features, in order to extract the right object of interest. In this paper we adopt a novel underwater object detection algorithm based on multi-scale covariance descriptor (MSCOV) for the image description and feature extraction, and support vector machine classifier (SVM) for the data classification. This approach is evaluated in pipe detection application using MARIS dataset. The result of this algorithm outperforms existing detection system using the same dataset. Computer vision in underwater environment suffers from absorption and scattering of light in water. Despite the work carried out so far, image preprocessing is the only solution to cope with this problem. This step creates a waste of time and requires hardware and software resources. But the proposed method does not require pretreatment so it accelerate the process.

Keywords

Object detection; pipe detection; underwater imaging; descriptor; classifier.

1. INTRODUCTION

The investigation of the underwater environment is a concern in many sectors such as oceanographic research[Eri01a], military applications [Baz00a] and recently the offshore wind energy[Vol17a] with the desire to exploit natural resources for more than 1000 meters deep. Automatic detection of manmade object laying on the seafloor is an important project for international marine research. A great attention is paid to this area. Autonomous Underwater Vehicles (AUV), used for this kind of project, are equipped by sonar system.

Compared to sonar, vision is not widely used in underwater research. This is due to degradation of image quality caused by absorption and scattering of light in water. But sonar suffers from several problems like cost resolution and complexity of use. Therefore there is a need for additional investigation to assess the actual potential of visual perception in underwater environments. Actually, the underwater video is increasingly used as a complementary sensor to the sonar especially for detection of objects or animals. However, the underwater images present some particular difficulties including natural and artificial illumination, color alteration and light attenuation [Bat10a]. Therefore in the detection of underwater objects it is impossible to take only color as a detection criterion, but location, shape and color information must be combined.

Object detection is based on the extraction of discriminative features. This extraction is done by the meaning of a descriptor which describes the image through a characteristic vector using specific features which differ from one descriptor to another.

In this paper the detection algorithm that we will adopt, take into account structural and content features such as pixel coordinates, intensity, gradients, etc. Feature extraction is the first step in which MSCOV [Aye12a] descriptor combines multi-scale features into a covariance matrix. Then, the classification of those matrixes is done by the SVM classifier to generate the detection model. The experiments are performed on MARIS [Ole15a] datasets and compared with object detection algorithm based on multi-scale graph-based segmentation (MGS) [Kal15a] and pixel-feature clustering (PFC) [Kal14a] method.

The rest of the paper is organized as follows. Section 2 reviews recent solution. In section 3, we describe the method proposed for better underwater object detection. The experimental setup and experimental results are presented in section 4. The paper concludes in section 5.

2. RECENT SOLUTIONS

Computer vision in underwater environment

In recent years, the interest of the scientific community in underwater computer vision has increased, taking advantage of the evolution of sensor technology and image processing algorithms'.

In [Weh14a], Wehkamp and Fischer described a workflow for stereoscopic measurements for marine biologists by providing instructions on how to assemble an underwater stereo-photographic system with two digital consumer cameras with underwater calibration. However their study didn't take into account the degradation of underwater images, and absorption and scattering phenomena in water. Artificial vision applications in underwater environments include detection and tracking of submerged artifacts. In [San13a] an embedded stereo-vision system for underwater object detection was presented based on FPGA technology. The system achieved a throughput of 26.56 frames per second (800x480 pixels). It is a high performance system in the hardware and a software levels. A standard in-air calibration procedure was adopted. Ortiz et al. proposed in [Ort02a] a single-camera vision system for real-time underwater cable tracking to detect power cables laid on the seabed.

Underwater object recognition using computer vision is difficult due to the lighting condition of such environment. Kim et al. [Kim12a] present a vision-based object detection method based on template matching and tracking for underwater robots using artificial objects. Results are not performed in submarine context, but only in a swimming pool. An important difficulty in the processing of underwater images came from the problem of attenuation of light in water. Bazeille et al. [Baz00a] cope with this problem and discuss the color modification in underwater environments and experimentally assess the performance of object detection based on color.

According to the pipe detection in underwater environment we will detail two approaches with which we will compare our work: Pixel-Feature Clustering (PFC) and Multi-scale Graph-based Segmentation (MGS). PFC algorithm performs clustering on the pixels of the image according to local pixel features and next selects the connected components according to the shape [Kal14a]. Features used by PFC, which are extracted from each pixel, consists of the color channels of HSV space, respectively hue , saturation and value , and of the gradient response to a Sobel filter. MGS algorithm belongs to the popular graph-cut approach representing the image as a grid graph. It first exploits color uniformity to enable better partitioning of the image into homogeneous regions, and then finds the shape searching for regular contours [Kal15a]. PFC and MGS algorithms reach a good detection rate using a small dataset. However they must their performance must be proved in a larger benchmark.

Image descriptors

2.1.1 Global descriptor

Global description consists on describing the whole image by their characteristics taken from each pixel. The color histogram is the best-known descriptor in this context. It represents the distribution of intensities or color components of the image. The most used global descriptors are statistics descriptors. They are determined following a frequency filtering, starting from the co-occurrence matrices, or from the first-order or high-order statistics.

Global descriptors are known for their speed and simplicity of implementation. The combination of several global characteristics can achieve good results. However, global description suffers from several problems. It implicitly assumes that the entire image is related to the object. Thus, any incoherent object would introduce noise affecting the description of the object. This limitation encourages the use of local descriptors, or even regions.

2.1.2 Local descriptor

Local description is based on the identification of local points of interest with a vector of attributes, and on the use of local descriptors which characterize only a small part of the image. SIFT[Par06a] is the most popular local descriptor.

The most interesting property of this descriptor is its robustness to the image transformation. The problem is that the objects are represented by a variable number of points of interest, whereas the classifiers require a vector with a fixed size as input. As for the descriptors by region, the feature vector is fixed, which is more suitable for classifiers.

2.1.3 Region descriptor

This approach consists in decomposing the image into a set of fixed or variable size regions and then characterizing each of these regions. The decomposition is done in a predictable way in order to make the regions' characteristics homogeneous with each other. These descriptors have recently been successful in several applications. Covariance descriptor [Tuz06a] is mainly used in human detection and re-identification. On the other hand, this descriptor has some limitations. Indeed, it implicitly assumes that the whole region is connected to the object to be modeled while the latter may have an incoherent shape.

MSCOV came to improve this descriptor by the adjustment of the trade-off between the local and the global description of the objects. This descriptor will be detailed in the next section.

3. PROPOSED SOLUTION

Object detection approach

Detecting object in underwater environment requires much attention. The purpose is identifying the region of interest which contains the target. Objects are characterized by the variation of textures, colors and shape. Those features will be adopted to create a model. The descriptor makes it possible to transform an image into a characteristics vector by extracting discriminative information of the class sought, in order to train the classifier.

The classifier is created from the database of positive and negative samples to define a detection model. Positive samples are images containing object and Negative samples are those which do not contain an object.
Then the image of the scene is scanned to find the candidates that will be compared to the database through the classifier.

The structure of an object detector is defined by the following figure.

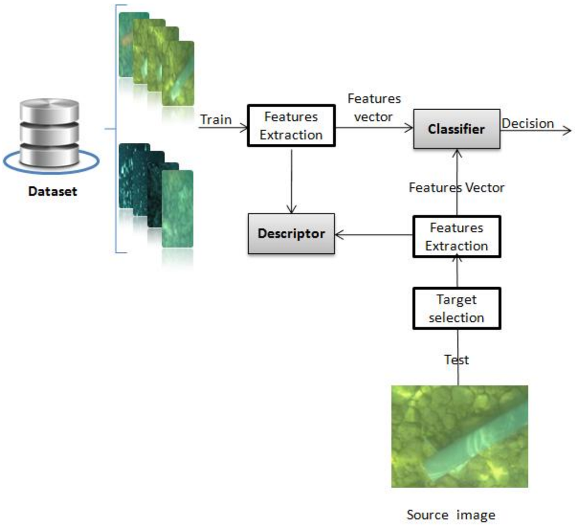

Figure 1. Detector approach

Multi-scale covariance descriptor

The descriptor adopted for this work is the multi-scale covariance descriptor [Aye12a]. It is based on the quadtree structure which explains the multi-scale aspect.

This structure is widely used for image representation in computer vision applications [Kim00a], [Lin01a], [Mal99a]. It is also used to store and index image characteristics and region of interest.

The quadtree represents a hierarchical structure constructed by recursive divisions of the image in four disjoint quadrants with the same size, according to homogeneity criterion, until a stop condition is reached. Fig 2.a presents a quadtree applied to a frame of the dataset used in this work.

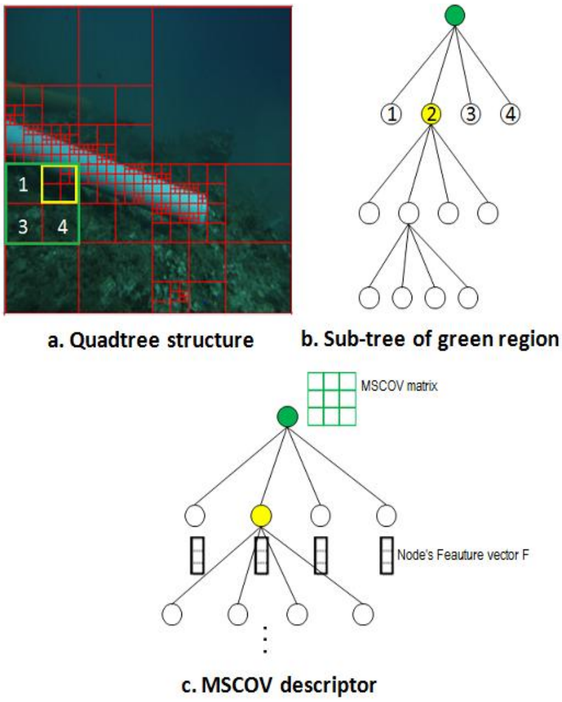

Figure 2. MSCOV and quadtree representation

Each image quadrant is represented by a quadtree node and the root node represents the whole image. Fig2.b represents a sub-tree of the green region in the pipe image.

As described in Fig2.c the MSCOV descriptor characterizes a quadrant image through the characteristics stored in its associated node. Each node stores a features vector defined by:

$$F = \left[x, y, Y, I, C_r, C_b, I_x, I_y, grad, mag \right]$$

Where x is the x location of the corresponding image quadrant, y is the y location of the corresponding image quadrant, Y is the node level, I is the I grayscale intensity value (the luminance component), C_r is the red chrominance component, C_b blue chrominance component, I_x is the norm of the first order derivatives in x, I_y the norm of the first order derivatives in y, grad is the gradient and mag is the magnitude.

Features are arranged in two groups. The structural characteristics which are related to image data location (x, y and Y), and the content characteristics that are derived from the color information (I, Cr, Cb, mag, grad).

Feature vectors in each node are combined into a covariance matrix defined by:

$$C_r = \frac{1}{N_r - 1} \sum_{c=1}^{N_r} (F_c - m)(F_c - m)^T$$

Where Nr is the number of the node in the sub-tree of r, m the mean of the nodes features and Fc the feature vector of the node c descendant of r. This structure is nominated as ‘‘Image Quadtree Features’’ (IQF).

The multi-scale covariance descriptor provides two main advantages. In fact, the decomposition into a quadtree makes it possible to capture the region of interest of the image (points of interest), and consequently it reduce the impact of noise and background information on the description of the object. Therefore the pre-treatment step is canceled. In addition, quadtree is used as a multi-level structure to extracts image features from different scales. It thus makes it possible to optimize the compromise between the local and the global description of the object.

Fig 3 and 4 show the difference between the ordinary description based on pixels of the whole region, and the multi-scale description based on the nodes. The latter is based on both the global information and the local information of the object, and focuses implicitly on the object described.

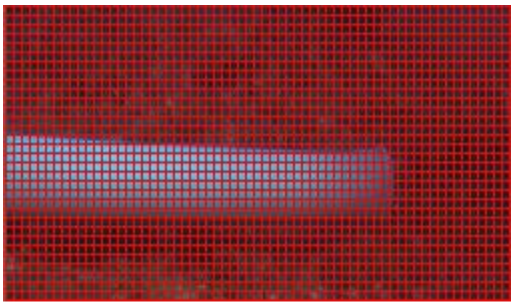

Figure 3. Global description from pixel

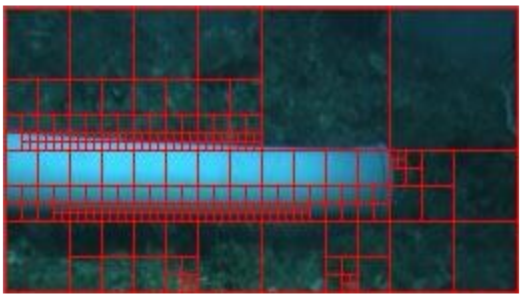

Figure 4. Local description from nodes

4. EXPERIMENTAL RESULT

This section is organized as follows. First we present the adopted evaluation metric. Then we describe the dataset used in the experimentation. Finally we compare our results to Fabjan et al[Kal15a] works using the same dataset.

Evaluation metric

Precision and recall and F-measure are the appropriate metric to evaluate the detection accuracy.

They have always been used for the evaluation of pattern detection algorithms. They are defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Where TP is the true positive, FN is the false negative and FP is the false positive. TP is the number of images real positive and predicted positive. FP is the number of image real positive and predicted negative. FN is the number of image real negative and predicted positive.

F-Measure is also a measure of a test's accuracy. It is the harmonic-mean of precision and recall. It’s defined by:

$$F_{Measure} = \frac{2.Reall.Precision}{Recall + Precision}$$

Maris dataset

Maris dataset [Ole15a] is used to evaluate the performance of the proposed approach.

This dataset is acquired using a stereo vision system near Portofino (Italy). It provides images of cylindrical pipes with different color submerged at 10 m deep. The dataset include 9600 stereo images in Bayer encoded format with 1292x964 resolution, and it include positive (frames containing a pipe) and negative (frames presenting only the background) frame Fig 5.

Figure 5. Maris dataset simples’

From this dataset we will use 305 frames, presented in Table 1, to train the classifier.

Graphics	Samples number	Positive simples	Negative simples
Training data	305	205	100

Table 1. Training dataset

This set of data presented previously contains exactly the same frames used by Fabjan et al in [Kal15a] in order to make subjective comparison. For our test we divide those frames in two classes. The first class contains positive frames, and the second one contains negative frames.

Results

Our work is compared to PFC and MGS algorithms using the same samples of Maris dataset.

This dataset is taken using two stereo camera a left and a right camera providing pairs of frame. We use only the left camera frames after been resized to350x250 pixels. Frames contain one or more pipe with different color. PFC and MGS take only one pipe in the foregrounds. But the proposed solution can take more than one. Therefore the performance will be grater.

Table 2 illustrates the performance parameters for the tree methods.

	MSCOV	PFC	MGS
TP	205	179	177
TN	96	91	94
FP	**4**	11	7
FN	**0**	24	27
Precision	98,08%	94.2%	96.2%
Recall	100%	88.2%	86.8%
Detection Accuracy	**98,68%**	89.2%	88.9%
1-FPRate	96%	89.2%	93.1%
F-Measure	99,03%	91.2%	91.2%

Table 2. Detection result for PFC, MGS and MSCOV algorithms on Maris dataset

The MSCOV detection accuracy is the best. There is an improvement of 10% in the detection accuracy. The Recall metric reach 100% for MSCOV, but never exceed 89% for the others. As a result the proposed approach outperforms PFC and MGS algorithms.

5. Conclusion

In this paper we have presented a novel underwater object detection algorithm based on multi-scale covariance descriptor for features extraction, and SVM classifier for data classification. This algorithm has been tested with Maris dataset and compared to PFC and MGS algorithms using the same frames. The experimental results show that it outperforms compared methods, and detection accuracy can reach 98%. In future work we will use a larger dataset in order to generalize the result and show that this adopted approach is valid in the submarine environment.

We will also work on descriptor parameters to improve the detection accuracy. There are two principal parameters: the homogeneity error tolerance ε and α the precision degree of description.

6. ACKNOWLEDGMENTS

Our thanks to RIZZINI Dario Lodi for allowing us to use Maris dataset and to compare our approach with theirs.

7. REFERENCES

[Aye12a] Ayedi, W., Snoussi, H., & Abid, M. A fast multi-scale covariance descriptor for object re-identification. Pattern Recognition Letters, 33(14), pp.1902-1907, 2012.

[Bat10a] Barat, C., & Phlypo, R. A fully automated method to detect and segment a manufactured object in an underwater color image. EURASIP Journal on Advances in Signal Processing, 568092, 2010.

[Baz00a] Bazeille, S., Quidu, I., & Jaulin, L. Identification of underwater man-made object using a colour criterion. In Conference on Detection and Classification of Underwater Targets p. xx,2007.

[Eri01a] Eriksen, C. C., Osse, T. J., Light, R. D., Wen, T., Lehman, T. W., Sabin, P. L., ... & Chiodi, A. M. (2001). Seaglider: A long-range autonomous underwater vehicle for oceanographic research. IEEE Journal of oceanic Engineering, 26(4), 424-436.

[Kal14a] Kallasi, F., Oleari, F., Bottioni, M., Rizzini, D. L., & Caselli, S. Object detection and pose estimation algorithms for underwater manipulation. In Advances in Marine Robotics Applications (AMRA)-in International Conference on Autonomous Intelligent Systems (IAS) pp. 1-7, 2014.

[Kal15a] Kallasi, F., Rizzini, D. L., Oleari, F., & Aleotti, J. Computer vision in underwater environments: A multi-scale graph segmentation approach. In OCEANS 2015-Genova pp. 1-6, 2015.

[Kim12a] Kim, D., Lee, D., Myung, H., & Choi, H. T. Object detection and tracking for autonomous underwater robots using weighted template matching. In OCEANS, pp. 1-5, 2012.

[Kim00b]Kim, H. K., & Kim, J. D. Region-based shape descriptor invariant to rotation, scale and translation. Signal Processing: Image Communication, 16(1), p. 87-93,2000.

[Lin01a] Lin, S., Ozsu, M. T., Oria, V., & Ng, R. An extendible hash for multi-precision similarity querying of image databases. In VLDB pp. 221–230,2001.

[Mal99a] Malki, J., Boujemaa, N., Nastar, C., & Winter, A. Region queries without segmentation for image retrieval by content. In International Conference on Advances in Visual Information Systems pp. 115-122. Springer Berlin Heidelberg, 1999.

[Ole15a] Oleari, F., Kallasi, F., Rizzini, D. L., Aleotti, J., & Caselli, S. An underwater stereo vision system: from design to deployment and dataset acquisition. In OCEANS 2015-Genova pp.1-6,2015.

[Ort02a] Ortiz, A., Simó, M., & Oliver, G. A vision system for an underwater cable tracker. Machine vision and applications, 13(3), pp.129-140 2002.

[Par06a] Park, U., Jain, A. K., Kitahara, I., Kogure, K., & Hagita, NVise: Visual search engine using multiple networked cameras. In *Pattern Recognition,. 18th International Conference on* IEEE Vol. 3, pp. 1204-1207,IEEE,2006.

[San13a] Sánchez-Ferreira, C., Mori, J. Y., Llanos, C. H., & Fortaleza, E. Development of a stereo vision measurement architecture for an underwater robot. In Circuits and Systems (LASCAS),IEEE Fourth Latin American Symposium on pp. 1-4, 2013.

[Tuz06a] Tuzel, O., Porikli, F., & Meer, P. Region covariance: A fast descriptor for detection and classification. In *European conference on computer vision* Springer Berlin Heidelberg, pp. 589-600, 2006.

[Vol17a] Voltaire, L., Loureiro, M. L., Knudsen, C., & Nunes, P. A. (2017). The impact of offshore wind farms on beach recreation demand: Policy intake from an economic study on the Catalan coast. Marine Policy, 81, 116-123.

[Weh14a] Wehkamp, M., & Fischer, P. (2014). A practical guide to the use of consumer-level digital still cameras for precise stereogrammetric in situ assessments in aquatic environments. Underwater technology, 32(2), pp.111-128, 2014.