Performance Evaluation of Multiscale Covariance Descriptor in Underwater Object Detection

Farah Rekik(✉), Walid Ayedi, and Mohamed Jallouli

Computer and Embedded System Laboratory,
National Engineering School of Sfax, Sfax, Tunisia
farah.rekik@enis.tn, ayadiwalid@yahoo.fr,
mohjallouli@gmail.com

Abstract. Object detection is the fundamental process for the majority of the investigation projects in the submarine environment, and object detection is mainly based on image description done by the appropriate descriptor. In this paper we select and optimize parameters of multi-scale covariance descriptor for object detection in the submarine context. We adapt the descriptor parameters to be suitable to cope with the degradation of image quality in underwater environment, working on the homogeneity error tolerance and the precision degree of description. We justify the use of specific parameters values and well defined features. To perform our work we use support vector machine for data classification and Maris dataset as a benchmark.

Keywords: Multi-scale covariance descriptor · Object detection · Classification · Underwater pipe detection · Autonomous Underwater Vehicles

1 Introduction

Automatic detection and recognition of underwater object laying on the seafloor is an important project for international marine research. A great attention is paid to this area. The inspection tasks automation, recognition, detection or physicochemical parameters measurement, is strongly justified in this vast environment. Remotely operated vehicles were the first developed technique where the human is involved in the decision-making chain.

Problems to be solved then were the domain of seafloor technology and the transfer of energy and information via an umbilical link. Since the late 80 years, many research programs have emerged in the United States, Europe and Asia, to provide a solution based on autonomous vehicles. They are called Autonomous Underwater Vehicles (AUV).

These devices therefore have the ability to go in inaccessible areas and to do what other submarines could not do and go where they could not go.

Because of the inherent danger and time-sensitive nature of such missions, the next urgent priority is to embed intelligence in the AUV so that it can immediately react to the data it collects. By adapting its survey route in situ and efficiently allocating resources the AUV can collect the most informative data for the task at hand while simultaneously reducing costs [1].

S. Battiato et al. (Eds.): ICIAP 2017, Part II, LNCS 10485, pp. 258–266, 2017.
https://doi.org/10.1007/978-3-319-68548-9_24

To achieve this goal, two major obstacles must be overcome. First, an algorithm is needed that can perform the detection and recognition of underwater object in near real-time onboard an AUV with limited processing capabilities. Second, a plan is needed for specifying how the information gleaned from the detection results can be exploited to intelligently adapt the AUV route. To accomplish their mission, 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. Currently, 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. 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.

Our detection system is based on Multi-scale covariance descriptor (MSCOV) [2] and in this paper we will try to define the best parameters for better object detection.

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

2 Autonomous Underwater Vehicle

Autonomous Underwater Vehicles (AUV), also known as unmanned underwater vehicles, can be used to perform underwater survey missions such as detecting and mapping submerged wrecks, rocks, and obstructions that pose a hazard to navigation for commercial and recreational vessels. The AUV conducts its survey mission without operator intervention. When a mission is complete, the AUV will return to a pre-programmed location and the data collected can be downloaded and processed in the same way as data collected by shipboard systems.

Among the companies active in the field of underwater drones, some have AUVs equipped with sonar and video cameras. These drones are designed to detect and identify objects.

Object detection is a fundamental process for several submarine missions. Underwater surveillance and tracking require vision based control. Collision and obstacle avoidance are the basis of a safe UAV, so it's important to use a high-performance object detection and recognition system to ensure the safety of the submarine.

In this paper we are interested in the detection process and we will evaluate the descriptor parameter used for data description.

3 Image Descriptors

3.1 Global Approach

The structure of an object detection system is based on image description and data classification. This approach is described in Fig. 1.

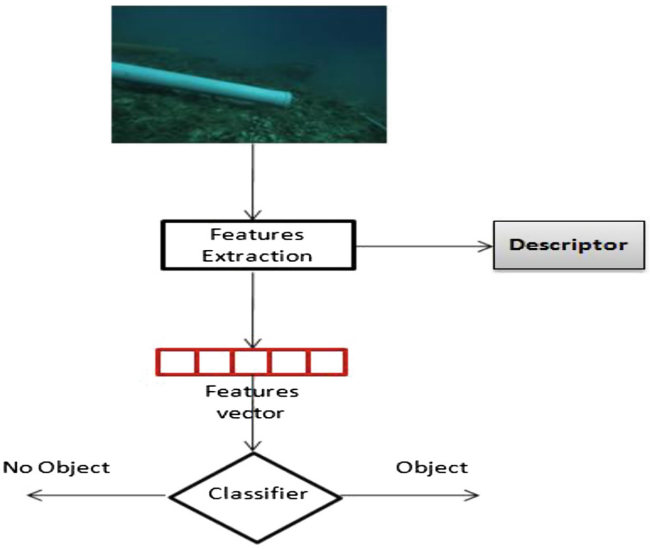

Fig. 1. Global approach for object detection

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. This vector is used to train the classifier.

3.2 Type of Descriptor

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.

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 [3] 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.

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 [4] 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.

3.3 Multiscale Covariance Descriptor

The descriptor adopted for this work is the multi-scale covariance descriptor. 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 [5–7]. 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. Figure 2 presents a quadtree applied to a frame of the dataset used in this work. Each image quadrant is represented by a quadtree node and the root node represents the whole image.

MSCOV descriptor characterizes a quadrant image through the characteristics stored in its associated node. Each node stores a features vector.

Feature vectors in each node are combined into a covariance matrix defined by (1).

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

where Nr is the node number 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

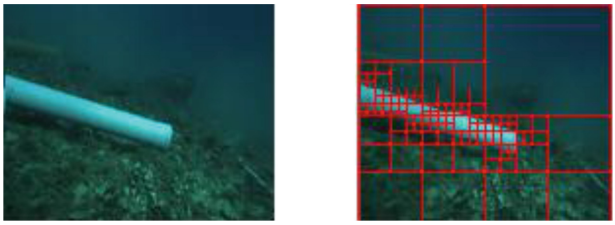

Fig. 2. Quadtree structure

"Image Quadtree Features" (IQF) [8]. Features are arranged in two groups. The structural characteristics which are related to image data location and the content characteristics that are derived from the color information as shown in Table 1.

Table 1. MSCOV features

	k	Features
Structure features	0	The x location of the corresponding image quadrant
	1	The y location of the corresponding image quadrant
	2	Node level
Content features	3	I grayscale intensity value (the luminance component)
	4	Cr color component value (the red chrominance component)
	5	Cb color component value (the blue chrominance component)
	6	Ix, the norm of the first order derivatives in x
	7	Iy, the norm of the first order derivatives in y
	8	Gradient, $o(x,y) = \arctan(\frac{I_y(x,y)}{I_x(x,y)})$
	9	Magnitude $mag(x,y) = \sqrt{I_x^2(x,y) + I_y^2(x,y)}$

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.

MSCOV depends on two principal parameters: ε the homogeneity threshold, and $\propto$ the precision degree of description. The selection of ε depends on the background complexity. Indeed, background with low intensity variation can be easily discarded from image description using low ε values. In contrast, high intensity variation needs high ε values to be discarded. It depends on the nature of the image and the object to be described. Hence a tradeoff must be determined for better object description and therefore for better object detection. By varying $\propto$, the object is described from fine

resolution level to coarse one. This parameter can considerably affect the object detection. The second factor that can also affect the detection rate is the choice of discriminative features for image description.

3.4 Support Vector Machine

Support vector machine (SVM) classifier is proposed by Vapnic [9]. It was very useful in pattern recognition. RBF is one of SVM kernel. We choose to adopt this kernel in our work with specific values of C and sigma parameter. In coming works we will compare it with other kernel in order to choose the best one with best parameters.

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 present experiments that conduce to select and optimize parameters of MSCOV descriptor in the context of underwater object detection, and justify the choice of features for pipe detection.

4.1 Evaluation Metric

Precision, 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 by (2) and (3).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

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:

$$F_{Measure} = \frac{2.\mathrm{Re}call.Precision}{Recall + Precision} \tag{4}$$

4.2 Maris Dataset

Maris dataset [9] 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 1292 × 964 resolution, and it include positive (frames containing a pipe) and negative frame (frames presenting only the background) Fig. 3.

Fig. 3. Maris dataset simples (Color figure online)

From this dataset we extract arbitrary two sets as shown in Table 2.

Table 2. Maris dataset

	Data splits	Pipes/split	Non-pipe/split
Train set	2	300	300
Test set	2	300	300

4.3 Result

To select the best combination of $\propto$ and ε we compute F-measure using Train set and Test set described in Table 2 with different values of $\propto$ and ε.

From Fig. 4 we can conclude that detection performance is better when ε range from 10 to 15 which can explain the complexity of image background. According to the histogram the best result of F-Measure is 99.83%. It is reached when ε = 15 and $\propto$ = 2.

To choose the best characteristic which walks with the submarine context and pipe detection, we evaluate the detection rate with different combination of content and structure features on Maris dataset. To reduce the number of combinations, it is possible to group the characteristics of the same context, as shown in Table 3.

The presented values in the following table are obtained after having made several tests by making random combinations of test and train subsets (pipe/non pipe) (Table 4).

Referring to the table we conclude that the best detection rate (99.91) is obtained when using F1, F2, F3, F4, F5 among all combination. It is the combination of all used structure features (location and level) and content features (shape and color).

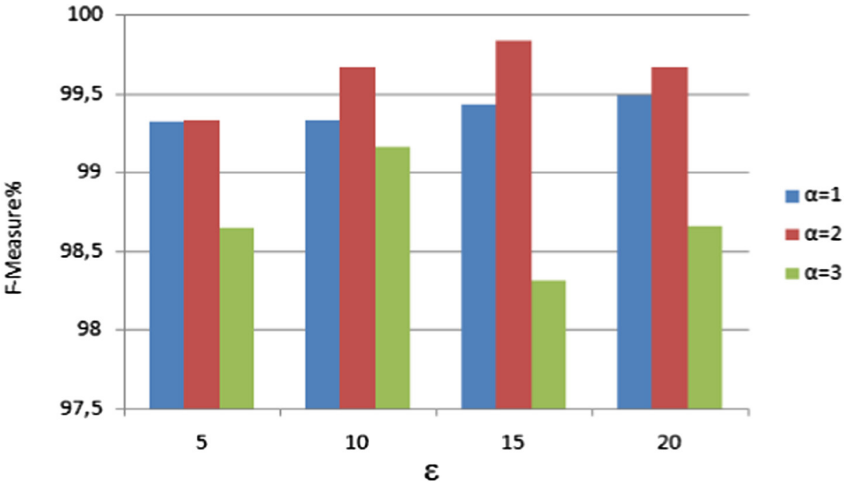

Fig. 4. F-measure for different combinations of ε and α

Table 3. New partition of MSCOV features

	F	k	Features
Structure feature	F1	0, 1, 2	Quadtree information's
Content features	F2	3, 4, 5	Luminance and chrominance components
	F3	6, 7	Norm of the first order derivatives in x and y
	F4	8	Gradient
	F5	9	Magnitude

Table 4. Detection rate and execution time for different features combination

Features	Detection rate %	Execution time ms
F1, F2	96.74	829630
F1, F3	92.58	782051
F1, F4	67.58	712023
F1, F5	66.33	645693
F2, F3	**99.08**	**309829**
F3, F4, F5	95.41	559557
F2, F3, F4, F5	98.16	463920
F3, F4, F5	93.41	429563
F1, F2, F3, F4, F5	**99.91**	**394016**

Although it has the best detection rate the execution time of F1 F2 F3 F4 F5 is not the optimal. The lowest execution time is obtained for F2 F3 combination with good detection rate (99.40). Therefore if we have a time constraint we can use this combination.

In previous work [10] we have compared this approach with PFC and MGS [11] algorithms using the same dataset. The experimental results show that it outperforms compared methods.

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. We have adopted the MSCOV descriptor for the pipe detection in the submarine context by choosing the suitable parameters and the suitable features in order to reach more than 99% as a detection rate. Those results are token using train and test sets from Maris dataset.

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 compare this result with hough transform, hog and covariance descriptors.

References

1. Williams, D.P.: On adaptive underwater object detection, April 2012
2. Ayedi, W., Snoussi, H., Smach, F., Abid, M.: The multi-scale covariance descriptor: performances analysis in human detection. In: 2012 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS), pp. 1–5. IEEE, September 2012. Park, U., Jain, A.K., Kitahara, I., Kogure, K., Hagita, N.: Vise: visual search engine using multiple networked cameras. In: 18th International Conference on Pattern Recognition, vol. 3, pp. 1204–1207. IEEE (2006)
3. Tuzel, O., Porikli, F., Meer, P.: Region covariance: a fast descriptor for detection and classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006). doi:10.1007/11744047_45
4. 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)
5. 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)
6. Malki, J., Boujemaa, N., Nastar, C., Winter, A.: Region queries without segmentation for image retrieval by content. In: Huijsmans, D.P., Smeulders, A.W.M. (eds.) VISUAL 1999. LNCS, vol. 1614, pp. 115–122. Springer, Heidelberg (1999). doi:10.1007/3-540-48762-X_15
7. Ayedi, W., Snoussi, H., Abid, M.: A fast multi-scale covariance descriptor for object re-identification. Pattern Recogn. Lett. **33**(14), 1902–1907 (2012)
8. 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)
9. Sain, S.R.: The Nature of Statistical Learning Theory. Springer, New York (1996). doi:10.1007/978-1-4757-2440-0
10. Rekik, F., Ayedi, W., Jallouli, M.: Evaluation of an object detection system in the submarine environment. In: WSCG 2017 (2017)
11. Kallasi, F., Rizzini, D.L., Oleari, F., Aleotti, J.: Computer vision in underwater environments: a multiscale graph segmentation approach. In: OCEANS 2015-Genova, pp. 1–6. IEEE, May 2015