International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Detection of Fishes in Underwater Videos Based on Signature Invariant to Scale and Rotation

Puneeth Kumar B. S[1], Suresha M.[2]

[1]Department of Computer Science, Kuvempu University, Shivamogga, Karnataka, India

[2]Department of Computer Science, Kuvempu University, Shivamogga, Karnataka, India

ABSTRACT

This paper introduces a novel shape descriptor approach for the automatic detection of fish in underwater video environment. The detection process is applied on isolated fish objects in underwater video and it requires segmentation of objects in video, segmentation is done by background subtraction method followed by extraction of representative shape signature and the similarity estimation of pairs of objects. In order to achieve an efficient object representation, a novel boundary-based shape descriptor invariant to rotation is introduced particularly for fish, formed by a set of one dimensional signals referred to as shape signature. During detection, the cross correlation metric is used to measure the degree of similarity between objects. The proposed method is invariance to scaling is performed when correlating shape signature. The proposed vision system is robust to scaling, rotation, translation. The detection performance has been examined using the Fish4Knowledge dataset.

Keywords: Cross-Correlation, Detection, Fish, Rotation, Scale, Signature.

I. INTRODUCTION

Recently in the domain of a video processing, an increasing emphasis on being able to extract the characteristics from a video. For example, in marine applications, it is essential to extract a fish objects shapes in underwater video for comparison with stored prototype fish shapes for representation. The benefits of video monitoring process over the manned videography or web, that it does now not affect marine lives and provides a large and accurate view by means of video knowledge. It is tough job to human to manually analyze huge range of video data. It is a time consuming and error prone procedure. This article is mainly focused on investigation of the fundamental principles and methods needed for an autonomous system which is able to recognize fishes in underwater videos.

Automatic fish recognition in videos is a hard video processing task consisting of numerous challenges and requirements. A robust fish recognition system should succeed to detect fish objects which are dynamically shifted, rotated, distorted in underwater video. In addition, the recognition system should recognize fish objects regardless of their mercurial scale in video environment. The fact that the shape of objects in real underwater video varies depending on the scale of observation has significant implications in portraying them. The aim of this work is to present a novel fish recognizable framework in videos that meets the above invariance prerequisites.

In this article, we present a novel vision approach for the recognition of actual fishes in underwater videos which is effective to image transformations like scaling, rotation, translation. The recognition system

is centered on the evaluation of assemblies of video regions with a previously stored view of a known prototype fish object. The process has two steps: pre-processing and recognition. The first involves the segmentation of objects from underwater video using background subtraction method [1], extraction of significant features from the objects, such as objects boundaries and the tracing of object contours. The second step takes into consideration a novel boundary-based descriptor long-established with the aid of a set of one- dimensional signals known as signatures, to allow an efficient object representation. During recognition, the cross-correlation and associated metrics of object signatures are used to measure the degree of similarity between two shapes yielding a set of similarity phrases. An experiment object shall be of no curiosity to us so long as compared with the prototype fish object, the measure of similarity stays below a suitable acceptance threshold.

II. Related Work

Object detection algorithms have various applications. For instance, computer vision techniques provide a great opportunity to make animal monitoring more accurate, less time consuming and fully automated. In particular, different approaches and methods of moving fish detection have been used for the task of fish detection, which is the fundamental step in building a fish observation system.

Palazzo et al.[2] proposed an approach for object detection that imply explicit modelling both the background and the foreground for each frame. This allows to avoid misdetection when the background is not stationary or when target objects have the same color as the background. The latter problem is also solved by introducing texture information in models as well as the color. The algorithm was evaluated on real underwater videos and showed high and stable performance.. In[3],an Expectation-maximization (EM) algorithm was proposed for fish detection. Given intensity images of fish a certain threshold is selected, such that pixels with higher intensities are assumed to belong to the fish. Then the sets of x and y locations corresponding to these pixels are defined the shape of each fish is assumed to be a multivariate Gaussian, then an image is modelled as GMM. And finally, the parameters of GMM, including the number of fish, are estimated using an EM algorithm. The method was tested on southern blue fin tuna fish, an individual fish was modelled by using a two-dimensional Gaussian distribution with x and y pixel locations as an input. An automated video processing system for underwater video surveillance is proposed [4], The system is able to solve the tasks of texture and shape analysis. In [5] compared different algorithms under extreme conditions like typhoon that somehow recall the ones present in underwater scenes such as erratic motion, sudden and global light change, presence of periodic and multimodal background, arbitrary changes in the observed scene, low contrast and noise.

In pilot studies to monitor length frequencies of fish in aquaculture cages [6]. Samples taken in aquaculture cages can approach 95% of the population and the measurement technique is noninvasive. Balance Guaranteed Optimized Tree is used for fish recognition by [7]. This algorithm is a hierarchical classification tree method which overcomes the problems and performs better than flat SVM classifier. Sparse coding for histograms of local binary patterns, applied for image categorization [8] and linear SVM is used in fish identification task of ImageClef in 2014 [9]. Background segmentation, key-points selection with an adaptive scale and description with Opponent-Shift and linear SVM is used in [10] to recognize fish. Their process reaches a good precision but a worse recall. A shortcoming of the flat classifier is that it uses the same features to classify all classes without considering that some classes have certain similarities and can be better separated by some customized features. Snout-to-tail and other body spans on the fish are measured from the video

recordings and, using a length-weight regression [11], the weights of the fish are estimated to an accuracy of a few per cent. Commercial systems such as VICASS [12] and the AQ1 AM100[13] are widely used in aquaculture to determine size distributions based on simple length and span measurements, and thereby deduce biomass from an estimated number of fish in the cage or tank.

III. Proposed Work

The flow diagram of proposed approach is shown in Figure 1.In the proposed method first segments the objects in the under video using Gaussian Mixture Model technique based on background modeling secondly A novel boundary based shape descriptor established which is invariant to Scale and Rotation and then made a similarity evaluation to estimate the degree of similarity .A test object will be detected as a fish object if measure of similarity with the prototype fish object stays above a suitable acceptance threshold T otherwise treat as futile object

A. Segmentation using background subtraction method

In order to find moving object, we extract the background of underwater video using Gaussian Mixture Model technique [14,15] based on background modeling. The simplest way to model the background is to obtain a background image which doesn't contain any moving object. The background information is obtained using a pre-pixel mixture of Gaussian. It consists of a weighted sum of Gaussian densities, which allow the color distribution for each pixel to be multimodal. Modeling the history of pixel values through a few normal distributions helps the system to be robust against occlusion and regional illumination alterations .Typically, k = 2 distributions are used, one component represents foreground pixels and the another component represents the weight (ω), mean (μ), and covariance (ρ) of background pixels, these are updated dynamically over time. The likelihood ρ

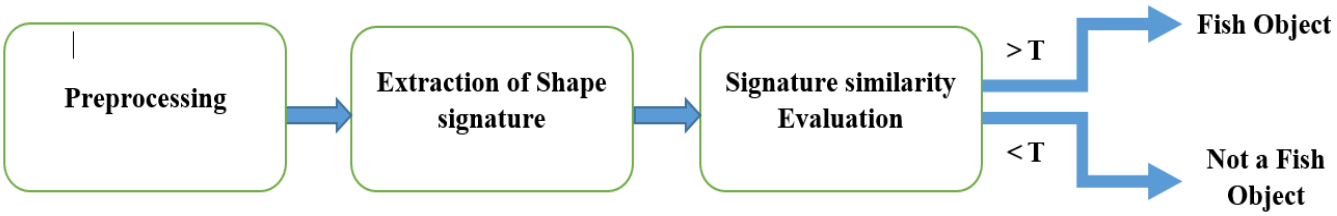

Figure 1. Flow diagram of Proposed Method

of occurrence of a color μ at current pixel i and time t is given as:

$$O(X_t = \mu) = \sum_{i=1}^{k} \left(w_{i,t}, \eta\left(X_t, \mu_{i,t}, \sum_{i,t}\right)\right), \quad (1)$$

where, $\eta\left(X_t, \mu_{i,t}, \sum_{i,t}\right)$ is the i^{th} Gaussian and $\omega_{i,t}$ its weight. The covariance $\sum_{i,t}$ is assumed to be diagonal with $\sigma_{i,t}^2$ as its diagonal elements. For each pixel, the first step consists of determining the closest equivalent Gaussian, i.e. the Gaussian for which the intensity of the pixel is within T_σ variation of its mean. The parameters of the matched component are then updated by weight, mean and covariance with α being a user-defined learning rate.

$$\omega_{i,t+1} = (1-\alpha)w_{i,t} + \alpha \quad (2)$$

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1} \quad (3)$$

$$\sigma_{i,t+1}^2 = (1-\rho)\sigma_{i,t}^2 + \rho\left(X_{t+1} - \mu_{i,t+1}\right), \quad (4)$$

where, ρ *is computed as*

$$\rho = \alpha.\eta\left(X_{t+1}, \mu_{i,t}, \sum_{i,t}\right). \quad (5)$$

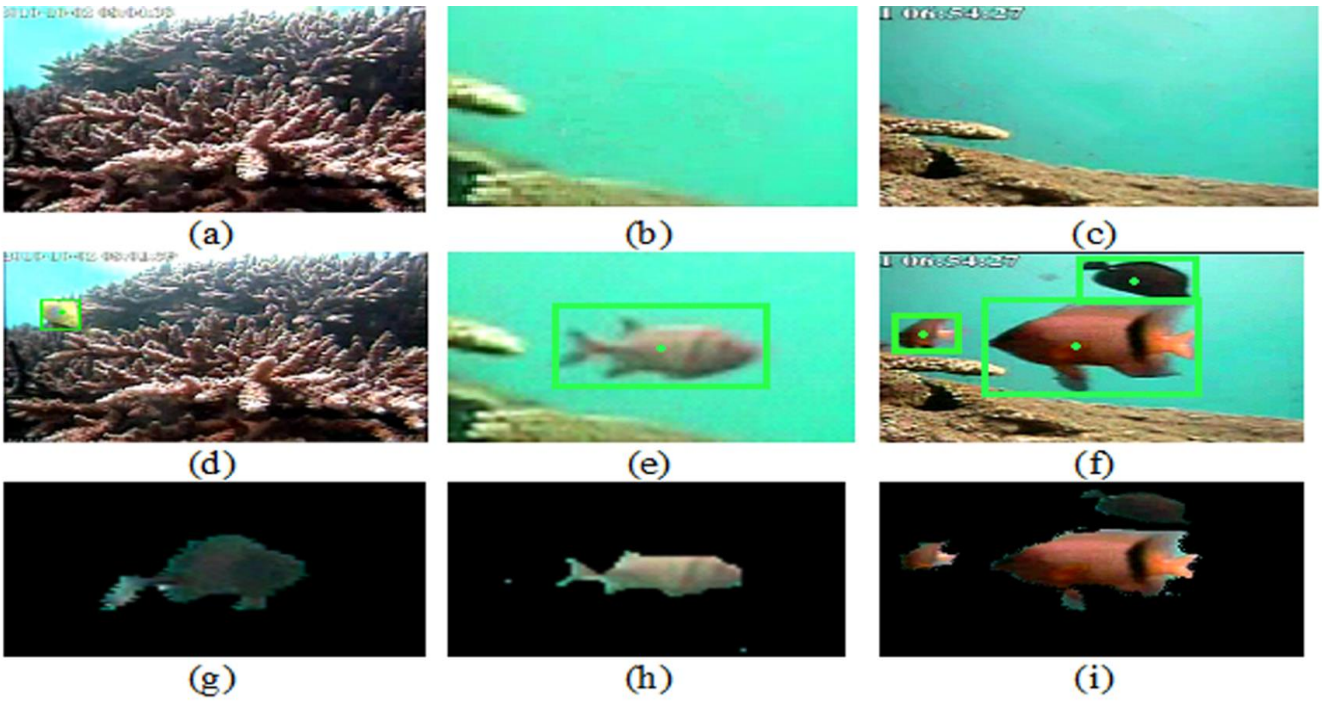

Figure 2. (a)-(c) shows the background of video, (d)- (f) shows detected moving objects, (g)-(i) shows the segmented moving object

The weight of unmatched components is updated using Eq. (6) with and $\mu_{i,t+1}, \sigma^2_{i,t+1}$ remain unchanged. If no component matched, a new Gaussian with mean X_t ,a large variance σ_0^2 and a small weight ω_0 is created to replace the existing Gaussian with the lower weight.

$$w_{i,t+1} = (1-\alpha) w_{i,t} \tag{6}$$

Once Gaussian are updated based on the value $\frac{\omega_{i,t+1}}{\sigma^2_{i,t+1}}$ only the most reliable distributions are chosen to represent the background (B), with pixels which are at more than T_α standard deviations away from any of those B distributions are labeled as foreground. Figure. 2 shows the background of the underwater video obtained using GMM technique based on background model.

$$B = \arg\min_k \sum_{i=1}^{n} \left(\omega_{i,t+1} < T_{bg}\right) \tag{7}$$

After background subtractions performed, morphological operations open are applied to the resulting foreground mask to eliminate noise. Then closing type morphological operation are used to merge

together the small features in the image that are close together and might have been separated during the background subtraction process. Then "Flood-Fill Operations" type morphological operators are used to fills the hole in the image, if there is any.

Finally, blob analysis detects groups of connected pixels, which are likely to correspond to moving objects.

In Figure. 2, the first row shows the background of underwater video obtained using GMM technique, second row represents the frames detecting moving objects in underwater video using background subtraction method, and third row shows the segmented moving objects.

B. Shape Signature

In this section, describes shape signatures based on two dimensional contour shape, propose a novel descriptor particularly for recognize the fish object, that captures the fish object's boundary information under the condition of translation, rotation and uniform-scaling invariance. It is required a number of preprocessing steps applied on object image for the object representation.

To begin with, the recovery of the shape descriptor requires the extraction of the edge map (collection of edge pixels) of the object. This can be easily obtained by applying morphological operation and edge detection process (for example the sobel operator [16]) to the image. The so called" Connected-Component Labeling" technique described in [17], is employed to represents the objects. Each connected component becomes input to a contour tracer. The boundary tracing approach employed by in this work is based on the one proposed in [18]. The output of the algorithm is an array say as *a* containing the co-ordinates of the boundary traced in a clockwise direction.

Let a fish object *F* represented by a set of traced contour points $\{(x_1, y_1) \ldots (x_n, y_n)\}$, where *x* and *y* stand for the *x* and *y*-coordinates of the *n* points on the image plane. The notation $F \equiv \{(x_1, y_1) \ldots (x_n, y_n)\}$ may be used. Then we compute the Euclidean distance between contour points and the centroid (x_c,y_c) of the fish object gives the signature S_c defined as below.

$$Sc = \left\{ \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \; for i = 1...n \right\} \quad (8)$$

aforementioned signature is varient to differently rotated fish object. To make the signature invariant to rotation, we propose a novel shape signature *S* defined as the set of the distances between the points of the traced boundary sequence and a reference point (x_k,y_k) that represents the particular position of fish. For the fish object *F*, the proposed shape descriptor *S* is defined as:

$$S = \left\{ \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2} \; for i = 1...n \right\} \quad (9)$$

From the above definition it is clear that the signature *S* is shown in equation (9) depends on the reference point (x_k, y_k); is defined by global maxima of *Sc* is shown in equation (10) and it is found that always refers the snout (mouth of a fish) as our experiments on different fish species is shown in figure.1 and we get the almost similar signatures for all fish bodies.

$$(x_k, y_k) = (x_i, y_i) \quad (10)$$

where (x_i,y_i) is the corresponding point for max(*Sc*)

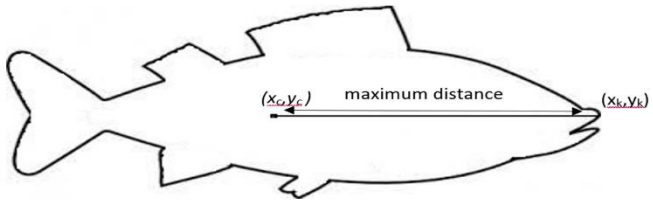

Figure.3 shows distance from centroid to snout is always maximum.

According to this representation, since each object will be described by a set, *S*, of shape signatures generated by calculating the distance functions *S* with respect to constant fish body part snout, the descriptor is invariant to changes in the location of the test object in the image and rotated object will include the same signatures as the descriptor of the original one shown in figure.4. It can be proven that uniform scaling of an object does not affect the shape of signature *S*, only its amplitude and its length are multiplied by a factor proportional to the scaling one. Therefore, as it will be shown later on, when combined with a suitable similarity measure, the proposed shape descriptor is also invariant to uniform-scaling of the Shape Signature Matching for Object Identification Invariant to test object.

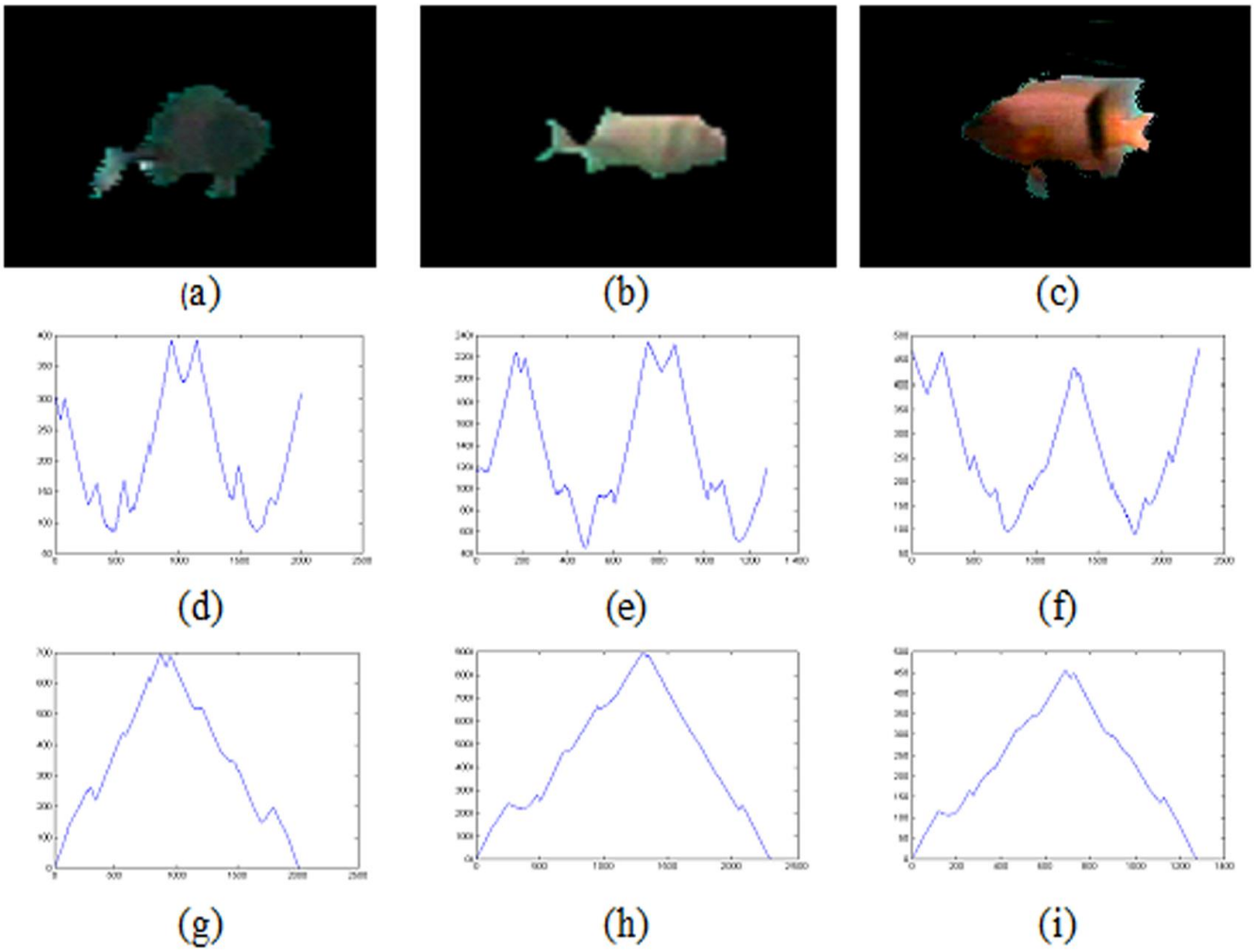

Figure 4. (a)-(c)are the segmented objects; (d)-(f) the signatures generated by considering reference point as centroid (x_c,y_c) of (a),(b) and (c) objects respectively; (g)- (i) are signatures generated with reference point(snout)(x_k,y_k) of (a)-(c) objects respectively.

C. Shape similarity Evaluation

As explained in the previous section, invariance to rotation and translation is intrinsic to the proposed shape descriptor. A special treatment is given to achieve invariance to uniform-scaling, while measuring shape similarity. In the aforementioned case (scaling) the length of the signature changes. Since the shape of a signature is not seriously affected by changes in the object's size, uniform-scaling could be handled in similarity estimation by normalizing the compared signatures and making them of the same length.

Let a prototype object consisting of N boundary points, be represented by the set of shape signatures $V \equiv \{E_i$, for $i = 1 \ldots N\}$, where V is the distance function with reference to the contour point (x_i, y_i) on the boundary of *prototype object*. Similarly, let a test object consist of M boundary points be described by the set $S \equiv \{Z_i$, for $j = 1 \ldots M\}$. Without loss of generality, let's assume that $M < N$. The similarity between prototype object and *test object* is estimated by comparing the $N \times M$ signature pairs(E_i,Z_i). In the first step of similarity extraction, the smaller of the compared objects that is *test object*, is treated as a scaled version of the bigger one that is *prototype object*, and the maximum of the Normalized Cross-Correlation (NCCC) is employed to gauge the degree of similarity between each signature pair (E_i, Z_j). Since the above measure requires the compared signatures to be of the same length, interpolation is

employed to stretch the Z_j signatures and make them of size N, yielding the signatures Z_j. The Normalized Cross-Correlation of a signature pair (E_i, Z_j) is defined as:

$$r_{ij} = \frac{\sum_n \left(E_i(n) - \bar{E}_i \right) \left(Z_j(n) - \bar{Z}_j \right)}{\sqrt{\sum_n \left(E_i(n) - \bar{E}_i \right)^2 \left(\sum_n Z_j(n) - \bar{Z}_j \right)^2}} \quad (11)$$

where, $n = 1 \ldots N$ is the signature point index and E_i, Z_j stand for the mean value of E_i and Z_j, respectively. The similarity of a signature pair (E_i, Z_j) is given by:

$$c_{ij} = (E_i, Z_j) = \max_i (r_{ij}) \quad (12)$$

The use of normalized cross-correlation is invariant to differences in the signatures' amplitude due to scaling. According to the above object comparison, for each contour point on the smaller object associated with a signature Z_j there will be a point on the bigger object associated with a signature E_i which bears the highest similarity to Z_j among all the signatures corresponding to the bigger object's contour points. Therefore, each contour point j on the smaller object is assigned a value (equal to $max_i\{C_{ij}\}$) indicating the maximum similarity of the two objects with reference to point j. Our main idea for similarity estimation stems from the fact that, when comparing two similar objects differing only in size, the $maxi\{C_{ij}\}$ values will be high (ideally equal to 1) for all the contour points.

D. Results and Evaluation

The public dataset Fish4Knowledge that includes seven video sequences (blurred sequence, Complex background sequence, crowded sequence, dynamic background sequence, and Hybrid sequence, Luminosity variation sequence, Camouflage Foreground Objects sequence) of 30 seconds video sequence was used to evaluate the performance of fish detection. Table I describes the video sequences of dataset.

TABLE I

Description Fish4Knowledge dataset use in evaluation.

Sequence	Notes
Blurred	Smoothed and low contrasted frames.
Complex Background	Background is featured by complex textures.
Crowded	With many occluding objects.
Dynamic Background	With background objects movements (e.g. plant movements due to the marine currents).
Luminosity variation	Transient and abrupt luminosity change.
Camouflage Foreground Objects	It is of camouflage.
Hybrid	It is a combination of the above conditions.

In order to illustrate the performance of fish object detection, the true detection rate TD and the false detection rate FD were adopted as defined below

$$TD = \left(\frac{TP}{N} \right) \times 100\% \quad (13)$$

$$FD = \left(\frac{FP}{TP + FP} \right) \times 100\% \quad (14)$$

where N is the total number of fish objects, True Positive(TP) is the total number of true detection fish objects, and False Positive(FP) is the total number of false detection fish objects that is some other background objects detected as fish objects.

Table II shows the results of the performance evaluation for the aforementioned Fish4Knowledge dataset. Experimental results show that the proposed method has the average true detection rate TD and false detection rate FD are 77.83 % and 12 %, respectively. The proposed method has good performance in terms of true detection rate and false detection rate.

Table II
performance Evaluation of Fish4knowledge dataset.

Sequence	N	TP	FP	TD (%)	FD (%)
Blurred	139	112	14	80	10
Complex Background	98	69	7	71	8
Crowded	210	151	19	72	9
Dynamic Background	60	41	11	68	18
Hybrid Sequence	123	95	14	77	11
luminosity variation	80	61	20	76	25
Camouflage Foreground Objects	135	110	12	81	8

IV. CONCLUSION

A new system for the identification of two-dimensional objects in underwater video scenes has been presented in this work. The identification method is a prototype-based one, where prior knowledge of the target is given by a prototype template consisting of a set of representative features. In order to make the proposed system invariant to rotation and translation, a novel shape descriptor is introduced as the set of shape signatures extracted from the object's contour and defined as the distance of the boundary points with reference to a fixed point represents the particular part of fish. To estimate similarity between two signature is generally use the cross correlating signature pairs included in their shape descriptors with that to achieve scale invariance we normalize the cross correlation and estimate the similarity

V. REFERENCES

[1] Zivkovic, Z. Improved adaptive Gaussian mixture model for background subtraction, International Conference on Pattern Recognition (ICPR-2004), 17th International Conference, vol. 2, page 28-31, 2004

[2] Palazzo, S., Kavasidis, I. and Spampinato, C., Covariance based modeling of underwater scenes for fish detection. Proceedings of 20th IEEE International Conference on Image Processing, 1481–1485, 2013.

[3] Fiona H Evans. Detecting fish in underwater video using the em algorithm. In Proceedings of the 2003 International Conference on Image Processing (ICIP) , volume 3, pages III–1029. IEEE, 2003.

[4] 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. ARTEMIS '10, pages 45–50, 2010

[5] Porikli F Achieving real-time object detection and tracking under extreme conditions. JReal Time Image Process 1 (1):33–40,2006.

[6] Harvey, E. S., Cappo, M., Shortis, M. R., Robson, S., Buchanan, J. and Speare, P. The accuracy and precision of underwater measurements of length and maximum body depth of southern bluefin tuna (Thunnus maccoyii) with a stereo-video camera system. Fisheries Research, vol.63: page 315-326.2003.

[7] P. Huang, B. Boom, R. Fisher. Underwater live fish recognition using a balance-guaranteed optimized tree Asian Conference on Computer Vision (2012).

[8] Paris S, Halkias X, Glotin H Sparse coding for histograms of local binary patterns applied for image categorization: Toward a bag-of-scenes analysis. In: 21st International Conference on Pattern Recognition (ICPR), pp 2817–2820.2012.

[9] T. Blank K,Lingrand D,Precioso F.Fish Species recognition from video using SVM classifier In: working Notes CLEF 2014 conference 2014.

[10] K. Blanc, D. Lingrand, F. Precioso, "Fish species recognition from video using svm classifier", Proceedings of the 3rd ACM International Workshop on Multimedia Analysis for Ecological Data, ACM, pp. 1-6, 2014.

[11] Pienaar, L.V. and Thomson, J. A. Allometric weight-length regression model. Journal of the Fisheries Research Board of Canada,vol. 26:page.123-131. 1969.

[12] Shieh, A. C. R. and Petrell, R. J. Measurement of fish size in Atlantic salmon (salmo salar l.) cages using stereographic video techniques. Aquacultural Engineering, 17(1):pp. 29-43,1998

[13] AQ1 Systems. http://www.aq1systems.com/products (accessed March 14, 2013)

[14] C. Stauffer, W. Grimson. Adaptive background mixture models for real-time tracking. Computer Vision and Pattern Recognition, vol. 2, pp. 246–252, 1999.

[15] T. Bouwmans, F. El Baf and B. Vachon. Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey. Recent Patents on Computer Science 1, 3, pp. 219-237, 2008.

[16] Sobel, I., Feldman, G., "A 3x3 Isotropic Gradient Operator for Image Processing", presented at the Stanford Artificial Intelligence Project (SAIL) in 1968.

[17] Haralick, R., Shapiro, L.G. (eds.): Computer and Robot Vision, vol. I, pp. 28–48. Addison Wesley, London, UK (1992)

[18] Carter, J.R.: Boundary tracing method and system. European Patent Application EP341819-A3 (1989)