Automatic Fish Classification for Underwater Species Behavior Understanding

Concetto Spampinato, Daniela Giordano, Roberto Di Salvo
Department of Informatics and Telecommunication Engineering - University of Catania
Catania, 95125, ITALY
{cspampin, dgiordan, rdisalvo}@diit.unict.it

Yun-Heh Chen-Burger, Robert B. Fisher, Gayathri Nadarajan
School of Informatics, University of Edinburgh
Edinburgh, EH8 9LE, United Kingdom
{jessicac, rbf}@inf.ed.ac.uk gaya.n@ed.ac.uk

ABSTRACT

The aim of this work is to propose an automatic fish classification system that operates in the natural underwater environment to assist marine biologists in understanding fish behavior. Fish classification is performed by combining two types of features: 1) Texture features extracted by using statistical moments of the gray-level histogram, spatial Gabor filtering and properties of the co-occurrence matrix and 2) Shape Features extracted by using the Curvature Scale Space transform and the histogram of Fourier descriptors of boundaries. An affine transformation is also applied to the acquired images to represent fish in 3D by multiple views for the feature extraction. The system was tested on a database containing 360 images of ten different species achieving an average correct rate of about 92%. Then, fish trajectories, extracted using the proposed fish classification combined with a tracking system, are analyzed in order to understand anomalous behavior. In detail, the tracking layer computes fish trajectories, the classification layer associates trajectories to fish species and then by clustering these trajectories we are able to detect unusual fish behaviors to be further investigated by marine biologists.

Categories and Subject Descriptors

I.5.4 [**Computing Methodologies**]: Pattern Recognition—*Applications Subjects: Computer vision*

General Terms

Algorithms

ARTEMIS'10, October 29, 2010, Firenze, Italy.

Keywords

Fish Species Description and Classification, Fish Behavior Understanding in Open Sea, Curvature Scale Space Transform, Fourier Descriptors

1. INTRODUCTION

Usually, marine biologists establish the existence and quantities of different types of fish using methods such as casting nets in the ocean, human underwater observation and photography [15], combined net casting and acoustic (sonar) [4] and, more recently, human hand-held video filming. These methods have two main shortcomings: first, they are invasive, hence not able to capture normal fish behaviors, and second, the quantity of collected data is not enough to describe the observed environment. In order to overcome such limitations in acquiring biological data of fish in their natural environment, in recent years embedded video cameras have been widely used. For instance, in the Taiwanese Ecogrid project[1], ten underwater cameras have been located at the Third Taiwanese Power Station for undersea coral reef and marine life observation. Currently, these videos are manually analyzed by marine biologists to find useful information. This procedure, of course, is very tedious, since it requires a lot of time and human concentration. It is also error prone since it is not realistic to fully investigate all the information in the acquired videos. In fact, an operational camera generates about 2 Terabytes of images (20 million frames) per year and it is estimated that one minute's video processing requires about 15 minutes for manual classification and annotation. Therefore, to fully analyze all existing videos, generated by the ten cameras over the past six years, would take approximately 900 man years. Hence, there is the necessity to develop automatic video processing methods to convert this huge bulk of data in accessible information for the marine biologists. In order to accommodate this need, a hybrid semantics - and planning-based approach within an integrated workflow framework has been developed by Nadarajan et al. in [12] and in [13]. In these works an automatic system supporting marine biologists' activity is developed by using video and image processing ontologies for representation and to provide a basis to enable an ontology-based planner to support workflow executions. Clearly, such

[1]http://ecogrid.nchc.org.tw

a system for intelligent underwater video analysis requires methods for fish detection, fish classification and behavior understanding. In this paper we propose an automatic system for fish species classification and tracking followed by a system for clustering fish trajectories and behavior understanding. As far as we know, so far, not much research has been carried out on fish species classification, especially in their natural environment. In fact, most of existing works for automatic fish image classification either use databases of dead fish, e.g. in [7] and [14], or work on constrained areas, e.g. in [11]. The other methods that operates directly in underwater environments classify only a few species of fish: for instance, Benson et al [3] proposed a Haar detector and classifier for the *Scythe Butterfly fish* achieving a performance rate of 89% on a database with about 3500 images, Edgington et at. in [5] developed a a classification system for the *Rathbunaster Californicus* using visual attention with an accuracy of 90% on a dataset with 6000 images, whereas Rova et al. in [16] developed an automatic classification system for the *Striped Trumpeter* and the *Western Butterfish* based on 2D textural appearance obtaining an average performance of 90% on a database with 320 images.

Abbasi and Mokhtarian [1] developed a system for retrieval of marine animal images, introducing for the first time curvature scale space analysis combined with affine transformation, with an accuracy of about 85% using a database with 1100 images.

All these systems share the use of global appearance shape descriptors, which are negatively affected by affine transformation and also leave out the use of texture features when shape descriptors are used.

The challenges that differentiate our proposed work when compared with other traditional methods are: 1) we combine texture features with shape descriptors preserved under affine transformation, 2) the images are taken in the fish natural environment and 3) fish trajectories are automatically extracted. Concerning automatic fish behavior understanding, so far, only one system has been proposed by Soori in [17], who proposes an automatic video segmentation technique to study fish schooling characteristics.

In addition to fish classification, fish trajectory dynamics analysis is carried out in order to understand fish behavior and to detect rare events, which may represent the cue of interest for marine biologists. Fish behavior understanding is performed by means of 1) a tracking system to extract fish trajectories, 2) a classification system that associates fish species to these trajectories and finally, 3) a system for trajectory clustering for each fish species (see fig.1). The analysis of trajectory's clusters, therefore, allows us to highlight unusual events, potentially of interest for marine biologists, as, for instance, one or more trajectories that don't belong to the clusters that describe normal behaviors or two clusters of trajectories where one of them contains only one trajectory may point to interesting events.

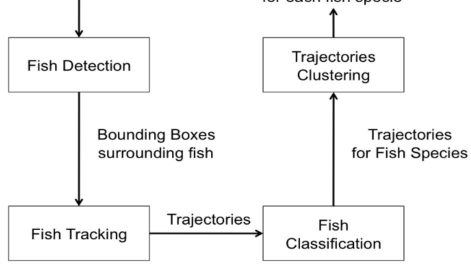

Figure 1: Fish Trajectories Analysis System

The remainder of the paper is as follows: section 2 briefly describes the tracking algorithm for fish trajectories extraction, whereas section 3 shows the texture and shape features used for fish description. Section 4 describes the classification system and its performance and section 5 describes the fish trajectory analysis system. Finally, concluding remarks are reported.

2. FISH TRACKING SYSTEM

The first step of the proposed system aims at extracting trajectories by tracking fish over consecutive frames. The tracking system, proposed by the authors in [18], first automatically detects fish (see fig. 2) by means of a combination of the Gaussian Mixture Model and Moving Average algorithms, then tracks fish by using the Adaptive Mean Shift Algorithm. The obtained accuracy for both fish detection and tracking is about 85%. The output of the detection system is shown in fig.2, where a bounding box is drawn around the fish.

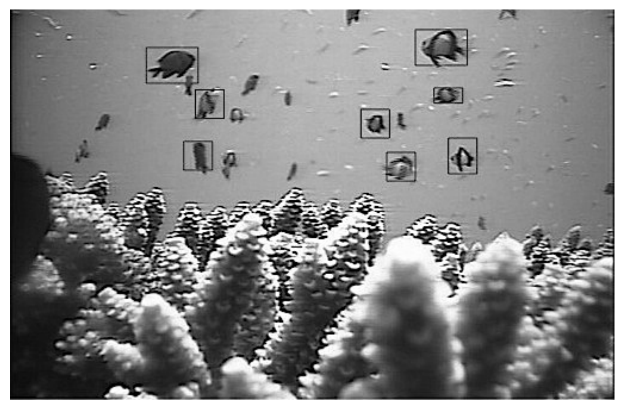

Figure 2: Output of the detection system

The tracking system allows us to extract fish trajectories. Fig. 3 shows the trajectories of two fish for a video sequence of 30 sec.

3. FISH DESCRIPTION

In order to associate fish species to the computed trajectories, a fish species classification system that works on the bounding box identified by the detection module, is carried out by a 2D affine object recognition method using invariant information from the boundary and from the texture of fish. We use affine invariants features, such as affine curvature [2] and Fourier descriptors, because they are independent of the object's position, orientation, scale and slant and usually fish can be at any position and orientation relative to the camera.

3.1 Fish Affine Transformation

In order to build a reliable classification system we use a fish affine transformation since it is necessary to represent

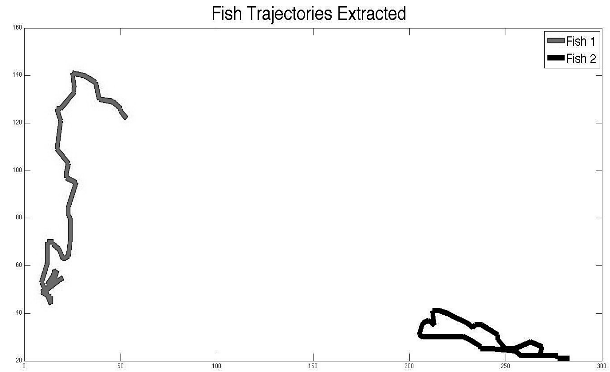

Figure 3: Output of the tracking system

3D fish shape that describes the different views while fish move in the water. An affine transformation between two planes π and π' can be described as:

$$\begin{pmatrix} x_a(\pi') \\ y_a(\pi') \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x(\pi) \\ y(\pi) \end{pmatrix} + \begin{pmatrix} h \\ k \end{pmatrix} \quad (1)$$

where a, b, c, d, h, k are real numbers that determine space rotation and translation. By choosing $a = d$ and $c = -b$ we obtain a similarity transformation. Moreover, a degree of affine deformation γ is introduced, so that each parameter can range between $1 - \gamma$ and $1 + \gamma$.

Fig. 4 shows some affine transformations with $\gamma = 0.3$ for the *"'Pseudochilinus Hexataenia"'* specie.

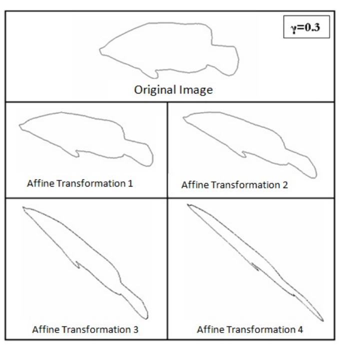

Figure 4: Affine transformation of *Pseudochilinus Hexataenia* **species contour.**

In our system, we use 18 affine images for each fish to describe the arbitrary views that a fish may assume in a video sequence. To describe each fish we use two types of affine invariant features: 1) texture features and 2) boundary features.

3.2 Texture Features

The first step for describing a fish image consists of texture feature extraction. In detail, we use features derived from the grey level histogram, from the Gabor filters and from the grey level co-occurrence matrices (GLCM). The 8 features extracted from the grey level histogram are: mean, standard deviation, third moment and fourth moment (that respectively describe slope and flatness of the histogram), Contrast C, Correlation Cr, Energy E and Homogeneity

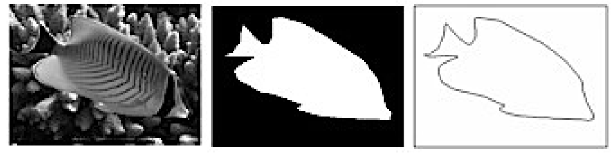

Figure 5: Fish Contour

H [6]. Afterwards, Gabor filters are applied to obtain the $G - Maps$. A two dimensional Gabor function $g(x, y)$ can be described by the following formula:

$$g(x, y; \lambda, \psi, \sigma, \gamma) = e^{-\frac{x^2+\gamma^2 \cdot y^2}{2 \cdot \sigma^2}} \cdot cos(2\pi \frac{x}{\lambda} + \psi) \quad (2)$$

where $\lambda, \psi, \sigma, \gamma$ are respectively the orientation, the scale, the mean and the standard deviation of the considered Gabor filter. Given an image $I(x, y)$, the Gabor transform is obtained by a convolution between the image I and the function g. We use 6 scales and 4 orientations, thus obtaining 24 complex images. Then, we compute the mean and the standard deviation of the magnitude of each of these complex images. Finally, gray level co-occurrence texture features are extracted by using the grey level co-occurrence matrix (GLCM), which describes the frequencies at which two pixels occur in the image. Once the GLCM has been created, we extract the following statistics: Energy, Correlation, Inertia, Entropy, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Difference Average, Difference Variance, Difference Entropy, Information measure of correlation 1, Information measure of correlation 2, Maximal Correlation Coefficient. Therefore for each identified fish we obtain a vector consisting of 70 elements (24x2 from the Gabor Filters, 8 from the gray level histogram, and 14 from the GLCM) that describe the texture features.

3.3 Boundary Features

This step aims at extracting the information about contour features from the detected fish. First of all, contours of fish have been extracted (see fig.5) by means of morphological operations to adjust the contour where it appears interrupted, jagged or thicken.

The extracted boundary can be expressed as a sequence of coordinates $s(k) = [x(k), y(k)]$, where $x(k)$, $y(k)$ are the coordinates of the points of the boundary. Each pair of coordinated can be considered a complex number, i.e. $s(k) = x(k) + j \cdot y(k)$. The discrete Fourier transform (DFT) is:

$$a(u) = \frac{1}{K} \sum_{k=0}^{K-1} s(k) \cdot e^{-j2\pi uk/K} \quad (3)$$

for $u = 0, 1, 2...K$, where K is the number of points belonging to the identified boundary. The complex coefficients $a(u)$ are the Fourier descriptors and provide a means for representing the boundary of a two-dimensional shape. Since it is not feasible to use all the Fourier descriptors in the classification step due to their high number, in order to describe the frequency variability of the shape we use a histogram of their modulus with 30 values. The histogram of Fourier descriptors is invariant to affine transformation as shown in fig.6.

After the Fourier descriptors computation, curvature anal-

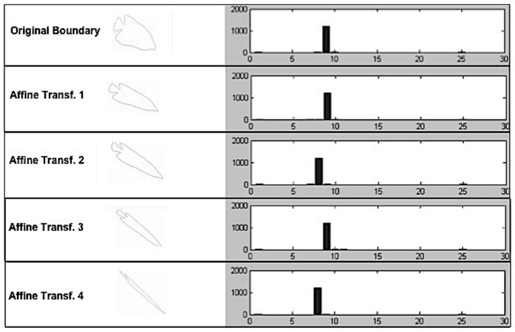

Figure 6: Histogram of Fourier Descriptions when Affine Transformations are applied

ysis is carried out. We compute the CSS (Curvature Scale Space) image, according to [10], by iteratively smoothing the curve until the number of points where the curvature is zero (zero crossing points) is equal to zero. The CSS image represents curvature zero crossings during shape evolution in the plane (u, σ), where u is the normalized arc length between consecutive zero crossing points and σ is the width of Gaussian kernel used for shape smoothing. The curvature is defined as the changing rate of curve slope, according to the formula (the notation is the same of the original approach proposed in [2]):

$$\kappa(u) = \frac{\dot{x}(u)\,\ddot{y}(u) - \ddot{x}(u)\,\dot{y}(u)}{\sqrt{(\dot{x}(u)^2 + \dot{y}(u)^2)^3}} \tag{4}$$

where u is the curve formed by the computed boundary. To find the CSS image, we iteratively smooth the extracted boundary. Let $g(u, \sigma)$ be a $1 - D$ Gaussian kernel of width σ, then the components of the evolved curve Λ_σ may be represented by $X(u, \sigma)$ and $Y(u, \sigma)$ according to the properties of convolution:

$$X(u, \sigma) = x(u) * g(u, \sigma)$$

$$Y(u, \sigma) = y(u) * g(u, \sigma)$$

where $(*)$ is the convolution function. The derivatives are:

$$X_u(u, \sigma) = x(u) * g_u(u, \sigma)$$

$$X_{uu}(u, \sigma) = x(u) * g_{uu}(u, \sigma)$$

where $g_u(u, \sigma)$ and $g_{uu}(u, \sigma)$ are, respectively, the first and the second derivative of the gaussian function. The same holds for $Y_u(u, \sigma)$ and $Y_{uu}(u, \sigma)$. The curvature of the evolved digital curve is:

$$\kappa(u, \sigma) = \frac{X_u(u, \sigma)Y_{uu}(u, \sigma) - X_{uu}(u, \sigma)Y_u(u, \sigma)}{(X_u(u, \sigma)^2 + Y_u(u, \sigma)^2)^{3/2}} \tag{5}$$

As σ increases, the shape of Λ_σ changes. Thus, we have to calculate several times the curvature zero crossing points of Λ_σ during the curve evolution, until when the number of such points will be zero. For each iteration (value of σ) the arc length between consecutive zero crossing points is plotted in the CSS image. Fig. 7 shows the CSS images for the contour of a fish when the affine transformation is applied. As feature vector we extract the first 20 local maxima (that are curve length values) of the CSS image, normalized by the fish contour length.

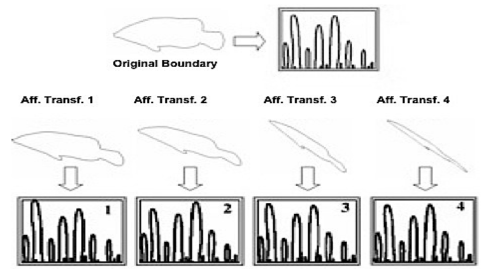

Figure 7: CSS images for the contour of *Pseudochilinus Hextataenia* species for 4 different affine transformations.

In conclusion the total number of extracted shape features is 50.

4. FISH SPECIES CLASSIFICATION

The texture and shape features extraction produce a vector consisting of many values, therefore principal component analysis to reduce the number of features is required. This process is used to distinguish planes that separate fish classes. Starting from 120 components, PCA allowed us to select only the 24 features creating new data for the further classification step.

4.1 Data set: Ecogrid Images

The images used for the evaluation were acquired from an ecological source in Taiwan and are made accessible to marine biologists via the Ecogrid project, coordinated by the National Center for High Performing Computing (NCHC), Taiwan. The project is a joint effort between NCHC and several local and international research institutes which provides a Grid-based infrastructure for ecological research.
These images are acquired using 10 cameras, located at 3 sites, live streaming of bioactivities goes from the reef communities to marine scientists' desktop and is available at the link *http://eco055.nchc.org.tw/lsi/*. Our system was applied on 10 different fish types: *Bodianus mesothorax, Chaetodon trifascialis, Chromis viridis, Dascyllus albisella, Dascyllus aruanus, Dascyllus reticulatus, Gomphosus varius, Hemigymnus fasciatus, Plectorhinchus lessonii* and *Pseudocheilinus hexataenia*. For each fish we used 14 images acquired from the live streaming and 18 images obtained by affine transformation. In total we used 32 images for each species. Therefore, the total size of our database was of 320 images.

4.2 Classification with Discriminant Analysis

The Discriminant Analysis consists of a sequence of methodologies that, given a k-dimensional set X partitioned into p subsets $X_1, ..., X_p$, assigns a generic observation x to one of the p subset. To calculate the expected risk when a large amount of data are considered, the K-fold cross-validation method was used. Using $N = 320$ labels for the data that represents the observations, we divide them into K subsets:

- K-1 subsets are used as training sets (learning);
- the remaining subset is used as the test set.

This operation is repeated leaving out each k subset, with $k = 1, 2, , K$ and the final risk is obtained from the combination of the k intermediate estimates. To evaluate the effectiveness of the classifier $K = 5$ was chosen, while Correct Rate (CR) and Error Rate (ER) was chosen to estimate the accuracy. Table 1 shows the obtained results in terms of CR and ER, while K varies from 1 to 5.

K-Iteration	CR (%)	ER (%)
1	92.34	7.66
2	98.01	1.99
3	99.05	0.95
4	91.21	8.79
5	96.30	3.70

Table 1: Obtained Results in terms of CR and ER

After that, we tested our system for each type of fish by using 10 test images (not included in the training set) from each species, obtaining the results shown in table 2 with an average performance of 92%.

Fish Name	Correct Rate (%)
Bodianus mesothorax	100
Chaetodon trifascialis	90
Chromis viridis	80
Dascyllus albisella	100
Dascyllus aruanus	100
Dascyllus reticulatus	90
Gomphosus varius	90
Hemigymnus fasciatus	90
Plectorhinchus lessonii	90
Pseudocheilinus hexataenia	90

Table 2: Test Results for each type of fish carried out on a test set that contains 10 images for each fish

For *Chromis viridis* we obtained a lower accuracy. This was mainly due to the smooth texture that characterizes the species, in fact, under some light conditions fish appears just white.

5. FISH TRAJECTORY ANALYSIS SYSTEM

The classification system allows us to associate fish species with the extracted trajectories; for example, fig. 8 shows the trajectories of fig. 3 associated with two species: *Chromis viridis* and *Dascyllus aruanus*.

The first step of the the fish trajectory analysis system is the trajectory preprocessing, which aims at producing a suitable trajectory representations for clustering. Indeed, the difficulty in this case is represented by the different lengths of the trajectories. In order to ensure a comparison between different trajectories we subsample the input vectors to reduce the number of points according to the Douglass-Peucker algorithm[8]. The next step is the trajectory clustering which is performed by means of I-kMeans algorithm

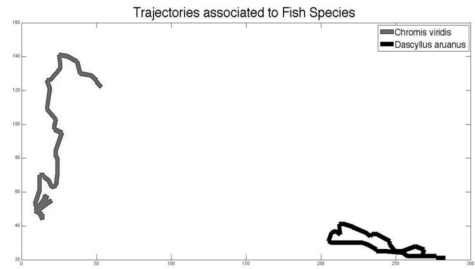

Figure 8: Trajectories of two fish (*Chromis viridis* and *Dascyllus aruanus*) in a video sequence of 5 sec.

[9] since it works in real-time and the choice of number of clusters is unsupervised. Once the scene model has been constructed, fish behaviors can be analyzed. In detail, we consider interesting events, cases when a clustering iteration produce clusters with few elements with respect the total number of trajectories for a fish species in a whole video.
For instance, let us consider the trajectories, shown in fig. 9-b, of the eight detected *Dascyllus reticulatus* (9-a) in the above video sequence. By applying the unsupervised clustering we find three clusters 9-c: the first with 6 elements, and the other two with one trajectory each. The two clusters with one element each represent, of course, a potential events of interest to be investigated. Therefore, we were able to detect that for a specific video sequence two *Dascyllus reticulatus* fish had a different behavior with respect to the others.

6. CONCLUDING REMARKS

In this paper we propose a classification system for underwater video analysis. The proposed classification system defines a new method to recognize a large variety of underwater species by using a combination of affine invariant texture and shape features, providing accurate results (more than 90% of 10 different fish species). Future work for fish classification will regard: 1) the use of color features for a better fish description, 2) the improvement of the performance in classifying fish with smooth texture by integrating fish behavior. Moreover, the classification system aims at supporting fish behavior understanding that work on fish trajectories providing as output the events where fish show an unusual behavior. Future work will aim at establishing a probabilistic model of typical behavior by analyzing trajectories and also at modeling the scene and at correlating the computed trajectories with this model. This, of course, will represent a challenging task in object behavior understanding since it will work on real unconstrained environments where objects can move along three directions. Furthermore, since the behavior of a fish is often correlated to the behavior of another fish (e.g. predator fish chasing small fish) we are working to integrate in our framework a level which correlates temporal and spatial events among different fish species.

7. REFERENCES

[1] S. Abbasi and F. Mokhtarian. Robustness of shape similarity retrieval under affine transformation. In *Proceedings of Challenge of Image Retrieval, Newcastle upon Tyne*, 1999.

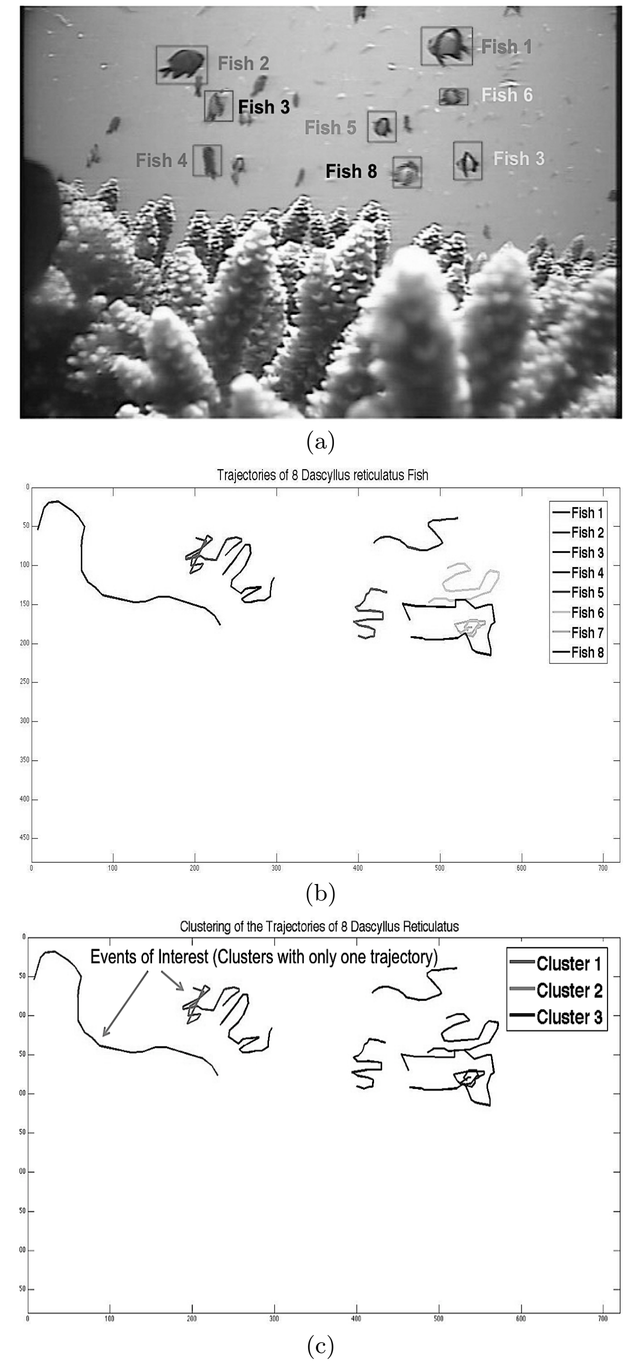

Figure 9: a) Eight *Dascyllus reticulatus* Fish Detected by the detection system , b) Trajectories of the detected fish extracted by the tracking system and c) Trajectories Clustering for Events Detection

[2] S. Abbasi, F. Mokhtarian, and J. Kittler. Curvature scale space image in shape similarity retrieval. *MultSys*, 7(6):467–476, 1999.

[3] B. Benson, J. Cho, D. Goshorn, and R. Kastne. Field programmable gate array based fish detection using Haar classifiers. In *American Academy of Underwater Science*, 2009.

[4] P. Brehmer, T. Do Chi, and D. Mouillot. Amphidromous fish school migration revealed by combining fixed sonar monitoring (horizontal beaming) with fishing data. *Journal of Experimental Marine Biology and Ecology*, 334:139–150, 2006.

[5] D. Edgington, D. Cline, D. Davis, I. Kerkez, and J. Mariette. Detecting, tracking and classifying animals in underwater video. In *OCEANS 2006*, pages 1–5, 18-21 2006.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[7] R. Larsen, H. Olafsdottir, and B. Ersboll. Shape and texture based classification of fish species. In *SCIA09*, pages 745–749, 2009.

[8] M.-Y. Liao, D.-Y. Chen, C.-W. Sua, and H.-R. Tyan. Real-time event detection and its application to surveillance systems. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pages 4 pp. –512, 0-0 2006.

[9] J. Lin, M. Vlachos, E. Keogh, and D. Gunopulos. Iterative incremental clustering of time series. In *IN EDBT*, pages 106–122, 2004.

[10] F. Mokhtarian and S. Abbasi. Shape similarity retrieval under affine transforms. *Pattern Recognition*, 35(1):31 – 41, 2002.

[11] E. F. Morais, M. F. M. Campos, F. L. C. Padua, and R. L. Carceroni. Particle filter-based predictive tracking for robust fish counting. *Computer Graphics and Image Processing, Brazilian Symposium on*, 0:367–374, 2005.

[12] G. Nadarajan. Planning for automatic video processing using ontology-based workflow. In *ICAPS 07*, 2007.

[13] G. Nadarajan, Y.-H. Chen-Burger, and R. B. Fisher. A knowledge-based planner for processing unconstrained underwater videos. In *STRUCK 09*, 2009.

[14] Y. Nagashima and T. Ishimatsu. A morphological approach to fish discrimination. In *MVA98*, 1998.

[15] W. Rouse. Population dynamics of barnacles in the intertidal zone. *Marine Biology Research Experiment*, 2007.

[16] A. Rova, G. Mori, and L. M. Dill. One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video. In *IAPR Conference on Machine Vision Applications*, 2007.

[17] U. Soori and M. Arshad. Underwater crowd flow detection using Lagrangian dynamics. *Indian Journal of Marine Science*, 38:359–364, 2009.

[18] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher. Detecting, tracking and counting fish in low quality unconstrained underwater videos. In *VISAPP (2)*, pages 514–519, 2008.