Automatic Fish Classification in Underwater Video

Clasificación Automática de Peces en Video Submarino

Classification Automatique de Poisson dans la Vidéo Sous-marine

MADHURI GUNDAM[1]*, DIMITRIOS CHARALAMPIDIS[1], GEORGE IOUP[1],
JULIETTE IOUP[1], and CHARLES THOMPSON[2]
[1]*University of New Orleans, 2000 Lakeshore Drive, New Orleans, Louisiana 70148 USA.*
**mgundam@uno.edu. dcharala@uno.edu. geioup@uno.edu. jioup@uno.edu.*
[2]*NOAA Southeast Fisheries Science Center, Stennis Space Center, Mississippi 39529 USA.*

ABSTRACT

Underwater video is currently being used by many scientists within NMFS to observe, identify, and quantify living marine resources. Processing of video sequences is typically a manual process performed by a human analyst. Partial automation of this time consuming and labor intensive analysis process will make data from underwater video more cost effective and available in a more timely fashion. This work introduces a technique for automatic fish classification in underwater video. The technique is based on a series of processing steps. Background processing is used to separate moving objects from the still background. Object tracking is used in order to associate different views of the same object found in consecutive frames. This step is especially important since successfully recognizing and classifying one of the views as a species of interest allows marking all views in the sequence as that particular species. Feature extraction using Fourier Descriptors is used to extract characteristic information from the shape of each identified object. Finally, a nearest neighbor classifier is used to classify identified objects as one of the species of interest. Results demonstrate the performance of the proposed technique in terms of correct classification and false alarms for three species, namely trigger fish, red grouper, and yellow tail snapper.

KEY WORDS: Computer classification, trigger fish, grouper, snapper, underwater video, background subtraction, tracking

INTRODUCTION

Marine biologists use underwater video sequences to study behavior and migration patterns of fish, to identify new species of fish, and to perform population analysis. Experts analyze videos manually, which is a tedious process consuming many hours to analyze one video. Automating this process, at least partially, will reduce the time and labor required to analyze the videos.

Most of the previous research done in this area has been in human controlled environments such as in fish tanks with adequate lights (Castignolles et al. 1994, Semani et al. 2002) and for fish taken out of water (White et al. 2006, Storbeck et al. 2000, Amer et al. 2011). Recently, a system to classify fish in their natural environment has been developed (Spampinato et al. 2010). This system extracts trajectories and associates fish with these tracks to study their behavior. In our system, tracking is used to classify fish in an uncontrolled environment. We assume that fish will be present in more than one frame. Individual fish are tracked to acquire multiple views of fish from consecutive frames. If one view of a fish is not appropriate for classification, it is expected that, as it moves, it may turn revealing a good side view and outline, and thus be classified correctly.

In this paper, we propose a system to classify fish in underwater video sequences. We consider three species, *Epinephelus morio*, *Ocyurus chrysurus* and *Balistes capriscus*, which are found frequently in the Gulf of Mexico. The first step in automating the process is to separate the fish regions from background, track the paths of all fish and finally classify each fish. The rest of the paper is as follows: Section 2 describes the background extraction, subtraction and thresholding of images to separate fish regions from surrounding background. Section 3 provides a brief description of Kalman filtering used for extraction of fish tracks. Section 4 presents Fourier descriptors to represent the shape of fish. Classification using a nearest neighbor classifiers is described in section 5. Section 6 presents classification results for the three species. Lastly, conclusions are presented in Section 7.

BACKGROUND PROCESSING

Background subtraction is used to distinguish the objects of interest (fish in this particular application) from their surroundings. The main steps involved are computation of the background image from multiple frames, subtraction of the background image from the current frame, and thresholding of the background-subtracted frame (Figure 1).

In order to calculate the background image, L consecutive frames, F_l , $l = 1,\ldots, L$, are considered. It was observed that the image contrast does not remain constant throughout the frame sequence. In order to alleviate this problem, the contrast of all frames is adjusted to that of the first frame, F_1. The median, M_{Fl}, of all pixel values in F_l , is a robust measure of the frame's intensity. On the other hand, the average can be affected by extreme pixel values caused due to the entering of new fish in the camera's field of view, or due to noise. Contrast adjustment is achieved by multiplying all pixels in F_l , with M_{Fl}/M_{Fl} .

For simplicity, in what follows, the term *frame* and the notation F_l are used to refer to the contrast-adjusted frames.

For the purpose of reducing the memory resources required for processing, the set of L frames is divided and processed in groups of N frames, such that $MN = L$. For each group, a partial background image, $B_m(x,y)$, m = 1,..., M, is obtained, where (x, y) represent the horizontal and vertical image coordinates. More specifically, each pixel in $B_m(x,y)$ is calculated as the median of corresponding pixels in all N frames:

$$(B_m(x,y) = \text{med}(F_l(x,y)) \ \forall \ (x, y) \quad (1)$$

The final background image, $B(x,y)$, is computed as the median image of the M partial background images:

$$B(x,y) = \text{med}(B(x,y)) \ \forall \ (x, y) \quad (2)$$

Usually $B(x,y)$ is obtained by frame averaging, which causes shadowing due to non-background objects found in some frames. However, if a fish occupies location (x, y) in less than half of the frames being processed, the median at (x, y) still corresponds to a background pixel. In eq. (2), $B(x,y)$ is a good approximation to the median computed using all L frames.

S_m^2 A variance-like measure, $S^2(x,y)$, is calculated for each pixel as follows:

$$(x, y) = \text{med}(((- B(x,y))^2) \ \forall \ (x, y) \text{ Over } l \text{ frames} \quad (3)$$

$$S^2(x,y) = \text{med}(B_m(x, y)) \ \forall \ (x, y) \quad \text{Over } m \text{ frame} \quad (4)$$

It can be observed that if the median is replaced by the average, $S^2(x,y)$, is identical to the sample variance of corresponding pixels located at position (x, y) in the L frames. A user-defined threshold parameter, T, is used to specify which pixels are significantly different from the background.

More specifically, if $F_l(x,y) - B(x,y) > S(x,y)$ T, $F_l(x,y)$ is considered to be associated with a fish. In the thresholded frames, fish and background pixels are marked as "white" and "black", respectively. Region growing is used to group white neighboring pixels. Each such group represents a candidate fish region. Regions smaller than a certain number of pixels are considered to be noise and are eliminated.

Tracking

An object is tracked from the point it enters until the point it exits the camera's field of view. Thus, information about the fish is collected from multiple frames. Tracking is helpful when a single fish view is unsuitable for classification. Having multiple views of the same fish increases the chances of obtaining at least one side-view, which can be successfully recognized. In this paper, the Kalman filter is used for fish tracking. The state vector includes information about the fish during tracking. This information is the fish center, (c_x,c_y), the coordinates of the top left and bottom right points, (bt_x,bt_y) and (bb_x, bb_y), of the rectangular bounding box enclosing the fish region, and the velocity of the center (v_x,v_y). In other words, the state vector of the fish in the *l-th* frame is $S_t = [c_x^{(l)}, c_y^{(l)}, v_x^{(l)}, v_y^{(l)}, bt_x^{(l)}, bt_y^{(l)}, bb_x^{(l)}, bb_y^{(l)}]^T$. The bounding box endpoints are chosen since the size of the bounding box is associated to the size of the fish area. Multiple fish regions are tracked simultaneously, each by a different Kalman filter. A constant velocity model is assumed.

The Kalman filter operations can be divided into two stages: prediction and correction (Li et al., 2010; Li et al., 2009). During the prediction stage, the filter obtains the *a priori estimate* for the current state, $\hat{s}_l^-$ using the previous state estimate, s_{l-1}, as shown in eq. (5). In this work, the vector of observations, z_l, includes the same variables as the state, but in z_l, these are not predicted or estimated, but calculated directly from the image data. The first time a fish region enters the field of view, its state is initialized according to the real observations, z_l, except the velocities which are set equal to zero. The state transition matrix, A, relates the previous state, s_{l-1}, with the present state. Moreover, w_l is the process noise, which is assumed to follow a zero-mean, white gaussian distribution.

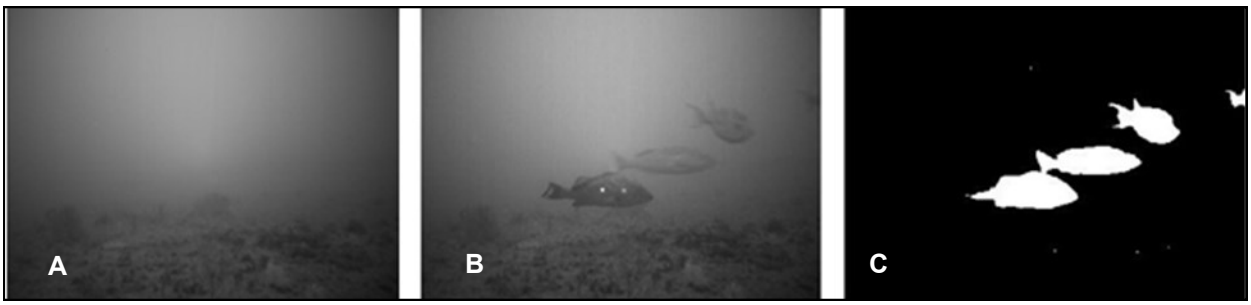

Figure 1. Background Processing Steps: (a) Background image, (b) an original frame, (c) thresholded frame

In order to associate the *i-th* fish region in F_{l-1}, namely $R_i^{(l-1)}$, to one of the regions in F_l, namely $R_j^{(l)}$, $j = 1,\ldots, J$, the Euclidean distances between the state estimate $\hat{s}_l^-$ of R_i $(l\text{-}1)$ and the real observations z_l of all $R_j^{(l)}$ are computed. The $R_j^{(l)}$ associated to the smallest distance, E_{min}, is associated to $R_i^{(l-1)}$ only if E_{min} and the size difference of the two regions are each smaller than a user-defined threshold. Otherwise, it is assumed that either a region splitting or a merging has occurred, and a new filter is assigned to track the new region.

During the correction stage, the filter corrects using the respective z_l to obtain the a posteriori estimate $\hat{s}_l^-$. The estimate is used as $\boldsymbol{s_{l-1}}$ in the next frame. The measurement matrix, H, in eq. (6) associates state predictions with observations, z_l, and K is the Kalman gain.

$$\hat{s}_l^- = As_{l-1} + w_{l-1} \tag{5}$$

$$\hat{s}_l = \hat{s}_l^- + K_l(Z_l - H\hat{s}_l^-) \tag{6}$$

As the fish regions are tracked, they are classified on-the-fly using the classifier presented in section 5. The next section discusses the features used for recognition of *Epinephelus morio* (EM), *Ocyurus chrysurus* (OC) and *Balistes capriscus* (BC).

Feature Extraction using Fourier Descriptors

Fourier descriptors (FD) use shape outline information to represent an object in the frequency domain (Zahn and Roskies, 1972). The FDs of a shape are computed using the coordinates of the shape's boundary points. Consider an P-point outline with (x_p, y_p) as the horizontal and vertical coordinates of the p^{th} point, and $c_p = x_p + iy_p$ as the complex representation of the point coordinates. The FDs are calculated as follows:

$$D(k) = \sum_{0}^{p-1} e^{(-i2\pi pk/P)}, \; k = 0, 1,\ldots, P\text{-}1 \tag{7}$$

The lower- and higher-frequency descriptors represent, respectively, the general shape and the finer details of the shape. Fish regions tend to have relatively smooth edges. Thus, the significantly high-frequency FDs can be ignored. The absolute FD values, $|D(k)|$ are invariant to fish rotation since the FD phases are ignored.

All points on the boundary are used for calculating the FDs. To make the FDs invariant to fish region size, the FDs magnitudes are normalized using the average of $D(1)$ and $D(P\text{-}1)$ (Lin and Chellappa, 1987). The FDs for k = 1, …, 128 and k = P-128,…, P-1 are used in this work since they correspond to lower-frequency FDs. Plots of the FDs corresponding to BC, EM, and OC are shown in Figures (2a), (2b), and (2c). respectively.

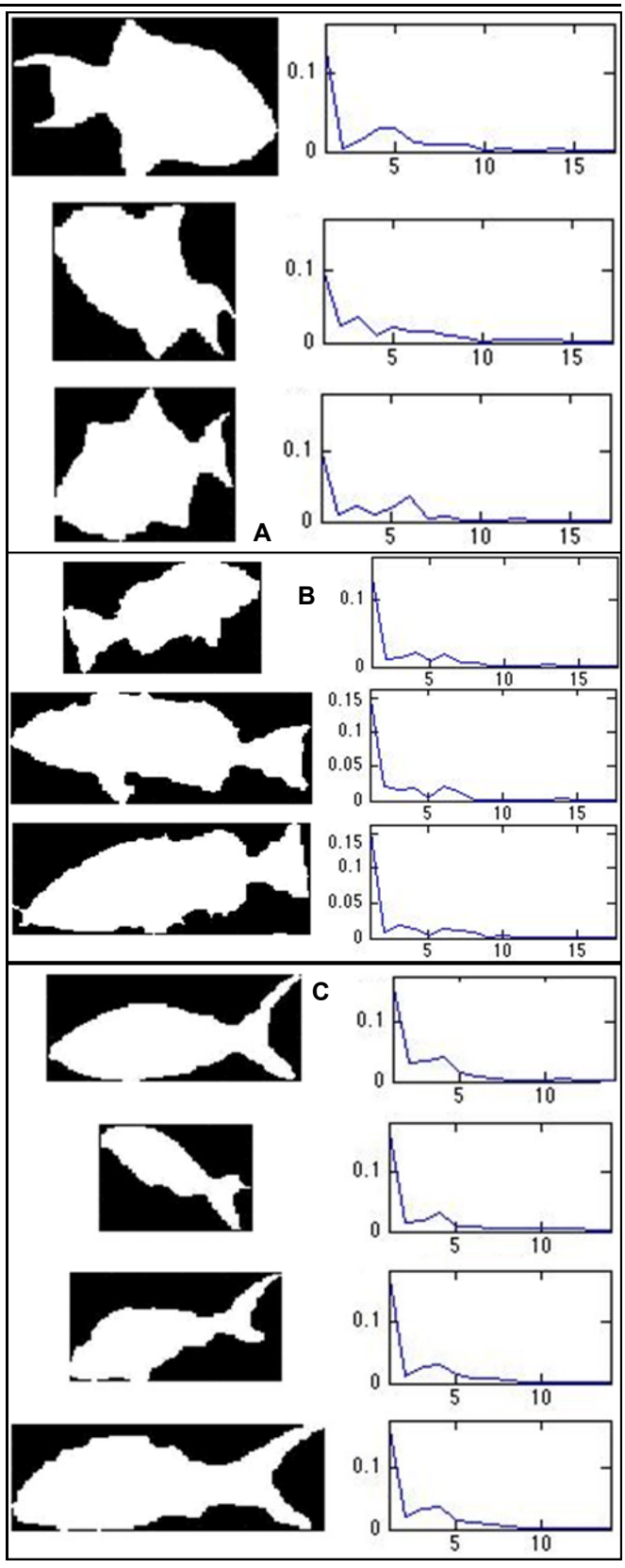

Figure 2. Templates and Fourier Descriptors of (a) *B. capriscus*, (b) *E. morio*, (c) *O. chrysurus.*

Classification

In this work, a Nearest Neighbor classifier (NNC) is used to classify feature vectors consisting of absolute normalized FD values. Based on the NNC algorithm, feature vectors with known classification (exemplars) are used to represent each of C classes. The *q-th* exemplar of the *c-th* class is defined as $D_{c,q}$, where $c = 1,\ldots, C$, and $q = 1,\ldots, Q$. The distance measure between a feature vector with unknown classification and each of the exemplars is used to classify the feature vector, and therefore its associated object, to one of the classes. The FDs of the EM, OC and BC templates shown in figure (2) are used as the exemplars. Therefore, currently, $Q = 3$ for EM and BC, and $Q = 4$ for OC. However, a larger number of exemplars per class may be necessary for effective classification of a larger number of species.

A weighted Euclidean distance is used in the NNC. The square of the distance is defined as follows:

$$dist^2\ (D_{c,q}\ ,\ D_t)\ =\ \frac{((D_{c,q},D_t)^T\ (D_{c,q},D_t))}{w_c} \qquad (8)$$

The weight, w_c, is the average square Euclidean distance between D_c and a few FD vectors extracted from subjectively "good" views of fish that are known to belong to class c.

Essentially, w_c is the sample variance associated with the multivariate feature distribution of class c, assuming that the covariance matrix of such distribution is diagonal with all diagonal elements equal to w_c. A fish region with corresponding vector D_t is temporarily assigned to the class c of minimum ($dist^2\ D_{c,q}\ ,\ D_t$). When NNC does not classify the region as EM, OC, or BC, the fish is labeled as 'Not Fish', which implies that the fish does not correspond to a fish or at least not to a fish of interest. Fish regions at the frame borders are not classified, since they are usually incomplete, and are marked as 'Not Fish' (Figures 3 - 6).

RESULTS

This section presents performance evaluation results for more than 3000 frames consisting of EM, OC, BC as well as other species of fish. All fish regions within a frame are automatically identified and segmented out of the overall frame.

In all experiments, the Kalman filter covariance matrices for the process noise and measurement noise were assumed to be diagonal, with diagonal values equal to 0.05 and 0.9, respectively. The threshold used for the square distance of eq. (8) was empirically chosen to be 0.36. As mentioned earlier, if the distance in eq. (8) is greater than the chosen threshold, then the region is classified as a fish of no interest. This threshold may be increased to be able to classify more shapes as one of the three species. In this case, all regions in the sequence are classified as this species. Figures 3 - 5 depict different example sequences. For each sequence, the top leftmost image shows the fish as detected for the first time. The black and white image to its right is the corresponding thresholded image obtained by the background subtraction process. Moving from left to right, and then from top to bottom, pairs of images show the same fish and its corresponding thresholded image as seen in consecutive frames. The identification of the bounding box enclosing the fish, as well as the association of the fish region from one frame to the next are performed automatically by the algorithm.

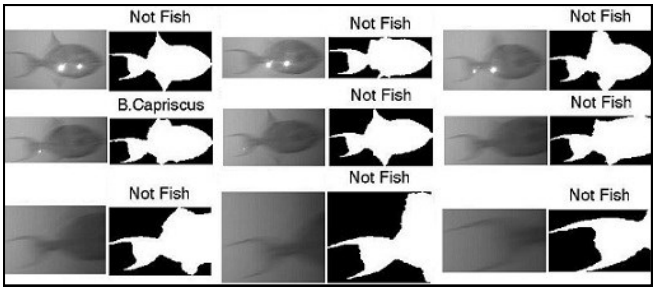

Figure 3. Classification results for *B. capriscus*.

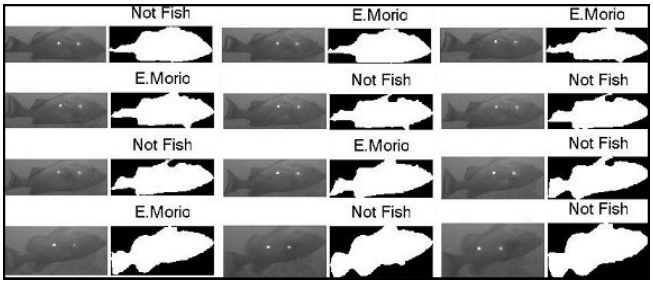

Figure 4. Classification results for *E. morio*.

Figure 5. Classification results for *O. chrysurus*.

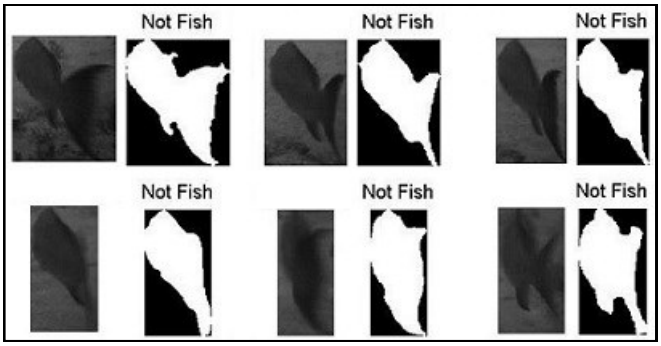

Figure 6. Classification results for other species.

It can be observed from Table 1 that the classification of individual fish regions based on the NNC produced a small percentage of false alarms. Although in the case of EM, OC, and BC, only a small percentage of individual fish regions were recognized for each sequence this was sufficient to classify the whole sequence as a particular species. This is illustrated in Tables 2-4. For instance, Table 2 presents classification results for ten different BC sequences (each row corresponds to a different sequence).

In 5 out of 10 sequences, the fish regions were recognized at least once, which implies that all fish regions in these 5 sequences were recognized. In total, BC was recognized in 122 out of 190 frames. The last column describes the main reason why fish regions in each sequence were not recognized. Similarly, Table 3 presents classification results for 13 EM sequences. The EM regions were recognized at least once in 6 out of 13 sequences, and in total, EM was recognized in 217 out of 513 frames. Table 4 presents classification results for 7 OC sequences. It can observed that OC has 100% recognition, i.e., all 7 OC sequences, were classified correctly.

It should be mentioned that in several cases when fish regions are not recognized, the reason is that there is no appropriate side view of the fish in the whole sequence. For example, in the first two sequences in Table 3, the fish faces the camera for 21 and 54 frames respectively. An example is shown in Figure 8. Therefore, it is expected that even an expert may not be able to identify the species from these sequences. However, these sequences are tracks of a single fish which the tracking algorithm failed to connect. In fact, in Table 3, rows 1 to 8 represent broken tracks of same fish. If the tracking algorithm is further improved, all these sequences would be combined and thus could be classified as EM. This would also improve the classification rate of our system. The fish shapes, shown in Figure 6, are not suitable for classification, since the fish starts turning away from the camera. Figure 7 presents false alarms. It can be noticed that shape of shark somewhat resembles OC and thus is classified as an OC.

CONCLUSIONS

This paper proposes a system to automatically classify fish in a non-controlled environment using underwater video cameras and image processing algorithms. Future work includes further improving the tracking algorithm in order to merge broken tracks of same fish sequence. Another technique that will be investigated is multiple thresholding in order to reduce the effect of using a fixed threshold for background subtraction.

It has been observed that most often, misclassifications occur due to lack of appropriate side views of the fish within the whole sequence, due to poor camera views, and due to fish region merging.

Table 1. Performance evaluation.

Species	% Correctly classified	% Miss-classified
Balistes capriscus	64.21%	0.005%
Epinephelus morio	42.3%	0%
Ocyurus chrysurus	65.26%	0.05%
Other Species	45%	55%

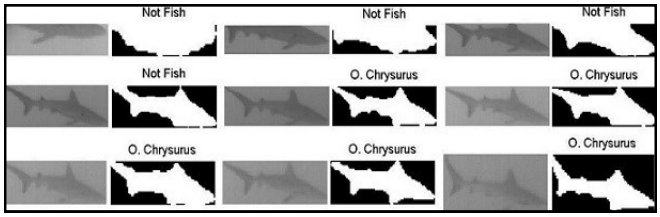

Figure 7. False alarms.

Table 2. Classification results for *B. capriscus*.

Number of frames	Classified as *B. capriscus* by NNC (# of fish classified as BC)	Miss-classified as other species by NNC (# of fish classified as EM or OC)	Explanation
31	Yes (3)	No (0)	-
17	Yes (1)	Yes (1 classified as OC)	The distance of this pattern is closer to BC than OC. Hence it will be classified as BC.
32	Yes (5)	No (0)	-
28	Yes (1)	No (0)	-
22	No (0)	No (0)	The resolution for this fish is not good enough for it to be detected from background.
14	Yes (2)	No (0)	-
7	No (0)	No (0)	This fish region stays merged with another fish region for entire track. Hence it cannot be classified.
11	No (0)	No (0)	The fish is always at the corners of frame. Hence it cannot be classified.
17	No (0)	No (0)	Fish region merged with another region and view of fish is bad.
11	No (0)	No (0)	This fish region stays merged with another fish region for entire track. Hence it cannot be classified.

Table 3. Classification results for *E. morio.*

Number of frames	Classified as E. Morio by NNC (# of fish classified as EM)	Miss-classified as other species by NNC (# of fish classified as BC or OC)	Explanation
21	No (0)	No (0)	Bad view - fish faces camera. (as shown in figure 8)
54	No (0)	No (0)	Bad view - fish faces camera.
3	No (0)	No (0)	Bad view - fish faces camera.
21	No (0)	No (0)	Bad view - fish faces camera.
10	No (0)	No (0)	Bad view - fish faces camera.
50	Yes (2)	No (0)	Bad view - fish faces camera.
70	No (0)	No (0)	Bad view - fish faces camera.
39	Yes (5)	No (0)	-
13	Yes (8)	No (0)	-
47	Yes (3)	Yes (1)	-
57	No (0)	No (0)	Bad view - fish faces camera.
61	Yes (4)	No (0)	-
7	Yes (1)	No (0)	-

Table 4.. Classification Results for *O. chrysurus*

Number of frames	Classified as O. Chrysurus by NNC (# of fish classified as OC)	Miss-classified as other species by NNC (# of fish classified as BC or EM)	Explanation
9	Yes (1)	No (0)	-
4	Yes (1)	No (0)	-
5	Yes (3)	No (0)	-
6	Yes (2)	No (0)	-
2	Yes (2)	No (0)	-
5	Yes (1)	No (0)	-
3	Yes (1)	No (0)	-

LITERATURE CITED

Storbeck, F. and B. Daan. 2000. Fish species recognition using computer vision and a neural network. *Fisheries Research* **51**(1):11-15.

Amer, M., E. Bilgazyev, S. Todorvic, S. Shah, I. Kakadiaris, and L. Cianelli. 2011. Fine-grained Categorization of Fish Motion Patterns in Underwater Videos. 13th International Conference on Computer Vision.

Castignolles, N., M. Catteon, and M. Larinier. 1994. Identification and couting of live fish by image analysis. SPIE. Vol 2182, Image and Video Processing II.

Semani, D., C. Saint-Jean, and C. Frelicot. 2002. Alive Fish Species Characterization from Video Sequences. SPPR & SPR, LNCS 2396, pp 689-698.

Spampinato, C., D. Giordano, R.D. Salvo, Y. Chen-Burger, R.B. Fisher, and G. Nadarajan. 2010. Automatic fish classification for underwater species behavior understanding. Pages 45-50 in: *Proceedings of the 1st ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams.*

Li, X., K. Wang, W. Wang, and Y. Li. 2010. A Multiple Object Tracking Method Using Kalman Filter. International Conference on Informations and Automation.

Li, J., C. Shao, W. Xu, and H. Yue. 2009. Real Time Tracking of Moving Pedestrians. International Conference on Measuring Technology and Mechatronics Automation.

Zahn, C. and R. Roskies. 1972. Fourier Descriptors for Plane Closed Curves. IEEE Transactions on Computers, C-21: 269-281.

Lin, C. C., Chellappa, R. 1987. Classification of Partial 2-D Shapes Using Fourier Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **9**(5).

White, D.J., C. Svellingen, and N.J.C. Strachan. 2006. Automated measurement of species and length of fish by computer vision. *Fisheries Research* **80**(2-3):203-210.

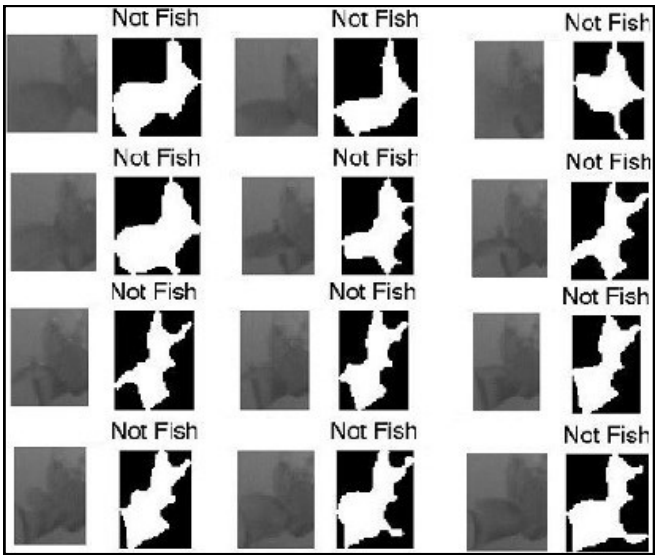

Figure 8. Classification results for bad view of EM.