Alive Fishes Species Characterization from Video Sequences

Dahbia Semani, Christophe Saint-Jean, Carl Frélicot,
Thierry Bouwmans, and Pierre Courtellemont

L3I - UPRES EA 2118
Avenue de Marillac, 17042 La Rochelle Cedex 1, France
{dsemani,csaintje,cfrelico}@univ-lr.fr

Abstract. This article presents a method suitable for the characterization of fishes evolving in a basin. It is based on the analysis of video sequences obtained from a fixed camera. One of the main difficulties of analyzing natural scenes acquired from an aquatic environment is the variability of illumination. This disturbs every phase of the whole process. We propose to make each task more robust. In particular, we propose to use a clustering method allowing to provide species parameters estimates that are less sensitive to outliers.

1 Introduction

Segmentation of natural scenes from an aquatic environment is a very difficult issue due to the variability of illumination [17]. Ambient lighting is often insufficient as ocean water absorbs light. In addition, the appearance of non-rigid and deformable objects detected and tracked in a sequence is highly variable and therefore makes identification of these objects very complex [13]. Furthermore, recognition of these objects represent a very challenging problem in computer vision. We aim at developing a method suitable to the characterization of classes of deformable objects in an aquatic environment in order to make their online and real-time recognition easier to a vision-based system. In our application, the objects are fishes of different species evolving in a basin of the *Aquarium of La Rochelle* (France). The method we propose is composed of the following tasks:

1. *scenes acquisition*: a basin of the aquarium is filmed by a fixed CDD camera to obtain a sequence in low resolution (images of size 384 x 288);
2. *region segmentation*: color images are segmented to provide the main regions of the each scene;
3. *feature extraction and selection*: different features (e.g. color, moments, texture) are computed on each region, then selected to form pattern vectors;
4. *species characterization*: pattern vectors are clustered using a robust mixture decomposition algorithm.

T. Caelli et al. (Eds.): SSPR&SPR 2002, LNCS 2396, pp. 689–698, 2002.

2 Segmentation

Image segmentation is a key step in an object recognition or scene understanding system. The main goal of this phase is to extract regions of interest corresponding to objects in the scene [9]. Obviously, this task is more difficult for *moving objects* as fishes or parts of fishes.

Under the assumption of almost constant illumination and fixed camera, the motion detection is directly connected to temporal changes in the intensity function of each pixel (x, y). Then, background substraction is usually applied to segment the moving objects from the remaining part of the scene [10][3]. By assuming that the scene background does not change over successive images, the temporal changes can be easily captured by subtracting the background frame $I_{back}(x, y)$ to the current image $I(x, y, t)$ at time t. The obtained image is denoted $I_{sub}(x, y, t)$. However, such detection of temporal changes are not robust to illumination changes and electronic noise of the camera. A solution consists in updating dynamically the background image by $I_{back}(x, y, t) = \sum_{s=1}^{t} I(x, y, s)/t$. Since obtaining a suitable background requires numerous images, $I_{back}(x, y, t = 1)$ is initialized off-line (from another available sequence). Then, thresholding the difference image provides the so-called binary difference picture:

$$I_{bin}(x, y, t) = \begin{cases} 1 & \text{if } |I_{sub}(x, y, t)| > \tau \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

When color images are available, e.g. in the three dimensional color space RGB (Red, Green and Blue), one can proceed for each color plane. Three corresponding binary difference pictures $I^R_{bin}(x, y, t)$, $I^G_{bin}(x, y, t)$ and $I^B_{bin}(x, y, t)$ are combined to compute the segmented image:

$$I_{seg}(x, y, t) = \begin{cases} 1 & \text{if } (I^R_{bin}(x, y, t) = 1 \text{ or } I^G_{bin}(x, y, t) = 1) \text{ or } I^B_{bin}(x, y, t) = 1) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

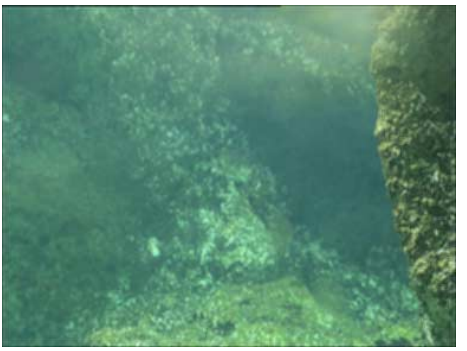
Thresholds are fixed empirically according to the sequence properties. Figure 1 shows: (a) an individual frame in the sequence, (b) the reconstructed background and (c) the resulting segmented image with $\tau_R = 40, \tau_G = 30$ and $\tau_B = 35$. Note that changes in illumination due to the movement of water induce false alarms as one can see at the top right part of (c).

3 Feature Extraction and Selection

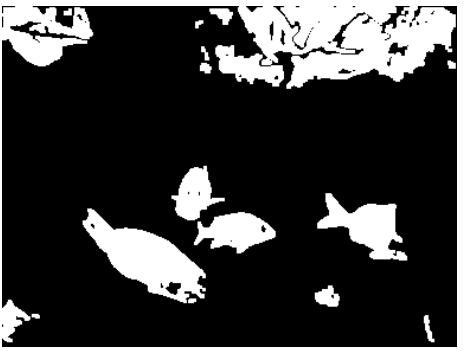
Regions issued from the segmentation process can be used as objects for the identification task. 38 features of different types, e.g. in [18], are extracted from each object:

– *Geometric* features directly relate to the objects' shape, e.g. area, perimeter, roundness ratio, elongation, orientation. Note that the wide variety of possible orientations of fishes to the camera focal axis makes geometric features inappropriate. A fish which is parallel to the image plane will exhibit its main shape while another one being orthogonal will not.

(a) An original frame $I(x, y, t)$



(b) Reconstructed $I_{back}(x, y, t)$



(c) Segmentation result $I_{seg}(x, y, t)$

Fig. 1. From an input image to segmented regions

- *Photometric* parameters are descriptors of the gray level distribution or the different color ones, e.g. maximum, minimum, mean and variance.
- *Texture* features are computed from the co-occurrence matrix, e.g. contrast, entropy, correlation.
- *Moments of Hu* which are known to be invariant under translation, scaling and rotation. Only the first four ones showed significant values.
- *Motion features* are computed from two consecutive frames within a sequence. Correspondence between regions from frame t and $t + 1$ are established with respect to geometric and photometric features. A classical *hypothesize-and-verify* scheme [2] is used to solve this correspondence problem which is similar to the correspondence problem in stereo [11] except that a geometric constraint (a disparity-window centred around the each region's centroid) replaces the epipolar one. The extracted features are the centroid displacement and the angle of this displacement. Note that some regions do not match because of occlusion, disappearance and appearance of objects.

Feature reduction is motivated by making the characterization process easier and speeding the recognition step up to achieve a real-time processing. In order to eliminate features which are either not useful or redundant, we have selected the most pertinent features in a two-stage process:

1. **Group-based clustering:** To make sure that every features group is represented in the reduced feature space, a hierarchical clustering algorithm is applied to each group with respect to the minimization of an aggregation measure, e.g. the increase of intra-cluster dispersion for Ward's method [1]. Cutting the hierarchy to a significant value leads to a partition of the features in clusters. Among the features within a cluster, the most discriminatory powerful one is selected and the others are discarded. We recall that the discriminatory power of a feature is its usefulness in determining to which class an object belongs.

2. **Global clustering:** In order to check whether some features from different groups are similar or not, the same clustering method is globally applied to the remaining features.

4 Species Characterization

From a statistical point of view, each extracted region being described by p features can be considered as a realization x of a p-dimensional random vector X [8]. We have then to estimate the Probability Density Function (pdf) $f(x)$ from a set of realizations $\chi = \{x_1, \ldots, x_N\}$, i.e. featured regions. In mixture model approach, $f(x)$ is decomposed as a mixture of C components:

$$f(x) = \sum_{k=1}^{C} \pi_k f(x; \theta_k) \tag{3}$$

where $f(x; \theta_k)$ denotes the conditional pdf of the k^{th} component and pairs (π_k, θ_k) are the unknown parameters associated with the parametric model of the pdf [12]. A priori probabilities π_k sum up to one. If a normal model is assumed, $\theta_k = (\mu_k, \Sigma_k)^T$ reduces to the mean μ_k and the covariance matrix Σ_k. Under the assumption of independent features of X, estimates of the model parameters $\Theta = (\pi_1, \ldots, \pi_C, \theta_1^T, \ldots, \theta_C^T)^T$ can be chosen such as the likelihood $\mathcal{L}(\Theta)$

$$\mathcal{L}(\Theta) = P(\chi|\Theta) = \prod_{i=1}^{N} \sum_{k=1}^{C} \pi_k f(x_i; \theta_k) \tag{4}$$

is maximized.

To solve this estimation problem, the EM (Expectation-Maximization) algorithm [5] has been widely used in the field of statistical pattern recognition because of its convergence. However, it is sensitive to outliers as pointed out

in [15]. This is a major drawback in the context of our application because incorrectly segmented regions can disturb the estimation process. Several strategies to robust clustering are available, including:

1. contamination models of data, e.g. fitting Student distributions [14],
2. influence functions of robust statistics, e.g. using an M-estimator [6],
3. adding a class dedicated to noise, e.g. Fuzzy Noise Clustering (FNC) [4].

We propose to use a robust clustering method (based on EM algorithm) that is a combination of the first two types [15]. Each component is modelled as a mixture of two sub-components:

$$f(x;\theta_k) = \underbrace{(1-\gamma_k)\mathcal{N}(x;\mu_k,\Sigma_k)}_{(A)} + \underbrace{\gamma_k\mathcal{N}(x;\mu_k,\alpha_k\Sigma_k)}_{(B)} \tag{5}$$

where $\mathcal{N}$ stands for the gaussian multivariate pdf.
First term (A) intends to track cluster kernel points while second term (B) allows to take into account surrounding outliers via multiplicative coefficients α_k. These γ_k and α_k control respectively the combination of the two sub-components and the spread of the second one by modifying its variance. Parameters of both sub-components are estimated through different estimators so that the conditional pdf is estimated by:

$$\hat{f}(x;\theta_k) = (1-\gamma_k)\mathcal{N}(x;\tilde{\mu}_k,\tilde{\Sigma}_k) + \gamma_k\mathcal{N}(x;\hat{\mu}_k,\alpha_k\hat{\Sigma}_k) \tag{6}$$

where $\tilde{\mu}_k, \tilde{\Sigma}_k$ are robust estimates whereas $\hat{\mu}_k, \hat{\Sigma}_k$ are standard ones. Among the possible M-estimators to be used, e.g. Cauchy, Tuckey, Huber, we have chosen the Huber M-estimator [7] because it performs well in many situations [19]. It is parametrized by a constant value h that controls the size of the filtering area. Such an estimator is an influence function $\psi(y,h)$, e.g. the Huber one:

$$\psi_{Huber}(y,h) = \begin{cases} y \text{ if } |y| \leq h \\ h\ sgn(y) \text{ otherwise} \end{cases} \tag{7}$$

This function allows to associate a weight $w(y,h) = \frac{\psi(y,h)}{y}$ as a decreasing function of y, e.g. the Huber one:

$$w_{huber}(y,h) = \begin{cases} 1 \text{ if } |y| \leq h \\ \frac{h}{|y|} \text{ otherwise} \end{cases} \tag{8}$$

We apply it to the distances between each point x_i and the cluster prototypes in order to compute a weight w_i associated with each x_i. According to the equation (8), all w_i belong to $[0,1]$ and outlying points are given a zero weight (see Figure 2).
Algorithm 1 replaces the parameters updating in the M-step of the EM algorithm. The more iterations, the less points are taken into account in the estimation process, so that one needs to use a stop criterion in order to ensure

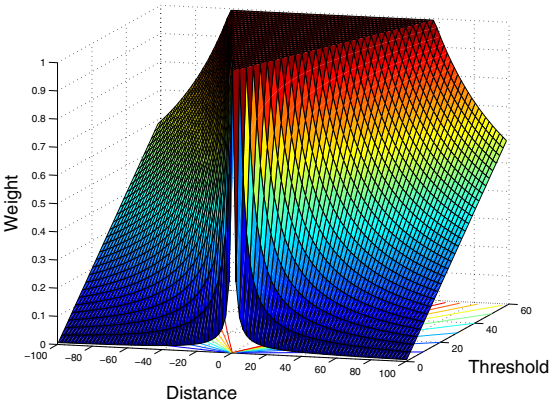


Fig. 2. Huber M-estimator weight as a function of distance y and threshold h

H 1: Iterative robust estimation of means and covariance matrices

Input: $\chi = \{x_1, \ldots, x_N\}$, $\hat{z}_{ik}$ current estimates of $P(C_k|x_i)$ from the E-Step, h the M-estimator threshold

$$\tilde{\mu}_k = \hat{\mu}_k = \frac{\sum_{i=1}^{N} \hat{z}_{ik}\ x_i}{\sum_{i=1}^{N} \hat{z}_{ik}}$$

$$\tilde{\Sigma}_k = \hat{\Sigma}_k = \frac{\sum_{i=1}^{N} \hat{z}_{ik}\ (x_i - \tilde{\mu}_k)(x_i - \tilde{\mu}_k)^T}{\sum_{i=1}^{N} \hat{z}_{ik}}$$

repeat

  for *i = 1 to N* **do**

    $d_i = (x_i - \tilde{\mu}_k)^T \tilde{\Sigma}_k (x_i - \tilde{\mu}_k)$ (Mahalanobis distance)

    $w_i = w_{huber}(d_i, h)$ (Huber M-estimator weight function - see Fig. 2)

$$\tilde{\mu}_k = \frac{\sum_{i=1}^{N} w_i\ \hat{z}_{ik}\ x_i}{\sum_{i=1}^{N} w_i\ \hat{z}_{ik}}$$

$$\tilde{\Sigma}_k = \frac{\sum_{i=1}^{N} w_i\ \hat{z}_{ik}\ (x_i - \tilde{\mu}_k)(x_i - \tilde{\mu}_k)^T}{\sum_{i=1}^{N} w_i\ \hat{z}_{ik}}$$

until *Stop Criterion*;

sufficient statistics. We use a combination of maximum number of iterations and maximum elimination rate (proportion of sample having a quite zero weight). It can be shown that the property of monotonous increase of log-likelihood of the EM algorithm no more holds because the iterative estimation process yields

an approximated realization of the maximum log-likelihood estimator. However, relaxing the maximum likelihood estimation principle allows to obtain more accurate estimates.

5 Experiments and Discussion

A sequence of 550 images was acquired in the *Aquarium of La Rochelle*, the filmed basin comprising 12 species. After segmentation and false alarms discarding, 5009 regions were obtained and labelled according to the different species. The first feature selection step (group-based clustering) allowed to reduce the 38 original attributes to 22 ones while the second step (global clustering) allowed to keep only 18 of them (see Table 1 for details), representing a compression rate greater than 52%.

Table 1. Summary of features selection

Number of features	Before selection	Group-based clustering	Global clustering
Geometric	10	4	4
Photometric	14	7	5
Texture	7	5	4
Moments of Hu	4	3	2
Motion	3	3	3
Total	38	22	18
Compression rate (%)		42.11%	52.63%

At least two features of each group are present in the final set of 18 selected features:

- *Geometric feautures:* width, elongation, roundness ratio and orientation.
- *Photometric feautures:* gray-level mean, minimum and variance ; blue average and minimum of the color.
- *Texture features:* entropy, contrast, homogeneity, and uniformity.
- *Moments of Hu:* second and third moments of Hu.
- *Motion features:* vector and angle of displacement.

During the labelling, we have noticed that different species were indeed subspecies members of which look very similar, e.g. subspecies *Acanthurs bahianus* and *Acanthurs chirurgus* shown in Figure 3. We decided to merge such subspecies decreasing the number of classes to 8. This choice was validated by the BIC (*Bayesian Information Criterion*) using unconstrained normal classes [16]. As labels were available and under the assumption of gaussian classes, class parameters $\theta^D = [\mu_1^D, \Sigma_1^D, \ldots, \mu_c^D, \Sigma_c^D]$ were computed directly from training 5009 samples.

Our goal was to provide as accurate as possible class parameters estimates with an unsupervised technique in order to characterize the fish species. We applied our clustering algorithm several times under random initializations. Parameters γ_k, α_k and h were fixed empirically and identical for each class ($\gamma_k = \gamma$ and $\alpha_k = \alpha$). According to the semantics of theoretical model of classes, only robust estimates $\tilde{\theta}_k = [\tilde{\mu}_k, \tilde{\Sigma}_k]$ were considered ($k = 1, c$). In order to assess the species characterization, a distance between θ^D and the final estimated parameters provided by the algorithm was calculated. Because of possible labels switching in class numbering, optimal permutation σ^* was obtained by computing the minimum over all possible permutations σ:

$$A(\theta^D, \tilde{\theta}, \sigma^*) = \min_{\sigma} \left(\sum_{k=1}^{C} dist_{\mathcal{M}}(\theta_k^D, \tilde{\theta}_{\sigma(k)}) \right) \tag{9}$$

where $dist_{\mathcal{M}}$ is the Mahalanobis distance between two normal distributions:

$$dist_{\mathcal{M}}(\theta_k, \theta_l) = dist_{\mathcal{M}}(\mu_k, \Sigma_k, \mu_l, \Sigma_l) = (\mu_k - \mu_l)^T (\Sigma_k + \Sigma_l)^{-1} (\mu_k - \mu_l) \tag{10}$$

A value of 15.07 was obtained for $A(\theta^D, \tilde{\theta}, \sigma^*)$. Using the EM algorithm, standard estimates $\hat{\theta}$ are involved, so (9) becomes $A(\theta^D, \hat{\theta}, \sigma^*)$. In this case, we obtained a value of 20.30. This clearly shows the advantage of including robust estimators as well as a contamination model.

(a) Acanthurs bahianus (b) Acanthurs chirurgus

Fig. 3. Specimens from different subspecies to be merged

6 Conclusion

In this paper, we address the problem of characterizing moving deformable objects in an aquatic environment using a robust mixture decomposition based clustering algorithm. Despite several difficulties in our application, particularly

changes in illumination conditions induced by water, preliminary experiments showed that our approach provides better estimates than the EM algorithm. Further investigations will concern the automatic selection of the different coefficients involved in the model and the test of non normal models.

Acknowledgements

This work is partially supported by the region of Poitou-Charentes.

References

1. M. R. Anderberg. *Cluster Analysis for Applications*. Academic Press, 1973. 692
2. N. Ayache. *Artificial Vision for Mobile Robots : Stereo Vision and Multisensory perception*. MIT Press, Cambridge, MA, 1991. 691
3. A. Cavallaro and T. Ebrahimi. *Video object extraction based on adaptative background and statistical change detection*, pages 465–475. In Proceedings of SPIE Electronic Imaging, San Jose, California, USA, January 2001. 690
4. R. Davé and R. Krishnapuram. Robust clustering methods: A unified view. *IEEE Transactions on Fuzzy Systems*, 5(2):270–293, 1997. 693
5. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society series B*, 39:1–38, 1977. 692
6. H. Frigui and R. Krishnapuram. A robust competitive clustering algorithm with applications in computer vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(5):450–465, May 1999. 693
7. P. J. Huber. *Robust Statistics*. John Wiley, New York, 1981. 693
8. A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–38, 2000. 692
9. A. Jain and P. Flynn. Image segmentation using clustering. *in Advances in Image Understanding, K. Bowyer and N. Ahuja (Eds), IEEE Computer Society Press*, pages 65–83, 1996. 690
10. R. Jain and H. Nagel. On the analysis of accumulative difference pictures from image sequences of real world scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):206–214, 1979. 690
11. R. Jain, R. Kasturi and B. G. Schunck. *Machine Vision*. McGRAW-HILL Inc., 1995. 691
12. G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley and Sons, 2000. ISBN 0-471-00626-2. 692
13. E. Meier and F. Ade. Object detection and tracking in range image sequences by separation of image features, stuttgart, germany. *In IEEE International Conference on Intelligent Vehicles*, pages 176–181, 1998. 689
14. D. Peel and G. McLachlan. Robust mixture modelling using the t distribution. *Statistics and Computing*, 10(4):339–348, October 2000. 693
15. C. Saint-Jean, C. Frélicot, and B. Vachon. *Clustering with EM: complex models vs. robust estimation*, pages 872–881. In proceedings of SPR 2000: F. J. Ferri, J. M. Inesta, A. Amin, and P. Pudil (Eds.). Lectures Notes in Computer Science 1876, Springer-Verlag, 2000. 693

16. G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978. 695
17. Z. Tauber, Z. Li, and M. S. Drew. *Local-based Visual Object Retrieval under Illumination Change*, volume 4, pages 43–46. In Proceedings of the 15th International Conference on Pattern Recognition, Barcelona,Spain, 2000. 689
18. S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press Inc. - ISBN 0-12-686140-4, 1999. 690
19. Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. Technical Report RR-2676, Inria, 1995. 693