2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery

Automated gelatinous zooplankton acquisition and recognition

Lorenzo Corgnati[1], Luca Mazzei[1], Simone Marini[1], Stefano Aliani[1], Alessandra Conversi[1,2], Annalisa Griffa[1], Bruno Isoppo[3], and Ennio Ottaviani[4]

[1] ISMAR - Marine Sciences Institute in La Spezia, CNR - National Research Council of Italy, Forte Santa Teresa, Loc. Pozzuolo, 19032 Lerici (SP), Italy,
[2] Marine Institute, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK,
[3] SVM srl, Via Turini 27 - 19030 Lerici (SP), Italy,
[4] On AIR srl, Via Carlo Barabino 26/4B, 16129 Genova, Italy

Abstract. Much is still unknown about marine plankton abundance and dynamics in the open and interior ocean. Especially challenging is the knowledge of gelatinous zooplankton distribution, since it has a very fragile structure and cannot be directly sampled using traditional net based techniques. In the last decades there has been an increasing interest in the oceanographic community toward imaging systems. In this paper the performance of three different methodologies, Tikhonov regularization, Support Vector Machines, and Genetic Programming, are analyzed for the recognition of gelatinous zooplankton. The three methods have been tested on images acquired in the Ligurian Sea by a low cost underwater standalone system (GUARD1). The results indicate that the three methods provide gelatinous zooplankton identification with high accuracy showing a good capability in robustly selecting relevant features, thus avoiding computational-consuming preprocessing stages. These aspects fit the requirements for running on an autonomous imaging system designed for long lasting deployments.

Keywords: pattern recognition, gelatinous zooplankton, underwater imaging, feature selection, underwater camera, GUARD1, autonomous vehicle

1 Introduction

Invasions of macro gelatinous zooplankton, jellies hereafter, have been reported as possible causes of major ecosystem changes and regime shifts with lasting ecological, economic and social consequences [9], as in the case of the invasion of the ctenophore *Mnemiopsis* in the Black Sea [3]. Monitoring jellies is certainly of importance for both marine ecologists and managers. Classical sampling with towed plankton nets is not appropriate for these delicate organisms, and is usually expensive. To overcome these shortcomings, imaging techniques can be used for monitoring this group. Underwater images acquisition quality is affected by the environment (e.g. water turbidity, light reflection or lack of natural light, presence of non-relevant objects as fishes, litter, algae, mucilage), the species

978-0-7695-5318-4/14 $31.00 © 2014 IEEE
DOI 10.1109/CVAUI.2014.12

characteristics (e.g. transparence, size) and the adopted technologies (e.g. instruments sensibility, sensor noise, field of view, lighting systems). Another crucial issue concerns the kind of platform hosting the imaging system (e.g. fixed, towed, mobile). In this paper a low cost underwater stand-alone system for image acquisition and elaboration is presented. GUARD1 has onboard image elaboration capability for the recognition of jellies and it is designed to autonomously operate on both fixed and mobile platforms. Its low cost, low volume and low power consumption make it an ideal system for long lasting deployments. Figure 1 shows the developed system GUARD1. Experiments and performance comparison of three methodologies for the recognition of gelatinous zooplankton, to be run onboard GUARD1, have been carried on the collected datasets, namely Tikhonov Regularization (TR), Support Vector Machines (SVM) and Genetic Programming (GP). The goal is to compare TR and GP, which are well established in literature but not commonly used in underwater images classification, with SVM, which is a benchmark in the field. All the three approaches select and use the most relevant image features in order to optimize the recognition performances and the computational cost. The methods have been validated within a cross-validation framework based on a ground-truth set of images. The experiments results prove that GUARD1 is a valid support for underwater imaging of jellies and guarantees high recognition performances.

2 Imaging acquisition device

The GUARD1 system is a low-cost stand-alone instrument endowed with a long-life battery pack and designed for installation on different platforms (e.g. fixed, towed, autonomous) [7]. It is fully programmable for being effective in a large range of applications. The image acquisition frequency is programmable and the battery pack life is preserved by a stand-by status between image acquisitions. The acquisition parameters are programmable as well (e.g. ISO, exposure time, focal length, iris aperture). The acquired images are analyzed (onboard) for extracting relevant information. The communication system is still under investigation and it will be released in the next months. The system consists of five modules. The acquisition component (i) is based on a programmable consumer camera. It is endowed with a lighting system (ii) that is turned on only if the natural light is not sufficient for the specific acquisition purposes. The elaboration and storage module (iii) consists of a CPU board running the algorithms for image elaboration (relevant features extraction) and pattern recognition (identification of relevant image content). The image elaboration algorithms run at scheduled time intervals on groups of images (not on single images), for saving up the battery pack (iv), whose capacity is designed to adapt to the specific deployment characteristics (e.g. duration, acquisition frequency). The control module (v) manages the operational workflows of acquisition, image elaboration and communication blocks, and can be programmed through a remote controller. GUARD1 has been tested on fixed and floating platforms so far, always achieving good performances in terms of stability, robustness and long endurance.

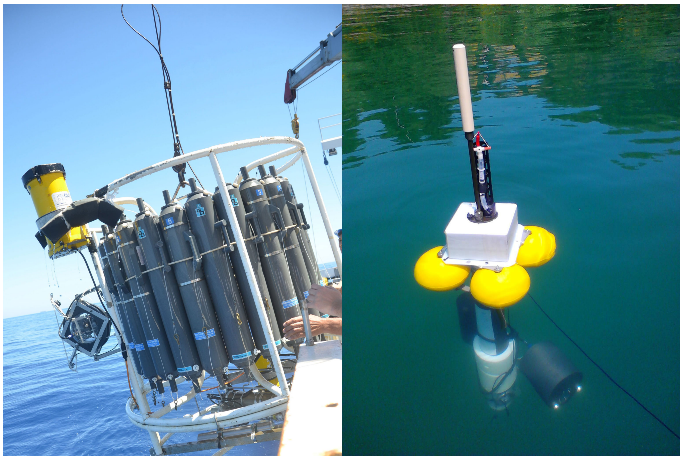

Fig. 1: The GUARD1 autonomous imaging system in two different configurations: mounted on a rosette (left) and onboard an Arvor (right).

3 Image processing and recognition

The dataset used for the experiments consists of 640x480 pixels images acquired by GUARD1 every 5 minutes in coastal waters at 5 m depth, from May to August 2013, during several ctenophores blooms in the Ligurian sea. It has 211 positive examples containing gelatinous zooplankton specimens and 211 negative examples containing only water, fishes, suspended particulate, litter and algae.

3.1 Feature extraction

Every image has been processed in order to extract the features used by the pattern recognition task, as shown in Figure 2. Histogram adjustment based on the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [8] is performed, for improving contrast between background and foreground items. The foreground is then segmented from the background using a box-shaped moving average filter with an area of size comparable with the size of expected objects (20 pixels), implemented through the integral image approach [10]. The foreground binary map is post-processed by opening/closing morphology operators to remove small dots and fill small gaps. All the connected foreground regions (blobs) are further filtered to select those with prominent edges along the boundary. Edge detection is performed with a filtering process based on the Sobel operator [11] combined with the spatial analysis of the filter response, in order to select an adaptive threshold for labelling edge pixels. The validated blobs have a minimum percentage of edge pixel along the boundary. The computational complexity of image enhancement, binarization and blob segmentation is $O(n)$, where n is the number of pixels in the image. The segmentation process is tuned in order to identify all foreground objects potentially containing jellies specimens, allowing an unavoidable incidence of false alarms. At this stage of the process, high false positive rates are acceptable to keep the detection rate as

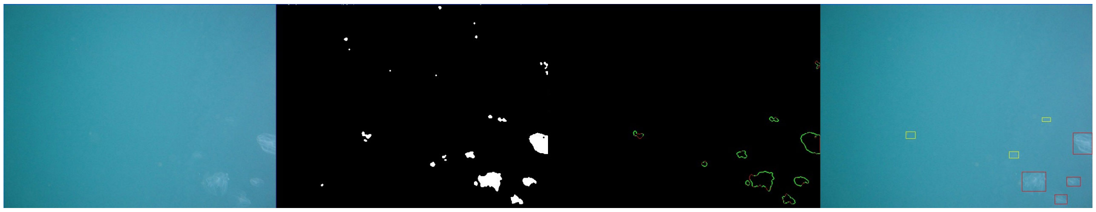

Fig. 2: Feature extraction steps: original image (extreme left), binarized image (center left), segmented image (center right), labeled image (extreme right).

high as possible. The false positive rate is greatly reduced by the classification step based on feature extraction. The extracted features belong to two groups: geometrical, based on the shape of the blob, and textural, based on the grey levels distribution inside (and outside) the blob. The geometrical features are: length of the minor semiaxis ($sAxm$), bounding box minor dimension (axm), bounding box major dimension (axM), eccentricity related to the semiaxis ratio (ecc). All these features are extracted in constant time ($O(1)$) once the blobs are identified. Other geometrical features, computed in $O(n)$, are the blob solidity defined as the area ratio between the blob and its convex hull (sol), area ($areap$), perimeter (per), radius histogram shape index defined as the ratio between the standard deviation and the mean value of the boundary ($hstI$), enthropy (ent). The textural features, extracted in $O(n)$, are: exterior-interior contrast defined as absolute difference between the averaged grey levels inside/outside the blob ($ctrs$), grays level standard deviation ($stdg1$), contrast index defined as the ratio between standard deviation and mean of the grey levels ($stdg$), gray levels enthropy ($entg$). The cost of feature extraction is modest with respect to the global cost, as most of the features are obtained by counting simple pixel attributes (e.g.(x, y) position of gray value).

3.2 Image Recognition

The recognition problem is a binary classification problem, where the classifier returns 1 if the blob identified in Section 3.1 contains a jellies specimen, 0 otherwise. The methods compared in this work are Elastic Net based on Tikhonov regularization [4], Support Vector Machines [2], and Genetic Programming [5]. Particular focus has been put on the feature selection performance of each method, in order to identify the most suitable features capable to discriminate jellies from other floating objects present in the images. In the following paragraphs, the set $E = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ consisting of n examples $x_i \in \mathcal{X} \subseteq \mathbb{R}^p, i = 1, \ldots, n$, each one characterized by p features and by a label denoted with $y_i \in \mathcal{Y} = \{0, 1\}$ will be considered for training and validation of the methods.

Tikhonov regularization (TR). The presented TR method formulation is described in [4] and it gives stable results even in presence of low cardinality datasets. According to the problem delineation above, the relation between x

and y is modeled as $y = \beta \cdot x$. Under these assumptions, the empirical risk is estimated through the least square as $(\beta \cdot x - y)^2$. The aim of the TR method is to determine a sparse model β^* of cardinality much smaller than p for which the expected risk is small. The core of the method is the minimization of the objective function defined by Zhou and Hastie [12]:

$$\frac{1}{n}\|Y - X\beta\|_2^2 + \mu\|\beta\|_2^2 + \tau\|\beta\|_1 \tag{1}$$

where X is the matrix containing the examples x_i and Y is the vector containing the labels y_i. The first term in (1) expresses the empirical risk. The second and third terms enforce the stability and uniqueness of the minimizing solution by penalizing respectively the l^2-norm and the l^1-norm of the model vector β. The non-negative parameters μ and τ are called the regularization parameters. The model selection procedure is based on a K-Fold cross validation (CV) scheme [1] defined on the set of n examples. The validation error is evaluated as the average error over different subsets for each regularization parameters pair. The optimal parameter pair (τ_{opt}, μ_{opt}) is selected as the one minimizing the validation error. Each classifier resulting form the CV returns a test error and a list of selected features. TR method produces stable solutions with good generalization performances by selecting groups of relevant correlated features [12].

Support Vector Machines (SVM). Complete description of this method can be found in [2]. SVM provide a supervised learning approach that permits to separate high dimensional data in both linear and non-linear classification tasks. Separation derives from the search for an optimal hyperplane maximizing the margin between positive and negative examples. In the case of linear separable data, the optimal hyperplane is searched in data space. In non-linear separable data scenarios, kernel based functions are used in order to perform classification in the feature space. The presented SVM method provides a classification based on Radial Basis Functions (RBF) Gaussian kernels. Considering two samples x_i and x_j, the kernel function κ is defined as:

$$\kappa(x_i, x_j) = \exp(\gamma||x_i - x_j||_2^2) \tag{2}$$

RBF based classifier involves generally two parameters: C, the soft margin parameter of the SVM common to all kernels, and γ the kernel key parameter. The tuning of C and γ is performed through a K-fold CV process selecting the values achieving the best correct detection score. SVM also allows for Recursive Feature Elimination (RFE) in order to collect a subset of key features. RFE experiments have been performed with a linear kernel SVM, in order to achieve an overall view of the features relevance rather than high performances.

Genetic Programming (GP). GP is an evolutionary computation methodology capable of learning how to accomplish a given task. GP generates the task solutions starting from an initial population of randomly generated functions, based on a set of mathematical primitives, constants and variables. The initial solutions are improved by miming the selection processes that occur naturally in biological systems through the Selection, Crossover and Mutation genetic operators [5]. In the proposed method, the set of mathematical operators $\mathcal{S} = \{+, -, *, /, sqrt, log, sin, cos, tan, atan\}$ is used to generate binary classifiers expressed as mathematical functions, whose variables correspond to the features discussed in Section 3.1. An initial population of randomly computed binary classifiers is created. Each generated classifier C is evaluated on the set of examples E. The classifier evaluation is obtained through the fitness function

$$F(C) = \frac{1}{|E|} \sum_{(x,y)\in E} J_C(x) - y, \quad J_C(x) = \begin{cases} 1 & \text{if } eval(C(x)) > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where $eval(C(x))$ returns a real number obtained by instantiating the variables of the classifier C with the features $x \in \mathcal{X}$ corresponding to the example $(x, y) \in E$. Classifiers better fitting the examples in E have higher probability of generating the new classifiers, i.e. the next generation of functions. New classifiers are generated through random mutation and crossover of the fittest classifiers. The process of forming new offspring populations of classifiers ends when a specified number of generations is reached. The more the procedure iterates through the subsequent generations, the higher is the probability to have evolved classifiers better fitting the set E. The best classifier of the final generation is selected and the whole procedure is repeated within the CV framework. A statistic analysis of the variables occurring in the classifiers resulting from the CV process identifies the most relevant features, as described in [6]. Within the statistical analysis, it is assumed that all the features have the same probability to appear in the classifiers (null hypothesis). Features for which the null hypothesis is rejected with p-value smaller than a selected value are deemed to be relevant, i.e. they appear in the evolved classifiers more times than by chance.

4 Recognition results

The experiments have been performed within a K-fold cross-validation (CV) framework [1] in order to estimate the generalization performance of the three methodologies. The structure of the CV scheme is based on a *10*-fold stratified cross validation and a nested random-sub-validation procedure, where 75% of the fold items is randomly selected for ten times. The set E of positive and negative examples discussed in Section 3 has been used as ground truth within the CV framework. The performance of the three investigated methods has been estimated by computing the average and standard deviation of Accuracy $ACC = \frac{TP+TN}{TP+FN+FP+TN}$, True Positive Rate $TPR = \frac{TP}{TP+FN}$, False Positive Rate $FPR = \frac{FP}{FP+TN}$ and False Negative Rate $FNR = \frac{FN}{FN+TP}$, where TP, FP, TN and FN represent True Positive, False Positive, True Negative and False Negative recognitions respectively. The CV framework has been also used to estimate the reliability of the relevant features identified by the three methods. A summary of the feature selection results is shown in Table 1. Each entry of the table is the percentage with which the corresponding feature has been selected in the experiments runs. The last row of the table shows only the features whose p-value is smaller than 10^{-10}. The detection results are shown in Table 2.

Table 1: Occurrence percentage of the features selected by the methods.

PR method	*sAxm*	*entg*	*stdg*	*axm*	*axM*	*ecc*	*sol*	*areap*	*per*	*hstI*	*ent*	*ctrs*	*stdg1*
TR	100	97	79	100	100	100	100	100	100	26	91	62	80
RFE SVM	100	100	34	100	100	80	100	100	100	0	100	32	54
GP	63	-	-	50	31	50	-	-	-	-	-	-	-

Table 2: Average and standard deviation (*in brackets*) of the performance indicators for each recognition methods.

PR method	**ACC**	**TPR**	**FPR**	**FNR**
TR	0.859 (*0.056*)	0.835 (*0.074*)	0.116 (*0.069*)	0.165 (*0.074*)
SVM	0.847 (*0.061*)	0.844 (*0.084*)	0.149(*0.090*)	0.155 (*0.084*)
GP	0.856 (*0.045*)	0.846 (*0.089*)	0.135 (*0.059*)	0.154 (*0.089*)

5 Discussion and Future Work

The three methods do not show significant performances differences in terms of prediction accuracy and performance indicators. The results of the three approaches are satisfactory, as they strongly enhance the precision of the simple blob analysis, and provide good generalization capability. On the contrary, differences are evident in terms of selection of the relevant features. The capability of selecting a small and robust set of relevant features is crucial for avoiding computational-consuming pre-processing tasks, according to the requirements of the autonomous imaging system presented in Section 2. In this study, as shown in Table 1, GP selects the smallest set of features by providing an accuracy and a false positive rate similar to the other two methodologies, as shown in Table 2. In order to improve the effectiveness of the three recognition approaches and to understand more deeply the behaviour differences, an accurate analysis of the correlation among features and between features and labels is now in progress. In this framework, a study on the influence of the different extracted features on the recognition rate will be conducted, together with the analysis of the incremental cost of including extra features in the classifiers. Moreover, a richer set of image features will be investigated. A further step for improving the overall efficacy of the system will be the implementation of a multi-class classifier for discriminating among different taxa of gelatinous zooplankton, where biometric features will be involved instead of only geometric and textural features.

Acknowledgment

The authors would like to thank the projects 3DStereoMonitor, GUARD1, TISANA and RITMARE.

References

1. Bartlett, P.L., Boucheron, S., Lugosi, G.: Model selection and error estimation. Machine Learning 48(1-3), 85–113 (2002)
2. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press (2000)
3. Daskalov, G.M., Grishin, A.N., Rodionov, S., Mihneva, V.: Trophic cascades triggered by overfishing reveal possible mechanisms of ecosystem regime shifts. Proceedings of the National Academy of Sciences 104(25), 10518–10523 (2007)
4. De Mol, C., Mosci, S., Traskine, M., Verri, A.: A regularized method for selecting nested groups of relevant genes from microarray data. Journal of Computational Biology 16(5), 677–690 (2009)
5. Koza, J.R.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA (1992)
6. Marini, S., Conversi, A.: Understanding zooplankton long term variability through genetic programming. In: Giacobini, M., Vanneschi, L., Bush, W. (eds.) Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, Lecture Notes in Computer Science, vol. 7246, pp. 50–61. Springer Berlin / Heidelberg (2012)
7. Marini, S., Griffa, A., Molcard, A.: Prototype imaging devices for jelly zooplankton a low-consumption, stand-alone system for image detection,analysis. SEA TECHNOLOGY Magazine 54, 44–48 (December 2013)
8. Reza, A.M.: Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology 38(1), 35–44 (2004)
9. Richardson, A.J., Bakun, A., Hays, G.C., Gibbons, M.J.: The jellyfish joyride: causes, consequences and management responses to a more gelatinous future. Trends in ecology & evolution 24(6), 312–322 (2009)
10. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. vol. 1, pp. I–511–I–518 vol.1 (2001)
11. Walther, D., Edgington, D.R., Koch, C.: Detection and tracking of objects in underwater video. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. vol. 1, pp. I–544. IEEE (2004)
12. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301–320 (2005)