# ID20- USING ORB, BOW AND SVM TO IDENTIFY AND TRACK TAGGED NORWAY LOBSTER NEPHROPS NORVEGICUS (L.)

JOSE A. GARCIA[108], DAVID MASIP[221], VALERIO SBRAGAGLIA[107], JACOPO AGUZZI[99]

*Abstract – Sustainable capture policies of many species strongly depend on the understanding of their social behaviour. Nevertheless, the analysis of emergent behaviour in marine species poses several challenges. Usually animals are captured and observed in tanks, and their behaviour is inferred from their dynamics and interactions. Therefore, researchers must deal with thousands of hours of video data. Without loss of generality, this paper proposes a computer vision approach to identify and track specific species, the Norway lobster, Nephrops norvegicus. We propose an identification scheme were animals are marked using black and white tags with a geometric shape in the center (holed triangle, filled triangle, holed circle and filled circle). Using a massive labelled dataset; we extract local features based on the ORB descriptor. These features are a posteriori clustered, and we construct a Bag of Visual Words feature vector per animal. This approximation yields us invariance to rotation and translation. A SVM classifier achieves generalization results above 99%. In a second contribution, we will make the code and training data publically available.*

*Keywords - Computer vision, object identification, video analysis, object tracking, ORB, SVM, BoW, behaviour, Nephrops norvegicus*

## I. INTRODUCTION

The Norway lobster, Nephrops norvegicus is a burrowing decapod representing a major target in crustacean European fishery [1]. The animals are caught by trawl nets only during burrow emergence, the timing of which is set upon the day–night cycle. Emergence is also modulated by social interaction in a fashion that is to date not clarified. Doubts on real stock size are reported by comparing field sampling data from trawling with more direct observations on individual behaviour in the laboratory [2]. Under isolating controlled conditions each individual expresses neat locomotor activity. Anyway, the analysis of catch samples by sex and size during different periods of the year suggests a modification of emergence during different stages of the growth or the reproductive cycle. Emergence is also apparently modulated by the close proximity of other co-specifics (as presence-absence close to the burrow), being this specie territorial [3], [4].

Behavioural animal video recording generates a huge number of videos with a large quantity of recorded hours. The human annotation of these videos requires trained people that cost large amounts of time and economical resources. Video-image analysis can be an efficient tool for microcosm experiments portraying the modulation of individual behaviour based on social interactions. Videoimage analysis is increasing its applicability to the biological research, both in the laboratory and in the field, due to the progress in frame processing for object recognition [5]. Differently from actography, hardware settings are easier, since they do not require the use of infrared barriers [6] or wheels[7], and it's not orientated to analyse social behaviour.

The analysis of social behaviour presents major limitations in the discrimination and tracking of the movement of single individuals within a group. This can be overcome with the design of particular individual tags [8], [9] to make possible the differentiation among individuals. Also it is possible to mark individuals using electronic devices like RFID chip [10] applied to Norway lobsters, or a combination of both technologies [11] (in this particular case applied to house mice). When using Computer Vision methods, the tag geometry or image quality become the central issues that condition the performance of video-image analysis and tracking with multiple individuals. In [12], authors used background subtraction techniques with flight path characteristics to identify up to 40 fruit fly (Drosophila melanogaster) individuals.

In this paper, we present a computer vision method for feature extraction and object recognition, in the context of an application to marine animal tracking. This study is a prerequisite to the posterior automated behaviour analysis, which is based on the location and recognition of the tags with different shapes placed on the top of animal's cephalothoraxes.

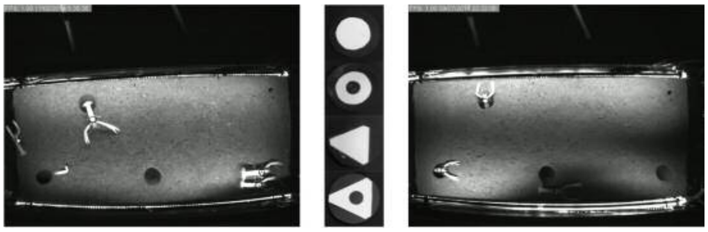


***Fig 1. Two different frames of distinct experiments. Notice the high variability in the illumination and the appearance of one claw on the bottom of the tank (left frame), which is a result of a fight between two animals. In the middle of the figure we depict the designed tags photographed out of sea water (in perfectly controlled conditions).***

## II. MATERIAL AND METHODS

In order to track individual animals, we designed four different tags in the experimental recordings. Tags are composed of a circle of black colour, and a white figure in centre of the circle with an approximate diameter of 45 mm. Figures are circle, holed circle, triangle and holed triangle and then are glued on the cephalothorax top. Figure 1 shows original form examples and animals with glued tag.

A fiberglass social tank of 150cmx70cmx30cm was constructed in order to simulate selected environmental features of N. norvegicus habitat (see an example in Figure 1), and include: the presence of four burrows (entrance and tunnel diameters of 10 and 7 cm, respectively; tunnel length of 25 cm; angular inclination of burrow entrance of 20o) and substratum simulating the sediment (made by synthetic acrylic glued to the tank base).

An USB 2.0 monochrome high-quality CMOS sensors digital camera (UI-1545LE-M, IDS) of 1280x1024 pixels resolution (SXGA/1.3 MP) took a frame each 1s. during 15 days through a software application (i.e. iSpy an open source surveillance software). That application stored each 24 hours a video record, naming it with the progressing date and time of acquisition. The video camera was endowed with a wide-angular objective of 6.0 mm and F1.4 screw C 1/2 (IDS) and it was placed in zenith position.

The illumination of the experiments was made with LED tubes of blue light (472 nm) and infrared (IR) light (860 nm), located in longitudinal position along the tank. We used blue light to simulate light conditions at deep sea [13], and IR light to allow recording the animals in darkness conditions. Finally all recordings were made in grayscale, given that the illumination light spectrum is not suitable for colour recordings.

The proposed benchmark dataset consists of four video extracted from distinct experimental trials. A total of 17 biological experiments were conducted, lasting 15 days each, and recorded at 24 fps during 60 minutes per day (approximately 500Gb of disk space). Figure 1 shows some examples of the tank and the prototypical examples of the tags.

Depending on exact time, some of the animals are partially/globally occluded in the burrows. In a preprocessing step, we took benefit of the static tank's position and we computed the bounding box of each animal using a simple background subtraction algorithm. From each detected bounding box, we found the central region of the animal, and obtained the candidate tag image. A human annotator manually labelled each image (32x32 pixels), and erroneous detections were discarded. The final tag database contains 46027 images, and it consists of: 15212 images from circles, 13451 images from holed circle, 6369 images from triangles, and 10995 images from holed triangle. Notice that the database is not fully balanced, given that some animals remain occluded longer periods of time. Figure 2 illustrates some of the segmented tags under different acquisition conditions.

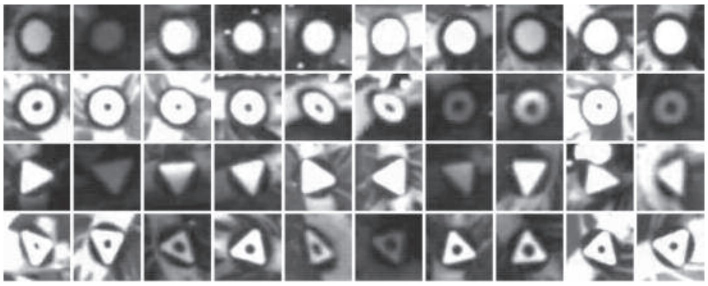
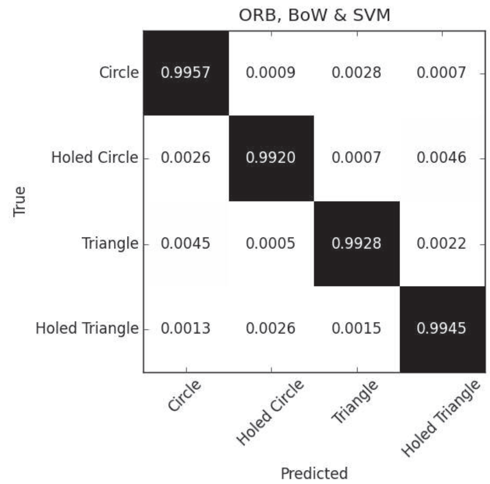


***Fig 2. Examples of several tags in a real situation, extracted from the same video recording. Notice differences in position, rotation and illumination.***




***(page before) Fig 3. Normalized confusion matrix of the implemented algorithm.***

Once the image has been segmented and the subject is located, we used only the bounding box of the tag location from the fixed position in the subject's back. The tag can appear to the classifier in any orientation, being the rotation invariant property critical for a successful classification process. Depending on the subject's position, we usually find slight variations in the scale and relevant out of plane rotations.

To classify the images, we used The Oriented, Fast and Rotated Brief (ORB) algorithm [14] for image feature extraction, and we used as a classification rules the Support Vector Machines classifiers [15].

The ORB algorithm is a fast visual descriptor based on the BRIEF (Binary Robust Independent Elementary Features) method [16]. BRIEF descriptors are a string of bits obtained performing simple random binary tests on the neighbourhood of each key point. In order to improve its robustness to in-plane rotation, ORB steers the key point neighbourhood with respect to its dominant orientation. In addition, the ORB algorithm improves BRIEF in the computation of the location of the binary tests. Instead of sampling random positions from a Gaussian distribution, ORB learns the best set of tests according to a training set, in a Greedy search for the tests with higher variance. In this paper we used the OpenCV implementation from [14], which has been successfully been applied to object detection and tracking.

In addition, we also implemented the Bag of Words model [17], given its strong success in the content based image retrieval literature [18]. Essentially we located relevant keypoints and computed the ORB local invariant features. Then, the obtained samples are clustered in 4096 bags, using the k-means algorithm. Per each image we construct a histogram according to the presence of the features with respect to the components of the bags. This histogram acts as a rotation invariant feature vector focused on the main features of each class. Finally, a SVM (RBF) is trained on these features as in [17]. The parameters from the SVM have been set automatically cross validating the training set.

The algorithm have been implemented using the out-ofthe- box code from the OpenCV library, and tests have been performed using the Python version of the OpenCV [19] and the Scikit-learn library [20].

**III. RESULTS**

We followed a 10-fold cross validation protocol. We randomly split the database in ten folds, and nine of them were used for training and one for testing. The experiments were repeated then times, each time with a different testing fold. Table 1, summarizes the mean accuracies along the ten iterations and the 95% confidence interval and CPU time consumed to classify one shape.

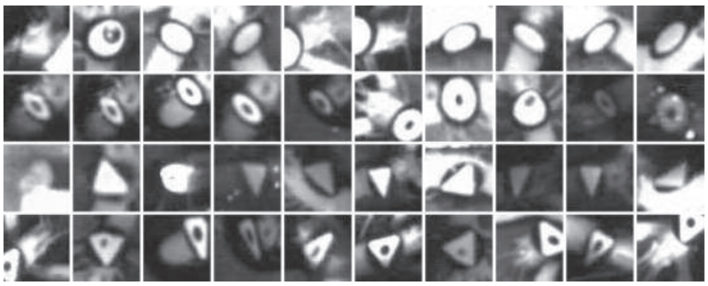


***Fig 4. Examples of misclassified shapes in multiple situations. Accuracy CPU TIME ORB, BoW & SVM 99.39 ± 0.06 0.02960849 s.***

| | **Accuracy** | **CPU TIME** |
|---|---|---|
| **ORB, BoW & SVM** | 99.39 ± 0.06 | 0.02960849 s. |

***Table 1. Mean accuracy and 95% confidence intervals of the proposed algorithm.***

**IV. DISCUSSION**

The ORB features are computed using specifically designed tests to differentiate the classes from the training set. This approximation obtains robust features with a strong degree of invariance to tags rotation. The algorithm's performance is similar across classes as shown in the normalized confusion matrix from Figure 3 Only residual confusions are found. In a qualitative analysis, Figure 4 illustrates several misclassified samples. Notice the strong out-of-plane rotations, deformations due to water flowing, and the extreme illumination conditions present in the images.

**V. CONCLUSIONS**

In this paper we introduce the use of a local descriptor in the automated monitoring of Nephrops norvegicus behaviour. We propose a complete set up to record and extract infrared images from an experimental set up. Our proposal evaluates the application of a computer vision method to the detection of especially designed tags placed in the animal's cephalothoraxes. The use of discriminant local descriptors (ORB) allows a real time detection of the tags with accuracy close to the human performance (above 99%). We plan as a future work to use more complex deep learning techniques to further improve the accuracies

on the tag detection, and extend the work to the detection of the position of the animal's limbs and head, as a previous stage to animal's interaction and behaviour modelling. In addition, we propose the possibility of changing tags shape and colours order, using the white colour to background and the black colour to the shape, given that the animal colour in IR light is white. We think that this fact could increase the visual differences between tags and it will make possible to increase their number to identify more than four individuals. The proposed code and database will be made publicly available.
Martech 2016.

***REFERENCES***


[*1] B. F. Phillips, Ed., Lobsters: biology, management, aquaculture and fisheries. Oxford, UK ; Ames, Iowa: Blackwell Pub, 2006.*

*[2] J. Aguzzi and F. Sardà, "A history of recent advancements on Nephrops norvegicus behavioral and physiological rhythms," Rev. Fish Biol. Fish., vol. 18, no. 2, pp. 235–248, May 2008.*

*[3] C. J. Chapman and F. G. Howard, "Field observations on the emergence rhythm of the Norway lobsterNephrops norvegicus, using different methods," Mar. Biol., vol. 51, no. 2, pp. 157–165, 1979.*

*[4] F. Sardà, "A review (1967-1990) of some aspects of the life history of Nephrops norvegicus," in ICES marine science symposia, 1995, vol. 199, pp. 78–88.*

*[5] S. Obdrzalek and J. Matas, "Object Recognition using Local Affine Frames on Distinguished Regions.," in BMVC, 2002, vol. 1, p. 3.*

*[6] J. Aguzzi, D. Sarriá, J. A. García, J. Del Rio, F. Sardà, and A. Manuel, "A new tracking system for the measurement of diel locomotor rhythms in the Norway lobster, Nephrops norvegicus (L.)," J. Neurosci. Methods, vol. 173, no. 2, pp. 215–224, Aug. 2008.*

*[7] S. H. Jury, C. C. Chabot, and W. H. Watson, "Daily and circadian rhythms of locomotor activity in the American lobster, Homarus americanus," J. Exp. Mar. Biol. Ecol., vol. 318, no. 1, pp. 61– 70, May 2005.*

*[8] J. Matas, L. M. Soh, and J. Kittler, "Object recognition using a tag," 1997, vol. 1, pp. 877–880.*

*[9] J.-Y. Kim, C.-J. Im, S.-W. Lee, and H.-G. Lee, "Object recognition using smart tags and stereo vision system on pantilt mechanism," Proc. ICCAS2005, 2005.*

*[10] J. Aguzzi, V. Sbragaglia, D. Sarriá, J. A. García, C. Costa, J. del Río, A. Mànuel, P. Menesatti, and F. Sardà, "A new laboratory radio frequency identification (RFID) system for behavioural tracking of marine organisms," Sensors, vol. 11, no. 10, pp. 9532–9548, 2011.*

*[11] A. Weissbrod, A. Shapiro, G. Vasserman, L. Edry, M. Dayan, A. Yitzhaky, L. Hertzberg, O. Feinerman, and T. Kimchi, "Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment," Nat. Commun., vol. 4, Jun. 2013.*

*[12] K. Branson, A. A. Robie, J. Bender, P. Perona, and M. H. Dickinson, "High-throughput ethomics in large groups of Drosophila," Nat. Methods, vol. 6, no. 6, pp. 451–457, Jun. 2009.*

*[13] S. Markager and W. F. Vincent, "Spectral light attenuation and the absorption of UV and blue light in natural waters," Limnol. Oceanogr., vol. 45, no. 3, pp. 642–650, May 2000.*

*[14] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in Computer Vision (ICCV), 2011 IEEE International Conference on, 2011, pp. 2564–2571.*

*[15] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.*

*[16] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in Computer Vision– ECCV 2010, 2010, pp. 778–792.*

*[17] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in Workshop on statistical learning in computer vision, ECCV, 2004, vol. 1, pp. 1–2. [18] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, vol. 2, pp. 524–531.*

*[19] J. Howse, OpenCV computer vision with Python: learn to capture videos, manipulate images, and track objects with Python using the OpenCV library. Birmingham: Packt Publ, 2013.*

*[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.*