Automated Identification and Tracking of *Nephrops norvegicus* (L.) Using Infrared and Monochromatic Blue Light

Garcia J. A. [a,b,1], Masip D. [a,2] Sbragaglia V. [b] and Aguzzi J. [b]

[a] *Universitat Oberta de Catalunya, Computer Vision Center, Spain*

[b] *Renewable Marine Resources Department, Institute of Marine Sciences (ICM-CSIC), Spain*

Abstract. Automated video and image analysis can be a very efficient tool to analyze animal behavior based on sociality, especially in hard access environments for researchers. The understanding of this social behavior can play a key role in the sustainable design of capture policies of many species. This paper proposes the use of computer vision algorithms to identify and track a specific specie, the Norway lobster, *Nephrops norvegicus*, a burrowing decapod with relevant commercial value which is captured by trawling. These animals can only be captured when are engaged in seabed excursions, which are strongly related with their social behavior. This emergent behavior is modulated by the day-night cycle, but their social interactions remain unknown to the scientific community. The paper introduces an identification scheme made of four distinguishable black and white tags (geometric shapes). The project has recorded 15-day experiments in laboratory pools, under monochromatic blue light (472 nm.) and darkness conditions (recorded using Infra Red light). Using this massive image set, we propose a comparative of state-of-the-art computer vision algorithms to distinguish and track the different animals' movements. We evaluate the robustness to the high noise presence in the infrared video signals and free out-of-plane rotations due to animal movement. The experiments show promising accuracies under a cross-validation protocol, being adaptable to the automation and analysis of large scale data. In a second contribution, we created an extensive dataset of shapes (46027 different shapes) from four daily experimental video recordings, which will be available to the community.

Keywords. computer vision, video analysis, object recognition, tracking, behaviour, social, decapod, *Nephrops norvegicus*

[1]Corresponding Author 1: Renewable Marine Resources Department, Institute of Marine Sciences (ICM-CSIC), Pg. Maritim de la Barceloneta 37-49, 08013 Barcelona, Spain; E-mail: jagarcia@icm.csic.es. Universitat Oberta de Catalunya, UOC, Rambla del Poblenou 156, 08018 Barcelona, Spain; E-mail: jagarco@uoc.edu.

[2]Corresponding Author 2: Universitat Oberta de Catalunya, UOC, Rambla del Poblenou 156, 08018 Barcelona, Spain; E-mail: dmasipr@uoc.edu.

Introduction

The study of rhythmic behaviour in deep-water marine organisms is, to date, limited both in the laboratory and in the field, due to the poor technological development of actography and telemetry [1]. Differently from land organisms, marine species present technical difficulties when their behaviour is monitored over prolonged periods of time, due to the presence of saline water [2]. In this context, the majority of studies dealing with marine species behavioural rhythms still often set the focus on single individuals, and avoid focusing on the influences that modulating intraspecific interactions may exert on the expression of activity itself. The biological clock regulation is then studied with techniques exported and adapted from the field of chronobiology, where individual actography still dominates the scene [3]. As a consequence, in front of the increasing awareness on neuronal processes regulating individual rhythmic behaviour, there is a concomitant and general lack of knowledge on the effects of sociality on its modulation.

The Norway lobster, *Nephrops norvegicus* is a burrowing decapod representing a major target in crustacean European fishery[4]. Animals can be captured by trawling only during burrow emergence, the timing of which is set upon the daynight 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 [5]. 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 [6][7].

Behavioral 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 costs 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. Video-image 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 [8]. Differently from actography, hardware settings are easier, since they do not require the use of infrared barriers [2] or wheels [9], and it's not orientated to analyze 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 [10][11] to make possible the differentiation among individuals. Also it is possible to mark individuals using electronic devices like RFID chip [12] applied to Norway lobsters, or a combination of both technologies [13] (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 [14], 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 comparative study of the most commonly used methods in computer vision 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 behavior analysis, which is based on the location and recognition of the tags with different shapes placed on the top of animal's cephalothoraxes.

1. Materials and methods

1.1. Animals, social tank and video settings

Animals were collected exclusively at night-time by a commercial trawler on the shelf area (100 m) of the Ebro delta (Tarragona, Spain). In order to avoid retinal damage [15] animals were never exposed to sunlight. Once captured (at night) all the operations on the deck of the trawler were performed under dim red light. Lobsters were immediately transferred to dark and refrigerated containers and then transported to the laboratory [2].

In the laboratory, specimens were transferred to acclimation tanks, hosted within a light-proof isolated chamber under the following conditions: (i) constant temperature of $13 \pm 1°$C, as reported for the western Mediterranean continental slope throughout the year [16]; (ii) random feeding time in order to prevent entrainment through food-entraining oscillators, as shown for crustaceans [17]; and (iii) LD blue monochromatic regime whose photophase duration matched the natural condition at the latitude of Barcelona (41° 23.0′ N). Also, light-ON and -OFF, were progressively attained and extinguished within 30 min in order to acclimate animal's eyes to light intensity change. The acclimation facility hosted individual cells ($25 \times 20 \times 30$ cm) made with a plastic net of different size in order to allow oxygenation, but not physical contact between animals. Acclimation was carried out at least 1 month prior to behavioural tests.

In order to track individual animals, we designed four different tags in the experimental recordings. Tags are composed of a circle of black color and a white figure in center 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 150cm×70cm×30cm 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 20°) 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 1280×1024 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 [18], 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 color recordings.

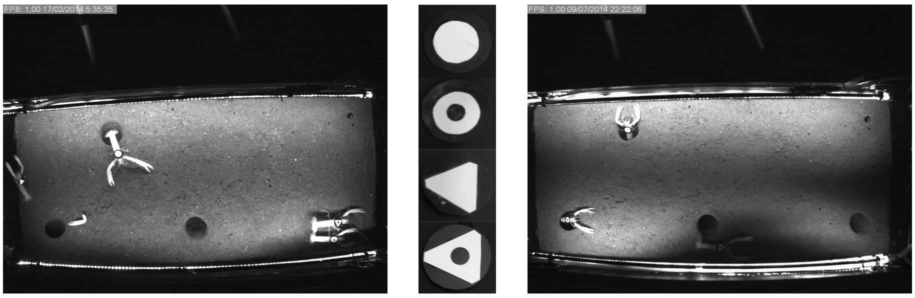

Figure 1. Two different frames of distinct experiments. Notice the high variability in the illumination and the appearance of one claw on the bottom of the tank (left frame), which is a result of a fight between two animals. In the middle of the figure we depict the designed tags photographed out of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

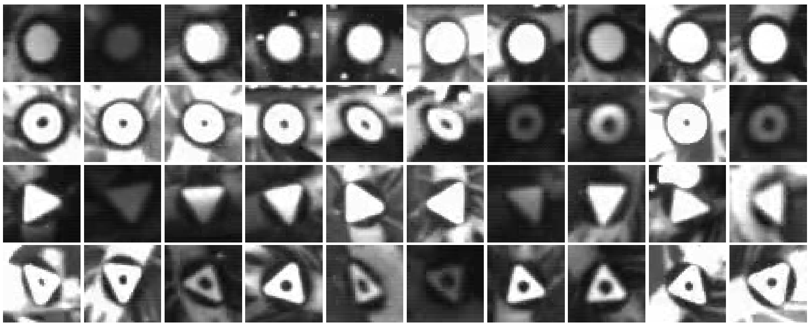

Figure 2. Examples of several tags in a real situation, extracted from the same video recording. Notice differences in position, rotation and illumination.

1.2. Construction of the benchmark database

The proposed benchmark dataset consists of four videos 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 (approximatelly 500*Gb* 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. Each image (32×32 pixels) was manually labelled by a human annotator, 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.

1.3. Image classification

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 evaluated 5 different feature extraction methods: the Scale Invariant Feature Transform (SIFT) [19], The Oriented, Fast and Rotated Brief (ORB) algorithm [20], the The Local Binary Patterns (LBP) algorithm [21], the classic Fisher Linear Discriminant Analysis (FLD) [22] and Principal Component Analysis (PCA) [23]. In this section we briefly overview the main details of the implementation of each feature extraction algorithm. We used as a classification rules the KNN Nearest Neighbors and the Support Vector Machines classifiers [24], depending on the method proposed by the referred authors.

The SIFT algorithm is a non linear feature extraction method based on two steps. First a salient point detection is performed. The most used method is to convolve the image with Gaussian filters and select local maxima and minima from different scales of the image. Then, per each key point, the magnitude and phase of the gradient are computed. An 8 bins histogram of the phase is computed and weighted by the magnitude of the gradient at each pixel. The histogram is ordered with respect to the dominant orientation, conferring to the algorithm with a strong robustness to image rotations. In this paper we followed the implementation from [19], which has been successfully applied to numerous object recognition problems.

The ORB algorithm is a fast visual descriptor based on the BRIEF (Binary Robust Independent Elementary Features) method [25]. BRIEF descriptors are a string of bits obtained performing simple random binary tests on the neighborhood of each key point. In order to improve its robustness to in-plane rotation, ORB steers the key point neighborhood 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 [20], which has been successfully been applied to object detection and tracking.

The LBP algorithm has been successfully applied to several computer vision tasks, such as object recognition [21] and facial expression classification [26]. The method also performs binary tests on the neighborhood of each pixel, in a dense manner. Nevertheless, the tests in LBP are performed clockwise with respect to the central pixel of its circular neighborhood. If a pixel value is greater than the central pixel, it gets a 1, and 0 otherwise. The eight surrounding pixels of each keypoint are encoded as 8 bits (8 neighbors), which constitute a number from 0 to 255. Finally, per each cell of the image the histogram of these numbers is computed and used as a visual descriptor.

Finally we used the well known Fisher Linear Discriminant Analysis (FLD) [22] and PCA [23] feature extraction algorithms as a baseline. PCA finds the projection matrix that reduces the data dimensionality minimizing the reconstruction error, or equivalently maximizing the data variance in the projected space. This linear transform is holistic and not invariant to scale, rotation or affine transforms. Similarly the FLD algorithm fins a

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

ORB + KNN	SIFT + KNN	ORB + SVM	LBP + KNN
86.03 ± 0.30	96.44 ± 0.20	99.24 ± 0.17	94.77 ± 0.09
Fisher + KNN	PCA + KNN	ORB + BOW + SVM	SIFT + BOW + SVM
97.20 ± 0.17	$98.67 \pm 0,10$	99.39 ± 0.06	99.40 ± 0.08

Table 2. Mean CPU time consumed in seconds to classify one shape.

ORB + KNN	SIFT + KNN	ORB + SVM	LBP + KNN
2.7447630	14.821450	0.004693985	1.4778730
Fisher + KNN	PCA + KNN	ORB + BOW + SVM	SIFT + BOW + SVM
0.0113380	0.1082115	0.02960849	0.02302241

projection matrix to perform a dimensionality reduction on the data. Nevertheless, FLD takes into account the class membership of each sample in the training data, obtaining the projection directions where samples from different classes are maximally scattered, and samples from the same class are projected as close as possible.

In addition, we also implemented the Bag of Words model [27], given its strong success in the content based image retrieval literature [28]. Essentially we used the best two feature extractors (ORB and SIFT) to locate relevant keypoints and computed local scale invariant features. Then, the obtained samples are clustered in 1024 bags for SIFT and 4096 bags for ORB (using the k-means algorithm), and we construct a histogram per image. 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 [27]. The parameters from the SVM have been set automatically crossvalidating the training set.

All these methods have been implemented using the out-of-the-box code from the OpenCV library, and tests have been performed using the Python version of the OpenCV [29] and the scikit-learn library.

2. Experimental results

In all the cases, 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 1 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 intervals.

The best performing feature extraction method is ORB, especially when used with the SVM classifier. The difference in mean accuracy from *ORB+SVM*, *ORB+BOW+SVM* and *SIFT+BOW+SVM* is not statistically significant, although the three methods significantly outperform the rest. Figure 3 graphically shows the confidence intervals overlapping. The *PCA* and *FLD* algorithms perform similarly using the classical KNN implementation. Notice that *ORB+KNN* underperformed these four approaches. We conjecture that the current OpenCV implementation uses approximate nearest neighbor search (FLANN [30]), which might obtain less accurate NNs.

Table 2 shows the mean time consumed by each method. Experiments were performed on a regular computer (intel i3, 8Gb RAM, Linux OS, Python 2.7). Notice that ORB features are the faster to compute, providing real time classification of the tags.

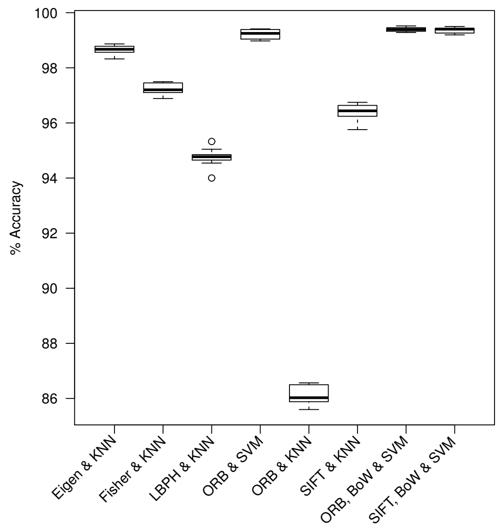

Figure 3. Boxplot of accuracy of each tested algorithm.

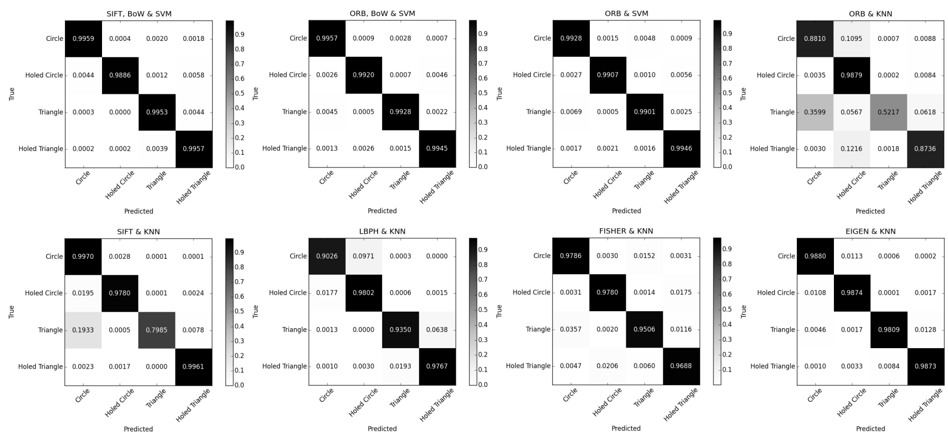

Figure 4. Normalized confusion matrix of each tested algorithm.

Approaches relying on the KNN algorithm tend to be slower, given the need to store and search on the full database of training samples. These methods could be considerably speeded up by using editing and prototype selection algorithms.

3. Discussion

Discriminant local descriptors as ORB or FLD outperform the rest in the classification task. Notice that ORB features are computed using specifically designed tests to differ-

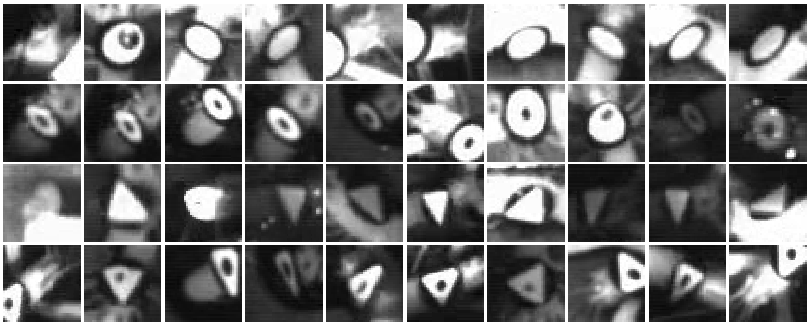

Figure 5. Examples of misclassified shapes in multiple situations.

entiate the classes in the training set. Figure 4 depicts the confusion matrices of each method, and ORB descriptors perform uniformly along the different classes. Only residual confusions are found among circle and triangle classes. SIFT and ORB with KNN tend to confuse *circle* and *triangle*. This fact is justified by the detection of key points in the border region of the tags and the absence of salient points on the straight edges, which might suggest a defective implementation design of the triangular shape. The black border in these tags is insufficient to facilitate the keypoints detection. Figure 5 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.

4. Conclusions

In this paper we introduce the use of local descriptors in the automated monitoring of *Nephrops norvegicus* behavior. We propose a complete set up to record and extract infrared images from a experimental set up. Our proposal evaluates the application of state-of-the-art computer vision methods to the detection of especially designed tags placed in the animal's cephalothoraxes. The use of discriminant local descriptors (particularly ORB) outperform the rest of the methods, and allows a real time detection of the tags with an 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 behavior modeling. In addition, we propose the possibility of changing tags shape and colors order, using the white color to background and the black color to the shape, given that the animal color 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.

Acknowledgements

Research supported by TIN2012-38187-C03-02 grant from the Spanish Ministry of Education Culture and Sports and TIN2015-66951-C2-2-R (MINECO/FEDER) grant from the Spanish Ministry of Economy and Competitiveness.

The data presented in this work were obtained during the RITFIM project (CTM2010-16274), which was funded by the Spanish Ministry of Science and Innovation (MICINN).

References

[1] D. Sarria, J. del Rio, A. Manuel, J. Aguzzi, J. A. Garcia, and F. Sarda, Actographic detection system based on infrared and computer vision technologies to measure the behaviour of species, in *16th IMEKO TC4 symposium: exploring new frontiers of instrumentation and methods for electrical and electronic measurements* (ISBN: 9788890314933), 2008.

[2] J. Aguzzi, D. Sarria, J. A. Garcia, J. Del Rio, F. Sarda, and A. Manuel, A new tracking system for the measurement of diel locomotor rhythms in the Norway lobster, *Nephrops norvegicus* (L.), *Journal of Neuroscience Methods*, **173-2** (2008), 215–224.

[3] Naylor, Ernest, Chronobiology: implications for marine resource exploitation and management, *Scientia Marina*, **69-1** (2005), 157–167.

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

[5] J. Aguzzi and F. Sarda, A history of recent advancements on *Nephrops norvegicus* behavioral and physiological rhythms, *Reviews in Fish Biology and Fisheries*, **18-2** (2008), 235–248.

[6] C. J. Chapman and F. G. Howard, Field observations on the emergence rhythm of the Norway lobster *Nephrops norvegicus*, using different methods, *Marine Biology*, **51-2** (1979), 157–165.

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

[8] S. Obdrzalek and J. Matas, Object Recognition using Local Affine Frames on Distinguished Regions., in *BMVC*, **1** (2002), 3.

[9] S. H. Jury, C. C. Chabot, and W. H. Watson, Daily and circadian rhythms of locomotor activity in the American lobster, *Homarus americanus*, *Journal of Experimental Marine Biology and Ecology*, **318-1** (2005), 61–70.

[10] J. Matas, L. M. Soh and J. Kittler, Object recognition using a tag, *Image Processing, 1997. Proceedings., International Conference* on, Santa Barbara, CA, **1** (1997), 877–880. doi: 10.1109/ICIP.1997.648106

[11] 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, *Proceedings of ICCAS2005*, 2005.

[12] J. Aguzzi, V. Sbragaglia, D. Sarria, J. A. Garcia, C. Costa, J. del Rio, A. Mnuel, P. Menesatti, and F. Sarda, A new laboratory radio frequency identification (RFID) system for behavioural tracking of marine organisms, *Sensors*, **11-10** (2011), 95329548.

[13] 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, *Nature Communications*, **4** (2013), DOI:10.1038/ncomms3018

[14] K. Branson, A. A. Robie, J. Bender, P. Perona, and M. H. Dickinson, High-throughput ethomics in large groups of Drosophila, *Nature Methods*, **6-6** (2009), 451457.

[15] E. Gaten, P. M. J. Shelton, C. J. Chapman, and A. M. Shanks, Depth related variation in the structure and functioning of the compound eye of the Norway lobster *Nephrops norvegicus*, *Journal of the Marine Biological Association of the United Kingdom*, **70-2** (1990), 343.

[16] J. Salat, Review of hydrographic environmental factors that may influence anchovy habitats in northwestern Mediterranean, *Scientia Marina*, **60** (1996), 21–32.

[17] F. F. De Miguel y H. Archiga, Circadian locomotor activity and its entrainment by food in the crayfish *Procambarus clarki*, *The Journal of experimental biology*, **190-1** (1994), 9–21.

[18] S. Markager and W. F. Vincent, Spectral light attenuation and the absorption of UV and blue light in natural waters, *Limnology and Oceanography*, **45-3** (2000), 642–650.

[19] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, **60-2** (2004), 91–110.

[20] 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*, 2564–2571.

[21] M. Heikkil, M. Pietikinen and C. Schmid. Description of interest regions with local binary patterns. *Pattern recognition*, **42-3** (2009), 425–436.

[22] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals Eugen*, **7** (1936), 179–188.

[23] Matthew A. Turk and Alex P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition. Proceedings CVPR'91., IEEE Computer Society Conference on 1991*, 586–591.

[24] C. Cortes and V. Vapnik, Support-vector networks, *Machine learning*, **20-3** (1995), 273–297.

[25] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, Brief: Binary robust independent elementary features, in *Computer VisionECCV 2010*, 778–792.

[26] Zhao Guoying, and Matti Pietikinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, **29-6** (2007), 915–928.

[27] 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 **1** (2004), 1–2.

[28] 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*, **2** (2005), 524–531. doi:10.1109/CVPR.2005.16

[29] G. Bradski, OpenCV Library, *Dr. Dobbs Journal of Software Tools*, 2000.

[30] M. Muja and David G. Lowe. Scalable Nearest Neighbor Algorithms for High Dimensional Data. *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, **36-11** (2014), 2227–2240.