16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

I^2N^2: A SOFTWARE FOR THE CLASSIFICATION OF BENTHIC HABITATS CHARACTERISTICS

Phooi Yee Lau[1], Paulo Lobato Correia[1, 2], Paulo Fonseca[3], and Aida Campos[3]

[1]Instituto de Telecomunicações, [2]Instituto Superior Técnico, [3]INRB/L-IPIMAR,
[1,2]Av. Rovisco Pais, 1049-001 Lisboa, Portugal, [3]Av. de Brasilia, 1449-006 Lisboa, Portugal
phone: + (351) 218418164, fax: + (351) 218418472, email: {laupy,plc}@lx.it.pt, {pfonseca,acampos}@ipimar.pt

ABSTRACT

Underwater video cameras mounted into towed platforms (e.g. sledges) have been increasingly used for the assessment of commercial crustacean stocks and also for more ecology-directed studies, including the impact of human activities in marine habitats. In this study a video camera was mounted on a trawl headline, to acquire footages at about 500 meters depth in Norway lobster (Nephrops norvegicus) fishing grounds, to automatically quantifying the species abundance and density of burrows, and assess the impact of fishing activities in these fishing grounds. Six complementary features are proposed to identify the lobsters and their burrows: average intensity, slant angle, run-length histogram, shape ranking, co-occurrence matrices and cross-counting. A prototype system, IT-IPIMAR Nephrops Norvegicus (I^2N^2) is presented and experimental results show that the proposed features, when used in combination, are able to effectively classify segmented regions as lobsters or burrows.

1. INTRODUCTION

Video technology is being increasingly used to assess the abundance of commercial stocks and monitor the effects of fishing on benthic ecosystems [1]. Huge amounts of images are being generated and highly trained staff needs to scan the video data for meaningful information. Often those videos display complex structure that appears with low contrast, making this task very demanding and its results dependent on the experience and concentration of the human expert. Also, manually processing video sequences to quantify species abundance is a lengthy and tedious task. Furthermore, an automated system may facilitate the interchange and comparability of video sequences from different institutes and fishing grounds to standardize counting.

Nowadays, underwater imaging devices are used by a number of scientific teams to study benthic habitats and monitor marine biodiversity [2-6]. In one of our previous works [3], a computer vision system was proposed by combining three visual features to detect and count Norway lobsters (*Nephrops Norvegicus*). In the present study, apart from detecting the presence of lobsters, as in [3], other visually distinguishable objects are also targeted. This is the case of Norway lobster's burrows, which are key elements for a more complete fisheries-independent evaluation of the species abundance. Burrows' counting gives a more stable species abundance estimate (even though they are not straightforward to observe), as only a reduced part of the lobster population is out of their burrows at a time. As such, a three-class detection problem is considered: 1) Norway lobsters; 2) burrows; and 3) others (not included in the first two classes, but may be of interest, e.g., marks of the trawl gear impact on the bottom).

In this paper, a framework is proposed to introduce a computer-based procedure to analyze benthic habitat characteristics based on image processing algorithms. Overall, it introduces several advantages:

1. Implementation of a computer-based method to analyze video sequences;
2. Minimize operator related errors that are dependant on the skill and concentration of the human operator;
3. Facilitate the standardization of counting among fishery research institutes.

The remainder of this paper includes: Section 2 outlining the characteristics of the underwater video sequences and the experimental set-up; Section 3 describing the proposed methodology; Section 4 discussing the I^2N^2 framework; Section 5 showing experimental results; and Section 6 concluding the paper.

2. UNDERWATER VIDEO SET-UP AND VIDEO SEQUENCE CHARACTERISTICS

Video sequences were acquired by the Laboratory of Fisheries Research of the Portuguese Institute for Biological Resources (INRB/L- IPIMAR) during fishing gear selectivity trials carried out within the scope of the EU Project 'NECESSITY', in July 2005. The images were obtained in *Nephrops* fishing grounds off the Portuguese southern coast, at about 500 meters depth. A Kongsberg Maritime OE1324 monochrome low-light SIT camera, with a light sensitivity (limiting) of 2e-4 lux, associated to a recording and powering system able to work up to 1500 meters depth, was used. This equipment, whose set-up is illustrated in Figure 1, was hung from the trawl's headline, angling down in the tow direction to register ground images. Tows were carried out during daylight at a towing speed of about 3.0 knots.

A sample image, extracted from the video sequences, is shown in Figure 2. The images, converted from Hi8 tape to AVI format, were collected at trawling speeds higher than desirable for this type of analysis, resulting in relatively low

The authors acknowledge the support from Fundação para a Ciência e Tecnologia (FCT).

quality seabed video sequences. Notably: 1) part of the sequences could not be used due to the presence of "marine snow", corresponding to re-suspended sediments that occlude the seabed; 2) low image contrast due to lack of natural illumination at high depths and exclusive dependence on an artificial illumination which is difficult to control; 3) camera motion due to being hung from a flexible structure - the trawl headline - leading to somewhat non-uniform surface areas being swept per time unit; 4) artefacts appearing in the borders of the image due to the camera container, used to withstand the high pressure.

(a) (b)

Figure 1: Video Acquisition: (a)trawl; (b) camera with light focus

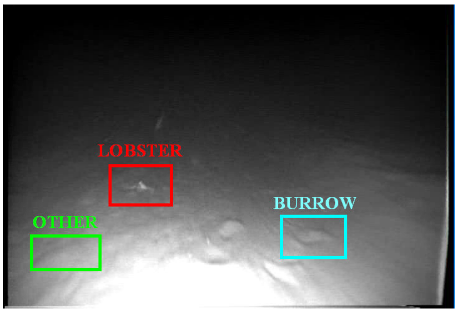


Figure 2: Sample input image

3. PROPOSED ANALYSIS METHODOLOGY

The block diagram shown in Figure 3 summarizes the proposed video analysis methodology. After acquisition, the video is pre-processed and a set of uniform regions that are candidates to be classified later, are segmented and processed. A set of features, selected for image analysis, is then extracted: average intensity (AI), slant angle (SA), run-length histogram (RH), shape ranking (SR), co-occurrence matrices (CM), and cross counting (CC). Finally, a classification decision into lobster (L), burrow (B) or other (O), is done.

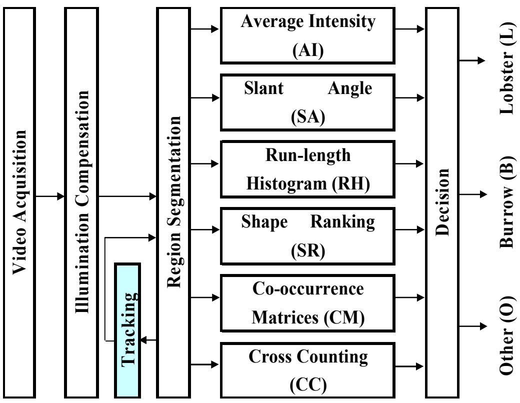


Figure 3: System architecture of the computer-based support system for benthic habitats analysis

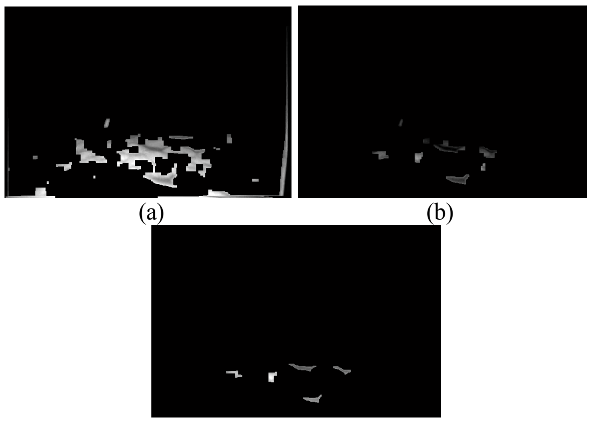
3.1. Pre-processing: Illumination Compensation

This module compensates the non-uniform artificial illumination that leads to the image part further away from the camera being too dark and to an over-illuminated area closer to the camera. Here, the image area near the camera is normalized by the average horizontal illumination (as applied in [3]), obtained using equation (1).

$$G_{lum}(h) = \frac{1}{W}\sum_{w=0}^{W-1} I(w,h) \qquad w \in \{0,1,2,\ldots,W\}, h \in \{0,1,2,\ldots,H\} \tag{1}$$

3.2. Region Segmentation

The main goal of the region segmentation module is to identify a set of relevant regions that may correspond to lobsters or burrows. Three types of candidate regions are identified: 1) Type *A* (*CRA*) – homogenous regions, with high contrast to the surrounding areas (either bright or dark regions), and spatial details corresponding to object borders and edges; 2) Type *B* (*CRB*) – *CRA* regions containing some background pixels are eliminated in order to extract more precise contours of large burrows (with high contrast and darker than surrounding areas); and 3) Type *C* (*CRC*) – *CRA* regions from which some foreground pixels are eliminated in order to extract lobsters' contours (with high contrast and brighter than surrounding areas). Figure 4 shows a segmentation result example.



(a) (b) (c)

Figure 4: Segmentation results: (a) homogeneous regions (*CRA*); (b) refined regions (*CRB*); (c) refined regions (*CRC*)

For *CRA* regions, at first, the gradient magnitude of the image is calculated, using the Sobel operator. Then, a thresholding operation is applied to select pixels with high gradient values, using a fixed threshold at 0.1 x *m*, whereby *m* is the maximum value of the gradient magnitude. 4-connected regions are detected. Regions that are connected to the image borders are removed, as they correspond to artifacts appearing due to the container used for the camera to withstand the high pressure. The remaining regions undergo a morphologic close operation, with a square-shaped structuring element of width 15 (large enough to group region fragments together), and a morphologic open with a disk-shaped structuring element of radius 3 to smooth the regions. The resulting 8-connected foreground regions are detected and labelled, after small regions (less than 0.03% of image area) elimination, constituting the candidate regions for subsequent average intensity (*AI*) and slant angle (*SA*) features analysis. These *CRA* candidate regions will undergo a temporal tracking.

CRB regions are obtained from thresholding *CRA* regions based on the regions' mean intensity value followed by the detection of a new set of 8- connected regions. Again, small regions (less than 0.03% of image area) are eliminated. These regions are taken as candidates for the co-occurrence matrices (*CM*) feature analysis.

CRC regions are obtained by iterating once more the threshold, small regions elimination and region labelling operations over *CRB* regions. The resulting regions are taken as candidate regions for run-length histogram (*RH*), shape ranking (*SR*), and cross-counting (*CC*) feature analysis.

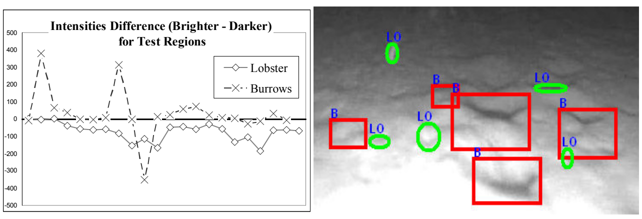
3.3. Tracking

The tracking module is used to ensure the stability of *CRA* regions along time, leading to the removal of inconsistent regions. This allows increasing the system efficiency, ensuring notably: 1) *CRA* candidate regions temporal consistency; 2) correspondence between two consecutive labelled regions; and, 3) removal of regions not related between consecutive frames. Tracking is implemented using change detection between consecutive frames, with pixels being labelled '0' (not changed) or '1' (changed). A *CRA* candidate region is successfully tracked when its new position leads to more than 5% (chosen to allow stable results at the considered vessel towing speed) of area overlap with the previous instant region position, else it is discarded.

3.4. Features used for Analysis

3.4.1. Average Intensity (AI)

Average intensity can be used to detect lobsters and burrows by comparing the number of relatively brighter (*Br*) and darker (*D*) pixels in *CRA* regions. Initial test data shows that lobsters contain less *Br* than *D* areas, while for burrows, the opposite is observed. This is illustrated in Figure 5 (a).



(a) (b)

Figure 5: Experimental results: (a) Difference between brighter and darker pixels in burrows and lobster regions; (b) Average Intensity

Brighter and darker pixel counting is done by comparing pixel gray levels against the region average value, Ave_R. Pixels with value above (or below) Ave_R are counted as brighter (or darker), as indicated in equation (2). Finally, classification of regions as burrows (*B*), or as lobsters (*L*) or others (*O*) is performed, according to equation (3). Figure 5 (b) shows experimental results, where regions classified as lobsters and burrows, are marked by circles and squares, respectively.

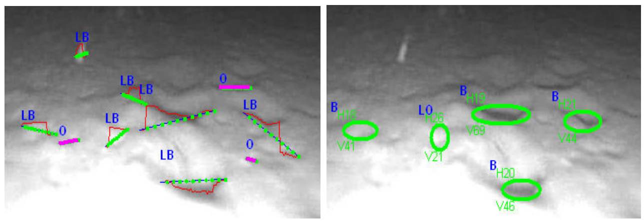
$$\begin{cases} Br = \sum_{(x,y)\in R} \left(I(x,y) > Ave_R \right) \\ D = \sum_{(x,y)\in R} \left(I(x,y) < Ave_R \right) \end{cases} \quad (2), \qquad \begin{cases} B : Br > D \\ L, O : Br < D \end{cases} \quad (3)$$

3.4.2. Slant Angle (SA)

The slant angle can be used to decide that a *CRA* candidate region is not a lobster or a burrow, as these regions are not expected to conform to a straight line. Slant line computation is done by finding the local minima pixel coordinates (marked as red lines in Figure 6 (a)), for each vertical line of the region, and connecting the first with the last pixel (shown as blue lines). For regions conforming to a straight line, the distance between the local minima line and the slant line should be small. Here, a distance threshold, *slt*, with default value of 10 is defined. The distance to the slant line is evaluated at 10 equally spaced positions (shown in Figure 6 (a) as green points), counting the cases where distance is not above *slt*. If this number is above 6, the region is assumed to conform to a straight line (marked as magenta line), not being a lobster or a burrow – see sample results in Figure 6 (a).

3.4.3. Run-length Histogram (RH)

The run-length histogram can be used to distinguish lobster and other regions from burrows, as burrows have an elongated appearance. Region's horizontal (*HR*) and vertical (*VR*) run length histograms are calculated. If R_i is the vertical width at column *i*, the horizontal run-length array consists of concatenating R_1, R_2 … R_m, where *m* is the *CRC* candidate region horizontal length – see Figure 6 (b). Experimental results show that burrows typically have average ratios (*VR*/*HR*) above 3 while the ratio for lobsters and other regions are around 1. This happens since lobsters present short and similar horizontal and vertical run-lengths, while burrows typically present long horizontal and short vertical run-lengths.



(a) (b)

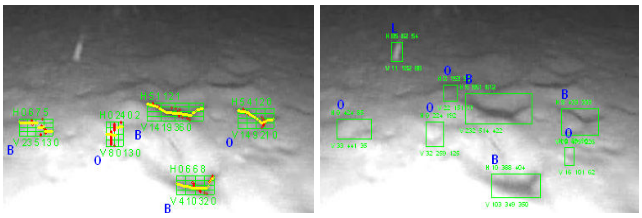
Figure 6: Experimental results: (a) Slant angle; (b) Run-length Histogram

3.4.4. Shape Ranking (SR)

The shape ranking consists of a one-dimensional string of codes that can be used to distinguish between lobsters, burrows and others, as lobsters present shorter and uniform codes, while burrows' codes are longer and have large variances and others' codes are short but not-uniform. The observation sequence is obtained by finding a region's middle pixel for each horizontal and vertical scan line, and considering each line divided into four blocks represented by a code that corresponds to a power of 2 – see example in Figure 7 (a). Each scan line is represented by the codes of the median on that scan line. Each *CRC* candidate region has its own observation sequence represented by *SR* = (*R*, *V*, *H*), where *R* is the region number, *V* and *H* are the codes for vertical and horizontal lines, respectively. Figure 7 (a) shows sample results.

3.4.5. Co-occurrence Matrices (CM)

The co-occurrence matrix is used to study the number of times a given pair of pixel characteristics occurs at a given distance and with a given orientation. Here, two directions (horizontal and vertical) and one distance (2 pixels) are used for analysing *CRB* candidate regions. Three types of co-occurrences are considered: *x-low*, referring to low pixel intensities in the given pair of pixels; *x-high*, referring to high pixels intensities, within a margin *m* (set to 5), of the region's median value; and *x-dhigh*, referring to extreme level changes in pixel intensity at each distance and orientation, within a margin *g* (set to 20). The associated classification for a region can be expressed as follows (according to the majority of observed pixel pairs): 1) lobsters: many *x-dhigh* pairs at horizontal and vertical direction as they have inconsistent pixel intensities, given the characteristics of lobsters, which are highly illuminated and present high pixel's intensity changes due to the lobster structure, 2) burrows: many *x-dhigh* at vertical but low in horizontal direction, given the characteristics of burrows, which are slightly rectangular and stably shaped, and 3) other: *x-high* and *x-low*, given characteristics of regions that are noisy (due to bio-disturbance). Experimental results are shown in Figure 7 (b).



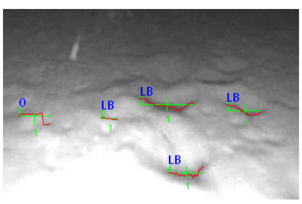
(a) (b)

Figure 7: Experimental results: (a) Shape Ranking: coded blocks and their classification; (b) Co-occurrence matrices

3.4.6. Cross Counting (CC)

Cross counting consists of counting the number of times local minima pixels in a region cross the centre baseline in the vertical and horizontal directions. It can be used to decide that a *CRC* candidate region is not a lobster or a burrow, as in both cases at least two crossings are expected, while 'other' regions typically do not present more than one crossing.

Item	Cross counting		
	1	2	more
Lobster	1	2	4
Burrow	0	20	19
Other	7	0	0



(a) (b)

Figure 8: Cross Counting: (a) Comparison for the three classes considered: *L,B* and *O*; (b) Experimental results

The horizontal and vertical centre baselines are obtained by finding the local minima intensity in the horizontal (or vertical) direction and calculating their mean location. Experimental results confirm that lobsters and burrows have two or more crossings – see Figure 8 (a). Further experimental results are shown in Figure 8 (b).

4. I^2N^2 FRAMEWORK

The analysis framework, *IT-IPIMAR Nephrops Norvegicus* (I^2N^2) has been developed in the Image Group Laboratory at Instituto de Telecomunicações, Instituto Superior Técnico. I^2N^2 is a research-oriented system with the capability to analyze benthic habitat characteristics taking as input video sequences in AVI format. The graphical user interface is shown in Figure 9. It was developed using the Matlab development environment. The current version of the system is able to analyze grayscale and color images. Currently the I^2N^2 framework is able to detect and classify regions as lobsters (L), burrows (B), or others (O).

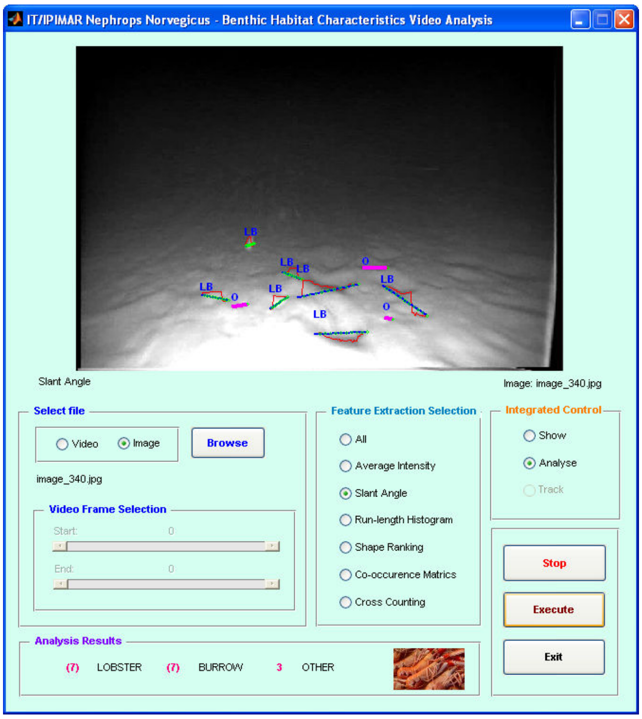


Figure 9: The I^2N^2 system graphical user interface

5. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithms, a selected video sequence with 430 frames, where Norway lobsters and burrows are visible, is used. First, the behaviour of each of the six considered features is analyzed, according to the classification rules summarized in Table 1. Then, the combined classification performance is evaluated.

Experimental classification results for each of the six extracted features (Table 2) indicate:

1. *AI* shows strength in burrows identification while *SA* can consistently classify regions belonging to 'others'
2. *CM* can identify all regions belonging to lobsters, but shows weaknesses for burrows identification.
3. *RH* shows a good potential for burrows classification, while *CC* shows strength in identifying 'other' regions.
4. SR can identify well defined lobster regions.

The individual feature results show that a clear classification decision (for a three-class problem) cannot be achieved, especially since some of the features only allow excluding one of the three targeted classes (e.g. *AI*/*SA*/*RH*/*CC)*.

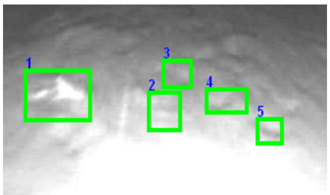
Table 1: Classification decision for each individual feature

	Struct. Statistics		Spatial Features			
	AI	SA	CM	RH	SR	CC
Lobsters (*L*)	Br < D	no	>x-dhigh (vertical/ horizontal)	ave ratio 1	short uniform	>2
Burrows (*B*)	Br > D	no	> x-dhigh (vertical)	ave ratio > 3	long variance	>2
Others (*O*)	Br < D	yes	x-low x-high	ave ratio 1	short not-uniform	1

Table 2: Experimental classification results using each of the individual extracted features

Item		AI	SA	CM	RH	SR	CC
Classification Decision		LO/B	LB/O	L/B/O	LO/B	L/B/O	LB/O
Region Type		CRA		CRB	CRC		
Total region(s)		1196		1171	924		
Classification	Lobster (L)	608	563	98	414	39	441
	Burrow (B)	588	563	766	510	339	441
	Other (O)	608	633	307	414	546	483
Significant (class)		(B)	(O)	(L)	(B)	(L)	(O)

As discussed above, a more reliable classification can be obtained by combining the partial results coming from the individual features. The final classification is done by fusion of the individual decisions, with each feature providing one or two votes. A larger weight (two votes) is given to *AI, SA, RH* and *CC* features, where one of the classes is clearly rejected. A majority voting decision is taken.



Region No.	*AI*	*SA*	*RH*	*SR*	*CM*	*CC*	***Decision***
1	LO	LB	LO	L	L	LB	***L***
2	LO	LB	LO	L	B	O	***L***
3	B	O	-	-	B	-	***B***
4	LO	O	-	-	-	-	***O***
5	B	B	LO	O	O	O	***O***

Figure 11: Experimental results based on votes

Table 3: Combined classification performance for each of the considered classes.

Item	Manual	I^2N^2			Wrong	Correct
		L	B	O	(%)	(%)
Lobster (L)	8	8	0	0	0 (0%)	8 (100%)
Burrow (B)	40	0	35	5	5 (13%)	35 (87%)
Other (O)	97	5	19	73	24 (25%)	73 (75%)

Sample experimental results, for two subsets of 20 contiguous frames each (40 frames in total) for which ground truth data is available, are included in Figure 11 and Table 3. In this case, a total of 145 regions are segmented. Results show that all lobster regions were correctly classified, only five burrows were classified as 'others', and among the regions that should be classified as 'others', 5 were wrongly classified as lobsters and 19 as burrows.

6. CONCLUSION AND FUTURE WORK

An automatic image analysis technique is proposed for the detection of benthic organisms and the structures resulting from their activity. A high commercially valuable crustacean, the Norway Lobster, is used as a case-study. Here, two central interconnected ideas are proposed: the first is to provide methods of image analysis for fisheries-independent abundance estimates of epibenthic species, as a reliable alternative to the currently used human operator based approach. The second idea is to implement those methods in a software platform to assist marine habitat research. Accordingly, six features and a method to combine their individual results are used to achieve a reliable classification, and implemented in an application with a user-friendly graphical interface.

Experimental results were encouraging. All lobsters were correctly detected as well as a significant portion of the burrows (87% in the small sequence sample for which manual analysis data were available). Nevertheless, there are some false positives, with some regions being incorrectly classified as lobsters or burrows.

The work here presented can be extended to additionally detect the impact marks of the fishing gears on the sea bottom, opening a further usage of this type of methodologies in the study of human impacts in the ecosystem. It must be stressed that for the specific case of the Norway lobster, and other burrowing species, where abundance is estimated by counting burrows instead of the individuals themselves, there is still a considerable amount of work ahead.

Future improvements to the segmentation algorithms are expected, by dealing with illumination inconsistencies.

REFERENCES

[1] J.S. Link, "Ecological Considerations in Fisheries Management: When Does It Matter," *Fisheries*, vol. 27 (4), pp. 10-17, 2002.

[2] I.D. Tuck, C.J. Chapman, R.J. Atkinson, "Population Biology of the Norway Lobster, Nephrops Norvegicus (L) in the Firth of Clyde, Scotland .1. Growth and Density", *ICES Journal of Marine Science*, vol. 54, pp. 125-135, 1997.

[3] P.L. Correia, P.Y. Lau, P. Fonseca and A. Campos, "Underwater video analysis for norway lobster stock quantification using multiple visual attention features," in *Proc. EUSIPCO 2007*, Poznan, Poland, pp. 1764-1768, Sep 3-7, 2007.

[4] P.Y. Lau, P.L. Correia, P. Fonseca and A. Campos, "Underwater image analysis to classify benthic habitat characteristics," in *Proc. Portuguese Conf. on Pattern Recognition*, Lisbon, Portugal, 2007.

[5] D. R. Edgington, D. Walther, C. Koch, "Automated Event Detection in Underwater Video," in *Proc. MTS/IEEE Oceans 2003*, San Diego, USA, 2003.

[6] D. Walther, D.R. Edgington, and C. Koch, "Detection and Tracking of Objects in Underwater Video," in *Proc. CVPR 2004*, Washington, USA, pp. 544-549, Jun 27 –Jul 2 2004.