

Mosaics For Nephrops Detection in Underwater Survey Videos

Ken Sooknanan*, Jennifer Doyle†, Colm Lordan†, James Wilson*, Anil Kokaram* and David Corrigan*

* Trinity College Dublin

Email: sooknank@gmail.com

† Marine Institute Galway

Abstract—Harvesting the commercially significant lobster, *Nephrops norvegicus*, is a multimillion dollar industry in Europe. Stock assessment is essential for maintaining this activity but it is conducted by manually inspecting hours of underwater surveillance videos. To improve this tedious process, we propose an automated procedure. This procedure uses mosaics for detecting the Nephrops, which improves visibility and reduces the tedious video inspection process to the browsing of a single image. In addition to this novel application approach, key contributions are made for handling the difficult lighting conditions in these kinds of videos. Mosaics are built using 1-10 minutes of footage and candidate Nephrops regions are selected using image segmentation based on local image contrast and colour features. A K-Nearest Neighbour classifier is then used to select the respective Nephrops from these candidate regions. Our final decision accuracy at 87.5% recall and precision shows a corresponding 31.5% and 79.4% improvement compared with previous work [1].

I. INTRODUCTION

Nephrops norvegicus, commonly known as the Dublin Bay prawn, is a slender, pink-orange species of lobster with estimated annual landings of some 60,000 tons [2]. To maintain this multi-million dollar [3] industry, stock assessment of this particular lobster is performed yearly throughout Europe. Marine scientists currently perform this assessment by inspecting hours of underwater surveillance videos manually. This is therefore tedious, time consuming and prone to error due to fatigue, which also confuses the repeatability of the process. To improve this situation, a system for automatic Nephrops detection is proposed in this paper.

To understand the challenges involved in this problem, the images in the left column in Figure 1 show example frames from typical underwater surveillance footage. This type of recording is made by a camera mounted on a sled and dragged across the sea floor. As can be seen, the image quality is poor due to the: i) uneven lighting, ii) narrow field of view, and iii) geometrically distorted scene. The first problem is due to the need for artificial lighting which causes vignetting, and the latter two problems are caused by the lens type and the position of the camera relative to the seabed.

This new area of research has recently been tackled in the literature by authors Lau et al. [1]. In their approach, lobster detection is addressed using video, where candidate regions are detected in the gray scale images via their edges, and then classified with a decision tree framework. The challenge of uneven lighting was addressed by performing object detection in a block-based manner using local contrast features. Although

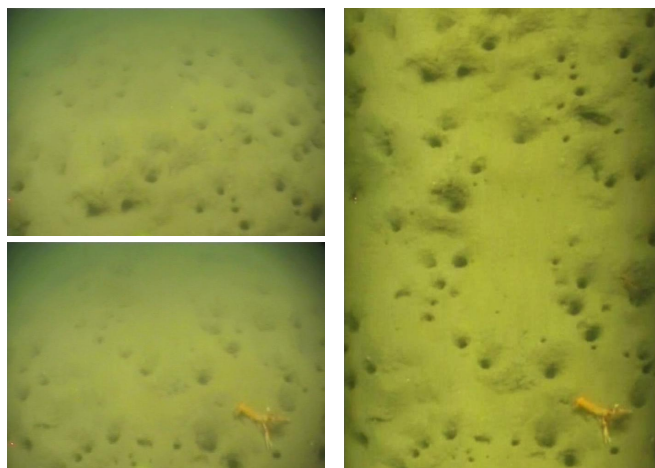


Fig. 1. Original Frames (Left) and corresponding mosaic generated using 100 frames (Right). Full test sequences and mosaics located at www.mee.tcd.ie/~sigmedia/Misc/ICIP2013

acceptable results were obtained from the test sets they used, their algorithm has four main drawbacks.

- 1) The object detection process which uses edges produces incomplete segmentations, and may not be effective on the blurry images observed in this work.
- 2) By processing the gray scale space only, many other objects on the sea floor are detected, which can decrease the overall efficiency of the system.
- 3) Using a strict set of rules, via their decision tree classification framework might restrict the system from generalizing well with other data sets.
- 4) To verify the automated results, scientists still have to inspect thousands of frames.

To improve on these drawbacks, four key contributions are introduced in this work:

- 1) Mosaics are used for detecting and summarizing the automated results, which reduces the tedious inspection process to the scanning of a single image. The mosaic generation process is specifically designed to cope with vignetting and geometric distortion in these types of underwater footage [4].
- 2) Segmentation based on the bright and pink-orange colour characteristics of the creature, are used for object detection. This procedure obtains more complete object regions than using edges only.

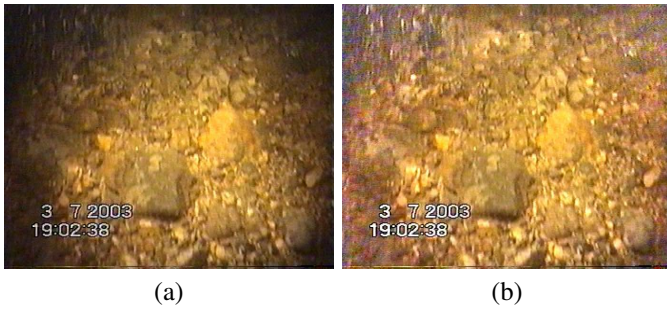


Fig. 2. a) Original, and b) Corrected images using method in [4].

- 3) A new feature set that is motivated by a current scientific description of Nephrops is proposed, which are easily understood by marine scientists. Some of these features include the diameter of the creature, which provides further statistical information relating to the size and population age of the species [5].
- 4) Supervised learning schemes are utilized to improve generalization on different data sets.

The proposed lobster recognition system is accomplished in four stages involving: i) Mosaic Generation, ii) Object detection and Grouping, iii) Feature Choice and Extraction, and iv) Classification. Each of these stages are discussed in the following sections, after which a comparison with the previous work by Lau et al.[1] is given.

II. MOSAIC GENERATION

The mosaics are created using the technique developed by Sooknanan et al. [4], as it was initially created for use in underwater videos, and is robust to these type of unevenly lit noisy images. This technique is a two stage process, involving image enhancement, followed by mosaic creation. In the image enhancement stage, the visibility is improved by correcting the radial illumination degradations in these images, as shown in Figure 2. To accomplish this task, these degradations are modeled with a two dimensional Gaussian shaped function, the parameters of which are estimated using point correspondences from consecutive frames throughout the sequence.

After this correction stage, the mosaics are created by mapping each frame, k , in the sequence to a common reference frame, r , using a transformation matrix, $T_{r,k} = \prod_{i=1}^{k-1} T_{i,i+1}$. Where $T_{i,i+1}$ is the affine global motion model between consecutive frames, which is estimated using a hybrid feature-based and exhaustive search algorithm. Once aligned, the overlapping regions among the respective frames are then rendered.

The rendering is accomplished using a two dimensional Gaussian-like weighting function that helps to further balance the uneven lighting problem by assigning higher weights to the well lit regions from each frame. This key function is estimated automatically using point correspondences, but for this application was fixed to select regions 1/5 of the bottom of the screen. This region was selected as it is the

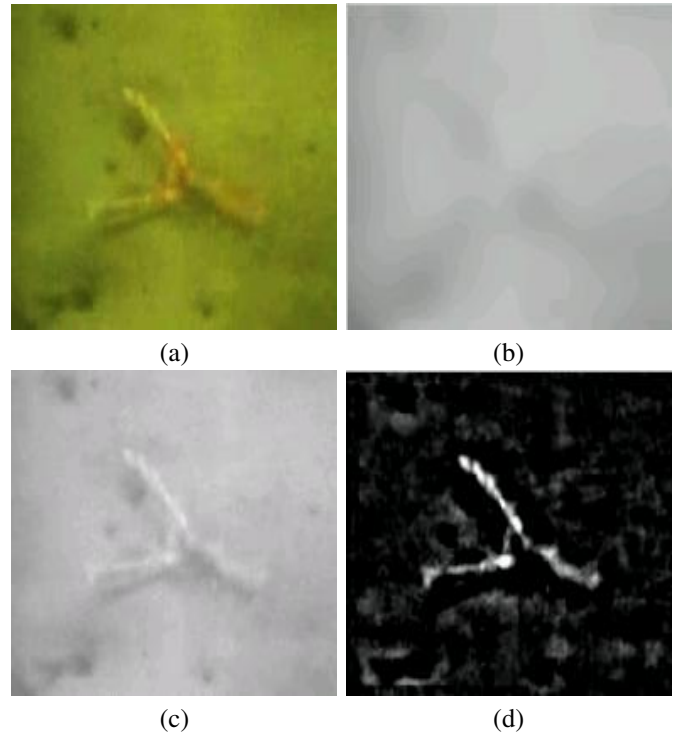


Fig. 3. Generating the bright region map of the original image (a), by subtracting a heavily blurred gray scale version (b), from a lightly blurred version (c). This highlights the local bright (candidate Nephrops) regions as local maxima regions, as seen in the bright map in (d).

designated analysis zone used by scientists in the current manual inspection procedure [6], and is well lit, thus making it ideal for generating high quality mosaics.

III. OBJECT DETECTION AND GROUPING

Once the mosaics are generated, the next step is to detect candidate Nephrop regions. A segmentation approach is used to accomplish this task because these images are usually very blurry, and detecting parts of the objects with techniques such as edges [1], might not be effective in some cases. Additionally, the scientifically important features such as the diameter of the creature is most effectively extracted from the entire region. The segmentation approach is performed by targeting the bright contrasting and pink-orange colour characteristics of Nephrops. To cope with any residual uneven lighting in these images, the bright contrasting characteristic is targeted in the Difference of Gaussians image or Bright Region Map, where the influence of absolute brightness has minimal effect. Whereas the characteristic pink-orange colour is targeted using the Hue and Saturation colour channels. This overall procedure has three main steps involving: i) generating the bright region map, ii) segmentation, and iii) grouping and labeling, which are explained in the following sections.

A. Bright Region Map Generation

To locate bright contrasting regions in the mosaic, I , a bright region map is generated as: $I_b = I * G_2 - I * G_1$. Where G_1 and G_2 are two dimensional Gaussian functions with 71 and 5 taps,

and corresponding variances of 30 and 2 respectively. Because of the large variance of G_1 , a homogeneous sandy background image is created, which when subtracted from a lightly blurred version ($I * G_2$), all of the locally bright (candidate lobster) contrasting regions are highlighted as local maxima regions. To obtain larger maxima values and hence improve detection, gamma correction is performed on the original image, $I = I^\gamma$, where $\gamma = 1.5$ is used, prior to the generation of I_b . Figure 3 illustrates the generation of this bright region map.

B. Segmentation

Candidate Nephrop regions are now obtained by performing segmentation on the i) bright region map, $I_b(\mathbf{x})$, and the ii) Hue, $I_h(\mathbf{x})$, and iii) Saturation, $I_s(\mathbf{x})$, colour channels. A two layer segmentation map $L(\mathbf{x})$ is estimated in which the labels are defined as:

$$L(\mathbf{x}) = \begin{cases} 1 & \text{Nephrop regions} \\ 0 & \text{Homogenous sandy background regions} \end{cases}$$

This segmentation is accomplished in a Bayesian framework where the Maximum A Posteriori (MAP) estimate for $L(\mathbf{x})$ is generated by maximizing the posterior: $p_o(L(\mathbf{x}) = \alpha | I_b(\mathbf{x}), I_h(\mathbf{x}), I_s(\mathbf{x}), \neg L(\mathbf{x}))$. Where $\neg L(\mathbf{x})$ is the respective 3×3 neighborhood pixel labels of image position \mathbf{x} . Factorizing this posterior term using Bayes Law [7], and dropping the notation \mathbf{x} for clarity, gives:

$$p_o(L = \alpha | I_b, I_h, I_s, \neg L) \propto p_b(I_b, |L = \alpha) p_h(I_h, |L = \alpha) \propto p_s(I_s, |L = \alpha) p_r(L = \alpha | \neg L) \quad (1)$$

where $\{p_b, p_h, p_s\}$ and p_r are the likelihood and prior terms. The likelihoods are assumed to be Gaussians as follows:

$$p_b(I_b | L = \alpha) \propto \exp - \left[\frac{(I_b - \hat{I}_{b,\alpha})^2}{2\sigma_{b,\alpha}^2} \right]$$

$$p_h(I_h | L = \alpha) \propto \exp - \left[\frac{(I_h - \hat{I}_{h,\alpha})^2}{2\sigma_{h,\alpha}^2} \right]$$

$$p_s(I_s | L = \alpha) \propto \exp - \left[\frac{(I_s - \hat{I}_{s,\alpha})^2}{2\sigma_{s,\alpha}^2} \right]$$

where $\alpha = \{0, 1\}$, and $\{\hat{I}_{b,0}, \hat{I}_{h,0}, \hat{I}_{s,0}\}$ and $\{\hat{I}_{b,1}, \hat{I}_{h,1}, \hat{I}_{s,1}\}$ are the characteristic mean brightness, hue and saturation values of the background and Nephrops regions respectively, and $\{\sigma_{b,\alpha}^2, \sigma_{h,\alpha}^2, \sigma_{s,\alpha}^2\}$ are their corresponding variances. To enforce spatial smoothness within these segmentations, a Gibbs energy function [8], with a 3×3 pixel neighborhood, is used for the prior, $p_r(\cdot)$, given by:

$$p_r(L(\mathbf{x}) = \alpha | \neg L) \propto \exp - \left[\Lambda \sum_{k=0}^7 \lambda_k |\alpha - L(\mathbf{x}_k)| \right] \quad (2)$$

where $\lambda_k = 1/||\mathbf{x} - \mathbf{x}_k||$, is a scalar weight that is inversely proportional to the distance between the current image site \mathbf{x}

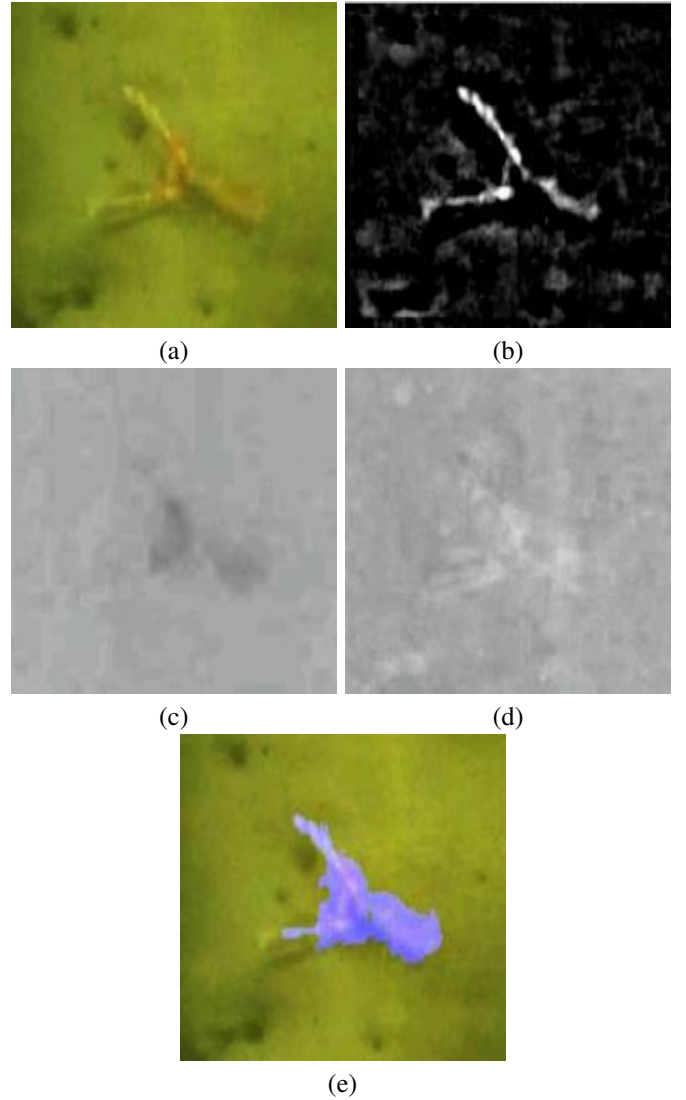


Fig. 4. Segmenting Nephrop in (a) Original image, by combining the (b) bright region map with the (c) Hue, and (d) Saturation channels, to obtain the final result (blue) given in (e).

and the respective neighbor \mathbf{x}_k in a 3×3 neighborhood, and Λ is a global weighting factor, that was set as $\Lambda = 1$ in these experiments.

Initialization. Good initial estimates for each parameter are obtained after analyzing the ground truth Nephrop regions. In detail, $\{\hat{I}_{b,0}, \hat{I}_{h,0}, \hat{I}_{s,0}\}$ and $\{\hat{I}_{b,1}, \hat{I}_{h,1}, \hat{I}_{s,1}\}$ were set to $\{100, 0.16, 0.65\}$ and $\{0, 0.17, 0.60\}$, and the corresponding variances were set as $\{\sigma_{b,0}^2, \sigma_{h,0}^2, \sigma_{s,0}^2\} = \{\sigma_{b,1}^2, \sigma_{h,1}^2, \sigma_{s,1}^2\} = \{1/|I_{b,0} - I_{b,1}|^2, 1/|I_{h,0} - I_{h,1}|^2, 1/|I_{s,0} - I_{s,1}|^2\}$. Using these settings, the posterior, p_o is optimized using the Iterated Conditional Modes scheme [9], where a checkerboard scan is utilized until there are no further changes in the labels or a maximum of 10 iterations are completed. Sample results obtained using this segmentation procedure are shown in Figure 4.

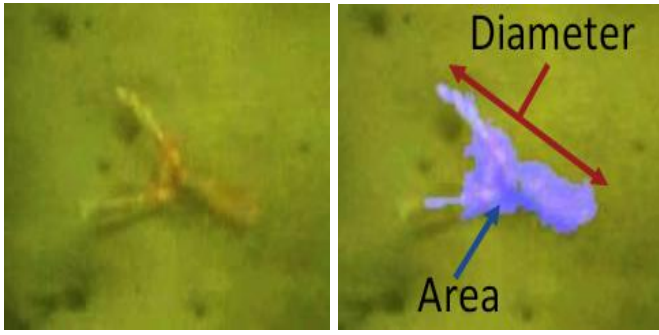


Fig. 5. Original (Left), and Segmentation (Right) showing the extraction of the Area and Diameter features.

C. Labeling

The candidate Nephrop regions, $L(\mathbf{x}) = 1$, that are locally connected are now grouped together and labeled with unique identification numbers. To perform this task, the Connected Component Analysis technique by Sammet et al. [10], with a 3×3 neighborhood, is used.

IV. FEATURE CHOICE AND EXTRACTION

In practice a large percentage of the objects detected are not Nephrops. Most of these other objects are attributed to the laser dots used in these surveys for calibration, which are also bright and orange-pink in appearance, as shown in Figure 6. To eliminate these false alarms, additional characteristic features of Nephrops (other than their colour and bright appearance used in the segmentation step) had to be examined. The choice of these features were derived from inspecting the ground truth data, which showed that most of the Nephrops regions have similar sizes and shapes. To use these similarities, along with the knowledge of the laser dot regions, four new features are developed for use in this application that explore the size, shape, and presence of laser dots in the segmented regions. Each of these features are described as follows:

Size

The size of the object are examined using the Area and Diameter features.

Area (a). This feature is important for eliminating small noisy regions. It is extracted as the number of pixels in the segmented region, and then scaled by 10 cm^2 (1000 pixels), to range typically between 0 and 1.

Diameter (d). Apart from its usage in classification, this feature provides further statistical information for scientists relating to the age of the creature [5]. It is extracted as the maximum distance between any two pixels in the segmented region and then scaled by 1 cm^2 (100 pixels), to range typically between 0 and 1.

Shape

For eliminating circular shaped objects such as the frequently detected laser dot regions, and for also describing the shape of the Nephrops regions, the Eccentricity shape feature is used.

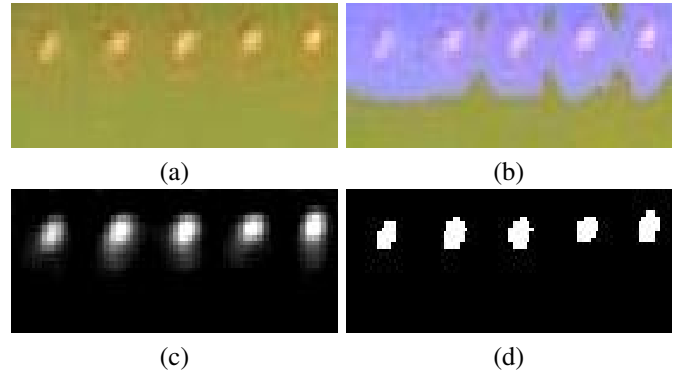


Fig. 6. Extraction of laser dots. (a) Original, (b) segmentation obtained from proposed system, (c) bright region map, and (d) detected laser regions.

Eccentricity (e). This value describes how elliptically shaped the region is, with a value of 0 representing a perfect circle and 1 corresponding to a straight line segment respectively. It is obtained by fitting an ellipse to the region. This is achieved using the relationship derived in [11] where the coefficients of an ellipse are equated to the first and second order moments of the respective region.

Presence of Laser Dots

Laser Dots (l_d). To identify the laser dot objects, each region is searched for the presence of these round intense dots, using two steps. First, to detect intense bright regions, k-means clustering using 3 clusters is performed on the bright region map of the particular object (initial centroid values of $\{0, 50, 120\}$ were used). Then small circular objects are identified as regions from the cluster with the largest intensity value that have eccentricity and diameter features within 50% of the characteristic laser dot values of $\{e = 0.4, d = 1 \text{ cm}(10 \text{ pixels})\}$. This feature is extracted as the quantity of these laser dots. Figure 6 illustrates this procedure.

V. CLASSIFICATION

The last stage in this recognition system is to classify the detected objects into lobster and non-lobster classes using the respective features extracted. To cater for the large diversity in size and shape features of Nephrops, the use of a well established supervised learning classification scheme, the K-Nearest Neighbor (KNN), is explored. This non-parametric classifier is used because of its simplicity. The key advantage this scheme offers in comparison to the previously used Decision Tree scheme developed by Lau et al. [1], is that it incorporates the use of training data into its classification procedure. The use of this data not only allows the system to identify a large variety of Nephrops, but also facilitates easy adoption to new data sets.

Ground truth data for these experiments were obtained from twenty mosaics, each generated from a 2,300 frame (576×712) sequence of actual underwater surveillance video. Each mosaic had approximately 2 Nephrops and 70 non-Nephrops, which were labelled manually by a trained expert. Ten of these mosaics were used as training data, and the other

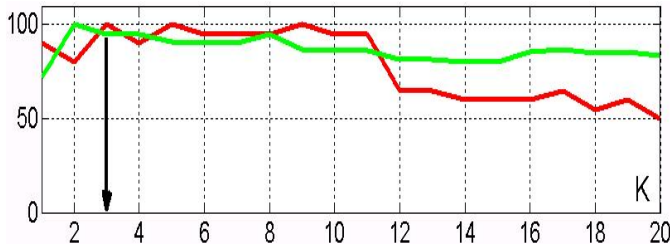


Fig. 7. Recall (red) and Precision (green) from KNN for different neighborhood, k values. Black arrow indicates highest recall and precision average

ten for obtaining the optimal neighborhood value, k . This optimization is accomplished by analyzing the performance of the KNN with k values ranging from 1 to 20. Figure 7 illustrates these results. For our experiments, the neighborhood value was set as $k = 3$, as this value gave the highest recall and precision average of 80%.

VI. RESULTS

The performance of the proposed system was compared to the previous state of the art video-based technique proposed by Lau et al. [1] using: i) video and ii) mosaics. The video analysis is performed by manually cross referencing the classified objects obtained in each frame with the corresponding ground truth mosaic. Ten of the twenty ground truth mosaics collected for this analysis (that were not used for training) were used to perform this comparison. These sequences contained a total of 22 Nephrops and 756 non-Nephrops regions. The recall and precision values obtained from each system are given in Figure 8.

Analysis of these results show the proposed system achieves superior results to the previous method using both video and mosaics. In detail, the proposed system obtained average recall and precision values of 87.5% and 87.5%. These values were 31.5% and 79.4% higher than the average recall and precision values obtained from the previous method using video, and 58.1% and 81.2 % higher when using mosaics. The degradation in performance in the previous method when using mosaics is mainly attributed to the absence of the four-frame object consistency step in their algorithm. This step could not be performed with mosaics as they are only single images, and as a result additional spurious objects due to noise were detected. These results verify that it is not only possible to use mosaics to detect Nephrops, but improved results can be achieved using this proposed technique compared to the previous method.

The abnormal results seen in test mosaic-4 of 0% in recall and precision values were obtained because this test video only had one Nephrop that was missed by both systems as it was too blurry. In these missed cases, the Nephrops bright and pink-orange colour characteristics were too low for detection, as shown in Figure 9. On the other hand, objects with similar appearance features as Nephrops were detected as false alarms, such as blurry laser dot regions and Sea pens, as shown in Figure 10. Analysis of the correctly detected Nephrops show

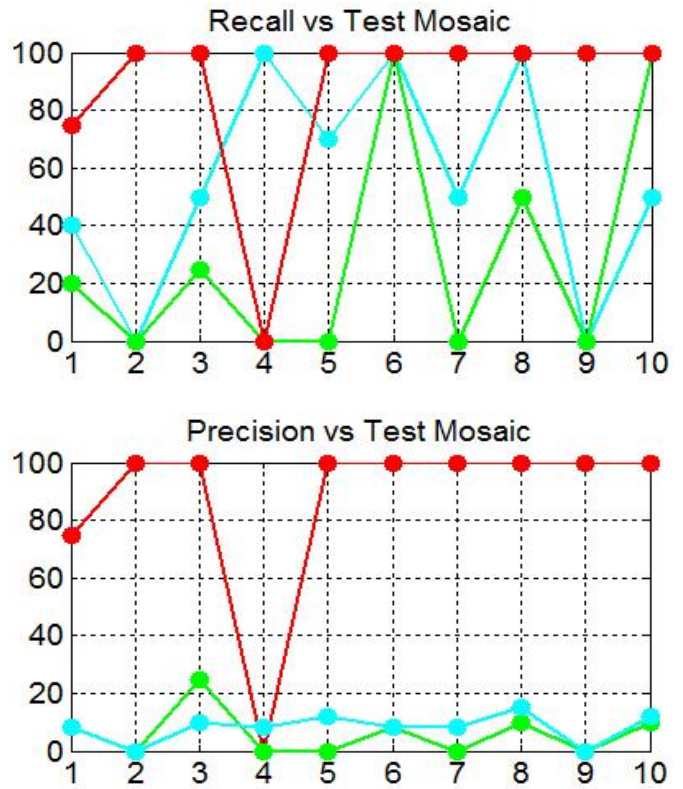


Fig. 8. Recall (top) and Precision (bottom) from the proposed (red) and previous system by Lau et al. [1], using video (blue) and mosaics (green)

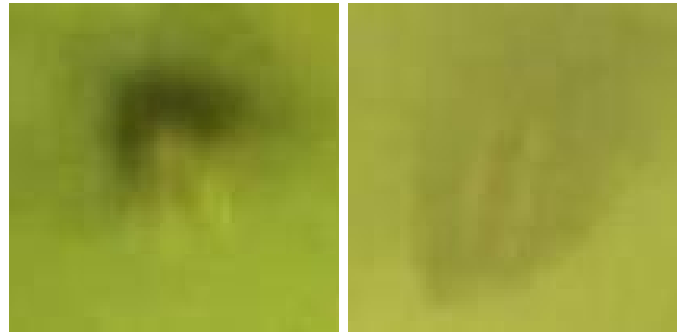


Fig. 9. Examples of missed Nephrops by both methods because of low contrast.

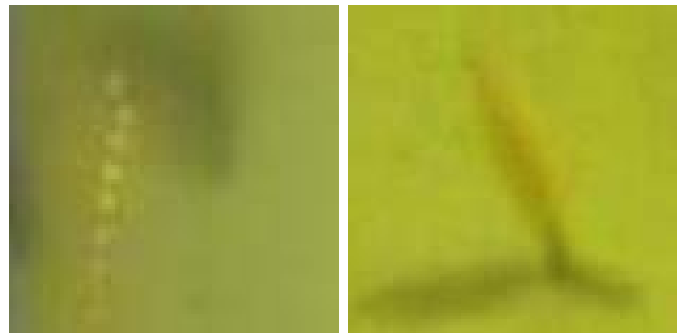


Fig. 10. False alarms obtained from both the previous method of Lau et al. [1], and the proposed method.

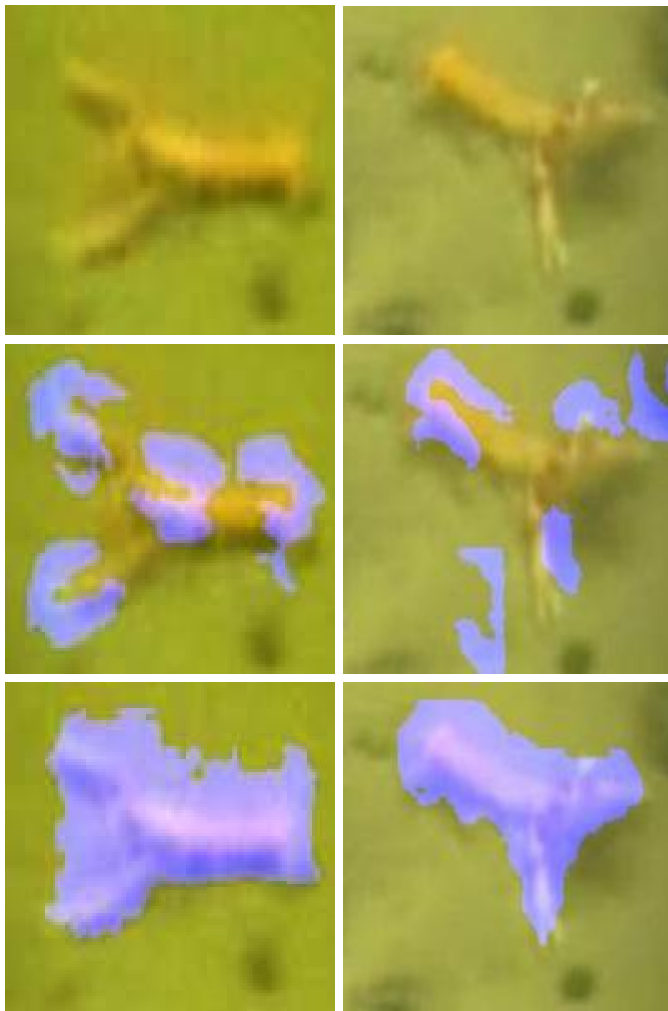


Fig. 11. Examples of correctly detected Nephrops. The top row are the original images, the middle row are the segmentations obtained by Lau [1], and last row are the segmentations obtained using the proposed method. As seen the proposed method obtains most of the object regions.

that more complete regions were obtained compared to the previous method by Lau et al. [1], which were very noisy, as shown in Figure 11.

VII. CONCLUSION

In this paper a new technique for detecting Nephrops in marine surveillance videos is presented. This technique improves substantially on the previous state of the art method introduced by Lau et al. [1], as a result of four key contributions:

- 1) Mosaics are used for performing object recognition, which improves visibility and reduces the tedious video inspection process currently performed by scientists to the browsing of a single image.
- 2) A novel segmentation technique is developed for object detection which targets bright and pink-orange regions that are characteristic of Nephrops.
- 3) New features are introduced, which are motivated by the current scientific description of Nephrops, and are thus

easily understood by marine scientists.

- 4) Supervised learning schemes are utilized to improve generalization on different data sets.

From the test data examined, the proposed system obtained a high recall and precision average value of 87.5%. This high value shows that it is possible to use mosaics to detect Nephrops in underwater surveillance videos. As a practical point of interest, when the scientists from the Marine Institute Galway were shown these results they agreed that this algorithm has the potential to assist with their current manual analysis procedure.

REFERENCES

- [1] P. Y. Lau, P. L. Correia, P. Fonseca, and A. Campos, "Estimating norway lobster abundance from deep-water videos: an automatic approach," *IET Image Processing*, vol. 6, pp. 22–30, 2012.
- [2] N. Campbell, H. Dobby, and N. Bailey, "Investigating and mitigating uncertainties in the assessment of scottish nephrops norvegicus populations using simulated underwater television data," *ICES Journal of Marine Science*, vol. 66, pp. 646–655, 2009.
- [3] Seafood Industry Value Chain Analysis of Cod, Haddock, and Nephrops, *KPMG AS, Centre for Aquaculture and Fisheries, and Sea Fish Industry Authority*, Fjordgaten 68, 7010 Trondheim, Finland, March 2004.
- [4] K. Sooknanan, A. Kokaram, D. Corrigan, G. Baugh, J. Wilson, and N. Harte, "Indexing and selection of well-lit details in underwater video mosaics using vignetting estimation," in *IEEE International Conference on Oceans (OCEANS 2012)*. MTS/IEEE, May 2012, pp. 1–7.
- [5] S. J. Marrs, R. J. A. Atkinson, C. J. Smith, and J. M. Hills, "Calibration of the towed underwater tv technique for use in stock assessment of nephrops norvegicus. final report to the european commission contract 94/069. study project in support of the common fisheries policy (xiv/1810/c1/94).," Tech. Rep., 1996.
- [6] Report of the Workshop and training course on Nephrops burrow identification, *International Council for the Exploration of the Sea (ICES) Living Resources Committee*, H. C. Andersens Boulevard 44-46, DK-1553 Copenhagen V., Denmark, ICES CM 2008/lrc:03 edition, Feb. 2008.
- [7] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley-Interscience, NY, 2nd edition, 2001.
- [8] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 6, no. 6, pp. 721–741, 1984.
- [9] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. Vol. 48, no. No. 3. (1986), pp. 259–30, 1986.
- [10] H. Samet and M. Tamminen, "Efficient component labeling of images of arbitrary dimension represented by linear bintrees," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, no. 4, pp. 579–586, 1988.
- [11] R. Haralick and L. Shapiro, *Computer and Robot Vision*, Addison-Wesley, Boston, 1st edition, pp. 620-660, 1992.