Mosaics For Burrow Detection in Underwater Surveillance Video

Ken Sooknanan*, Jennifer Doyle†, James Wilson*, Naomi Harte*, 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 the use of mosaics for the automated detection of burrows on the seabed. We present novel approaches for handling the difficult lighting conditions that cause poor video quality in this kind of video material. Mosaics are built using 1-10 minutes of footage and candidate burrows are selected using image segmentation based on local image contrast. A K-Nearest Neighbour classifier is then used to select burrows from these candidate regions. Our final decision accuracy at 93.6% recall and 86.6% precision shows a corresponding 18% and 14.2% 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. The assessment is based on the species burrow density distribution. Marine scientists estimate this by inspecting hours of underwater surveillance videos manually with mechanical tally counters. 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 and also increase repeatability, a system for automatic burrow detection is needed.

To understand the challenge involved, the images in the left column in Figure 1 show exampleS of frames from typical underwater surveillance footage. This type of recording is made by a camera mounted on a sled dragged across the seabed floor. The burrows are the dark circular holes in the otherwise smooth seabed floor. As can be seen, the image quality is poor due to uneven lighting, and the available field of view is quite narrow as well as geometrically distorted. The former problem is due to the need for artificial lighting which causes vignetting, and the latter problem is caused by the lens type and the orientation 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, burrow detection is addressed using video, where candidate regions are segmented, classified (with a decision tree) and tracked through frames. The challenge of uneven lighting was addressed by performing object detection in a block-based manner using local contrast features. Their framework enforces temproal consistency on the segmentations of individual frames to eliminate false alrams. Although acceptable results are obtained with this approach, its verification still involves manual inspection of many thousands of video frames.

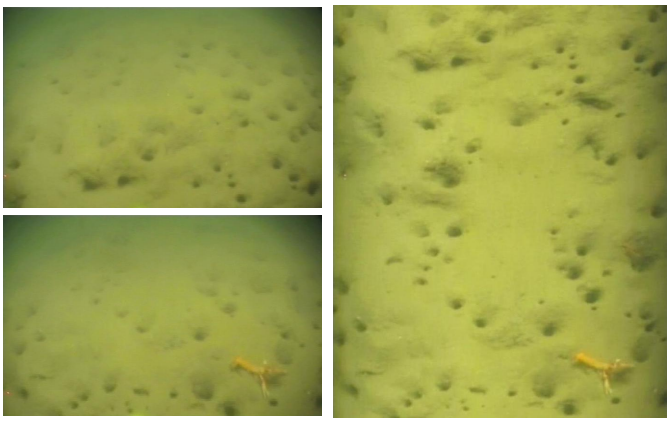


Fig. 1. Original Frames (Left) and corresponding mosaic generated using 100 frames (Right). More examples found at www.mee.tcd.ie/∼sigmedia/Misc/OCEANS2013

In our work, the problems of limited field of view and geometric distortions were indirectly resolved by using mosaics of several video frames. The mosaic generation process is specifically designed to cope with vignetting and geometric distortion of underwater footage and was presented previously in Sooknanan et al [4]. The right column of Figure 1 shows an example of a mosaic created from four seconds of footage. To cope with uneven lighting, our method eliminates the influence of absolute brightness by performing object detection in a contrast space created by taking the difference of two blurred versions of the image. The detected objects are segmented using a classical energy minimisation approach, and lastly are classified using a K-Nearest Neighbour classification framework.

Three key contributions are made in this paper. First, is the use of mosaics for this application, which improves visibility, reduces processing, and simplifies the tedious video inspection process to browsing of a single image. The second key

This work was supported by the Science Foundation Ireland PI Programme: SFI-PI 08/IN.1/I2112

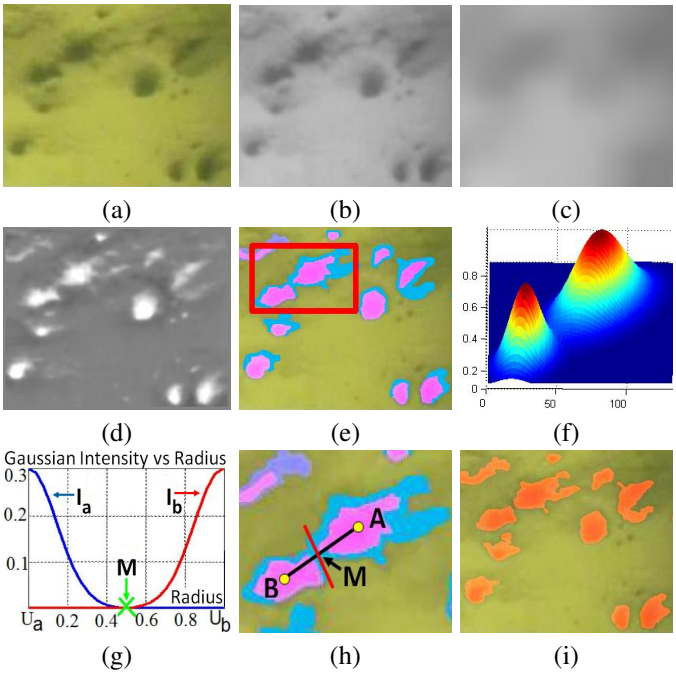


Fig. 2. a) Original, blurred b) lightly, and c) heavily grayscale images, d) DoG, e) dark entrances (pink) and scrapings (blue), f) Gaussian mixture model of composite region (red box), with g) component plots showing splitting point M, and h) separating line (red), and i) Our final segmentation.

contribution is a shape modelling step in our segmentation procedure for separating composite burrow regions. Lastly, we introduce a unique feature set for this application, which is motivated by a scientific description of lobster burrows [5]. As a result, marine biologists can easily relate to it.

In the next section we present a brief summary of the mosaic generation process and then describe our burrow segmentation, feature extraction and classification procedures. Comparison with previous work is shown in section 6, and conclusions in section 7.

II. Mosaic Generation

The mosaics are created by mapping frames to a common reference frame, Q_r, using the technique developed by Sooknanan et al.[4]. This mapping is done on a frame-by-frame basis where a transformation matrix, $T_{r,k}$, relating Q_r to frame k, is calculated as $T_{r,k} = \Pi_{i=1}^{k-1} T_{i,i+1}$. Where $T_{i,i+1}$ is the affine global motion model between consecutive frames, estimated using a hybrid feature-based and exhaustive search algorithm.

Overlapping regions among aligned frames are then combined using a two dimensional Gaussian-like weighting function that balances uneven lighting in the generated mosaic by selecting well lit regions from each frame. This key function is estimated automatically using point correspondences in [4], but for our application was fixed to select the lower 20% of the screen as this is the designated analysis zone used by scientists in the current inspection procedure [5]. This location also had minimal geometric distortion and was well lit, thus making it ideal for generating high quality mosaics.

III. Burrow Segmentation (BS)

The second stage of our application is to detect and segment candidate burrow regions in the generated mosaic, I. This was accomplished in four main steps of: i) generating the Difference of Gaussians (DoG) image, ii) performing segmentation, iii) Labelling, and iv) Splitting merged regions.

DoG Generation. The DoG image, I_d was generated as: $I_d = I * G_1 - I * G_2$, where G_1 and G_2 are two dimensional Gaussian functions with taps of 71 and 5, and corresponding variances of 30 and 2 respectively. Because of the large differences in G_1 and G_2, local minimum regions (i.e. burrows) appear as local maximum regions in I_d, even those in poorly lit areas, as shown in Figure 2 (a) to (d). To obtain larger maxima values and hence improve detection, gamma correction was performed on the original image, $I = I^{\gamma}$, where $\gamma = 1.5$ was used, prior to the generation of I_d. Since we are only interested in these maxima (candidate burrow) regions, I_d was further simplified by setting all negative values to zero.

Segmentation. Burrows have a characteristic dark entrance region ($L(\mathbf{x}) = 2$) surrounded by a lighter region ($L(\mathbf{x}) = 1$) created by the animal claw scrapings as it manoeuvres in and out of the hole. To capture useful features from these regions, they were each segmented and labelled differently from the homogenous sandy background regions ($L(\mathbf{x}) = 0$).

Proceeding in a Bayesian fashion we require to maximise the posterior, $p_o(L(\mathbf{x}) = \alpha | I_d(\mathbf{x}), \neg L(\mathbf{x}))$ w.r.t. $I_d(\mathbf{x})$ and $\neg L(\mathbf{x})$, the respective 3×3 neighbourhood pixel labels of image position $\mathbf{x}$. Factorising the posterior using Bayes' Law, and dropping the notation $\mathbf{x}$ for clarity:

$$p_o(L = \alpha | I_d, \neg L) \propto p_k(I_d | L = \alpha) p_r(L = \alpha | \neg L) \qquad (1)$$

where p_k and p_r are the likelihood and prior. The likelihood terms are modelled with gaussians as:

$$p_k(I_d | L = \alpha) \propto \exp - \Big[\frac{(I_d - I_\alpha)^2}{2\sigma_\alpha^2} \Big] \qquad (2)$$

where $\alpha = \{0, 1, 2\}$, and $\{I_0, I_1, I_2\}$ are the mean values of the background, scrapings and dark entrace regions respectively, and $\{\sigma_0^2, \sigma_1^2, \sigma_2^2\}$ are their corresponding variances. To enforce spatial smoothness within our segmentations, Gibbs energy functions [6], with a 3×3 pixel neighbourhood, are used for the priors, $p_r(.)$, as:

$$p_r(L(\mathbf{x}) = \alpha | \neg L) \propto \exp - \Big[\Lambda \sum_{k=0}^{7} \lambda_k |\alpha - L(\mathbf{x}_k)| \Big] \qquad (3)$$

where $\lambda_k = 1/||\mathbf{x} - \mathbf{x_k}||$, is a weight inversely proportional to the distance between the current site $\mathbf{x}$ and the respective neighbour $\mathbf{x_k}$ in a 3×3 neighbourhood, and Λ is a global weighting factor, set as $\Lambda = 1$ in our experiments. Initial estimates for these parameters and labels were obtained using k-means clustering on I_d, with 4 clusters to represent the different levels of burrow shaddings and the background.

The background and scrapings regions were labelled with cluster members associated with the first and second smallest

(a) (b) (c)

Fig. 3. a) Original, and segmentation showing extraction of b) Entrance Dark Area, Claw Marks, c) Burrow Diameter, Rectangularity, and Circularity Fit features.

centroid values respectively, and the dark entrance regions with the other two cluster members. The parameters $\{I_0, I_1, I_2\}$ were set using the three smallest centroid values, $\{w_1, w_2, w_3\}$ as $\{w_1, (w_1 + w_2)/2, (w_2 + w_3)/2\}$, while $\{\sigma_0, \sigma_1, \sigma_2\}$ were set as $1.5\{(I_1+I_0), (I_1+I_0), (I_1+I_2)\}$. Minization of p_o was performed using the Iterated Conditional Modes [7] scheme, where a checkerboard scan is utilized until there is no further change in labels or a maximum of 10 iterations completed.

Candidate Burrow Labelling: The locally connected scrapings and dark entrance regions, $L(\mathbf{x}) = \{1, 2\}$, of each candidate burrow are then labelled with unique identification numbers. The Connected Component Analysis technique by Sammet et al.[8], with a 3×3 neighbourhood, was used to perform this labelling. In practice however, burrows in close proximity to each were often segmented together, as illustrated in the red box in Figure 2 (c). These composite segmentations are identified as objects with multiple dark entrance regions.

Splitting merged regions: In the last stage of our segmentation algorithm these composite regions are split into their individual components. This was accomplished in two steps. First, the shape of the entire region is modelled with a mixture of Gaussians equal to the number of dark entrance regions. Then, each component is separated normally from its neighbour, at the point where their local mixing weights are equal, along the line joining their respective means. This procedure is illustrated in Figures 2 d) and e).

These mixture parameters are optimized using the Expectation Maximization algorithm [9]. In this algorithm, the mean, μ, and variance, σ^2, of each component are initialized as $\mu = \sum \mathbf{x} I_n(\mathbf{x}) / \sum \mathbf{x}$, and $\sigma^2 = \sum (\mathbf{x} - \mu) I_n(\mathbf{x})$. Where $I_n(\mathbf{x})$ is the normalized intensity of the corresponding dark entrance area region component, I_g, at image position $\mathbf{x}$, given by: $I_n(\mathbf{x}) = I_g(\mathbf{x}) / \sum I_g(\mathbf{x})$. This separation was performed on the DoG image, as in this domain the intensity profile of the holes are Gaussian-like i.e. their intensity profile decreases from the center, whereas in the raw image it is the opposite (see figure 2 b).

IV. Feature Extraction

In practice a large percentage of the segmented regions obtained were not burrows. To reject these false alarms, a unique feature set for distinguishing burrows was developed. This set was based on the characteristic burrow features which marine scientists use in the current identification process. Extraction of some of these features is shown in Figure 3, and can be explained as follows:

1) Entrance Dark Area (a): Burrows with caved-in entrances are deemed inactive and consequently not counted. Their key distinguishing characteristic is the absence of their entrance dark core region. This feature was extracted as the number of pixels in the entrance dark region, a (obtained from our segmentation step). However, surveys are recorded with different cameras, so to maintain consistency, this feature is scaled using the calibrated distance ($80cm$) between two laser dots present in these videos, to range typically between 0 and 1.

2) Burrow Diameter (d): The diameter of a burrow is defined as the largest diagonal along its opening [10], and helps scientists to identify the species of the occupant [5]. It was extracted as the maximum distance between any two pixels in the core dark region, and then converted to metric form (cm) using the same scheme as for a.

3) Claw marks (s): Displaced sediment due to species activity is commonly present in burrows. It manifests as a brigher region surrounding the entrance. This feature was extracted as the percentage of the region area outside the dark core region.

4) Entrance Dark Region Shape (r, M, e, c): Burrows of various species have characteristic shapes. Thus we propose four shape features for the classifier, which were extracted from the dark core region. First, the rectangularity, r, which was the ratio of region area and the area of its minimum bounding rectangle. Second, was the seven moments, $M = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7\}$, as detailed in [11]. Third was the eccentricity, e, describes how elliptical in shape the region is, with a value of 0 representing a perfect circle and 1 corresponding to a line segment respectively. It was obtained by fitting the region to an ellipse, as detailed in [12]. Lastly, the circularity fit, c, was the percentage of region area within a circle of radius, $r^2 = a/\pi$, positioned at the region center of mass (see Figure 3).

V. Classification

After the features are extracted, each segmented region is classified using a K-Nearest-Neighbour (KNN) classifer [13]. This classifier works by assigning query objects to the class of objects that occur most frequent among their respective k-nearest neighbours. The search metric used is the Euclidean distance, z, among the n features between the query object, $\mathbf{q}$, and each object from the training data set, $\mathbf{t}$, given by: $z^2 = \sum_n (\mathbf{q_n} - \mathbf{t_n})^2$.

Training data for the two KNN classes used, burrows and non-burrows, were obtained from two mosaics, each generated

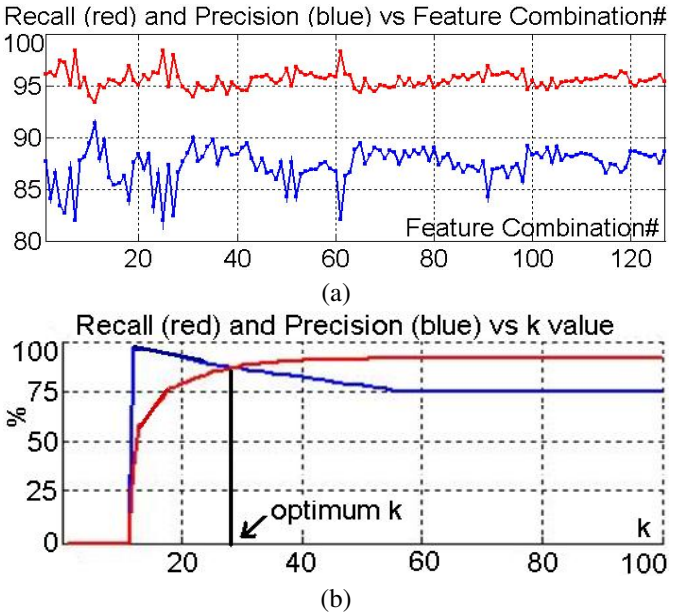


Fig. 4. Recall (red) and Precision (blue) results from training mosaic-2 for a) All Feature combinations, b) different k values

from a 2,300 frame (576×712) sequence of actual underwater surveillance video. Each mosaic had approximately 750 burrows and 1000 non-burrows, which were labelled manually by a trained expert. One of these mosaics was used asa training data for the KNN, while the other was used to obtain the optimal feature set and neighbourhood value. The optimal feature set was found by analysing the performance of the classifier with all 127 combinations of the seven features. This evaluation was performed using a Neighbourhood value of $k = \sqrt{n_b} = 27$ (as recommended by Duda et al.[13]), where $n_b = 750$, is the number of objects of interest (i.e. burrows) in the training set.

The recall and precision values obtained from this feature analysis is shown in Figure 4 (a). As seen, these results fluctuated narrowly around average recall and precision values of 96.1% and 86.4% respectively. As these fluctuations were so small, to aid in this analysis, results obtained were ranked in order of their recall and precision values averaged, $A_r = (recall + precision)/2$. Table I lists the results obtained from the individual features, along the three worst and ten best performing combinations, based on this ranking system.

The analysis of these results yielded three interesting points. First, from the individual feature analysis, the dark entrance area feature performed the best ($A_r = 91.9$), and the circulatity fit feature was the worst ($A_r = 90.0$). Secondly, when the circularity fit feature was however combined with dark entrance area feature, it performed the best ($A_r = 91.9$) among the 127 combinations, whereas the eccentricity and moments combination produced the worst ($A_r = 89.8$) results. Lastly, the dark entrance area feature seemed to be the most valuable from the entire set, as it appeared in each of the top seven combinations; and its performance was only 0.5 less

Analysis	Rank	Features	Recall	Precision	A_r
Individual	110	a	96.1	87.7	91.9
	4	s	96.2	84.1	90.2
	24	d	95.9	86.5	91.2
	11	e	97.4	83.4	90.4
	2	c	97.2	82.7	90.0
	18	r	95.0	86.9	90.1
	7	M	98.4	82.0	90.2
Worst	1	e,M	96.3	83.3	89.8
	2	c	97.2	82.7	90.0
	3	c,M	98.0	82.4	90.2
Best	118	a,s,d,e,c,r	95.3	88.7	92.0
	119	d,e	95.6	88.4	92.0
	120	a,s,d,e,c,r,M	95.4	88.7	92.1
	121	a,d,e,c,r,M	95.8	88.3	92.1
	122	a,d,e,r,M	95.7	88.4	92.1
	123	a,d,M	95.2	88.9	92.1
	124	a,d,c,M	95.2	89.0	92.1
	125	a,s,c	94.6	90.0	92.3
	126	a,M	94.8	89.7	92.3
	127	a,c	93.3	91.4	92.4

TABLE I
THE PERFORMANCE OF I) INDIVIDUAL FEATURES, ALONG WITH II) 3 WORST, AND 10 BEST COMBINATIONS FROM 4 (A).

in A_r value to the best combination. For our experiments, the combination, $\{a, c\}$, was used, as apart from having the highest A_r value, it also gave the best precision results, which is crucial in this application, as it usually has about 200% more non-burrow objects than burrows.

Using the best combination of (a, c), the optimal best neighbourhood value, k was then obtained by analysing the performance of the KNN classifier with k values ranging from 1 to 100. Figure 4 (b) illustrates these results. For our experiments, k was set as 26, for two reasons. First, it gave both high recall and precision values of 88.2% respectively. Secondly, smaller values resulted in a sharp drop in recall, whereas larger values gave a sharp drop in precision.

VI. RESULTS

We evaluated the performance of our algorithm with eight 1500-3000 frame (576×712) sequences of real underwater surveillance video. Each sequence was labelled manually in the corresponding mosaic by a trained expert. During this process the original frames were visually inspected to ensure the integrity of each mosaic, which proved to be accurate in all cases. Applying our technique to this test data (see table II), high recall and precision results were obtained, averaging 93.6% and 86.6% respectively, as illustrated in Figure 5. The small drop in recall values in the first three test sequences was obtained due to the significant amount of undetected blurry small burrows in these cases. These results show our application can detect burrows using mosaics with a high degree of accuracy.

TS	1	2	3	4	5	6	7	8
GT	524	712	632	715	954	1013	842	578

TABLE II
TEST SEQUENCES (TS) GROUND TRUTH (GT) BURROW NOS.

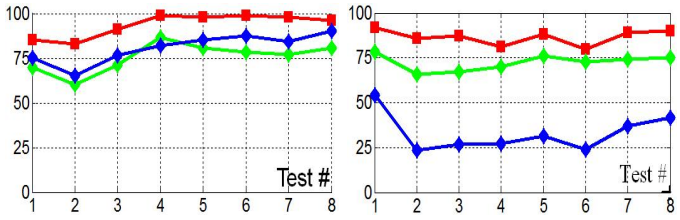


Fig. 5. Recall (Left) and Precision (Right) vs test mosaic number using our method (red), Lau[1] with video (green), and Lau[1] with mosaics (blue).

We compared results with the video-based technique proposed by Lau et al.[1] using: i) video and ii) mosaics. The video comparison was performed by cross referencing the detected burrows in each frame of the video with our ground truth mosaics. To perform the second comparison, we used the first stage of their algorithm applied directly to mosaics (the temporal constancy (i.e. 2nd) stage is not applicable in this case). Sample results obtained from test mosaic 1, are shown in Figure 6, while the recall and precision values obtained from each case are illustrated in Figure 5.

Analysis of these results show our method outperforming this previous technique in each case. For the video test, Lau et al.[1] obtained average values of 75.6% and 72.4% for recall and precision. For the mosaic tests, our modification of their technique obtained average recall and precision values of 80.75% and 33.1%. This drastic drop in precision values obtained in the mosaic tests with their algorithm highlights its high dependency on temporal information. In contrast our techique obtained high precision and recall values across all sequences tested.

VII. Conclusion

We have presented a novel tecnnique for burrow detection in video using mosaics that improves substantially on previous work. In addition, we have introduced an optimal segmentation technique and a novel feature set for this application and shown that it can achieve improved results without explicit temporal information. Our process also estimates diameter of burrows, which marine biologists agree is a useful tool for identifying [5], and analysing the size distribution of various species [10]. With initial detection rates as good as these obtained, this algorithm has the potential to significantly reduce the lengthy manual inspection time spent by the experts. In future work we would like to incorporate this method into a larger project geared towards identifying Nephrops burrows and their corresponding clusters [2].

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] Report of the Workshop and training couse 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.

[6] 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.

[7] 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.

[8] 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.

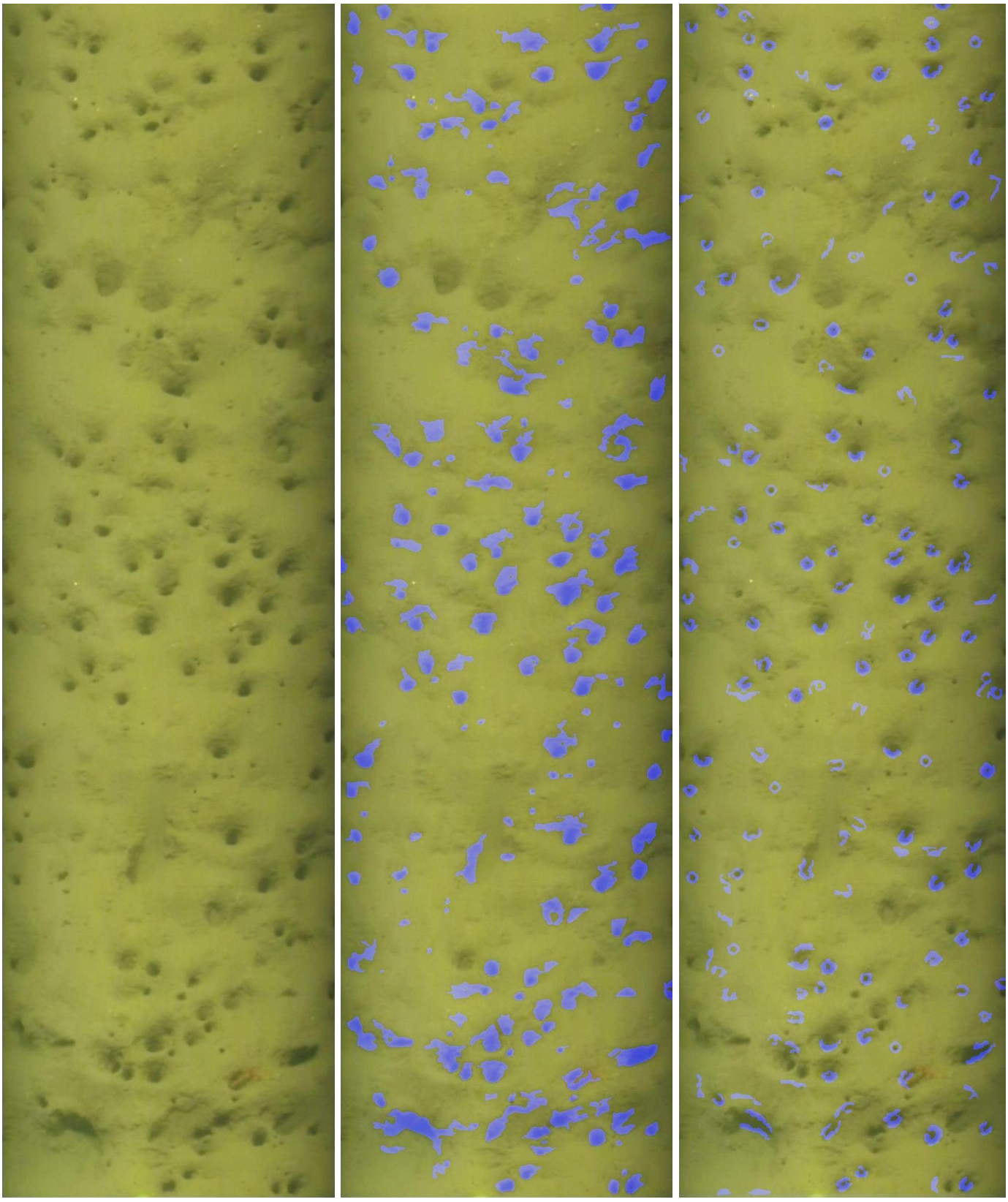
[9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, pp. 1–38, 1977.

[10] 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.

[11] M.-K. Hu, "Pattern recognition by moment invariants," *Information Theory*, vol. 49, pp. 179–187, 1961.

[12] R. Haralick and L. Shapiro, *Computer and Robert Vision*, Addison-Wesley, Boston, 1st edition, pp. 620-660, 1992.

[13] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley-Interscience, NY, 2nd edition, 2001.



(a) (b) (c)

Fig. 6. a) Mosaic generated from 360 Frames, and objects detected as burrows (blue) using b) our method, and c) Lau et al. [1] method.