

Detection of Stationary Animals in Deep-Sea Video

Marzieh Mehrnejad
Department of Electrical and
Computer Engineering
University of Victoria
BC, Canada
Email:marzmehr@uvic.ca

Alexandra Branzan Albu
Department of Electrical and
Computer Engineering
University of Victoria
BC, Canada
Email:aalbu@uvic.ca

David Capson
Department of Electrical and
Computer Engineering
University of Victoria
BC, Canada
Email: capson@uvic.ca

Maia Hoeberechts
NEPTUNE Canada
Victoria
BC, Canada
Email: maiah@uvic.ca

Abstract—Cabled observatory video data are a rich source of information for marine biologists. However, the large amount of recorded video creates a “big data” problem, which calls for automated detection techniques for sea life. Most of the related research has addressed detection based on animal motion. We propose a novel method for the detection of stationary animals, more precisely for crabs. Our approach integrates shape and color information for the automatic detection of crab and non-crab image patches. With these patches we train a feed-forward neural network, which is further used for classifying image patches into crab and non-crab classes. The experimental evaluation shows very promising results.

I. INTRODUCTION

Cabled observatory video data are collected by video cameras connected to power and communication systems via a seafloor network. These data allow for the remote observation of marine species and for the study of various topics of biological interest, such as biodiversity in the deep-sea environment, species behaviour, and community ecology. The 24/7 presence of cameras and the real-time connection to shore enables an unprecedented volume of data to be recorded. This ‘big data’ problem calls for the development of automated image processing tools to aid biologists in their visual search tasks. Motion is the most common cue used for the automatic detection of sea life in underwater videos. However, some animals such as sea stars, sea urchins, gastropods, crabs and sometimes fish exhibit very little motion in the video sequences collected at the Barkley Canyon site on the NEPTUNE Canada cabled observatory [1]. In this work, we describe a technique that does not rely on motion for detecting the presence of animals. We focus on identifying stationary crabs, without differentiating between species of crabs. Our approach integrates shape and color information for the automatic detection of crab and non-crab image patches. With these patches we train a feed-forward neural network, which is further used for classifying image patches into crab and non-crab classes.

Figure 1 shows an example of a grooved tanner crab (*Chionoecetes tanneri*) recorded at 834 m depth, 9 August 2006. As shown in Figure 2, the images in our database are not as high in quality as this example image.

The remainder of this paper is structured as follows. Section II discusses related work in the area of underwater sea life detection and, more generally, in stationary object detection. Section III explains in detail all steps of our proposed approach. Section IV discusses our experimental results. Section V concludes the paper and outlines future work directions.

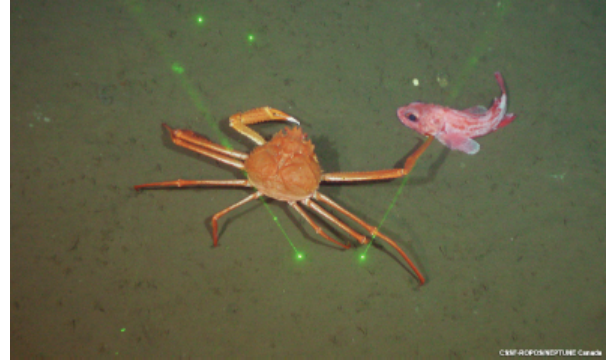


Fig. 1. Grooved tanner crab at 834 m depth.

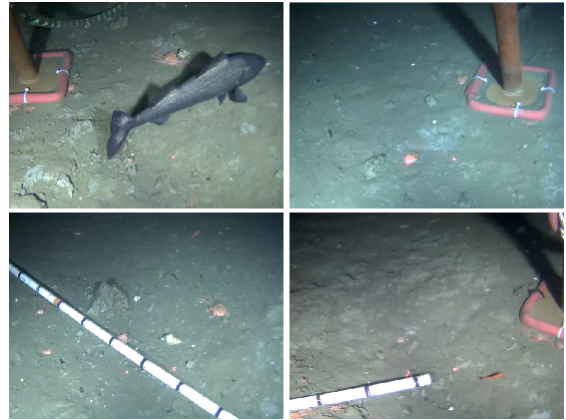


Fig. 2. Sample images from the videos in NEPTUNE's [1] database.

II. RELATED WORK

As previously mentioned, motion is the most common cue in sea life detection techniques. This is mostly due to the poor spatial resolution of underwater images and to optical phenomena such as back-scattering, which result in objects with ill-defined boundaries and variable appearance. In a recent survey article, Shortis et al [2] discuss various methods to automatically detect, count, measure, track or identify fish in underwater videos. Motion detection is relevant for applications such as underwater video summarization and detection of salient events, as shown by Gebali et al [3]. Since our focus is on detecting stationary or quasi-stationary sea life, motion-based detection techniques are not relevant. Very little

work has been done on the detection of underwater stationary animals, therefore we also consider methods for the detection of stationary objects from other computer vision application areas.

In video surveillance, stationary objects are usually detected via background subtraction techniques, as shown in a comparative study by Bayona et al. [4]. Background subtraction aims at generating a statistical model for the background that adapts quickly to background motion. For instance, Mathew et al [5] use Gaussian Mixture Models to characterize background and foreground regions; they detect stationary objects by observing transition states of distributions in GMM. Background subtraction techniques assume that illumination changes vary smoothly in the observed scenes. This assumption does not hold for the underwater video data in our database. Therefore, we attempt at using specific shape and appearance characteristics of the objects of interest (crabs) for their automatic detection. The following section describes our proposed approach.

III. PROPOSED APPROACH

Our approach is designed as a sequence of three main modules. Module A (Segmentation) identifies potential regions of interest (i.e. regions that might correspond to crabs) according to color and shape criteria. Module A consists of steps 1 and 2 in Figure 3 below. Module B (Feature extraction) is based on edge detection applied to the ROIs of crabs and non-crabs. Module B involves steps 3 and 4 in Figure 3. Module C (Neural networks) consists of the last two steps in Figure 3, which implement the training and test phase of a feed-forward neural network with a variable number of hidden nodes. The remainder of this section presents a detailed description of each module.

- | | |
|----|---|
| 1) | Perform color-based segmentation by thresholds derived from color histograms.
OUTPUT: Binary image. |
| 2) | Filter binary image according to shape criteria.
OUTPUT: Binary image of potential crab regions. |
| 3) | Perform local edge detection around potential crab regions.
OUTPUT: Crab and non-crab regions of interest. |
| 4) | Generate fixed-size crab and non-crab templates from their corresponding regions of interest.
OUTPUT: Fixed-size square matrices reshaped into vectors as neural network inputs. |
| 5) | Train feed-forward neural network with N nodes in hidden layer with vectors from training set.
OUTPUT: Trained classifier neural network for fixed-size segments. |
| 6) | Test feed-forward neural network with vectors from testing set.
OUTPUT Binary: 1 for crab and 0 for non-crab. |

Fig. 3. Main steps of the proposed approach

A. Segmentation

Given that the appearance of the color of crabs in the Barkley Canyon video data is very similar to Caucasian human

skin color, we have investigated related work on color-based human skin detection. As shown by Albiol et al [6], the choice of color space for describing color information is critical for the design of a successful skin detector. However, there is no clear consensus on which color space is optimal for skin detection applications. Albiol et al [6] recommend the use of the HSV space on the basis that it provides some robustness to variable illumination. Brand and Mason [7] work with RGB spaces and use the R/G ratio as the main color descriptor. Wand and Branstein [8] propose YIQ, a new color space derived from the standard NTSC. To investigate which color space works best for our task, we followed the recommendation of Elgammal et al [9] and collected a database of patches representing crab and non-crab regions. All patches were described in terms of their color histograms in each of the considered color spaces (RGB, HSV and YIQ). Figures 4, 5 and 6 show the histograms corresponding to each color space.

These histograms enabled the determination of specific threshold values for color-based thresholding, which was performed independently in each of the three color spaces:

- **RGB**

As shown in Figure 4, the R/G, B/G, B/R values for crab and non-crab patches are all confined to a limited range. The partial overlap in the B/G range of the two classes was solved by setting thresholds based on all three ratios. A similar solution was used for the other two color spaces. The thresholds for the ratios were manually determined from their corresponding histograms in order to maximize the ability to discriminate between crabs and non-crabs. The thresholds are presented in Table I.

TABLE I. THRESHOLDS OF CRAB AND NON-CRAB REGIONS IN THE RGB COLOR SPACE.

Class	R/G Range	B/G Range	B/R Range
Crab	0.98 - 1 1.01 - 1.095 1.109 - 1.21	0.85 - 1	0.8 - 1
Non-crab	0.62 - 0.64 0.66 - 0.89	0.87 - 1.08	1.019 - 1.52

- **HSV**

According to the histograms in Figure 5, the hue (H) and saturation (S) values of crab and non-crab patches are confined to a fairly limited range. There is significant variation for the V value for both classes; this is why the V value is not considered in the thresholding process. The thresholds are presented in Table II.

TABLE II. THRESHOLDS OF CRAB AND NON-CRAB REGIONS IN THE HSV COLOR SPACE.

Class	S Range	H Range
Crab	0.025 - 0.198 0.1595 - 0.215 0.225 - 0.251	0 - 0.2 0.94 - 1
Non-crab	0.12 - 0.4	0.345 - 0.54

- **YIQ**

As shown in Figure 6, the I and Q values of crab and non-crab patches are confined to a small range (with some overlaps in Q). The Y values are in a

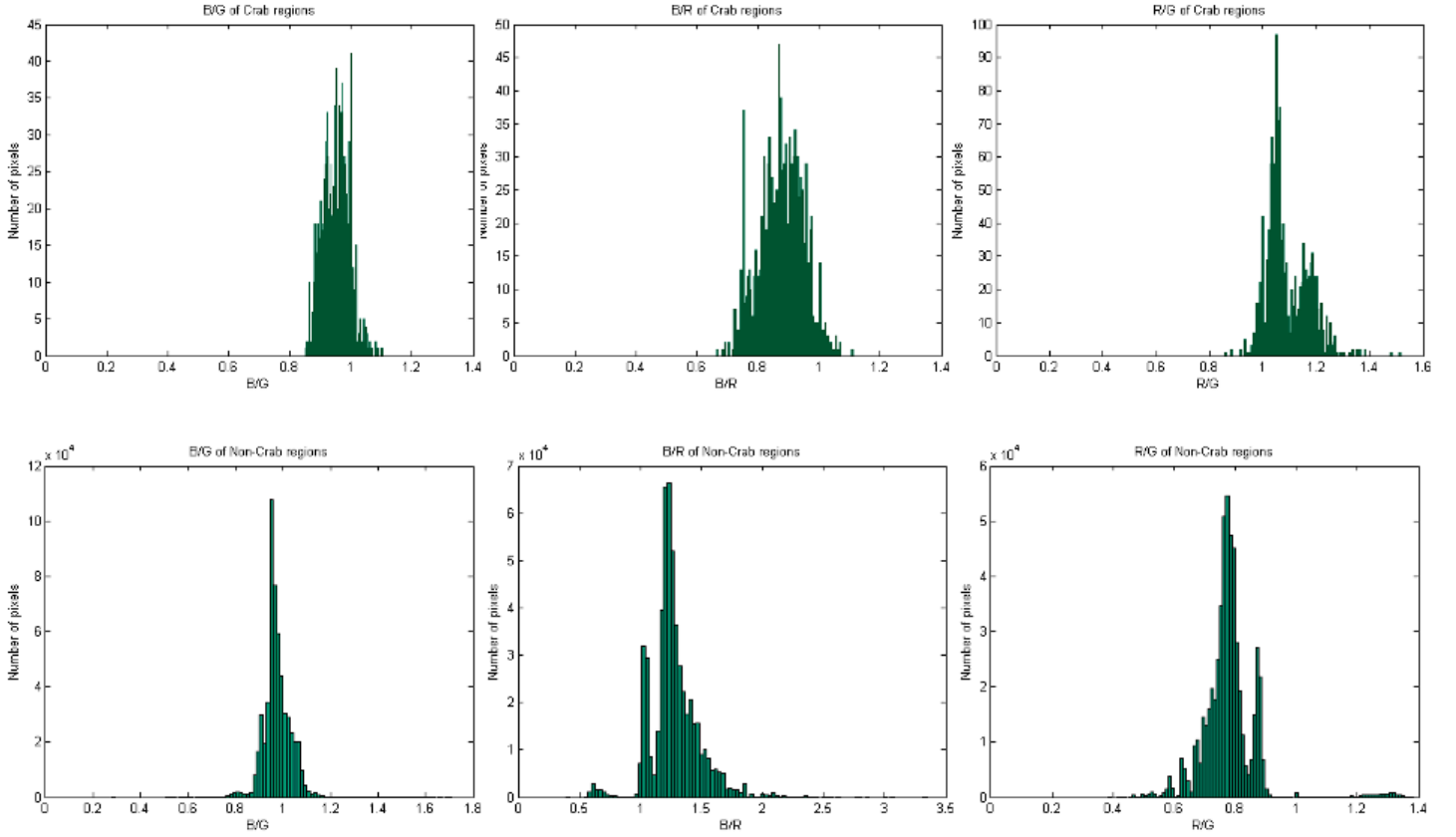


Fig. 4. The generated histograms in the RGB color space.

wide range of values, therefore the Y values were not considered for thresholding purposes. The thresholds are presented in Table III.

TABLE III. THRESHOLDS OF CRAB AND NON-CRAB REGIONS IN THE YIQ COLOR SPACE.

Class	I Range	Q Range
Crab	-0.01 - 0.096	-0.02 - 0.029
Non-crab	-0.088 - -0.021	-0.0435 - -0.01

The binary images resulting from color-based thresholding are postprocessed using shape considerations. We use an eccentricity measure in order to retain only compact and circular regions which are more likely to represent crabs.

At this stage, the performance of detecting crab and non-crab regions by color-based thresholding were evaluated for every color space. As shown in Table IV, the HSV space produces the best results. Therefore, the following steps work with images resulting from HSV color space thresholding.

TABLE IV. EVALUATION RESULTS OF THE THREE COLOR SPACES.

Color Space	Missed Detections	Correct Detections	False Detections	Precision	Recall
HSV	282	134	158	0.46	0.32
RGB	276	136	270	0.335	0.33
YIQ	260	176	503	0.26	0.40

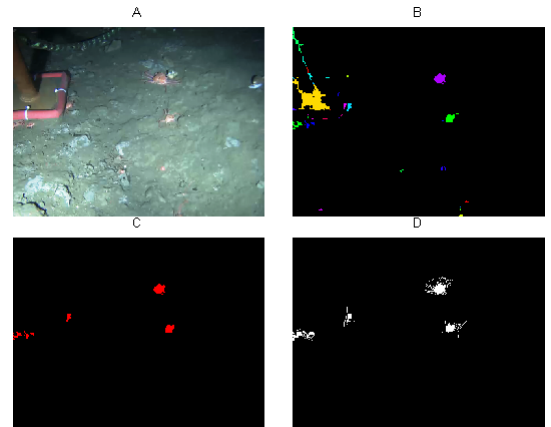


Fig. 7. Images illustrating the segmentation process in the HSV color space. A) Original image. B) Binary image output by color-based thresholding. C) Image after the elimination of regions with undesirable eccentricity values. D) Image after local edge detection around detected regions.

B. Feature Computation

Local edge detection using the Canny algorithm is performed around the detected regions from module A (Figure 7C). The local edge detection helped the crabs become more distinguishable, as their legs became visible. Figure 7D displays the resulting images after the operation explained above. The binary edge maps corresponding to each region are stored

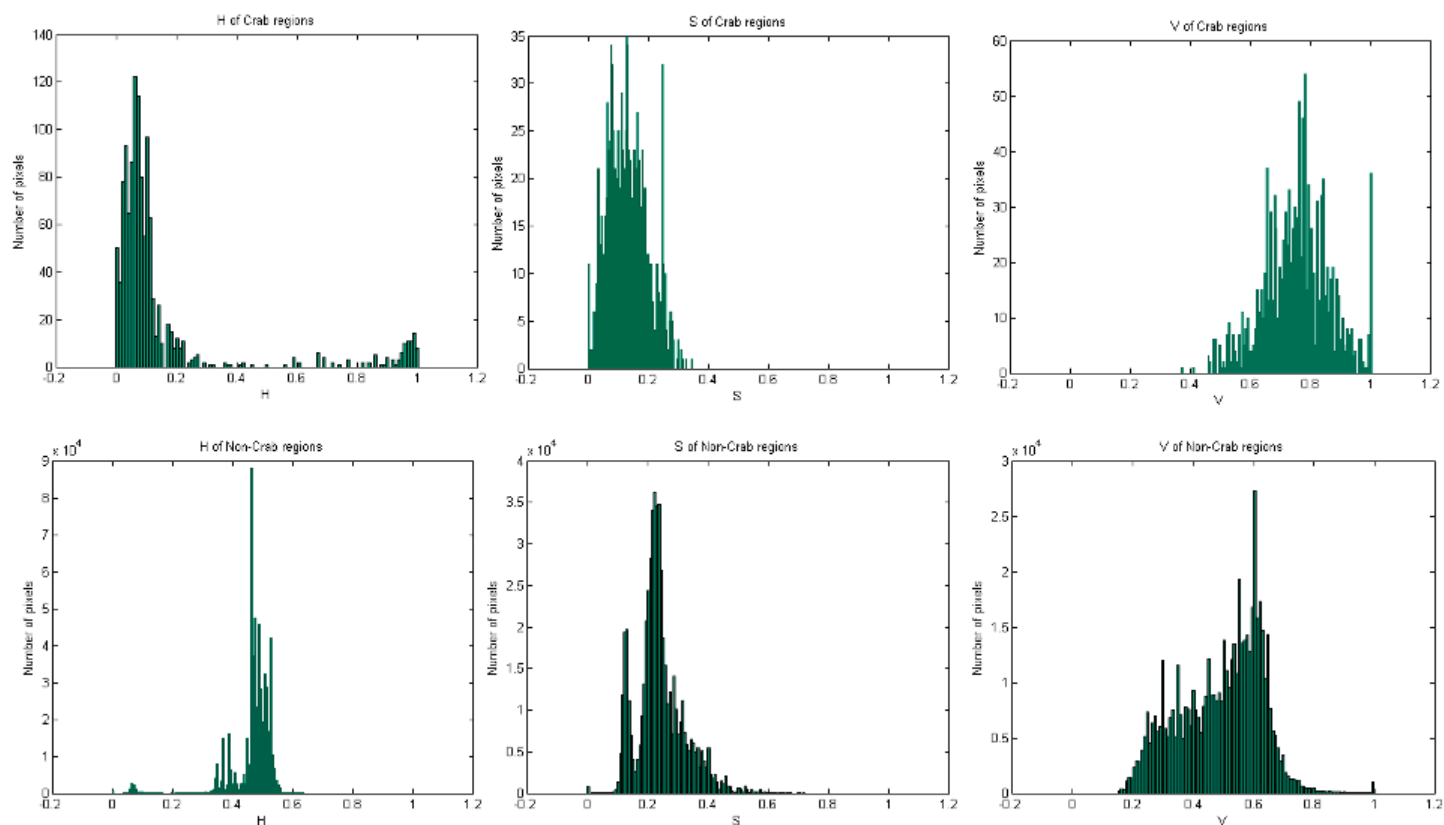


Fig. 5. The generated histograms in the HSV color space.

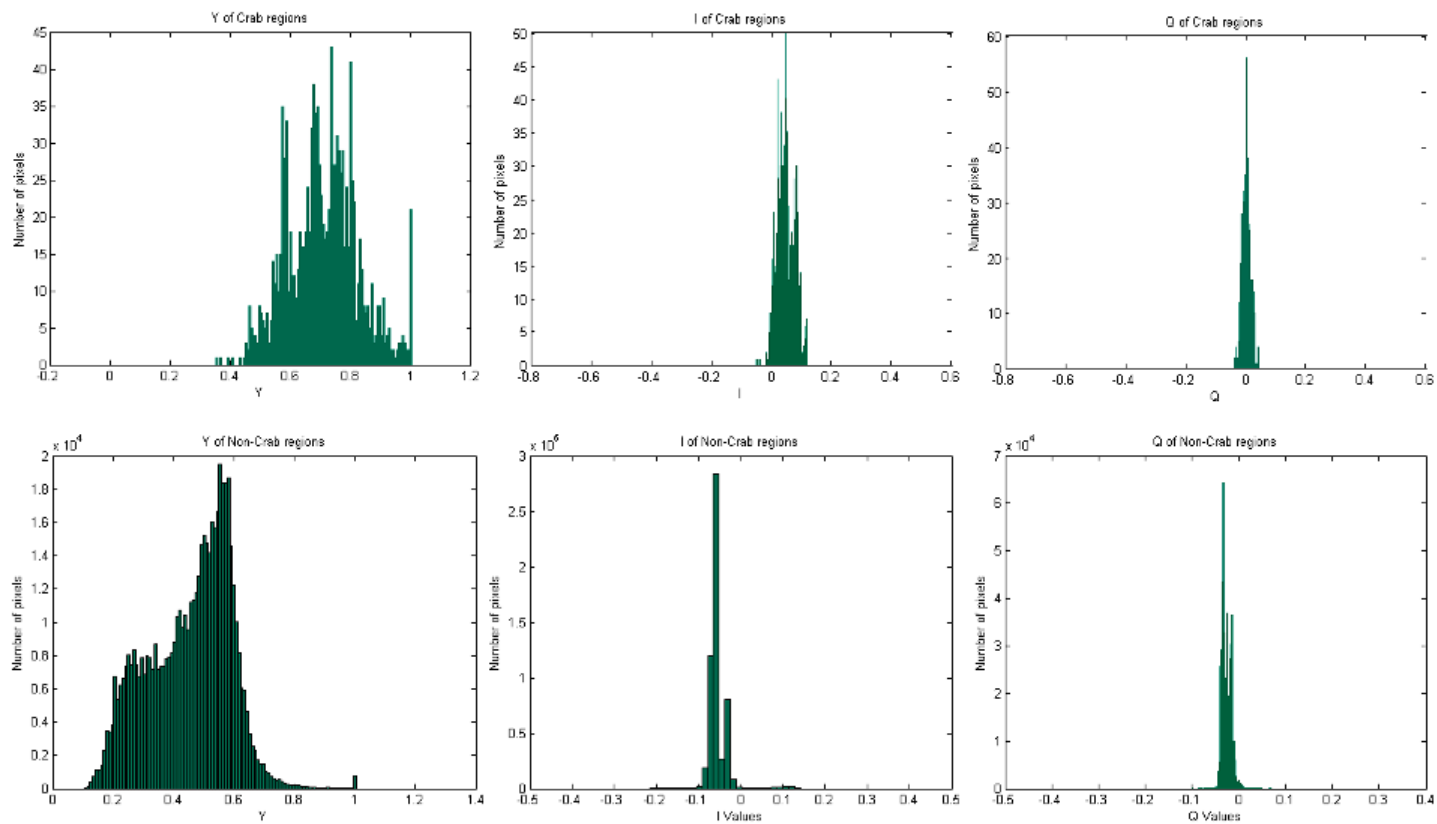


Fig. 6. The generated histograms in the YIQ color space.

as matrices composed of zeros and ones. These matrices are reshaped as vectors in order to form inputs to a neural network.

C. Neural Network Architecture

This study reports results on the detection of crabs with image sizes comprised between 30-60 pixels. The size constraint allows for a consistent training of the neural network with feature vectors of constant length. The binary matrices output by module B matrices were reshaped into column vectors, which represent the network's inputs. The entire collection of patches is divided into two sets, namely the training set and the testing set. Each set consists of a combination of crab and non-crab segments. Different training sets were used with the aim of finding the set that trains the network with minimum error in classifying a test set. Subsequently, the number of neurons in the hidden layer was varied in order to find the most appropriate network.

IV. EXPERIMENTAL RESULTS

The experimental database for this study consists of three 5 minute videos acquired at 30 frames per second from the Neptune Canada database, that is a total of 9000 images at 480×640 spatial resolution. One video was used for extracting the training set, which consists of 100 images. The image patches that were used to generate the color histograms and to determine the color thresholds were also taken from this set.

The test set, consisting of 200 images, was built with frames from the remaining two videos. The test set was used for finding the optimal color space, as well as for testing the neural network.

Table V presents the evaluation of the output of the proposed approach for crab detection. The performance of feedforward networks with one hidden layer and various number of nodes in the hidden layer are compared. These results are very promising as both precision and recall values are very high. During the evaluation process, it was observed that missed detections are mainly caused by uneven lighting across the image, while false detections occur mostly for man-made objects that are of similar color with crabs.

TABLE V. EVALUATION RESULTS OF THE CLASSIFIER NEURAL NETWORK

Number of nodes in hidden layer	Missed Detections	Correct Detections	False Detections	Precision	Recall	Error
25	0.3	0.7	0.07	0.94	0.75	0.15
30	0.1	0.9	0	1	0.9	0.05
35	0.15	0.85	0.105	0.89	0.85	0.15
33	0.1	0.9	0.14	0.86	0.9	0.125
32	0.2	0.8	0.06	0.94	0.8	0.125

V. CONCLUSION

This paper presents a novel approach for detecting stationary crabs in deep sea video. The experimental database was provided by NEPTUNE Canada. Our approach is modular. It first detects regions in the image that are candidates for crabs based on color and shape information. Next, crab and non-crab patches are used to train a feedforward neural network for the detection of large crabs. The experimental evaluation shows very promising results. Ongoing work focuses on the

relaxation of the size constraint, as we aim to be able to detect crabs of various sizes. Also, future work will investigate pre-processing techniques that will enhance the quality of input images prior to color-based segmentation.

REFERENCES

- [1] NEPTUNE Canada. (2013) Neptune canada. [Online]. Available: <http://www.neptunecanada.ca/>
- [2] M. R. Shortis, M. Ravanbakskh, F. Shaifat, E. S. Harvey, A. Mian, J. W. Seager, P. F. Culverhouse, D. E. Cline, and D. R. Edgington, "A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences," in *SPIE Optical Metrology 2013*. International Society for Optics and Photonics, 2013, pp. 87910G–87910G.
- [3] A. Gebali, A. Branzan-Albu, and M. Hoeberechts, "Detection of salient events in large datasets of underwater video," in *IEEE Oceans*. IEEE, 2012, pp. 1–10.
- [4] Á. Bayona, J. C. SanMiguel, and J. M. Martínez, "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques," in *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*. IEEE, 2009, pp. 25–30.
- [5] R. Mathew, Z. Yu, and J. Zhang, "Detecting new stable objects in surveillance video," in *Multimedia Signal Processing, 2005 IEEE 7th Workshop on*. IEEE, 2005, pp. 1–4.
- [6] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 1. IEEE, 2001, pp. 122–124.
- [7] J. Brand and J. S. Mason, "A comparative assessment of three approaches to pixel-level human skin-detection," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 1. IEEE, 2000, pp. 1056–1059.
- [8] C. Wang and M. S. Brandstein, "Multi-source face tracking with audio and visual data," in *Multimedia Signal Processing, 1999 IEEE 3rd Workshop on*. IEEE, 1999, pp. 169–174.
- [9] A. Elgammal, C. Muang, and D. Hu, "Skin detection-a short tutorial," *Encyclopedia of Biometrics*, 2009.