Automated Change Detection in an Undersea Environment using a Statistical Background Model

Zhi Qi
Centre for Intelligent Machines
McGill University
qizhi@cim.mcgill.ca

Jeremy R. Cooperstock
Centre for Intelligent Machines
McGill University
jer@cim.mcgill.ca

***Abstract*— Marine scientists are turning increasingly to underwater video cameras in their research. These provide enormous quantities of visual data that often overwhelm the manual processing abilities of the scientists. To cope with such large data sets, an automated change detection system is proposed that helps isolate the time periods in which *significant* activity is found in the video sequence. Unlike change detection algorithms in use in terrestrial environments, the system must account for the photometric complexity of underwater video, including interference from small floating particles ("sea snow"), the scatter of light as it propagates through water, and the non-uniform frequency decay of light intensity with distance. In addition, certain activity, such as the motion of swimming fish that are attracted by the use of artificial lighthing, is considered a distracter, and should, ideally, be ignored. These factors are addressed by our system, in large part through the use of Mixture-of-Gaussians background models.**

I. Introduction

Scientific observation of the undersea environment is a challenging problem, in particular at extreme depths where high pressure and the absence of light impose significant requirements on video equipment. While various deep sea dives conducted by robotic vehicles offer a limited glimpse into this incredible environment, there has been little opportunity for extended observation of particular areas of interest over a duration of weeks or months. With the deployment of the Victoria Experimental Network Under the Sea (VENUS) and the North-East Pacific Time-Series Undersea Networked Experiments (NEPTUNE) observatories now taking place, this situation is about to change dramatically.

From a scientific perspective, one of the most important aspects of observation is the ability to track both short-term phenomenon such as the feeding habits of a crab and long-term changes of undersea objects, such as gradual plant growth on the seabed or decomposition of a dead organism. However, it is infeasible for a limited number of marine scientists to view a continuous, live video feed from an undersea camera for more than relatively brief periods of time, outside of which, the events of interest may take place. This motivates the use of automated computer analysis of the video in order to detect key changes of relevance to the scientists.

Such processing is complicated by the photometric characteristics of the undersea environment [1], such as scatter of light and interference from small floating particles ("marine snow"), which absorbs or reflects light. As distance between the camera and observed objects increases, both contrast and clarity decrease significantly. These problems persist, regardless of the quality or resolution of the cameras employed.

Moreover, the use of artificial lighting, necessitated by the absence of natural light at significant depths, attracts fish to the observation area. These fish not only occlude other objects of interest, which we refer to as *significant* objects, but also create semantic ambiguity for change detection, as their motion is considered irrelevant for most scientific purposes. We thus denote such fish as *distracters*. More formally, given a selected time scale, *significant* objects may generally be in motion, but must remain static for certain periods, during which they can be observed as part of the background. *Distracters*, on the other hand, are seen to be in (nearly) constant motion. For most purposes, marine scientists would like to ignore these.

Additional factors also manifest underwater, such as non-uniform frequency attenuation that results in a "green-shift" of the resulting video. While this reduces the apparent colour range of the scene, it is not, however, an issue for change detection algorithms.

This paper introduces a novel change detection system that is tailored to the demands of this challenging undersea environment. First, Section II summarizes related work in change detection. Next, Section III introduces the supporting method of Mixture-of-Gaussians (MoG) models, which offers robustness to the complicating factors of noise and semantic ambiguity, noted above. Sections IV describes the details of our implementation and illustrate some of the experimental results. Finally, Section V discusses opportunities for further enhancements and applications.

II. Related Work

Early change detection system relied on the signed difference image [3][2]. After thresholding such a difference image, one obtains a change mask, as illustrated in Figure 1, indicating the regions that differ between one image and the next. Of course, this method cannot distinguish between significant objects, for example, the small movements of the crab, contained within the manually generated bounding box, and the remaining regions, which all correspond to distracters.

More sophisticated algorithms exploit spatial and temporal relationships between neighbouring pixels as predictions and use these to classify pixels as changed or not. Toyama *et al.* [5]

Fig. 1. Example of change mask generation. Sample frame from time t (left), sample frame from time $t + 50$ (middle), and change mask after thresholding the intensity difference between the two frames (right). Note that in this example, the significant objects are contained exclusively in the rectangular bounding box.

uses the linear combination of k previous values to predict the current value of a pixel, while Hsu *et al.* use spatial intensity prediction for change detection of surveillance images [4]. While more robust to noise than simple differencing, prediction methods are also incapable of separating significant and distracter objects, as these may have similar color, texture, or even short-term motion characteristics.

In recent years, statistical background modelling techniques were applied to change detection [6]. Stauffer and Grimson [7] model each pixel by a mixture of Gaussians and use an on-line approximation to update the model. Because significant objects are occasionally static, the background model will collect more samples corresponding to these than to distracters. This is a critical factor, is it provides a metric for distinguishing between the two. Furthermore, by updating multiple Gaussians in a gradual manner, their model is adaptive to nonstationary temporal information, such as background changes due to varying illumination or other photometric noise. Lee [8] further proposed an adaptive, rather than fixed, learning rate for each Gaussian to obtain improved convergence speed without sacrificing stability.

Relatively few change detection systems were designed explicitly for the underwater environment. Two prominent examples are those of Edington *et.al.* [13], and Lebart *et.al.* [14]. Edington *et.al.* first identify candidate object regions by analyzing the salient feature maps. After manually distinguishing interesting objects from noise objects (distracters), they track the former frame-by-frame to detect changes in the scene. Similarly, Lebart *et.al.* also begin with feature analysis, classify these into groups. A change is noted when the distance between similar feature groups in different frames exceeds a threshold. In both algorithms, the critical factor is the feature set. Unfortunately, these must be developed on a case-by-case basis to cope with the photometric complexities arising from variations in illumination, undersea location, and marine snow conditions. Further more, these algorithms require manual tuning in order to differentiate between the change of significant objects and distracters; neither can do so automatically. This observation motivated our adoption of statistical background modelling techniques, specifically, the MoG background model using adaptive learning rates, in order to support an automatic change detection algorithm.

III. Change Detection of Significant Objects

The operation of our change detection algorithm is shown in Figure 2. Given a set of images of the same scene, these are grouped into distinct time intervals. The images from each such interval are used to generate an MoG background model, from which we construct a *background image*, visually representing the static background. Finally, for each successive time interval, a *change mask* of significant objects is produced by comparing the background image with the MoG from a *different* time interval. This process is explained in further detail in the following sections.

A. Mixture-of-Gaussian model construction

Suppose all the pixels from the frames in a given time interval satisfy the distribution of our Mixture-of-Gaussian (MoG) background model. In this case, the probability that a pixel assumes a value X_t at time t is given by:

$$P(X_t) = \sum_{i=1}^{K} \frac{\omega_{i,t}}{\sqrt{(2\pi)^n |\Sigma_{i,t}|^{\frac{1}{2}}}} \times \exp\{-\frac{1}{2}(X_t - \mu_{i,t})^T \Sigma_{i,t}^{-1}(X_t - \mu_{i,t})\} \quad (1)$$

where at time t, $\mu_{i,t}$ is the mean of the i^{th} Gaussian, $\Sigma_{i,t}$ is the diagonal covariance matrix, and $\omega_{i,t}$ is its weight [7].

For our purposes, pixel values are represented in the YC_bC_r colour space, as this was found to offer greater robustness to photometric noise than either RGB or HSI. This result was also observed by Kristensen [9].

Any new pixel value is compared against the available models. If the value can be represented by an element of the MoG, it is used to update the model. Otherwise, the least-likely Gaussian element, *i.e.* with the smallest $\omega_{i,t}$, will be replaced by a new Gaussian initialized with the new pixel value. Our present implementation uses $K = 3$ Gaussians.

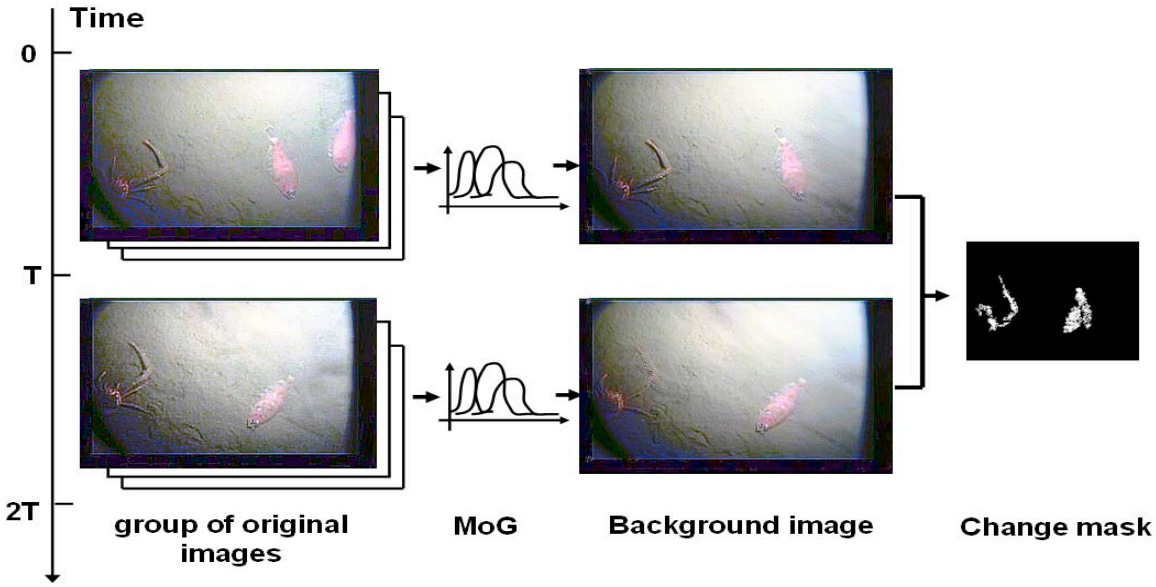

Fig. 2. Illustration of change detection algorithm.

After all the frames within a particular time interval have been processed as above, we begin the generation of the next MoG background model.

B. Background image generation

Next, a weighted summation of the means of each Gaussian element from the MoG generates a visual representation of the background model. This *background image* describes the static background during the given time interval, as shown in Figure 3.

A background image, *i.e.* the expected value, $E[X|B]$, of the observation X, assuming it to be background, is given by Lee [8] as the following weighted average:

$$\begin{aligned} E[X|B] &= \sum_{i=1}^{K} E[X|G_i]P(G_i|B) \\ &= \frac{\sum_{i=1}^{K} \mu_i P(B|G_i)P(G_i)}{\sum_{j=1}^{K} P(B|G_j)P(G_j)} \end{aligned} \tag{2}$$

Only the background and static significant objects are retained in the background image while distracter objects are effectively removed. Thus, subsequent comparisons between a background image and a model will not be biased by the semantic noise due to distracters. As is explained in the following section, these factors are critical to our change detection of significant objects.

C. Change detection

Change detection is closely related to the problem of foreground/background segmentation. For a given frame, we calculate the probability of each pixel belonging to the background according to the trained MoG background model:

$$\begin{aligned} P(B|X) &= \sum_{i=1}^{K} P(B|G_i)P(G_i|X) \\ &= \frac{\sum_{i=1}^{K} P(X|G_i)P(G_i)P(B|G_i)}{\sum_{i=1}^{K} P(X|G_i)P(G_i)} \end{aligned} \tag{3}$$

If this probability exceeds some threshold, the pixel is considered as an element of the background, and otherwise, as a foreground (or non-static) object.

IV. Experimental Results

The video data used in our experiments was taken from samples of the VENUS project database.[1] In this data set, we assume that the crab (at left) and the fish (middle) of Figure 2 are significant objects, while other moving objects are distracters that should be ignored.

A. Significant Object Change Detection

When detecting regions of significant object change, as described in Section III, thresholding[2] the probabilities of Equation 3 results in a single, raw change mask for the entire time interval ($[0, T]$ in the example of Figure 4). In this application, the values of X are taken from a background image and compared to the MoG model (the $G_i, i = [1, k]$ elements) from a different time interval ($[T, 2T]$ in Figure 4).

We apply binary image morphological operations to remove noise and then use the connected components algorithm [12] to separate the raw change mask into non-overlapping regions. These delineate the areas of the image in which significant

[1] `http://www.venus.uvic.ca/data/galleries/video/saanich_inlet/index.html`

[2] In our current implementation, we are using a threshold value of 0.11.

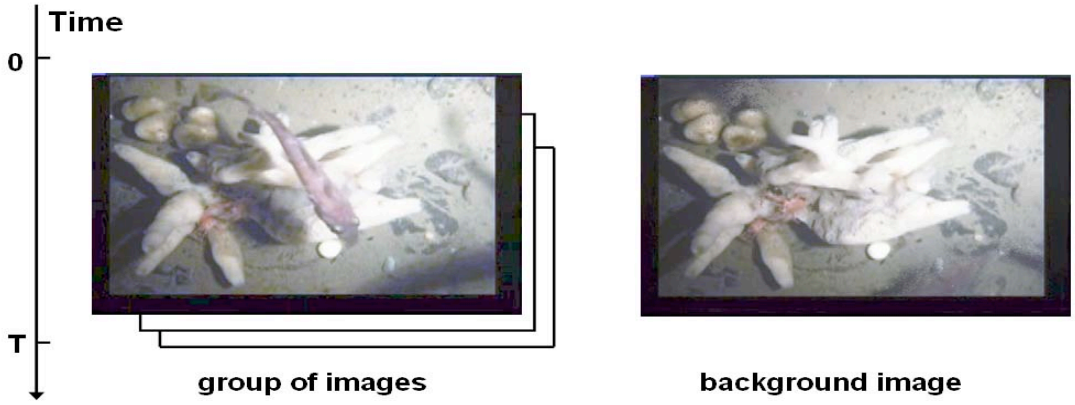

Fig. 3. Given a group of images over the interval $t = [0, 5]$ (left), we train the MoG background model, from which we generate a background image (right) that includes only the sea floor and stable significant objects seen within that interval.

object change has occurred between different time intervals, as shown in the example of Figure 5.

These results clearly reveal whether and where the changes of significant objects have occurred within the period covered by the associated time intervals ($[0, 2T]$ in Figure 4). However, we may achieve greater temporal accuracy by making use of *representative* frames,[3] selected from each interval as the frame that maximizes $P(B|X)$. In this case, the time at which the change occurred can be isolated to the smaller window of $[t_{x_1}, t_{x_2}]$.

It is difficult to quantify the accuracy of our algorithm, given that the only reasonable test video sequence available to us was extremely short and limited in variability. However, as one metric of performance, on the test sequence used for this initial study, we find that the system consistently distinguishes between the constantly moving fish (distracter) and both the periodically moving fish and crab (significant objects)). False positives only resulted from unpredictable background motion, as seen in the cloud of dust produced by a fish sweeping its tail along the sea floor. Another possible source of error is that the change detection algorithm may produce false negatives when significant objects exhibit similar colour to the background, just as camouflaged objects are difficult for humans to detect.

B. Distracter Object Removal

Once the regions of significant objects have been identified, we can, if desired, attempt to remove distracters from the video sequence, allowing scientists to focus their attention solely on the objects of interest.

This is accomplished by first comparing each frame in a given interval to the background model generated from that same time interval. Because the background image, by definition, is free of distracters, we cannot use it to detect such objects. Therefore, we use instead the observations X from individual frames in Equation 3. Thresholding $P(B|X)$, we obtain a raw change mask of this frame. Note that this result may contain both distracters and significant objects, as seen in Figure 6. The next step is to label these appropriately.

Referring to Figure 5, we note that significant objects appear only in some of the change masks, corresponding to the time intervals in which the objects were in motion. In order to separate these from distracters, we combine several successive significant object change masks in a logical OR manner to obtain a more inclusive aggregate representation, as seen in Figure 7.

Each isolated region of the aggregate change mask is enclosed by a bounding box. We then choose from a group of images one reference frame that satisfies the following two requirements:

- the sum of probabilities $P(B|X)$ within each bounding box exceeds some threshold, thus maximizing the likelihood that they do not contain distracter objects
- the number of *local features* contained within the bounding boxes, which correspond either to significant objects or background, as described below, is maximal, thereby increasing the chance of finding matching features in other frames

Pixel values within a circular neighbourhood are represented by a combination of SIFT features [10] and a 32-bin colour histogram. A local-feature based object recognizer [11] is then used to find matches between the reference frame and other frames. Any change regions that do not contain at least one matching feature from the reference frame are assumed to belong to distracter objects and can be replaced with the corresponding pixels from the background image, as illustrated in Figure 8. Repeating this process with every frame in a video sequence results in the effective *disappearance* of the distracters.

This approach is entirely dependent on the quality of the local features it is able to recover, and thus, suffers at times from the complicated photometric properties of the undersea environment. Additionally, motion blur degrades the quality of feature matching, resulting in the occasional unintended removal of significant objects. However, our distracter removal

[3]This frame is generally similar to the background image, but may exhibit photometric noise and contain distracters.

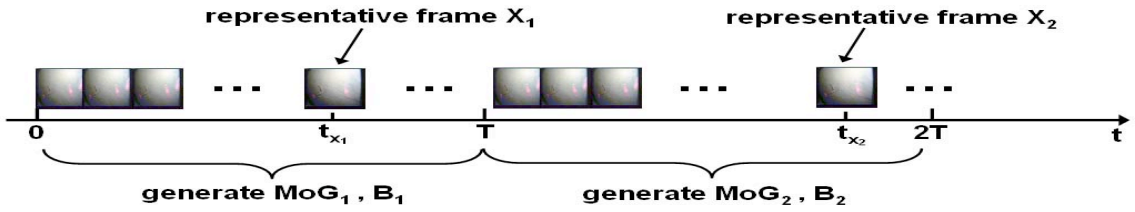

Fig. 4. In each time interval, $[0, T]$ and $[T, 2T]$, a background model, MoG_i, and background image B_i, are generated. A representative frame, X_i, maximizing $P(B|X)$ is then selected from each interval.

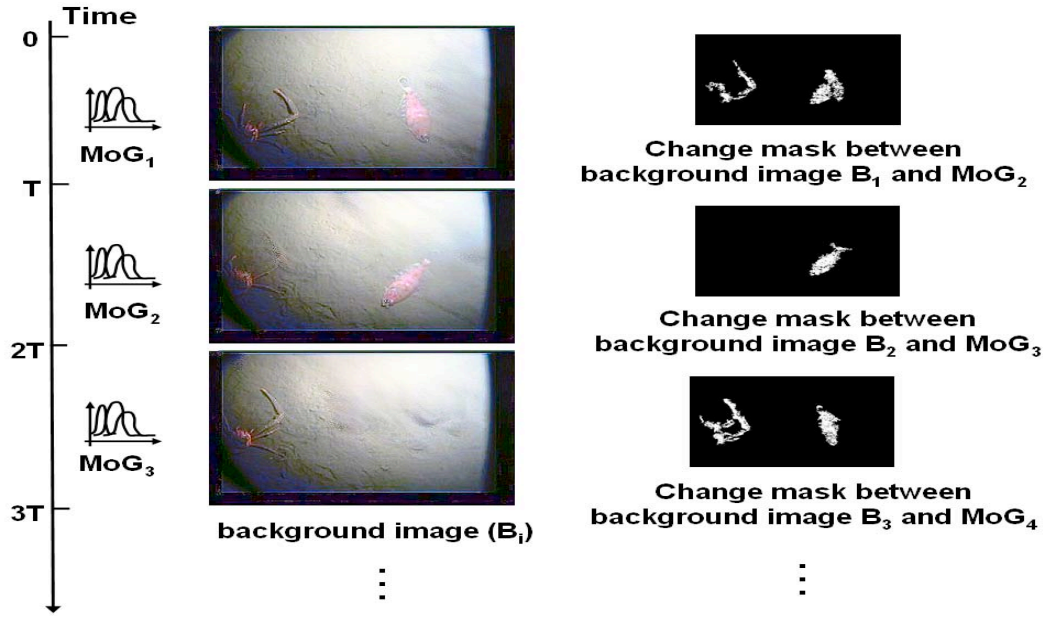

Fig. 5. In this example, each background image, B_i is compared to a successive background model, MoG_{i+1}. Thresholding the values of $P(B|X)$, we generate change masks of significant objects for each time interval.

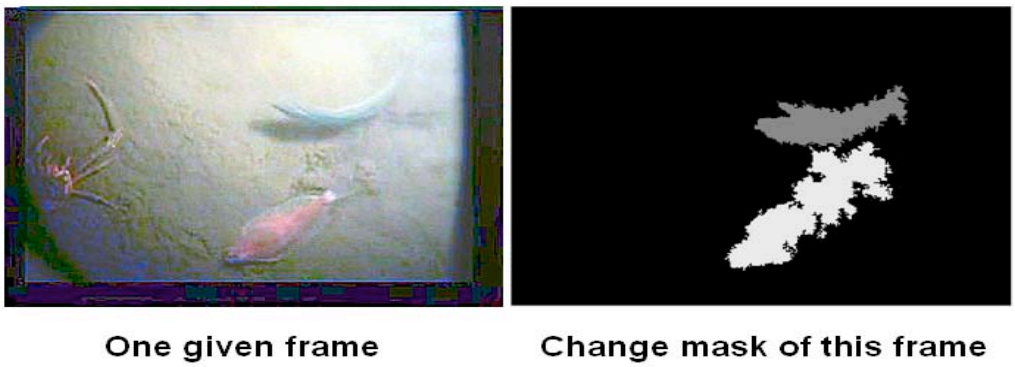

Fig. 6. A sample video frame (left) and the corresponding change mask (right). For the purposes of illustration, regions corresponding to significant objects (the lower fish) are shaded in a lighter color than the distracter.

method is notable for its low computational cost and avoidance of any manual segmentation operations.

V. Conclusions and Future Work

We introduced a novel change detection system for the challenges of scientific undersea observation. The system overcomes many of the problems resulting from the photometric complexities and semantic ambiguity associated with

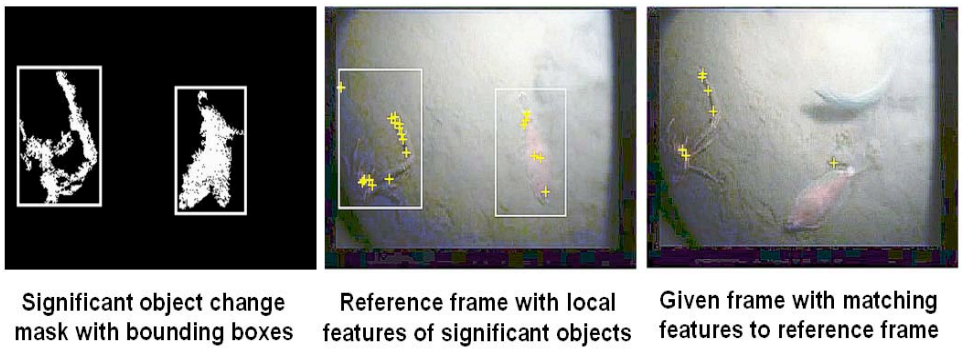

Fig. 7. For each region in the change mask (left), local features, marked as '+' symbols, are determined on the reference frame (middle). These features are then matched against other frames (right).

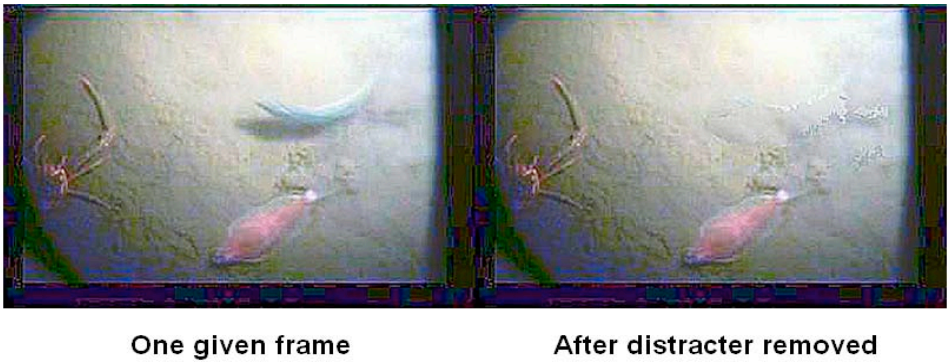

Fig. 8. A sample frame before (left) and after (right) removal of the distracter.

this environment. Under the conditions of colour dissimilarity between foreground and background and no occlusion between significant objects and distracters, the system performs very well on the test sequences we have used. Although our test data does not violate these conditions, we plan to combine temporal motion flow and spatial texture consistency with this algorithm to improve its robustness.

Additional future possibilities include the investigation of pre-processing or de-noising operations to improve image quality, as well as distance-sensitive frequency correction to account for the green-shift described in Section I. Perhaps more importantly, robust object tracking may be employed to improve the accuracy of region identification or to support higher-level operations. While it remains to be seen how well the initial results scale to the wide range of conditions we might encounter from live VENUS or NEPTUNE video data, we look forward to future experiments as the observatories become available.

REFERENCES

[1] A. Arnold-Bos, J.P. Malkasse, and G. Kerven: A Pre-processing Framework for Automatic Underwater Images Denoising. European Conference on Propagation and Systems, Brest, France. (2005)

[2] W.A. Malila: Change Vector Analysis: An Approach for detection Forest Change with Landsat. Proc. 6th Annu. Symp. Machine Processing of Remotely Sensed Data. (1980) 326–335.

[3] A. Singh: Digital change detection techniques using remotely-sensed data. Int. J. Remote Sensing. **10** (1989) 989–1003.

[4] Y.Z. Hsu, H.H. Nagel, and G. Rekers: DNew likelihood test methods for change detection in image sequences. CVGIP. **26** (1984) 73–106.

[5] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers: Wallflower: Principles and practice of background maintainance. ICCV. (1999) 255–261.

[6] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y.H. Tsin, D. Tolliver, N. Enomoto, and O. Hasegawa: A System for Video Surveillance and Monitoring. Robotics Institute, Carnegie Mellon University. CMU-RI-TR-00-12 (2000).

[7] C. Stauffer, and W.E.L. Grimson: Adaptive Background Mixture Models for Real-time Tracking. CVPR. **II** (1999) 246–252.

[8] D.S. Lee: Effective Gaussian Mixture Learning for Video Background Subtraction. IEEE Trans. PAMI. **27** (2005) 827–832.

[9] F. Kristensen, P. Nilsson, and V. Owall: Background Segmentation Beyond RGB. ACCV. **II** (2006) 602–612.

[10] D.G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints. IJCV. **60** (2004) 91–110.

[11] D.G. Lowe: Object Recognition from Local Scale-Invariant Features. ICCV. (1999) 1150-1157.

[12] W.K. Pratt: Digital Image Processing. New York, John Wiley & Sons, Inc. (1991).

[13] D.R. Edgington, K.A. Salamy, M. Risi, R.E. Sherlock, D. Walther, and K. Christof: Automated Event Detection in Underwater Video. Proceedings of OCEANS 2003.

[14] K. Lebart, E. Trucco, D.M. Lane: Real-time automatic sea-floor change detection from video. OCEANS 2000 MTS/IEEE Conference and Exhibition