Towards Detecting Changes in Underwater Image Sequences

Olivier Delaunoy, Nuno Gracias and Rafael Garcia
Computer Vision and Robotics Group
University of Girona, 17071, Spain
{delaunoy, ngracias, rafa}@eia.udg.edu

Abstract—Detecting changes between images of the same scene taken at different times is of great interest for monitoring and understanding the environment. It is widely used for on-land application but suffers from different constraints. Unfortunately, Change detection algorithms require highly accurate geometric and photometric registration. This requirement has precluded their use in underwater imagery in the past.

In this paper, the change detection techniques available nowadays for on-land application were analyzed and a method to automatically detect the changes in sequences of underwater images is proposed. Target application scenarios are habitat restoration sites, or area monitoring after sudden impacts from hurricanes or ship groundings. The method is based on the creation of a 3D terrain model from one image sequence over an area of interest. This model allows for synthesizing textured views that correspond to the same viewpoints of a second image sequence. The generated views are photometrically matched and corrected against the corresponding frames from the second sequence. Standard change detection techniques are then applied to find areas of difference. Additionally, the paper shows that it is possible to detect false positives, resulting from non-rigid objects, by applying the same change detection method to the first sequence exclusively. The developed method was able to correctly find the changes between two challenging sequences of images from a coral reef taken one year apart and acquired with two different cameras.

I. INTRODUCTION

The study of benthic areas benefits from recent progress in underwater technology, allowing the deployment of optical cameras for systematic surveying. Detecting the changes between images of the same scene taken at different times is of great interest for monitoring and understanding the environment. Several application fields would benefit from this technology, e.g. benthic habitat monitoring, deep water geological exploration or mapping of archeological sites. During the last decades, change detection algorithms were widely studied and have achieved significant progress in on-land applications [1]–[13]. The main requisite of these algorithms is a highly accurate geometric and photometric registration of the images. Therefore they are still restricted to aerial or satellite imagery.

The underwater medium can significantly degrade the quality of the acquired images. Scattering effects and non-uniform lighting result in differences in intensity levels between the images of the same area. Moreover, because of the rapid attenuation of light and the scattering effects, the images have to be taken at short range, and the area of interest cannot generally be acquired in one single view. Therefore, it is necessary to acquire a sequence of images covering the interest area and align them into a photomosaic to gain a global perspective of the site. Unfortunately, acquiring images at short range also emphasizes the parallax effects of the 3D relief resulting in local misregistrations.

Despite of the poor quality of underwater images, methods for detecting the changes can be useful in several underwater applications such as habitat restoration sites, assessing impacts from hurricanes [14] or ship groundings, coral monitoring and supervision of geothermal and volcanic activities.

The objective of this work is to develop automated tools able to deal with the constraints of the underwater sequences for detecting temporal changes in the appearance of benthic structures.

A. Related work

Several change detection algorithms were proposed [1]–[3]. Most of them require a highly accurate alignment between the images [15], [16]. They were widely used in remote sensing for the monitoring of fire, flooding, deforestation, urbanization among others [4], [5], [17]. The images used for this kind of application are taken from very high altitude (satellites or planes) in order to attenuate the effect of the 3D relief on the registration. Normally, the images to be compared are acquired at the same period of the year and on the same time of the day to avoid the effect of the shadows coming from the sun light.

Change detection techniques are also widely used in video surveillance [6]. In most cases, the camera is static or only rotating around its optical center. The absence of translation prevents parallax, making registration easier and improving its accuracy. A known background is generally used as a base for the comparisons.

In the medical imaging context, standard change detection techniques are also used to analyze MRI, X-ray or ultrasound imaging [7], [8]. In this context, they have to deal with the specific motion of the body in order to align the images.

Some techniques transform the original images by reducing the redundant information among bands [10], [11]. These methods use multi-band images such as multi-spectral, color or polarized synthetic aperture radar (SAR) imaging. However, they are generally used for finding only one typical type of change. Other methods are based on classified images [12], [13] and are then invariant to the change in illumination

but require an accurate classification of the images. These methods present the advantage to provide change direction information *(from-to)*. Their inconvenient is that they require a long training in order to provide an accurate classification. In [9], the changes are detected through a neural network. An image is compared to a sequence of images of the same scene and the network stabilizes in a change mask. This method is less sensitive to noise and to misregistrations but generally requires a long sequence.

Black et al. [18] have proposed a technique that takes profit of optical flow in order to reject changes detected due to misregistration.

Within the underwater imaging field, previous work in change detection is almost non-existent. A notable exception is the work of *Lebart et al.* [19] where a method was developed to find seafloor transition (e.g. from rocks to sandy seafloor) in a sequence of underwater images.

In this work, we propose a method that is composed of six steps as illustrated in Fig. **??**. The method aims to compare two image sequences of the same area possibly acquired with two different cameras and at different times. In the first step, a mosaic representing all the images aligned in a common frame is created in order to find overlapping pairs of images [20], [21]. Secondly, the relief of the ocean floor and the camera positions and orientations (camera *poses*) of one sequence are estimated using robust computer vision methods that allow recovering the accurate camera poses and a large set of 3D points [22]. This step requires high overlap among the images and a good calibration of the intrinsic parameters of the camera. Then, by matching the images of the second sequence with their overlapping correspondent in the first one, their camera positions are estimated. The fourth step consists in generating a set of synthetic images for each image of the second sequence. These synthetic images are the modeled images warped in such a way that their textures are seen from the camera position of the images in the second sequence. This is useful in order to reduce the effect of the 3D relief and to provide better registration. Next, photometric matching techniques are used to correct the differences in illumination and/or differences caused by the image sensors between the two datasets. Finally, once an image viewing the same area and from the same location as the image of the second sequence is generated, standard change detection algorithms can be applied. This process is applied to every image of the sequence and its corresponding set of synthetic images. As a result, we have a *binary change mask* for each pair of images that says whether there is or not a change between the two images. Those change masks can be merged into a general mask representing the changes of the surveyed area. A mosaic of the change mask can be easily obtained as we already know the position of the images corresponding to the change mask.

For testing the method, two sequences of images from a coral reef survey in the Bahamas were used. These images were acquired ten months apart (in December 2005 and October 2006) by two different sensors under different illumination conditions. The radial distortion of the images was corrected

and the intrinsic parameters of the cameras were accurately computed.

The paper is organized as follows. Section **??** explains in detail the steps to produce the synthetic images of the model that will be compared. Section **??** presents the results obtained. Conclusions and future work are discussed in section I.
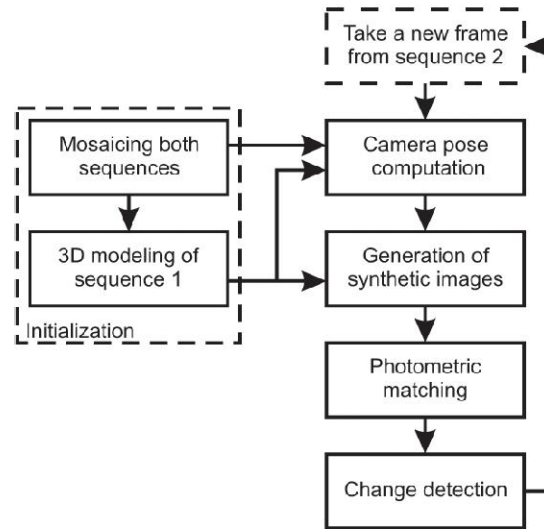

Fig. 1. Flowchart of the proposed method.

II. METHODOLOGY

The initialization of the proposed methodology is composed of 2 steps: (1) The mosaicing of the whole set of images and (2) the 3D reconstruction of one of the sequences (hereafter called reference sequence). After this initialization, for each image of the other sequence (hereafter called comparison sequence), four additional steps are needed: (1) computation of the camera pose (position and orientation); (2) Synthesizing a set of images from the 3D model; (3) Adjusting the intensity values of the set of synthetic images; (4) applying change detection techniques between the image and its set of synthetic images. Finally, a change mask is obtained for each image of the comparison sequence (see Fig. **??**).

A. Finding overlapping pairs

Aligning images that overlap in a common frame is known as mosaicing. By knowing the position of every image in a common frame called mosaic frame, it is possible to extract only the area of interest in order to reduce the number of images to process in the following steps.

Automatic algorithms for mosaicing are widely available [20], [21], [23]. They generally use feature-based motion estimation methods. These methods are composed of four main steps: (a) Feature detection [24]–[26], (b) Feature matching [25], [26], (c) Motion estimation with outlier rejection [27], [28] and (d) global alignment [29].

B. 3D reconstruction

The main problem of change detection algorithms is the alignment of the images. Since images are acquired at short range, the 3D relief of the scene introduces misregistration between the images. Change detection algorithms may take profit of knowing the 3D structure of the scene. This 3D model can be computed from one of the sequences. See [30] for a survey on 3D reconstruction methods.

In this paper, we used structure from motion [22] to estimate the model. The 3D structure is composed of a set of 3D vertices characterized by the local image descriptors resulting from feature tracking. Building this 3D model is achieved in two steps. The first step consists of initializing the structure of the scene. This is done by computing the camera motion between the first two frame and by determining the 3D positions of a set of matched points. In the second step, each subsequent camera pose is computed by matching image features with the 3D points of the model using a robust method. In order to update the 3D model, new 3D points are added by computing their positions from tracked features using the estimated camera poses (Fig. ??).

C. Computing the camera poses

For each image of the comparison sequence, a set of synthetic images are generated using the 3D model obtained from the reference sequence. These images are generated to simulate the view point of the comparison camera pose. The camera poses of the images in the comparison sequence are estimated by associating 2D features from these images with the 3D vertices in the model.

It is possible, especially when the sequences of images come from different cameras, that the matching method does not provide enough correspondences in order to compute an accurate camera position. In this case the number of 3D points to be matched with the comparison image can be reduced by taking only those that are present in its overlapping pairs of the modeled sequence. This reduces the number of potential wrong matches. There are cases in which additional correspondences have to be set manually in order to compute the camera pose accurately.

D. Synthesizing the images

Two images viewed from the same camera position do not suffer from any parallax effect. The main idea of the proposed method consists then in warping the images of the model in such a way that they are viewed from the same camera pose as that of the comparison image. This is possible by knowing the position and orientation of the camera that has to be simulated (see section ??).

The 3D points of the model are interpolated using Delaunay triangulation [31] in order to have a surface (Fig. ??). Knowing the intrinsic parameters of the camera in addition to its pose allows to compute the directions of the ray vectors of each pixel of the synthetic image. Their intersection with the 3D surface is then calculated generating a 3D position for each pixel. The intensity values of these pixels are extracted by back-projecting their 3D positions into the reference images. A bilinear interpolation of the pixel values is performed. This generates, for each image of the model, a synthetic version that represents the same View of the scene as that of the comparison image (Fig. ??). In this way, the two images have the same size and are already aligned up to the accuracy of the 3D model.

E. Photometric matching

The different datasets are likely to have different illumination conditions and different sensor responses resulting in different intensity values of the same scene points. A photometric matching technique is then used to correct these differences.

We applied a local histogram specification between the comparison image and its corresponding synthetic ones. The local histogram of a sub-window of the comparison image is extracted and this histogram is imposed in the same region of the synthetic images. The sub-windows have overlapping areas, to avoid abrupt changes in the histogram specification. By computing the average of the results over these windows, a set of photometrically matched synthetic images are obtained (Fig. ??).

F. Change detection Methods

For each image to compare, a set of synthetic images were generated from the reference ones. In this set, the parallax effects caused by the 3D relief has been reduced (see section ??). Their intensity values are matched with the comparison image (see section ??). Change detection techniques can then be applied between the comparison image and its synthesized reference images.

Change detection algorithms often generate a change image that represent a quantification of the changes for each pixel. In order to convert it to a change mask, a threshold must be applied. As one change mask is obtained for each synthetic image, they are combined in order to generate a general one. This can be done either by applying a voting scheme on the change masks or by averaging the change images before applying the threshold. This final mask is more accurate than the individual ones. Effectively, some false changes are triggered in the synthetic images because of moving objects such as fishes or local misregistrations due to the inaccuracies of the 3D model. These kinds of changes are generally detected at different positions of the synthetic images. Therefore, while generating the general mask, as the change is detected at the same position in few images, their contributions are less important and they are not taken into account.

III. RESULTS

In order to test the method described in the previous section, two sequences of images from a coral reef taken 10 months apart were used. The sequences were acquired with different cameras. Both cameras were accurately calibrated and the images were corrected for radial distortion. The scene presents an important 3D relief that produces parallax effects and
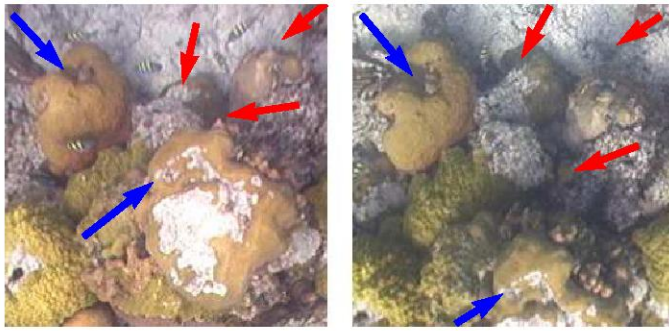
Fig. 2. Parallax effect due to 3D relief in two images of the same sequence: some features are occluded from one image to another (red arrows) making impossible the detection of changes with a 2D image alignment. Blue arrows show the same locations.

makes it impossible to detect changes directly from 2D image alignment (Fig. **??**).

Two 2D mosaics of these sequences of images were computed independently by matching SURF features (Fig. **??**).

The area around the same coral colony was extracted from both image sequences. The sequence from October 2006 (see Fig. **??**) was chosen arbitrarily as the reference, to generate the 3D structure of the scene. The traced SURF features were used to compute the positions of the 3D vertices and cameras poses. More than 15,000 3D vertices were extracted from the images (Fig. **??**).

Fig. **??** shows a sample image from the sequence of 2005. Local SURF features of this image are matched with a reduced set of the 3D points from the model built from the 2006 sequence in order to estimate its camera pose with respect to the model (see section **??**).

Next, the intersections of the pixel rays of the synthetic images with the 3D surface obtained by Delaunay triangulation are computed. The 3D position of each pixel (Fig. **??**) is then back projected in each image of the model in order to obtain its intensity values. A set of synthetic images is then generated (Fig. **??**). These synthetic images are photometrically matched to the comparison image that was selected (Fig. **??**).

The change detection method used for these results is an operation between two images that results in a "change image". Each pixel of the change image is obtained by summing the squared differences of the input images over a 7×7 pixel window centered in that pixel. Due to the limited accuracy of the 3D model, only three synthetic views were used to detect changes. These views were selected as having the highest overlap with the comparison image. The three different change images were averaged and thresholded to produce a final mask (Fig. **??**).

As it can be seen from Fig. **??**, the algorithm detects some changes in the bottom left of the image. Those changes are triggered because of algae that could not be modeled in the 3D structure as they are moving with the waves.

This kind of changes can be discarded by finding the changes between the synthetic images themselves. Effectively, as they are not modeled, these algae are also moving in the

synthetic images and a mask can be obtained. This change mask specify the regions in which there exists feature that could not be modeled such as moving algae or shadows.

As for finding the change between the two sequences, it is also possible to use the combination of change images coming from different pairs of synthetic images in order to have a better accuracy of the change mask. In this example, as three synthetic images were selected in order to detect the changes, it is possible to get three change masks between images (1-2, 1-3, 2-3). Fig. **??** shows the regions in which we can discard the detected changes since they have a high probability of being false positives. Fig. **??** shows, in red, the changes detected in the comparison image; in green, the changes detected within the sequence of synthetic images that can be discarded and, in yellow, the changes detected in the comparison image that are actually discarded. Finally, Fig. **??** shows the comparison image and one of its synthetic versions with the detected changes outlined in red and, in green, the areas where the change detection can be discarded.

It can be seen that the red areas which are not in close proximity to the green areas correspond to real changes in the benthos, such as the numbered tiles and the coral colony which were bleached (whiter) in the reference sequence but not in the comparison one.

IV. CONCLUSION AND FUTURE WORK

We have developed a method to deal with the constraints of the underwater environment in order to detect changes in sequences of images. Although preliminary, this paper presents promising results which validate the approach.

Synthesizing images from a 3D model presents several advantages. The main one is that it minimizes the impact of the relief on the registration, which is the main problem for change detection algorithms.

The proposed algorithm incorrectly detects changes in the areas of the images that could not be correctly modeled or that could not be modeled at all. This can happen when moving objects like fishes or algae are present in the scene. In some sequences of images the light source is also moving, resulting in moving shadows. Fortunately, it is possible to detect those areas by finding the changes within the sequence of synthetic images, and thus remove them from the final "change mask".

There are several ways to improve the results of this method. One direction is to improve the accuracy of the 3D model since it has a direct impact on the accuracy of the synthetic images. The 3D modeling algorithm can be improved by using a dense matching approach such as match propagation [32].

In this paper we used a standard change detection algorithm which produced good results for the test sequences. However, other more sophisticated algorithms such as the neural network algorithm proposed in [9] need to be implemented and tested.

Another direction is in combining the change masks obtained for each image of the comparison sequence into one single view. If the combination of masks is done in a 2D mosaic, misalignment of changes will appear. As the 3D position of every pixel of the synthetic images is known, it is
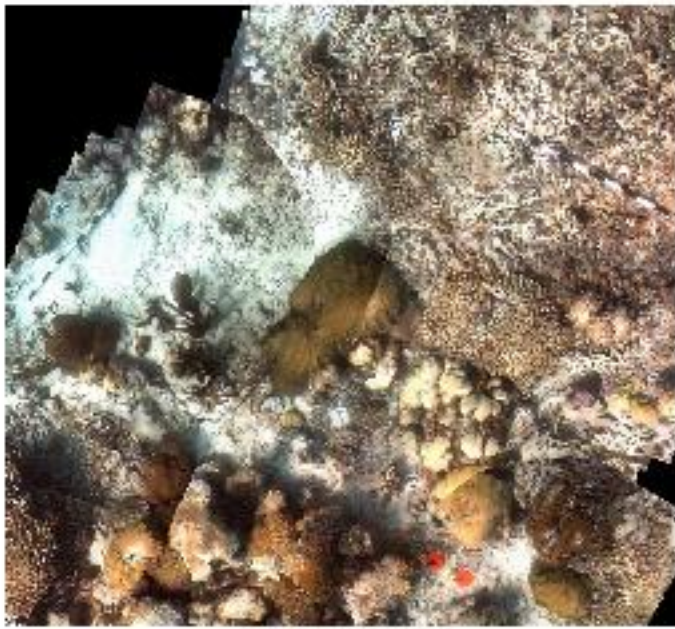
(a) Mosaic from 2005
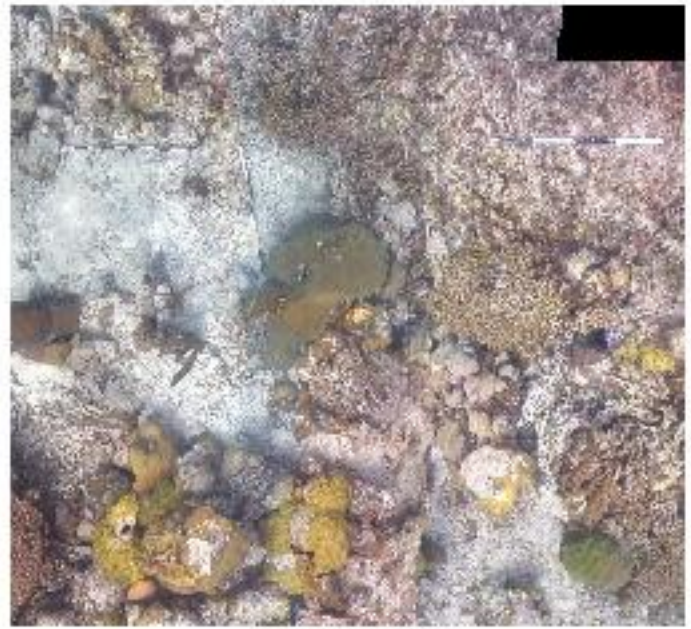
(b) Mosaic from 2006

Fig. 3. Cropped versions of the mosaics of image sequences from December 2005 (a) and October 2006 (b). Note the significant 3D relief of the scene and the changes in illumination and image textures between the two sequences.
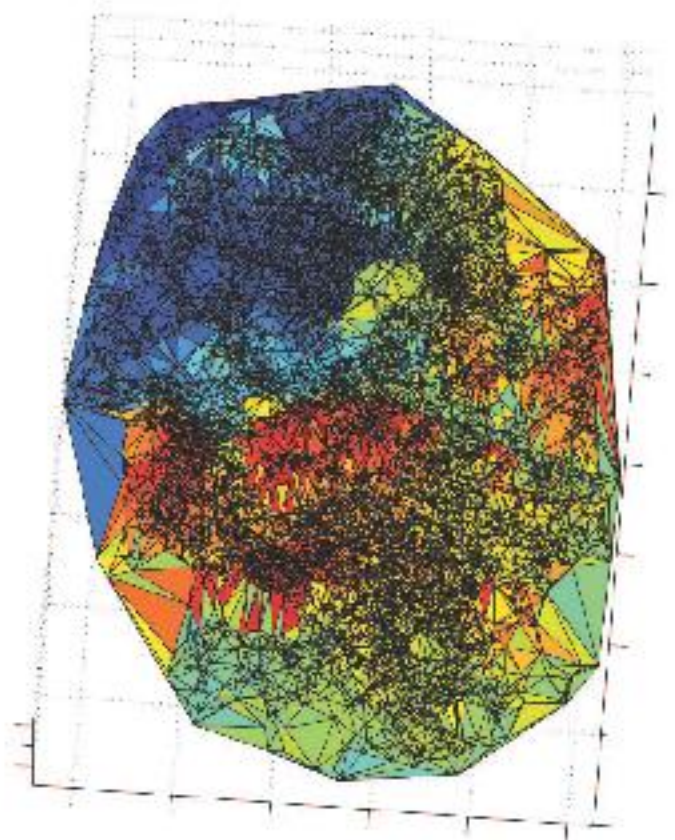

Fig. 4. 3D model of the image sequence from October 2006. It contains more than 15,000 vertices and the surface is obtained by Delaunay triangulation. Note the important 3D relief. Red corresponds to higher elevation and blue to lower one.

Fig. 5. Image that is compared to the sequence used to generate the 3D model

better to combine the change mask into a 3D mask. Moreover, if there is high overlap between the images of the comparison sequence, there is redundant information in each mask and the final 3D mask should be more accurate.

Finally, in the current implementation, several parameters such as the size of the sub-windows for the photometric matching or the threshold in the change detection techniques had to be manually adjusted. Optimal parameters should be investigated.

(a) Comparison Image   (b) Synthetic Image   (c) Synthetic image photometrically corrected

Fig. 6. This triplet of images shows the different steps in order to obtain the images to compare. One image from the sequence of 2005 (a) is selected. Its camera pose is computed in order to generate a set of synthetic images. Each image of this set is then photometrically matched. (b) shows one of the synthetic images and (c), the same image photometrically matched. The change detection algorithms are applied on (a) and (c).
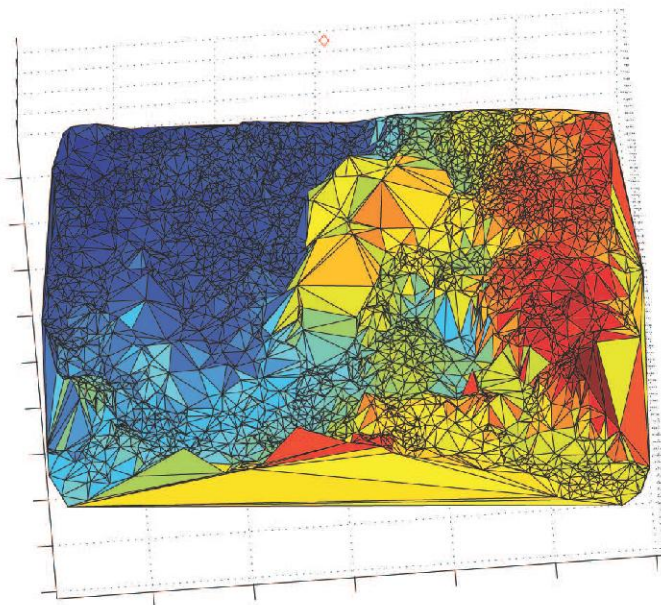
Fig. 7. 3D model of the image to synthesize. Note the similarity with the image to compare (figure ??) and the inaccuracy of the model.

REFERENCES

[1] J. Mas, "Monitoring land-cover changes: A comparison of change detection techniques," *International Journal of Remote Sensing*, vol. 20, pp. 139–152, Jan 1999.

[2] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," *International Journal of Remote Sensing*, vol. 25, pp. 2365–2407, Jun 2004.

[3] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," *IEEE Transactions on Image Processing*, vol. 14, pp. 294–307, Mar 2005.

[4] L. Ulander, M. Lundberg, W. Pierson, and A. Gustavsson, "Change detection for low-frequency sar ground surveillance," *IEEE Proceedings on Radar, Sonar and Navigation*, vol. 152, pp. 413–420, Dec 2005.

[5] P. Gamba, F. Dell'Acqua, and G. Lisini, "Change detection of multitemporal sar data in urban areas combining feature-based and pixel-based techniques," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, pp. 2820–2827, Oct 2006.

[6] L. Marcenaro, F. Oberti, and C. Regazzoni, "Change detection methods for automatic scene analysis by using mobile surveillance cameras," in *IEEE International Conference on Image Processing*, vol. 1, Sep 2000, pp. 244–247.

[7] Y. Matsuo, A. Shimizu, and H. Kobatake, "An interval change detection method for two chest x-ray images with different rotation angles of the human body and its performance evaluation," *Syst. Comput. Japan*, vol. 36, no. 10, pp. 30–42, 2005.

[8] D. D. Gomez, C. Butakoff, B. Ersboll, and J. Carstensen, "Automatic change detection and quantification of dermatological diseases with an application to psoriasis images," *Pattern Recognition Letters*, vol. 28, no. 9, pp. 1012–1018, 2007.

[9] G. Pajares, "A hopfield neural network for image change detection," *IEEE Transactions on Neural Networks*, vol. 17, pp. 1250–1264, Sep 2006.

[10] J. Sun, T. Aib, C. Zha, and H. Yan, "Assessing vegetation degradation in loess plateau by using potential vegetation index," *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pp. 1794–1797, Jul 2007.

[11] B. Qiu, V. Prinet, E. Perrier, and O. Monga, "Multi-block pca method for image change detection," in *IEEE International Conference on Image Analysis and Processing*, vol. 00. Los Alamitos, CA, USA: IEEE Computer Society, 2003, p. 385.

[12] C. Munyati, "Wetland change detection on the kafue flats, zambia, by classification of a multitemporal remote sensing image dataset," *International Journal of Remote Sensing*, vol. 21, no. 9, pp. 1787–1806, Jun 2000.

[13] F. B. F.B. Sanli and C. Goksel, "Defining temporal spatial patterns of mega city istanbul to see the impacts of increasing population," *Environmental Monitoring and Assessment*, Dec 2007.

[14] A. Gleason, D. Lirman, D. Williams, N. Gracias, B. Gintert, H. Madjidi, R. Reid, G. Boynton, S. Negahdaripour, M. Miller, and P. Kramer, "Documenting hurricane impacts on coral reefs using two-dimensional video-mosaic technology," *Marine Ecology*, vol. 28, no. 2, pp. 254–258, Jun 2007.

[15] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, pp. 1566–1577, Sep 1998.

[16] J. Townshend, C. Justice, C. Gurney, and J. McManus, "The impact of misregistration on change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, pp. 1054–1060, Sep 1992.
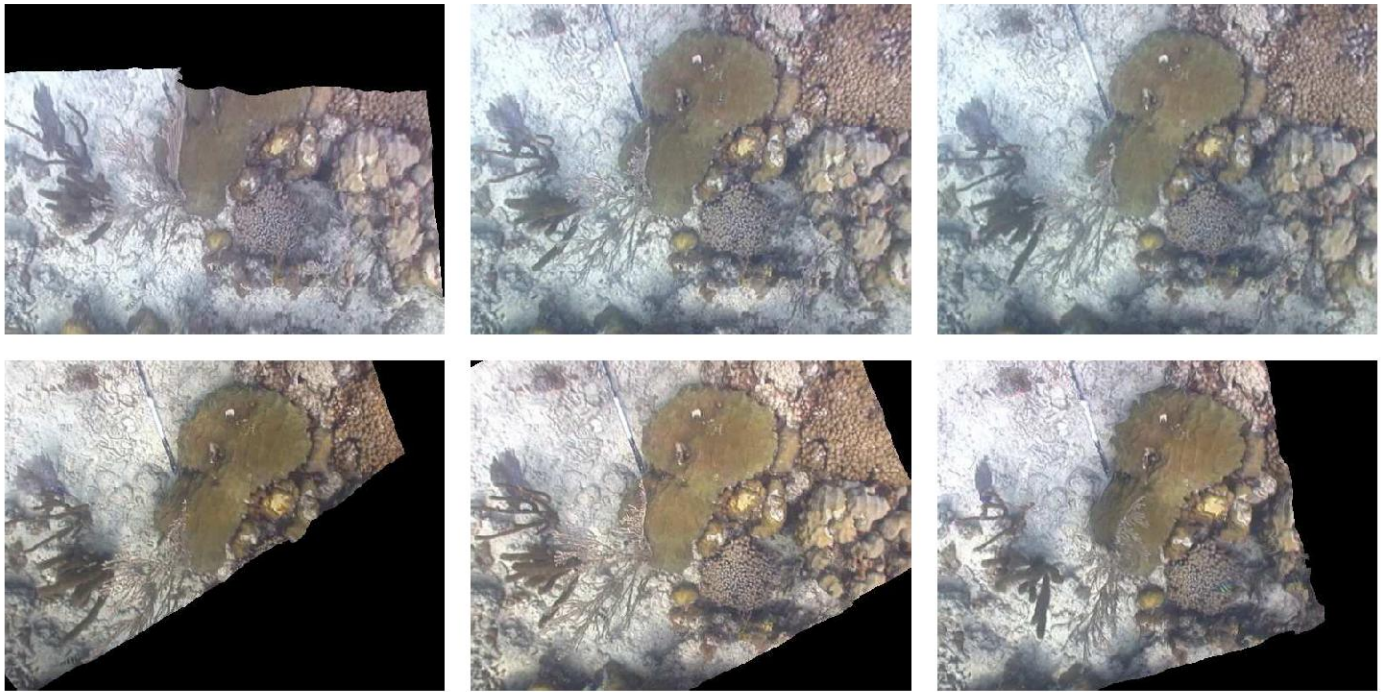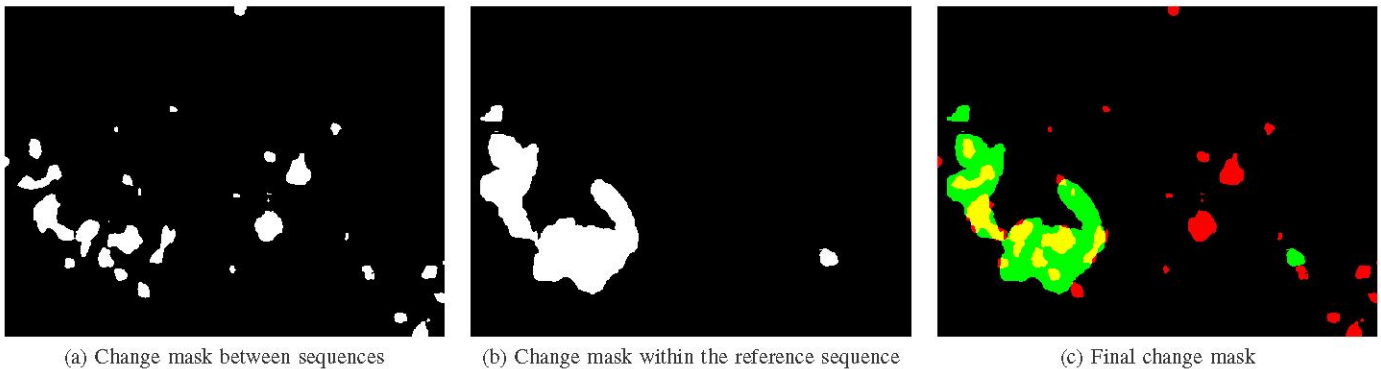
Fig. 8. Examples of synthetic images obtained from the reference sequence used to generate the 3D structure of the scene.

(a) Change mask between sequences   (b) Change mask within the reference sequence   (c) Final change mask

Fig. 9. (a) Change mask obtained by applying a threshold on the average of the change images between the image to compare and its synthetic images. Note that some wrong matches are detected in the bottom left of the mask. (b) Change mask obtained within the sequence of synthetic images. This mask shows the areas where changes can appear but should not be taken into account. (c) Change mask showing in red the areas where change has been detected, in green, the areas that can be discarded and, in yellow, the areas that where initially detected as "change" but finally discarded.

[17] J. Stamm and R. Briggs, "Change detection in digital orthophotos," in *GIS/LIS Conference Proceedings*, Nov 1998.

[18] M. Black, D. Fleet, and Y. Yacoob, "Robustly estimating changes in image appearance," *Computer Vision and Image Understanding*, vol. 78, pp. 8–31, 2000.

[19] K. Lebart, C. Smith, E. Trucco, and D. Lane, "Automatic indexing of underwater survey video: Algorithm and benchmarking method," *IEEE Journal of Oceanic Engineering*, vol. 28, pp. 673–686, Oct 2003.

[20] R. Szeliski and H. Shum, "Creating full view panoramic image mosaics and environment maps," in *Special Interest Group in GRAPHics*, vol. I, Los Angeles, CA, USA, August 1997, pp. 251–258.

[21] A. Leone, C. Distante, A. Mastrolia, and G. Indiverr, "A fully automated approach for underwater mosaicking," in *MTS/IEEE OCEANS Conference*, September 2006, pp. 1–6.

[22] T. Nicosevici and R. Garcia, "Online robust 3d mapping using structure from motion cues," in *MTS/IEEE OCEANS Conference*, Kobe, Japan, 2008.

[23] D. Lirman, N. Gracias, B. Gintert, A. Gleason, R. P. Reid, S. Negah-daripour, and P. Kramer, "Development and application of a video–mosaic survey technology to document the status of coral reef communities," *Environmental Monitoring and Assessment*, vol. 159, pp. 59–73, 2007.

[24] C. G. Harris and M. J. Stephens, "A combined corner and edge detector," in *Alvey Vision Conference*, Manchester, U.K., 1988, pp. 147–151.

[25] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[26] T. T. H. Bay and L. V. Gool, "Surf: Speeded up robust features," in *European Conference on Computer Vision*, Graz, Austria, may 2006, pp. 404–417.

[27] P. Meer, D. Mintz, and A. Rosenfeld, "Analysis of the least median of squares estimator for computer vision applications," in *IEEE Conference on Computer Vision and Pattern Recognition*, September 1992, pp. 621–623.

[28] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated

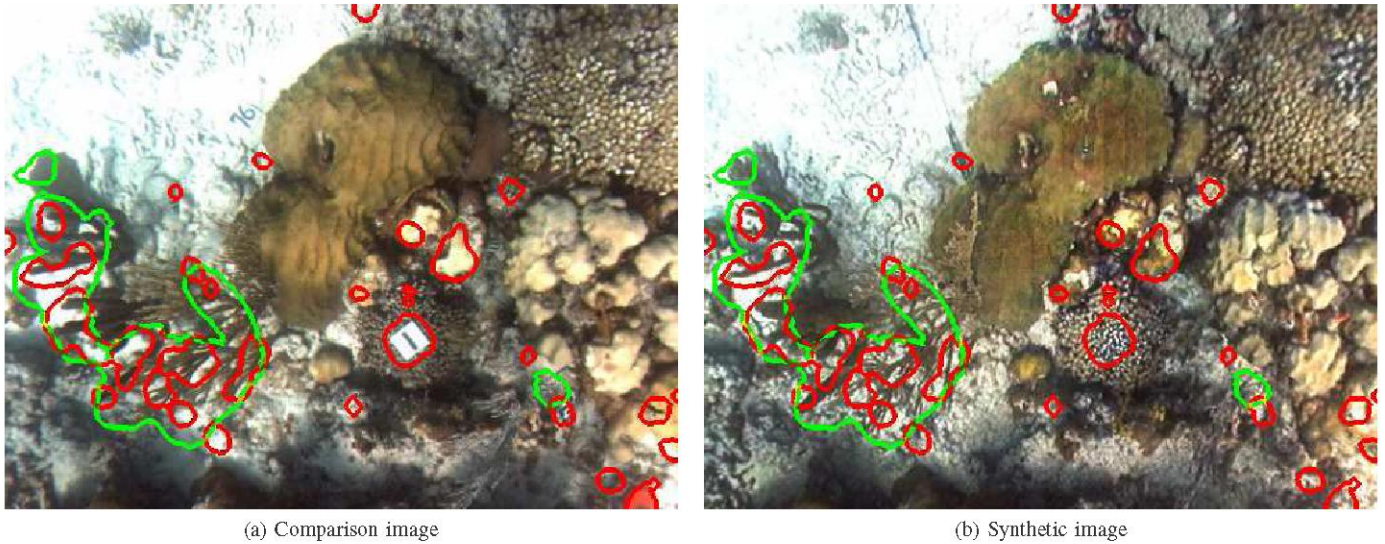(a) Comparison image   (b) Synthetic image

Fig. 10. The comparison image (a) and one of its synthetic version (b) with the changes detected outlined in red and the areas where changes detected can be discarded outlined in green

cartography," *Comm. Assoc. Comp. Mach.*, vol. 24, no. 6, pp. 381–395, 1981.

[29] H. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," in *European Conference on Computer Vision*, vol. II, Freiburg, Germany, June 1998, pp. 103–119.

[30] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[31] J. Foley, A. V. Dam, S. Feiner, J. Hughes, and R. Phillips, *Introduction to Computer Graphics*. Addison-Wesley Professional, 1993.

[32] M. Lhuillier and L. Quan, "Match propagation for image-based modeling and rendering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 8, pp. 1140–1146, Aug 2002.