ROBUST VIDEO MOSAICING FOR BENTHIC HABITAT MAPPING

Hiêp Quang Luong and Wilfried Philips
Ghent University, Department of Telecommunication and Information Processing, Image Processing and Interpretation Research Group
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

Anneleen Foubert
Ghent University, Department of Geology and Soil Science, Renard Centre of Marine Geology
Krijgslaan 281 S.8, 9000 Ghent, Belgium

Keywords: Video registration, mosaicing, M-estimators, outlier rejection.

Abstract: Nowadays remotely operated vehicles (ROV) have become a popular tool among biologists and geologists to examine and map the seafloor. For analytical purposes, mosaics have to be created from a large amount of recorded video sequences. Existing mosaicing techniques fail in case of non-uniform illuminated environments, due to the presence of a spotlight mounted on the ROV. Also traditional image blending techniques suffer from ghosting artifacts in the presence of moving objects. We propose a general observation model and a robust mosaicing algorithm which tackles these major problems. Results show an improvement in visual quality: noise and ghosting artifacts are removed.

1 INTRODUCTION

Still in recent past, benthic sampling techniques, using box corers, Van-Veen grabs or dredges, were the most common tools for biologists and geologists to examine the seafloor and to groundtruth geophysical datasets (sidescan sonar, multibeam, seismic). However, during the last decade the use of ROV's and submersibles became more and more widespread in marine research. The increasing ROV-based exploration and visualization of deep-water environments revealed a large number of new insights in already long studied environments. These new technologies producing a large amount of visual data (formerly not available as such) are claiming for new analytical methods, both qualitative and quantitative.

Image mosaicing is the process that warps a collection of overlapping images into a common coordinate system and that merges the overlapping regions of the warped images into a single image which covers the entire visible area of the scene. The merged output image is called the mosaic or the panorama. Registration is finding the appropriate transformation of an input image or a set of input images with respect to a reference image. Many registration methods require users interaction through selecting ground control points (GCP) in the reference image and their corresponding points in the input image. GCP's are a set of selected pixels (or regions) that contains important features like intersection of roads or coastlines. However manual registration requires a lot of time and labour and is furthermore not accurate due to human mistakes (Luong et al., 2004). The huge amount of incoming video data from new missions mandate the need for automatic registration.

Several automatic registration techniques do exist, they can roughly be divided into two categories: area-based (e.g. minimizing the overlapped intensity differences, Fourier-based methods, etc.) and feature-based methods (Zitova and Flusser, 2003; Lowe, 2004; Tuytelaars, 2000). Intensity-based registration methods require a uniform illumination throughout the video sequence and is consequently not suitable for our application. Most blending methods use an averaging scheme which results in ghosting effects when dealing with moving objects. These two major issues require novel techniques. In the next chapters, we propose a general observation model, we describe our algorithm and we discuss the results.

2 THE OBSERVATION MODEL

In our specific case, we propose an extension of the generic model used in (Pires and Aguiar, 2005). Each pixel of the ith video frame I_i can be modeled as a noisy sample of the panorama P which can additionally be occluded by a moving object O. The moving

Quang Luong H., Philips W. and Foubert A. (2006).
ROBUST VIDEO MOSAICING FOR BENTHIC HABITAT MAPPING.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 306-310
DOI: 10.5220/0001368603060310

objects in this situation are benthic animals (see figure 1). The local illumination changes due to the spotlight of the ROV are modeled by the filter H, also the end of a floating rope from the ROV (which is visible on a somewhat static position throughout the video sequence) will be incorporated by H (see figure 1). Since a video frame I_i has only a limited view of the panorama, we truncate its view with a binary region of interest mask M_i. If the region is observed by the ith image, then $M_i(\underline{x_i})$ will become 1 otherwise 0. The generic observation model will become

$$I_i(\underline{x_i}) = [[(1 - \delta_{\underline{u}-\underline{u_t}})P(\underline{u}) + \delta_{\underline{u}-\underline{u_t}}O(\underline{u_t})] \cdot H(\underline{x_i}) + N(\underline{x_i})]M_i(\underline{x_i}) \quad (1)$$

where N denotes the sum of the noise generated with and without the spotlight filter H. In the rest of this paper we assume that the noise has a zero-mean Gaussian distribution. The image coordinates $\underline{x_i} = (x_i, y_i)$ are expressed in the coordinate system of the image I_i, while the coordinates $\underline{u} = (u, v)$ are expressed in the coordinate system of the panorama P, which is in our case the same as the coordinate system of the reference image I_0. Since the objects O are moving, the coordinates $\underline{u_t}$ are related to time. The function $\delta_{\underline{u}-\underline{u_t}}$ is the Dirac delta function, which yields 1 if $\underline{u}$ equals to $\underline{u_t}$ and zero otherwise. If the delta function is 1 than the panorama is occluded by the moving objects and we deal with an unoccluded panorama otherwise. Note that the delta function is only defined on the discrete grid, which means that no subpixel coordinates are used here.

The relationship between the reference coordinate system of the panorama P and the coordinate system of the images I_i is denoted for example by a global parametric mapping model. Perspective projection models (8 degrees of freedom) and polynomial models, such as translation, affine or biquadratic transformations (with respectively 2, 6 and 12 degrees of freedom), are very common in use (Zitova and Flusser, 2003). The coordinates $\underline{x_i}$ and $\underline{u}$ are related by

$$\underline{x_i} = m(\underline{\theta_i}; \underline{u}) \quad (2)$$

We have to estimate first the parameters of the mapping model, this is known as the registration problem. Our final goal is then to recover the panorama P in good lighting conditions as much as possible without the moving objects, which is related to robust background estimation.

3 THE MOSAICING SYSTEM

As we have mentioned in the previous chapter, we have to estimate the parameters $\underline{\theta_i}$ of the mapping model m first. Since the parameter space, for direct mapping between the image I_i and the panorama P, is very large, we register the images sequentially in order to reduce the computation time. This produces good initial estimations, but it still can lead to propagation errors, which is clearly visible for non-consecutive video frames covering the same region of the panorama. In the second step, the estimation of the parameters is then corrected by registering the images I_i with a temporary build mosaic. Deriving a global optimal solution for the parameters $\underline{\theta_i}$ as in (Pires and Aguiar, 2005) is very difficult because of the presence of the spotlight. Because we a priori know how the spotlight filter H approximately behave, we can use a predefined weight map W to model the behaviour of H. The weights $W(\underline{x_i})$ are low if we want to exclude a specific region. In the last step, the final mosaic P is estimated by combining the overlapped frames.

Figure 1: An original image from the video sequence. The poor illumination conditions and the presence of benthic animals (and the floating rope visible at the up middle part of the image) make mosaicing much more difficult.

3.1 Robust Image Registration

Because of the spotlight, we can not use featureless registration methods. So we build a robust feature point matching algorithm in order to register subsequent images I_i and I_{i+1} and estimate parameters $\underline{\theta'_{i+1}}$. Since areas of uniform intensity,i.e. with no structural information, do not give us reliable registration information, we need to find well textured blocks. We apply the Noble corner detection (Noble, 1988) as the feature point detector on image I_i. Since we only use the feature point detector to distinguish areas with uniform intensity from ares with interesting structural information, there is no point using more complex (in-

variant) detectors (and descriptors) as in (Lowe, 2004; Tuytelaars, 2000). We define the (squared) blocks B_1 on image I_i around the detected feature points. These blocks are matched with blocks B_2 from image I_{i+1} using the weighted zero-mean normalized cross-correlation (CC):

$$CC = \frac{\sum_i w_i(B_{1,i} - \overline{B_1})(B_{2,i} - \overline{B_2})}{\sqrt{\sum_i w_i(B_{1,i} - \overline{B_1})^2 \sum_i w_i(B_{2,i} - \overline{B_2})^2}} \quad (3)$$

where $\overline{B_1}$ and $\overline{B_2}$ are denoted as the mean values of respectively blocks B_1 and B_2. This correlation measure is illumination invariant, i.e. blocks with a biased illumination change will yield the same correlation as blocks with no biased illumination change. The weights w_i are chosen to favour the central part of the window (for example with a Gaussian function). Higher (subpixel) accuracy is obtained by fitting the neighbourhood of the highest correlation coefficient to a second degree polynomial model.

In the next step, we have to estimate the parameters of our transformation model m using the matched pairs. The influence of the worst matches (outliers) should be minimized. A robust estimate of these parameters can be achieved with Hough transforms, RANSAC, LMeds, M-estimation, bootstrap methods, etc. (Rousseeuw and Leroy, 1987). Based on the generalized maximum likelihood and least squares formulation, we will use M-estimators. In particular, the M-estimate of $\underline{a}$ is

$$\widehat{\underline{a}} = \arg\min_{\underline{a}} \sum_i \rho(r_{i,\underline{a}}) \quad (4)$$

where ρ is a robust loss function and r_i is the scale normalized residual. A good (robust) initialization is crucial for the success of M-estimation, otherwise it would yield poor results due its low breakdown point (Stewart, 1999). A robust initialization is achieved using a coarse-to-fine multiresolution framework. In the coarsest level, we can use temporal information from the registration between I_{i-1} and I_i, which also additionally reduces the computation time. Using Kalman or particle filtering could result in a better prediction (Doucet et al., 2000). But in this case, we keep it simple: we use the previous estimation as the new prediction. Solving this robust regression problem leads to W-estimators and the iterative reweighted least squares (IRLS) algorithm (Stewart, 1999). In each iteration, the weights of each pair are adapted in function of their residuals and a weighted least squares (WLS) algorithm is applied until convergence is reached. In order to recover numerical stable parameters, singular value decomposition is used to solve the linear system in the WLS algorithm. We initialize the weights of the IRLS algorithm with CC information: if a matched pair has a high correlation (hence is more reliable), then it should have more influence on the parameter estimation. After applying IRLS, we do not only have an estimate for parameters $\underline{\theta'_{i+1}}$, but also the final output weights which represent the importance of each contributing pair. With this information we can exclude bad registered regions (typically caused by moving objects) in all levels of the hierarchical framework.

The combination of the transformation parameters $\underline{\theta'_{i+1}}$, which are obtained from the registration between subsequent images, and the parameters $\underline{\theta_i}$, which are obtained between the previous image I_i and the panorama P, form a good initial estimation for the parameters $\underline{\theta_{i+1}}$ from the registration between I_{i+1} and P. We correct the parameters $\underline{\theta_{i+1}}$ using the same previously described algorithm and update the provisional mosaic with image I_{i+1}. Since the next image I_{i+2} has the most similar features as image I_{i+1} (taking the spotlight into account), more weights are assigned to the last image when blending it into the provisional mosaic using an averaging scheme. The whole process is now repeated for image I_{i+2}.

3.2 Robust Image Fusion

After transformation and resampling of the images I_i (using the 8-point windowed Blackman-Harris sinc function), we have a vector of candidates for each pixel of the panorama P. Simple averaging will create severe artifacts due to non-uniform illumination conditions, moving objects and possible misregistration. We can tackle this illumination problem by assigning weights to each candidate pixel proportional to the weights $W(\underline{x_i})$. Since we are interested in a panorama in good lightning conditions, the weights $W(\underline{x_i})$ for dark regions will tend to zero.

Moving objects can be modeled as a non-zero mean Gaussian distribution and we classify misregistration to the noise N. With these considerations, each candidate vector is observed as a weighted mixture of Gaussians. Since we are only interested in the single Gaussian density which represents the background, we want to suppress the influence of other densities by lowering the weights of the candidates which are part of the moving objects. Similar to background subtraction techniques (Radke et al., 2005), we calculate the (weighted) average of all candidates. Afterwards we compare this average to all candidates, if the absolute difference exceeds a certain threshold (typically a number of standard deviations from the mean background model), then the candidate belongs most likely to an object and its weight is set to zero. The new weighted average is a good first estimate,

which we further refine to the background peak using robust M-estimation. We recall the fact that we can not recover the panorama P perfectly if an object covers the same region on every image. The candidates are represented by their luminance component in the CIELab colour space. Afterwards, their output weights, retrieved from the IRLS algorithm, are used to combine the separate channels in the RGB colour space.

4 RESULTS

In this paper we have described our algorithm using a generic transformation model m. In our implementation, we use the perspective projection model (8 parameters):

$$u = \frac{\theta_{i,0} + \theta_{i,1}x_i + \theta_{i,2}y_i}{1 + \theta_{i,6}x_i + \theta_{i,7}y_i} \quad (5)$$

$$v = \frac{\theta_{i,3} + \theta_{i,4}x_i + \theta_{i,5}y_i}{1 + \theta_{i,6}x_i + \theta_{i,7}y_i} \quad (6)$$

Equation 4 can be solved using IRLS with a weight function $w(r) = \rho'(r)/r$ (W-estimator). After testing several robust loss functions, we find the logistic function $\rho_{logistic}$ and the Cauchy function ρ_{cauchy} give the best performance respectively to registration and to image fusion. The corresponding weight functions are

$$w_{logistic}(r) = \frac{\tanh r}{r} \quad (7)$$

$$w_{cauchy}(r) = \frac{1}{1 + r^2} \quad (8)$$

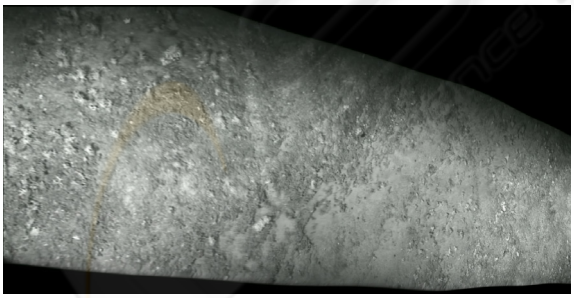
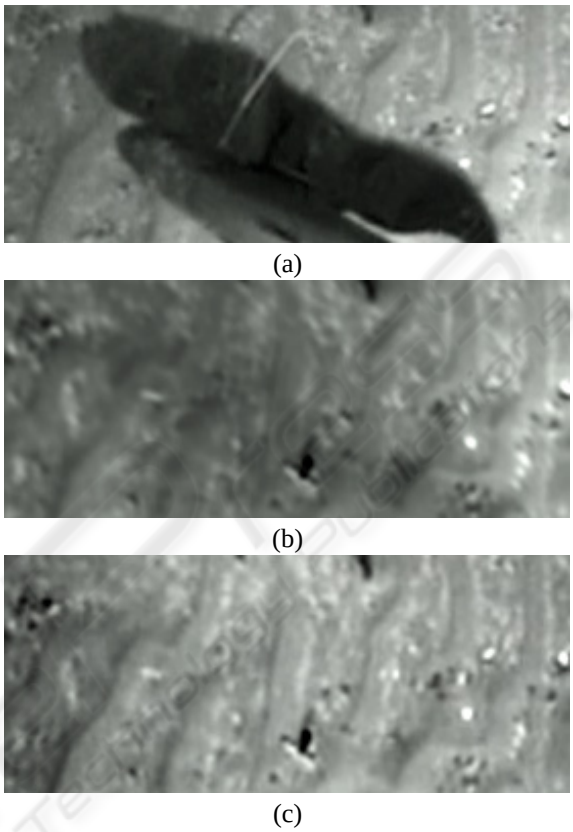


Figure 2: A part of a mosaic showing typical dead coral facies and sediment clogged dead coral.

In figure 2, a part of a mosaic is shown. The panorama is recovered with less illumination artifacts. In figure 3 we see a region where a moving object is removed. Both results are created using the same image registration parameters. Our proposed image fusion outperforms traditional blending and additionally it also improves the image quality (compared to the original image): noise and compression artifacts are reduced. Also ghosting effects (due to moving benthic fauna) are removed.



(a) (b) (c)

Figure 3: A contrast enhanced detailed region: (a) original image, merging using (b) an averaging scheme and (c) our proposed method.

5 CONCLUSION

We have proposed a generic mosaicing model and we have presented a mosaicing algorithm which can handle video sequences recorded in a non-uniform illuminated environment. Additionally our algorithm can deal with moving objects. Robust M-estimation is used in the image registration as well as in image merging. Our proposed algorithm reconstructs the mosaic in good lighting conditions. Results show also an improvement in visual quality: noise and ghosting artifacts are removed.

ACKNOWLEDGEMENTS

The authors would like to thank the captain, crew, scientific party and especially the VICTOR-team from IFREMER on board of R/V Polarstern (2003) during the ARK-XIX/3a cruise. The ROV imagery is courtesy and copyright of IFREMER. Anneleen Foubert is PhD student funded through an FWO-fellowship.

REFERENCES

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10:197–208.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Luong, H., Gautama, S., and Philips, W. (2004). Automatic registration of synthetic aperture radar (sar) images. In *Proc. of IEEE International Geoscience and Remote Sensing Symposium 2004*, pages 3864–3867.

Noble, J. (1988). Finding corners. *Image and Vision Computing*, 6(2):121–128.

Pires, B. and Aguiar, P. (2005). Featureless global alignment of multiple images. In *Proceedings of IEEE International Conference on Image Processing 2005*, pages 57–60.

Radke, R., Andra, S., Al-Kofahi, O., and Roysam, B. (2005). Image change detection algorithms: A systematic survey. *IEEE Trans. on Image Processing*, 14(3):294–307.

Rousseeuw, P. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley Series in Probability and Mathematical Statistics.

Stewart, C. (1999). Robust parameter estimation in computer vision. *SIAM Reviews*, 41(3):513–537.

Tuytelaars, T. (2000). *Local Invariant Features for Registration and Recognition*. PhD Dissertation KULeuven.

Zitova, B. and Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000.