Proceedings of the 10th WSEAS International Conference on COMPUTERS, Vouliagmeni, Athens, Greece, July 13-15, 2006 (pp1141-1146)

Underwater archaeological mosaicing

F.Bellavia, G.Gagliano, D.Tegolo, C.Valenti
Dipartimento di Matematica e Applicazioni,
Università degli Studi di Palermo, Italy
{tegolo,cvalenti}@math.unipa.it

Abstract: - Archaeological mosaicing is one of the challenges of the computer vision community and it can be faced in a 2D or 3D approach. This contribution regards a methodology to do a mosaic of an underwater bi-dimensional scene. A number of problems arise from the acquisition of images by a remote operated vehicle. Radial distortion, poor luminosity, cloud water, presence of artefacts are part of the issues that can occur; for instance, the radial distortion has been corrected to improve the quality of the input images. Keypoints detection (through SIFT transform), Singular Value Decomposition, Random Samples Consensus are some of the techniques applied in our method.

Key-Words: - Radial distortion, SIFT, RANSAC, underwater mosaicing.

1 Inroduction

Mosaicing is the process of merging overlapping photos into a single view, usually by assuming planar textures. We are interested in the underwater exploration of sites with wrecks together with their shipments. This task is useful to verify the condition of the area and to roughly estimate the relative positions of the findings, in order to better prepare following recovery campaigns. A further advantage is represented by the fact that the data can be analyzed offline in different laboratories, museums and schools by using low cost personal computers.

A Remotely Operated Vehicle (Comex Pro Super Achille 2000) has been equipped with two high sensitivity black and white cameras (Hitachi KP). It must be pointed out that actually the results presented here have been obtained by taking into account just one of these cameras.

In underwater environments, due to poor visibility conditions, a lot of small areas need to be close up photographed. Though each image overlaps the successive one, it is very hard to combine by hand all these photos to get a huge mosaic that consists of the whole site; in our case the site measures 11×4 meters.

It is possible to formulate this task as a global minimization problem, which needs to compute the homographies that map every photo onto the mosaic image. Each new frame is related to the previous one to find corresponding points. Surveys about the state of the art techniques on this research field can be found in [1, 2].

We have developed an unsupervised technique that compose the frames coming from the camera via a process that can be sketched as: image normalization (described in Section 2), extraction and matching of features (Section 3), computation of the homography and warping of consecutive images (Section 4). Conclusions are reported in Section 5.

2 Correction of radial distortion

Usually cameras introduce a significant amount of distortion which can be of great impact on large mosaics. This aberration depends on the difference in refractive index between water and air, on the particular lenses and it is more evident faraway the center of the image. Our model assumes tangential distortion as irrelevant with respect to radial distortion which is approximated to the first order coefficient:

$$\begin{cases} x' = x + kx(x^2 + y^2) + O(x^2, y^2) \\ y' = y + ky(x^2 + y^2) + O(x^2, y^2) \end{cases}$$

where the center lies on the origin of the system, (x', y') indicate the coordinates in the corrected image and k represents the amount of radial distortion [3].

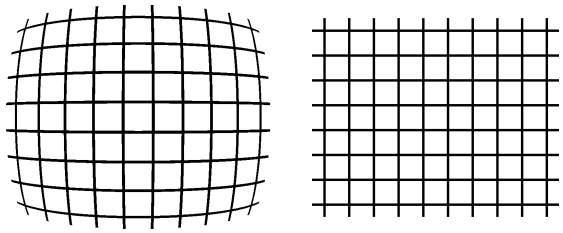

Figure 1: Example of radial distortion correction.

The algorithms already present in literature to correct geometric distortions can be divided into two main classes.

In the first case, the internal parameters (e.g. focal distance, pixel ratio, image center, scale factor, radial and tangential aberration) and the external parameters (i.e. roto-translation matrix) of the acquisition tool are extracted thought the correspondences of a set of known points in the three-dimensional space [4, 5]. The drawback of this method relies on the intrinsic dependence of these parameters [6].

In the second case, a "pin-hole" camera model is assumed to compute these parameters so to mantain the aspect of straight lines. We have slightly modified the error function used by the recursive algorithm presented in [7]. That method avoids the problems typical of analytical approaches (e.g. Gauss-Newton, Levenberg-Marquardt) which require the definition of a differentiable error function. In particular, iterative methods seem to return more accurate results than those obtained by a least square approach (e.g. Moore-Penrose), as indicated in Figure 2. Moreover, our implementation is unsupervised because it does not require a manual selection of the edges already extracted by a Canny operator.

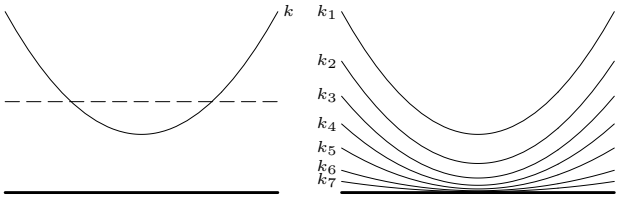

Figure 2: The least squares approximate the regression line, dashed (*left*). On the contrary, successive steps tend to the real line, bolded (*right*).

This edge detector is applied on a real pattern to locate the lines for verifying the response of the algorithm (see Figure 3). The measure of the error introduced during each step is computed as the mean of the squared differences between the horizontal (vertical) position of the points of the lines in the pattern and the horizontal (vertical) position of the corresponding points computed by linear regression. The final choice of the horizontal or vertical direction depends on the slope of the least squares fitting regression line that implies the smallest error value.

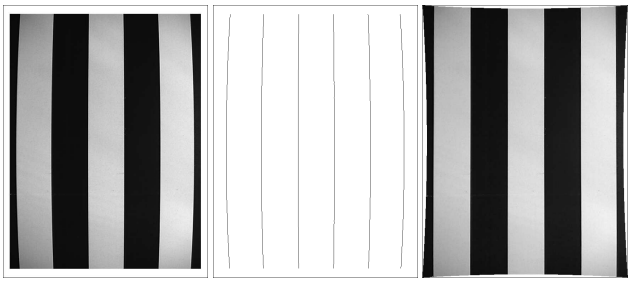

Figure 3: Vertical pattern (*left*), its edges (*center*) and corrected version (*right*). The use of both the vertical and horizontal patterns returns a better value of k.

It can be proved that the error function is convex and that it can be expressed in terms of the parameter k. The algorithm divides the range of k into at least 3 intervals and centers the new range on the position of k that gave the minimum error during the previous iteration. This procedure ends and returns the last value of k when the width of the newly computed range is smaller than a fixed threshold. Figure 4 shows that a few iterations were sufficient to reach a radial distortion error almost equal to 0.

Figure 4: Error versus the width of range of k.

3 Keypoints detection and matching

Due to the lack of luminosity at a depth of just a few meters, it is necessary to observe the seabed in the downwards direction, putting the camera close to the bottom. This introduces severe prospective errors when the frames of the video have to be composed. In order to properly superimpose the images, a set of anchorage points must be detected on each of them. A successive module will single out all common keypoints between pairs of consecutive frames and will calculate the coefficients of the function, called homography, to modify the images so as to reduce the effects of prospective.

The "scale-space" model, based on the computation of a set of Difference of Gaussian maps, to locate promising keypoints, has been introduced in [8, 9, 10]. This algorithm performs edge detection by blurring the input image I through a couple of Gaussian filters centered in the origin and with predefined standard deviations $\sigma_1 = \sqrt{2}$ and $\sigma_2 = 2$:

$$L_i = G_i \otimes I,$$

where

$$G_i(x, y) = \frac{e^{-\frac{x^2+y^2}{2\sigma_i^2}}}{2\pi\sigma_i^2}.$$

Each convolution eliminates structures with different sizes, which correspond to different frequencies in the scale-space. A difference between these Gaussian maps puts in evidence the zero crossing zones which focus at higher frequencies that are usually due to edges and pointlike noise (pixels that have some variation in their surrounding neighborhood):

$$DG = L_1 - L_2.$$

In order to locate the maximum and minimum values of the resulting map, each pixel is compared to its three-dimensional neighbors (8 pixels in DG and 9 pixels in each scale L_i). The computational cost of this control is reasonably low because most candidates are eliminated soon after a few comparisons. Moreover, this technique is scale invariant and very fast due to the separability of the variables in the convolution kernels.

The algorithm described in [11, 12] has been used to search for local features, invariant to noise, scale and luminosity (named SIFT). The DG image is further processed to remove false keypoints, that are usually due to noise within very low contrast zones. The positions of the remaining candidates are selected by:

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2}\frac{\partial D}{\partial \mathbf{x}},$$

where D represents the first two terms of the Taylor series expansion of DG:

$$D(\mathbf{x}) = DG(\mathbf{x}) + \frac{\partial DG^T(\mathbf{x})}{\partial \mathbf{x}}\mathbf{x} + \frac{\mathbf{x}^T}{2}\frac{\partial^2 DG(\mathbf{x})}{\partial \mathbf{x}^2}\mathbf{x}.$$

A digital signature for each keypoint is needed in order to align two overlapping frames. It has been proved that a vector of descriptors containing 128 quantized values of orientation coming from both L_1 and L_2 contains enough information to properly identify a given keypoint. In particular, the contribution of the luminosity gradient of the pixels weighted within a Gaussian window, centered in the keypoint, is accumulated in 16 sub-samples (see Figure 5). The 8 possible directions of the resulting quantized vector are normalized so to let the signature be invariant in case of uniform change of luminosity.

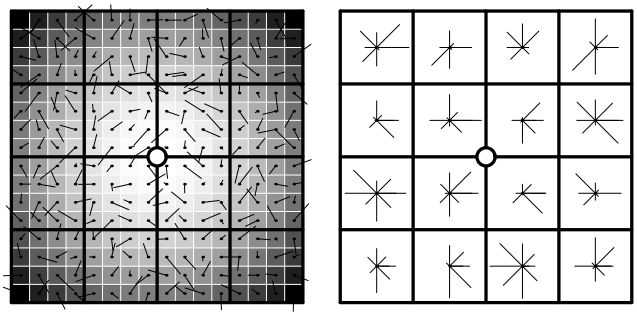

Figure 5: The weighted gradients within a Gaussian window (*left*) are sub-sampled to obtain a vector of quantized values (*right*) that identify each keypoint, represented as a circle.

This method is also the underlying idea of optical flow, applied to roughly predict the movements of the camera and to better estimate the parameters of the homography function. Assuming that the ROV does not turn, but scans the wreck in a straight line, we can use the norm of the optical flow vector to dynamically fine tune the step between consecutive images of the movie, that is how many frames must be skipped because they are too similar to each other. It is noteworthy that the optical flow can be processed on the fly, while looking for the keypoints.

4 Mosaicing

To increase the robustness of the matching an outlier rejection module is also required. The RANdom SAmple Consensus (RANSAC) is a probabilistic model to estimate, through a random greedy strategy, the parameters that describe the homography when passing from frame to frame [13]. The signature $\underline{v}_{i_1}$ of each keypoint k_{i_1}, located on the first frame F_1, is compared, in turn, against all signatures $\underline{v}_{i_2}$ of k_{i_2} on the second frame F_2. In order to couple these keypoints, the alignment of their signatures, expressed as the

arc cosine of the angle between $\underline{v}_{i_1}$ and $\underline{v}_{i_2}$, must be minimized:

$$\arg\min_{i_2} \cos^{-1}\langle \underline{v}_{i_1}, \underline{v}_{i_2} \rangle.$$

Of course, the set of matched keypoints, together with the correctness of the global result, depend on the random choice of the initial i_1; it has been experimentally proved that the signatures have enough measurements to reach a high specificity. Figure 6 shows an example of matched keypoints.

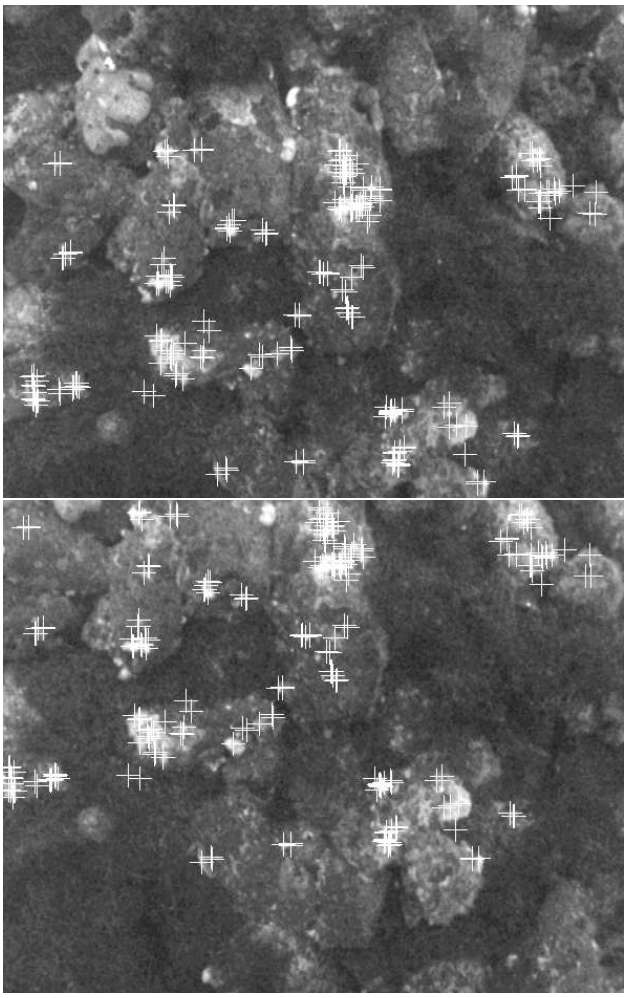

Figure 6: An example of matched keypoints (represented by crosses) in two consecutive frames.

A generic projective model has been preferred to take into account not only occlusions in the scenes, but also pitch, yaw and roll movements of the ROV. Therefore, at least 8 parameters are needed to correctly normalize the images, and thus 4 random couples of keypoints are selected to apply the Direct Linear Transform with Singular Value Decomposition [14] to calculate the homography between F_1 and F_2. The RANSAC module repeats at least 10 times the whole matching process and, when no more than 1% of the keypoints are considered as outliers (i.e. keypoints that imply a movement in gross disagreement with the overall motion), returns the homography that minimizes the re-projection error.

The warping of the whole landscape is done by mapping the homography to the mosaic, already obtained, against the new frame. This blending procedure returns a bigger and bigger image which distributes and reduces the effect of point-like noise on the overlapped images. An example of mosaic is presented in the following Figure 7.

Underwater images are often acquired when natural light is scarcely sufficient, thus the "brightness constancy constraint" is not valid anymore. The ROV, while moving, illuminates the scene in a non-uniform way, producing shadows and a bright spot in the center of the image with darker areas surrounding it. A high sensitivity camera avoids the use of artificial lights, reduces many of these lighting artifacts and lets simplify the blending module, which requires a simple averaging on the border of the images.

5 Conclusions and remarks

Unsupervised mosaicing of an underwater scene is a very hard task and during the last years a number of methodologies have been reported. They can be divided in side-scan sonars techniques, forward-looking sonars methods, acoustic cameras acquisition, multibeam echo-sounders data management and cameras with high sensitivity to low levels of luminosity. In this last case, the greater the distance of the subject is, the more degraded the visibility becomes [15].

This paper faces with the mosaicing of seabed landscapes and, in particular, of archaeological sites. Although our vehicle is equipped with additional sensors, only the video information coming from a digital camera has been used to realize the mosaic.

Degradations caused by polarization has been recently reduced by means of a polarizer at different orientations, but its application to deep-water inspections is still subject for further research [16]. We have preferred to study naturally illuminated photos because of the significant artifacts introduced by the lights of the ROV.

Underwater currents make the vehicle contin-

uously correct its position and mosaicing can be effectively used to keep track of the camera with respect to the environment.

Quadranocular video acquisition [17] has been already used to automate the ROV trajectory [18] and we believe that the use of the information coming from the second camera onboard our vehicle should improve the precision of the concatenation of motion estimates.

Moreover, if the path of the ROV describes a loop, the images already processed could be realigned to obtain a better valuation for the set of local homographies [19]. We expect that such a cycle of topology refinements will enhance the correctness of wider mosaics.

We are going to include in our method the improvements above described, but preliminary results are already satisfactory, particularly because they do not refer to almost superficial findings and the system is completely unsupervised.

Acknowledgements

We are grateful to ArenaSub Company for providing the ROV and the input video sequence, studied here. Besides, this work has been partially carried out together with Centro Oceanologico Mediterraneo.

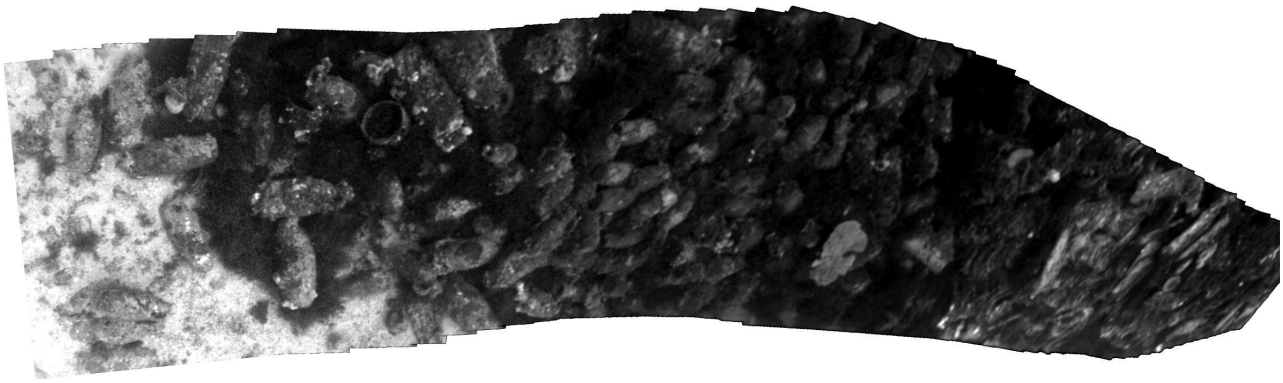

Figure 7: This mosaic has been assembled with 53 out of 925 frames (37 seconds of real video at 25 fps). Each frame has 592×470 pixels and the final result has 2457×698 pixels. These findings are located at a depth of about 70 meters.

References

[1] O.Pizarro, H.Singh, "Toward Large-Area Mosaicing for Underwater Scientific Applications", IEEE Journal of Oceanic Engineering, 28(4):651–672, 2003.

[2] R.García Campos, "A proposal to estimate the motion of an underwater vehicle through visual mosaicing", University of Girona, Spain, Ph.D. thesis, 2001.

[3] J.Weng, P.Cohen, M.Herniou, "Camera calibration with distortion model accuracy evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(10):965–980, 1992.

[4] R.Y.Tsai, "An efficient and accurate camera calibration technique for 3D machine vision", Proceedings of Conference on Computer Vision and Pattern Recognition, 364–374, IEEE Computer Society Press, 1986.

[5] Z.Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330–1334, 2000.

[6] F.Devernay, O.Faugeras, "Straight lines have to be straight", Machine Vision and Applications 13:14–24, 2001.

[7] R.Cucchiara, C.Grana, A.Prati, R.Vezzani, "A Hough Transform-based method for Radial Lens Distortion Correction", Proceedings of the 12th International Conference on Image Analysis and Processing, 182–187, 2003.

[8] A.Witkin, "Scale-space filtering", Proceedings of the International Joint Conference on Artificial Intelligence, 1019–1022, 1983.

[9] J.Koenderink, "The structure of images", Biological Cybernetics, 50:363–396, 1984.

[10] T.Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales", Journal of Applied Statistics, 21(2):224–270, 1994.

[11] D.Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60(2):91–110, 2004.

[12] M.Brown, D.Lowe, "Recognising panoramas", Proceedings of the 9th International Conference on Computer Vision, IEEE Computer Society Press, 1218–1225, 2003.

[13] M.A.Fischler, R.C.Bolles, "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography", Communications of the ACM, 24(6):381–395, 1981.

[14] R.Hartley, A.Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2004.

[15] N.Gracias, S.Negahdaripour, "Underwater mosaic creation using video sequences from different altitudes", Proceedings of MTS/IEEE Oceans, 1:1234–1240, 2005.

[16] Y.Schechner, N.Karpel, "Recovery of Underwater Visibility and Structure by Polarization Analysis", IEEE Journal of Oceanic Engineering, 30(3):570–587, 2005.

[17] H.Zhang, S.Negahdaripour, "Epiflow quadruplet matching: enforcing epipolar geometry for spatio-temporal stereo correspondences", Proceedings of the 7th IEEE Workshops on Application of Computer Vision 2005, 1:481–486, 2005.

[18] N.Gracias, J.Santos-Victor, "Underwater Mosaicing and Trajectory Reconstruction using Global Alignment", Proceedings of MTS/IEEE Oceans, 4:2557-2563, 2001.

[19] S.D.Fleischer, S.M.Rock, R.L.Burton, "Global position determination and vehicle path estimation from a vision sensor for real-time video mosaicking and navigation", Proceedings of MTS/IEEE Oceans, 1:641–647, 1997.