

Piecewise Planar Underwater Mosaicing

Fabio Bellavia, Marco Fanfani,
Fabio Pazzaglia and Carlo Colombo
CVG, University of Florence
Florence, 50139, Italy
email: name.surname@unifi.it

Riccardo Costanzi, Niccolò Monni,
Alessandro Ridolfi and Benedetto Allotta
MDM Lab, University of Florence
Florence, 50139, Italy
email: name.surname@unifi.it

Abstract—A commonly ignored problem in planar mosaics, yet often present in practice, is the selection of a reference homography reprojection frame where to attach the successive image frames of the mosaic. A bad choice for the reference frame can lead to severe distortions in the mosaic and can degenerate in incorrect configurations after some sequential frame concatenations. This problem is accentuated in uncontrolled underwater acquisition setups as those provided by AUVs or ROVs due to both the noisy trajectory of the acquisition vehicle – with roll and pitch shakes – and to the non-flat nature of the seabed which tends to break the planarity assumption implicit in the mosaic construction. These scenarios can also introduce other undesired effects, such as light variations between successive frames, scattering and attenuation, vignetting, flickering and noise. This paper proposes a novel mosaicing pipeline, also including a strategy to select the best reference homography in planar mosaics from video sequences which minimizes the distortions induced on each image by the mosaic homography itself. Moreover, a new non-linear color correction scheme is incorporated to handle strong color and luminosity variations among the mosaic frames. Experimental evaluation of the proposed method on real, challenging underwater video sequences shows the validity of the approach, providing clear and visually appealing mosaics.

I. INTRODUCTION

Video mosaicing [1] is a quite popular tool in underwater vision exploration and autonomous navigation [2], [3] since it gives a global, immediate, detailed and handy overview of the seabed, without requiring more expert knowledge in the 3D manipulation of the scene. However, despite the recent progress in the field [4], [5], obtaining good mosaics is still a challenging and not fully solved task. This is mostly due to the theoretical requisites of the mosaics, assuming input data with sufficiently distance from the scene or images acquired by camera rotations only [6]. These assumptions are not always met in practice, resulting in image misalignments and ghosting artefacts for which different blending techniques have been proposed [1], [7], [8], some of which tailored for the environments [9].

Furthermore, a commonly ignored problem, yet often present in practice, is the selection of a reference homography reprojection frame to which to attach the various mosaic images. The most common and trivial choice is to use the first image frame or, often supported by geo-referential camera positions, a user predefined one. A bad choice for the reference frame can lead to severe distortions in the mosaic and can degenerate in incorrect configurations after some sequential frame concatenations (see Fig. 1). This problem is accentuated in an uncontrolled underwater acquisition setup. For

example videos acquired from AUVs or ROVs due to the noisy trajectory of the acquisition vehicle – with roll and pitch shakes – and, in most cases, the non-flat nature of the seabed, tend to break the mosaic requirements. Moreover, these scenarios can introduce other undesired effects, such as light variations between successive frames, scattering and attenuation, vignetting, flickering and noise. In order to overcome these issues, recent underwater mosaic techniques also integrate 3D Structure-from-Motion and acoustic data [10], [11] to improve the results. However, these methodologies require more controlled environments or more hardware than a single camera.

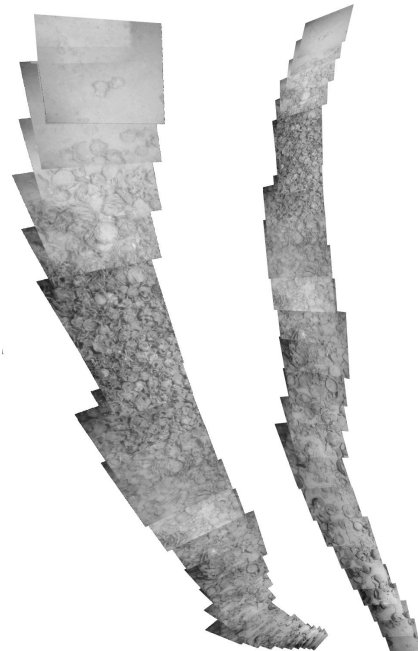


Fig. 1. A distorted mosaic due to a wrong reference homography selection (left), opposing to the proposed method (right). In both cases no post-processing color correction or blending have been applied to highlight the image frames.

This paper presents a novel mosaic pipeline which includes a strategy to select the best reference homography in planar mosaics from video sequences. This reference homography globally minimizes the distortions induced on each image frame by the mosaic homography itself. Additionally, a new non-linear color correction scheme is presented which robustly handles strong color and luminosity variations among the mosaic frames.

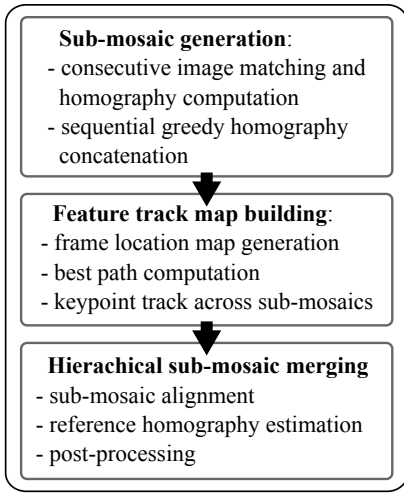


Fig. 2. A schematic view of the mosaic pipeline, see text for more details.

The proposed approaches were developed by the UNIFI CVG in the framework of the THESAURUS project for Typhoon class AUVs [12], designed and built by the UNIFI MDM Lab. The Typhoon AUV, which is also one of the AUVs of the heterogeneous fleet of the ARROWS European project [13], proved to be a useful tool for archaeologists and performed many underwater missions. It is a middle-sized class AUV with a maximum reachable depth of about 300 m, more than 10 hours of autonomy and a maximum speed of 6 knots. The vehicle can house onboard suitable payload, including an optical stereo-camera system used to acquire the video input to be processed for mosaicing.

The robust pipeline developed is presented in Sect. II. It is specifically designed for underwater environments, since in this kind of scenarios it is difficult to correctly match and track the keypoints required to compute inter-frame transformations. No user interaction is needed, as all the steps of the process are fully automatic. Next, Sect. III describes the proposed color correction scheme used to blend the mosaic frames. The method is experimentally evaluated on real challenging underwater video sequences in Sect. IV, showing the validity of the approach, which provides clear and visual appealing mosaics. Conclusions and future work are presented in Sect. V.

II. THE MOSAIC PIPELINE

The overall pipeline of the method is schematically described in Fig. 2 – more details can be found in [14]. The approach starts by splitting the video sequence into successive piecewise planar sub-mosaics with small deformations with respect to the original image frames.

A feature track map is then generated to match keypoints between image frames of sub-mosaic pairs in order to compute the homography between the sub-mosaics. This is required to reduce the propagation of homography estimation errors due to the non-planar real nature of the scene and to obtain robust matches between image keypoints features on non-consecutive image frames.

Finally, sub-mosaics are merged hierarchically, according to their overlap. In choosing which sub-mosaics must be

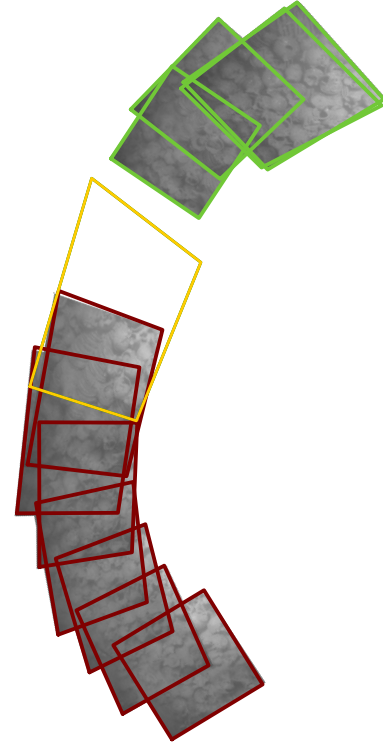


Fig. 3. Sub-mosaic generation, see the text for further details (best viewed in color).

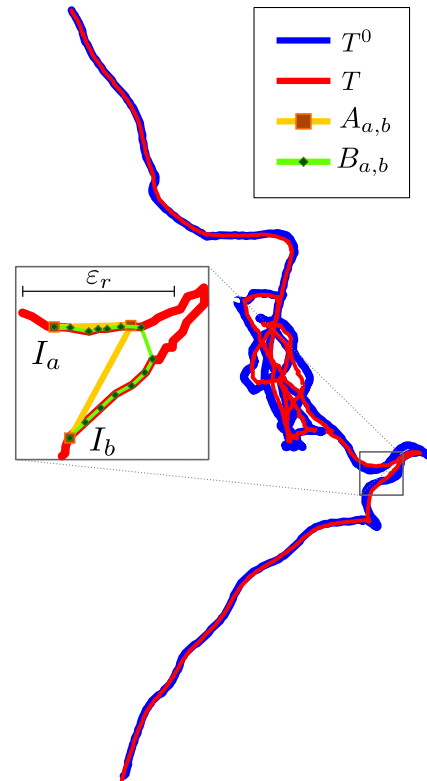


Fig. 4. The track map used to match keypoints on different sub-mosaics, see text (best viewed in color).

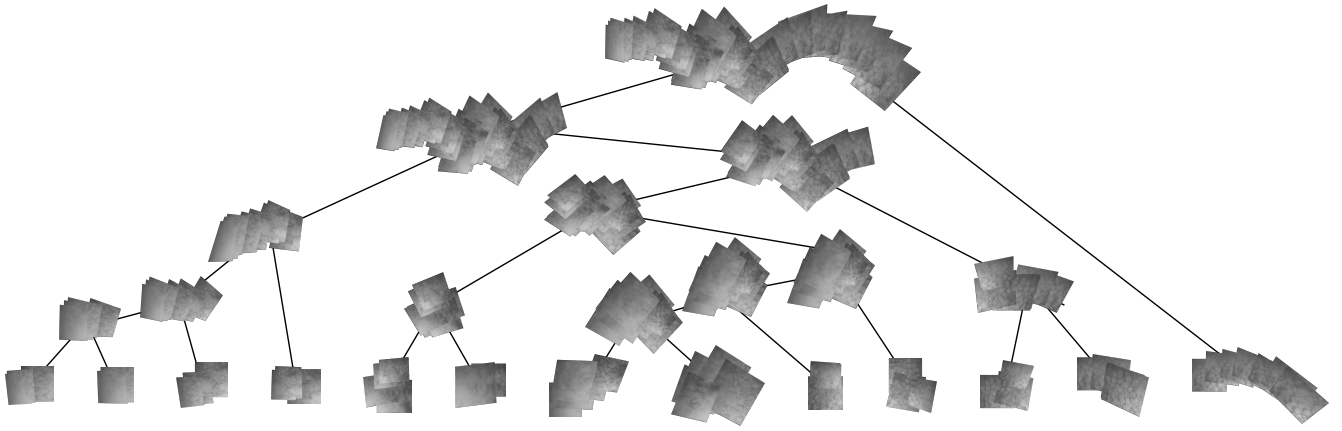


Fig. 5. Hierarchical merging of the sub-mosaics according to their overlap on the average homography.

merged, their overlap when both are projected into the best reference homography, minimizing the frame distortions, is evaluated. This is achieved by exploring the search space of the *average homographies* (defined later in this section) between the two sub-mosaics. Color correction is then eventually applied to refine the results as described in the next section.

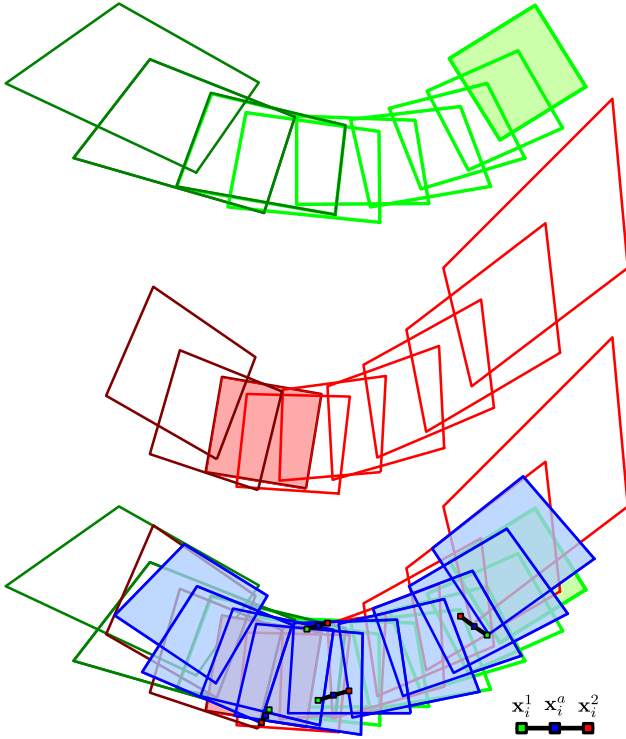


Fig. 6. The average homography which minimizes the frame distortion between two sub-mosaics, see text for more details (best view in color).

More in detail, referring to Fig. 3, sub-mosaics are grown by adding the consecutive image frames of the video sequence according to the standard mosaic pipeline [1] (red frames) by considering the first frame attached as the reference frame (red filled frame), until the distortion inducted on the last attached frame is too high (yellow frame). The frame distortion is measured in terms of area and semi-diagonals ratios, which cannot exceed a predefined range. In this case the current

sub-mosaic stops and a new one (green) is started using the last frame as reference homography (green filled frame). Note that in order to improve the computation and avoid final noisy effects, only frames with a low overlap with the current sub-mosaic are actually added to this.

The feature track map, employed for tracking keypoints across non-consecutive frames of the different sub-mosaics to be merged, is then computed. This map gives a 2D distance between the different frames of the video sequence. Referring to Fig. 4, it is initialized by considering the average key-point displacement between consecutive frames (blue line). Each frame location on the map is then iteratively refined by adjusting its position according to the distance from its neighbourhood (red line).

Still referring to Fig. 4, in order to establish a match between the frames I_a and I_b , keypoints are tracked on consecutive frames (green dots) belonging to the best path between the frames I_a and I_b . The best path (green line) is constrained in length with respect to the minimum path (yellow line) and minimizes the maximum edge inside the path. This property reflects the fact that a small edge in the track map implies that the corresponding frames are close to each other, so that robust matches can be obtained – see [14] for more details.

Finally, sub-mosaics, on which keypoint correspondences have been established as described above according to best paths between frames, are merged hierarchically in order to form the final mosaic as shown in Fig. 5. The order by which the sub-mosaic pairs are merged depends upon their overlap; sub-mosaic pairs with higher overlap are merged first. Since the overlap between two sub-mosaics depends on the reference homography used when merging them, the homography which minimizes the frame distortion between the two sub-mosaics, named *average homography*, must be computed as follows.

Referring to Fig. 6, the two sub-mosaics (dark and light color) are merged according to their reference frame (filled frames) and aligned. Using a RANSAC approach, four corresponding points are sampled on each sub-mosaic (green, red dots) and their midpoints are computed (blue dots), the homography between the points of first sub-mosaic and the midpoints is computed and the distortion error on the new projected frames given by the midpoints is computed. After a

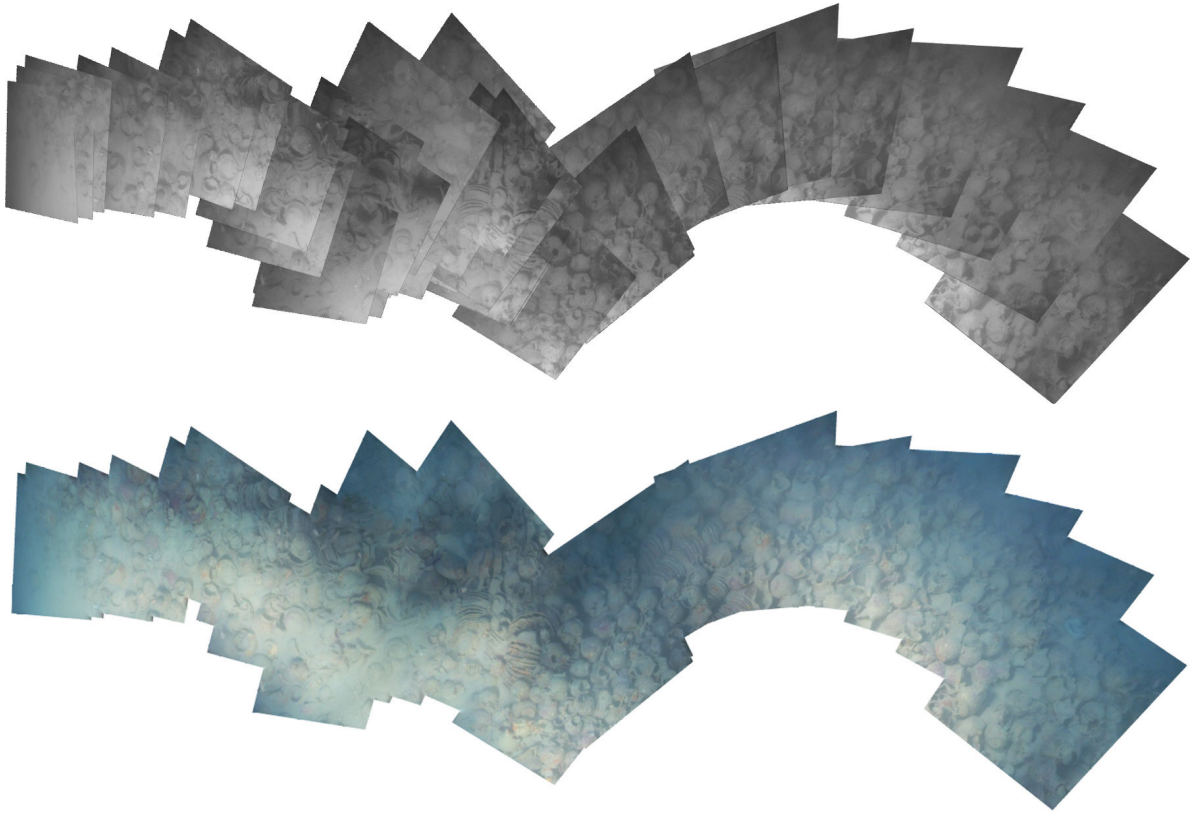


Fig. 7. Final mosaic obtained with the hierarchical merging (top row) and after the color correction and blending post-processing (bottom row, best viewed in color).

fixed number of iterations, the homography minimizing the distortion error is selected as the average homography (blue filled frames). The distortion error is computed by considering the ratios between successive and consecutive sides of the original and projected frames, their areas and the angle differences – see [14] for further details. The final result is shown in Fig. 7.

III. COLOR BLENDING

The visual analysis of underwater scenarios is quite challenging. Indeed, scattering and attenuation effects, vignetting, flickering and noise result in abrupt luminosity changes in the mosaic frames, thus requiring post-processing blending and color corrections. A new color correction scheme [15] using a spline model is used to adjust and reduce the color variations among the mosaic frames, followed by pyramidal blending [16] to smooth geometric inconsistencies. Referring to Fig. 8, the proposed color correction approach tries to find the best monotone cubic spline (green curve), parametrized by its two free knots, which minimizes the image similarity error between the overlapping frames in terms of intensity and gradient values. In order to reduce the spline search space, the two knots are sampled according to pre-defined grids (blue and black crosses) and the several heuristics are used to further reduce the computational time – see [15] for more details. The spline color map, computed block-wise is then smoothly propagated to the non-overlapping part of the image. The final results is shown in Fig. 7.

Experimental evaluations presented in [15] show that the proposed method gives better results as compared to the stan-

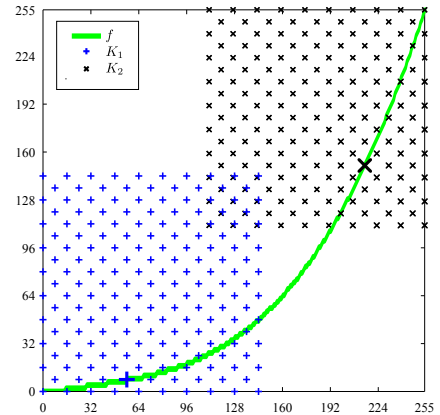


Fig. 8. Spline color mapping knot search space (best viewed in color).

dard Reinhard's [7] and gain compensation [1] methodologies. Figure 9 shows a comparison of the proposed method, with respect to the others; the target image to be corrected is superimposed onto the source image. As it can be noted, the original images (Fig. 9(a)) present strong color variations on the overlapping boundaries due to vignetting effects; the proposed method (Fig. 9(b)) behaves better in term on smooth color variations on the boundaries, followed by the Reinhard's method (Fig. 9(c)).

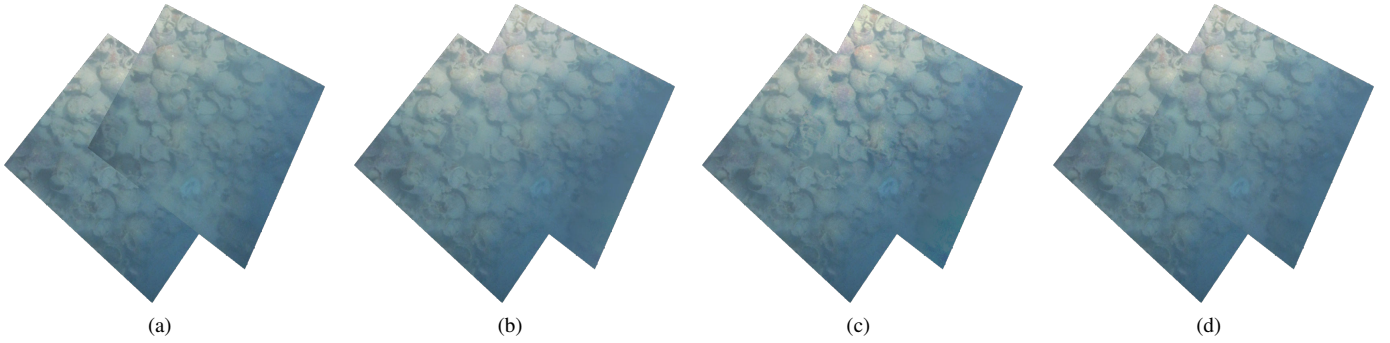


Fig. 9. Target image superimposed on the source image without color correction (a), using the proposed spline model (b), the Reinhard's method (c) and the gain compensation (d). No blending is applied, the reader is encouraged to zoom on the figure to better appreciate the differences between the various methods (best viewed in color).



Fig. 10. A video frame for the mosaic in Fig. 11 (best viewed in color).

IV. EXPERIMENTAL RESULTS

We tested the proposed pipeline on several underwater video sequences. Two different video sequences were taken from the archaeological site of the Scoglietto area, near the isle of Elba (Italy) – a snapshot of the video sequence frames is shown in Fig. 10. Figures 7, 11 show different mosaics obtained from these videos. Differently from Fig. 7, the mosaics of Fig. 11 are obtained from the same video sequences, automatically split by the proposed mosaic pipeline, since the filmed area strongly violated the planarity assumption of the mosaic, producing too distorted frames which could not be recovered. In this case, the method cut the nodes of tree obtained hierarchically as for Fig. 5, according to the frame distortion error, producing a cluster of planar mosaics.

Another video sequence was taken in the Caesarea harbour in Israel by the Typhoon class AUVs [12] – a video snapshot is reported in Fig. 12. This video is more challenging than the previous ones, due to a large slant with respect to the seabed and more flickering effects, so that a cluster of mosaics is obtained. An example of the mosaics inside the cluster is shown in Fig. 13.

As it can be noted, the resulting mosaics are good, with no evident misalignment glitches or strong frame deformations. Underwater scenes are very challenging, due to their high in-

tensity changes and repeated patterns, which make the feature tracking difficult, so that the quality of the results strengthen the validity of the proposed method. Indeed, the presented method allowed us to get in a completely automatic way as good-looking mosaics as those obtained with strong expert user intervention.

V. CONCLUSIONS AND FUTURE WORK

This paper proposes a new mosaic pipeline designed for underwater videos. The method is robust and minimizes the frame distortions in the case of planar mosaics by employing a hierarchical approach, supported by a novel color correction scheme to resolve strong color variations among the mosaic frames. Experimental results show the ability of the proposed method to handle complex underwater scenes, yielding to high quality unsupervised mosaics.

Future work will include more evaluation tests as well incorporating in the pipeline new stitching algorithms [4], [5] with the aim to improve results in the case of strong 3D content which break the planar mosaic assumptions. Further research will also address the integration of image segmentation methodologies to improve the spline color correction model.

ACKNOWLEDGEMENT

Thanks to Pamela Gambogi of the “Soprintendenza per i Beni Archeologici della Toscana”, Italian Ministry of Culture, and to Kobi Sharvit of the “Marine Archaeological Unit”, Israel Antiquities Authority, for providing the input video sequences.

This work has been supported by the European ARROWS project, that has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no 308724.

REFERENCES

- [1] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.
- [2] N. Gracias and J. Santos-Victor, “Underwater video mosaics as visual navigation maps,” *Computer Vision and Image Understanding*, vol. 70, no. 1, pp. 66–91, 2000.

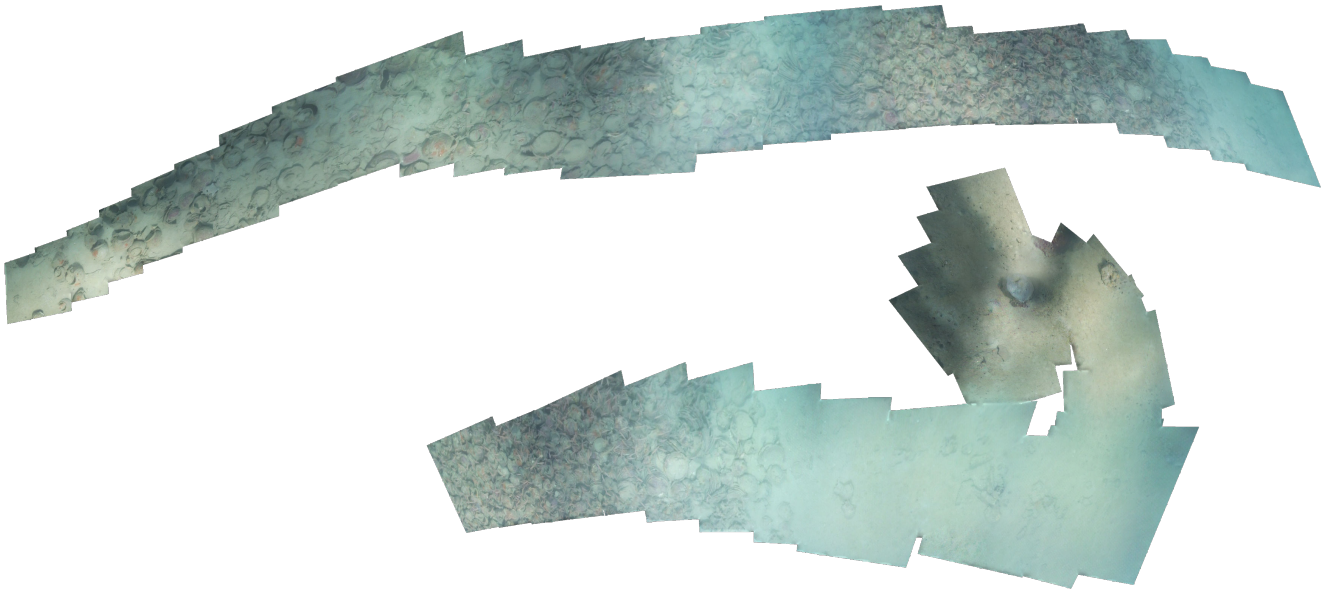


Fig. 11. Mosaic cluster from the Scoglietto are obtained with the proposed method (best viewed in color).



Fig. 12. A video frame for the Caesarea harbour video sequence (best viewed in color).

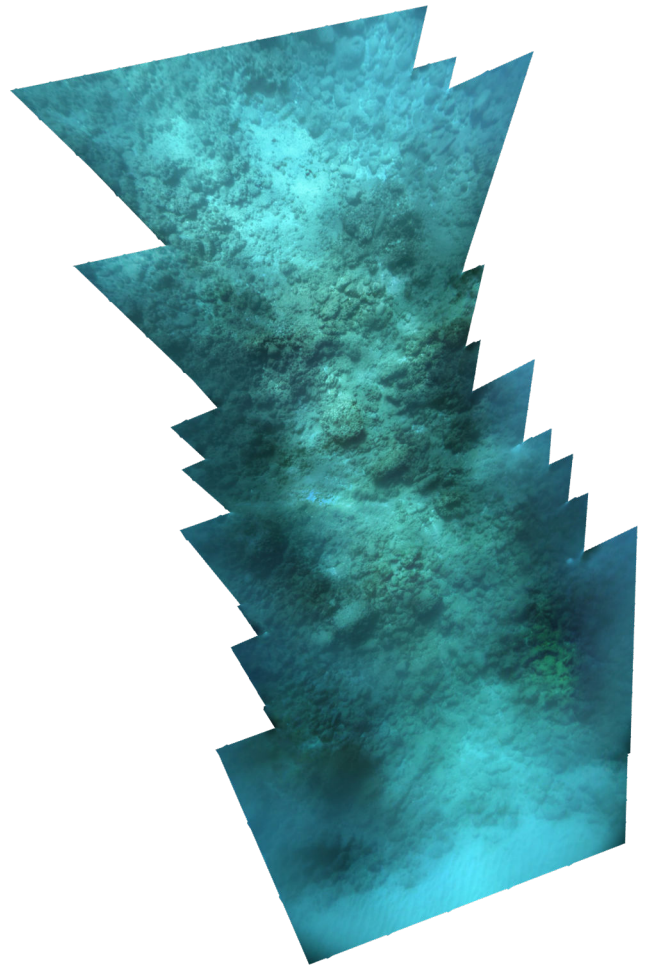


Fig. 13. A mosaic example for the cluster of the Caesarea harbour video sequence (best viewed in color).

- [3] O. Pizarro and H. Singh, "Towards large area mosaicing for underwater scientific applications," *IEEE Journal of Oceanic Engineering, Special Issue on Underwater Image and Video Processing*, vol. 28, no. 4, pp. 651–672, 2003.
- [4] J. Zaragoza, T. J. Chin, Q. H. Tran, M. Brown, and D. Suter, "As-Projective-As-Possible image stitching with moving DLT," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1285–1298, 2014.
- [5] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3262–3269.
- [6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
- [7] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 34–41, 2001.
- [8] M. Uyttendaele, A. Eden, and R. Szeliski, "Eliminating ghosting and exposure artifacts in image mosaics," in *Computer Vision and Pattern Recognition*, 2001, pp. 509–516.
- [9] R. Prados, R. Garcia, N. Gracias, J. Escartín, and L. Neumann, "A novel

- blending technique for underwater giga-mosaicing,” *IEEE Journal of Oceanic Engineering*, vol. 37, no. 4, pp. 626–644, 2012.
- [10] H. Singh, C. Roman, O. Pizarro, R. Eustice, and A. Can, “Towards high resolution imaging from underwater vehicles,” *International Journal of Robotics Research*, vol. 26, no. 1, pp. 55–74, 2007.
 - [11] T. Nicosevici, N. Gracias, S. Negahdaripour, and R. García, “Efficient three-dimensional scene modeling and mosaicing,” *Journal of Field Robotics*, vol. 26, no. 10, pp. 759–788, 2009.
 - [12] B. Allotta, L. Pugi, F. Bartolini, A. Ridolfi, R. Costanzi, N. Monni, and J. Gelli, “Preliminary design and fast prototyping of an autonomous underwater vehicle propulsion system,” *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 2014.
 - [13] B. Allotta et al., “The ARROWS project: adapting and developing robotics technologies for underwater archaeology,” in *IFAC Workshop on Navigation Guidance and Control of Underwater Vehicles (NGCUV 2015)*, 2015 (to appear).
 - [14] F. Bellavia and C. Colombo, “Estimating the best reference homography for planar mosaics from videos,” in *International Conference on Computer Vision Theory and Applications (VISAPP 2015)*, 2015, pp. 512–519.
 - [15] —, “Color correction for image stitching by monotone cubic spline interpolation,” in *7th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2015)*, 2015 (to appear).
 - [16] P. J. Burt and E. H. Adelson, “A multiresolution spline with application to image mosaics,” *ACM Transactions on Graphics*, vol. 2, no. 4, pp. 217–236, 1983.