Robust Underwater Image Stitching Based on Graph Matching

Xu Yang[1], Zhi-Yong Liu[1,2,3(✉)], Chuan Li[1,4], Jing-Jing Wang[1], and Hong Qiao[1,2,3]

[1] State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
{xu.yang,zhiyong.liu,lichuan2013,wangjingjing2014,hong.qiao}@ia.ac.cn
[2] Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, People's Republic of China
[3] University of Chinese Academy of Sciences, Beijing 100049, China
[4] Research Institute of Chengdu, Beijing Jingdong Century Trade Co., Ltd., Chengdu 610046, China

Abstract. Image stitching is important in intelligent perception and manipulation of underwater robots. In spite of a well developed technique, it is still challenging for underwater images because of their inevitable appearance ambiguity. For the feature based underwater image stitching, robust feature correspondence is the key because most other algorithmic parts are less directly associated with the characteristics of underwater images. Structural information between feature points may be helpful for robust feature correspondence, and based on this idea the paper proposes a robust underwater image stitching method by incorporating structural cues as additional information, whose effectiveness is validated on real underwater images. Specifically, the appearance information and structural cues are integrated by a labeled weighted graph, and the underwater image correspondence is formulated by graph matching. After geometric transformation estimation, the underwater images are finally blended into a wider viewing image.

Keywords: Underwater image · Image stitching · Feature correspondence · Graph matching · Structural information

1 Introduction

Image stitching aims at the combination of two images, or more images, with overlapped areas into a wide viewing composite, or even a panorama. It plays a key role in those robot tasks in places presenting a difficult access for human beings, such as some tasks by underwater robots, e.g. the remotely operated underwater vehicle (ROV) [1], the autonomous underwater vehicle (AUV) [2], or the autonomous remotely operated vehicle (ARV) [3]. It is because once a robot is equipped with a camera, the visual perception of its operating environment

Y. Sun et al. (Eds.): IScIDE 2017, LNCS 10559, pp. 521–529, 2017.
DOI: 10.1007/978-3-319-67777-4_46

is usually of interest. Since the common camera equipped to the robots usually has a limited field of view, image stitching is thus useful to obtain a larger field of view over the operating area.

Image stitching itself has long been an important topic in image processing and computer vision. Many methods have come out, which can be roughly categorized into two types, i.e. the region based methods and the feature based methods. Early image stitching usually adopts the region based methods, which usually find a common region in two images through the region appearance information, e.g. pixel intensity. The ideas of these methods are usually straightforward and easy to implement. But at the same time these methods usually suffer from illumination changes, occlusions, geometric distortions in different images, and therefore inappropriate for real world images, especially those images obtained on a mobile robot platform. Recently most researchers in the computer vision and robot vision communities tend to use the feature based image stitching methods, because of their robustness to changing factors, such as the above mentioned changing illuminations, geometric distortions, etc. Particularly, the emergences of many excellent local features such as SIFT [4] and SURF [5] in the last fifteen years have promoted the success of the feature based image stitching, which is used in many real world camera applications, including those on intelligent mobile phones. It can be said image stitching is a solved problem for many types of images, especially the images on land.

Different from the images on land, it is still a challenging problem to stitch underwater images. The main obstacle lies in the inevitable appearance ambiguity of underwater images. It is because the limited light refracted into the water or shot from the main robot body, would further be scattered or absorbed by water molecules, plankton, or sands. Such a condition would significantly deteriorate the performance of general image stitching methods, even the feature based methods, because the ambiguous appearance of underwater images often leads to a poor discriminant ability, or effect lost, of the feature descriptor.

From the algorithmic perspective, the feature based methods mainly consist of three steps, i.e. feature correspondence, transformation estimation, image blending [6], where the combination of feature correspondence and transformation estimation is also known by the term image registration. It can be noticed that the latter two steps are less directly related with the discriminant ability of the feature descriptor, which implies that once the feature point are successfully corresponded, the stitching of the underwater images are almost the same with the stitching of the images on land. In other words, robust feature correspondence of the underwater images is the key for their stitching. Since only the appearance information is inadequate, introducing additional information or incorporating additional constraints is a intuitive way to improve the robustness of feature correspondence on underwater images. The structural information between feature points may be an effective choice [7]. Because by incorporating the structural constraints it requires that the structures extracted from the feature point sets should be consistent while maintaining the appearance similarity, which may help to avoid abnormal feature assignments.

Based on these understandings, this paper proposes a roust underwater image stitching method by introducing additional structural information. Specifically, the appearance information and structural information are integrated by a labeled weighted graph model, and the feature correspondence between two underwater images is formulated and solved by graph matching, which is followed by outlier assignment refinement. After the estimation of geometric transformation from the inlier assignments, the underwater images are finally blended into a wider viewing image.

The remaining manuscript is organized as follows: After the discussion of related works in Sect. 2, the proposed underwater image stitching method is introduced in Sect. 3, which is followed by the experimental evaluation in Sect. 4. Finally Sect. 5 concludes the paper.

2 Related Works

In this section, we first give some discussions on recent image stitching algorithms, and then introduce their applications to underwater images.

2.1 Image Stitching Algorithms

In recent years, the feature based methods have overtook the area based methods as the most common image stitching algorithms. Their success can be attributed to the robustness to changing factors, such as the geometric distortions. A benchmark algorithm is the famous scale invariant feature transform (SIFT) feature [4] based image stitching method. Then researchers have applied many types of local features to the image stitching task, of which some representative features include the speeded up robust feature (SURF) [5] feature, the binary robust independent elementary features (BRIEF) [8], the shape context feature [9], etc. Most of these algorithms are dedicated to image stitching tasks with general purposes, and get superior performance on common natural images, especially images on land. Only a few researchers generalize them to the underwater images, as introduced below.

2.2 Underwater Image Applications

In the image stitching method proposed by Leone et al. [10] for underwater environment, the Harris corner point detector with certain specific improvements is used to extract the feature points, and the texture information is used to built the feature point descriptor. Then the correspondence between two feature point sets which represents two underwater images is established by matching the feature point descriptors. The homography transformation, i.e. the translations, rotations and scaling effects, between two underwater images is estimated based on the correspondence, and then stitched image is obtained. A similar scheme is used by Elibol et al. [11] in their underwater image stitching work, or called by underwater optical mapping in their work. Differently, they adopted the SIFT feature point and descriptor extracted from the underwater images, of which the outlier assignments are refined by the famous random sample consensus (RANSAC) technique. In the real time image stitching method proposed by Ferreira et al. [12], after the binary robust independent elementary features (BRIEF) based motion estimation, the SURF is used in the feature correspondence step. Garcia-Fidalgo et al. [13] in their underwater image stitching method used a feature which is a variant of BRIEF in the framework of bags of words (BoG).

Generally, the proposed method follows a similar scheme with the above methods, but incorporates structural constraints in the feature correspondence step, which is useful against the ambiguous appearance of underwater images.

3 Underwater Image Stitching

The proposed underwater image stitching method is introduced in this section. As mentioned above, robust feature correspondence is particularly important for underwater images, which is realized by incorporating structure cues beyond the appearance information in this paper.

A feature point extracted from an underwater image is represented by a weighted labeled graph $\mathcal{G}$ to integrate the appearance information and the structural relations between feature points. Almost any of the well known local features could be adopted as the feature extractor and descriptor, e.g. Harris corner detector, SIFT extractor and descriptor, SURF extractor and descriptor, BRIEF extractor and descriptor, etc., which implies that the incorporation of structural cue lowers the demand of discriminant feature extractor and descriptor. Then it is straightforward to represent the feature set by a labeled weighted graph by representing each feature point by a graph vertex, representing the link between a pair of feature points by a graph edge, describing the vertex by a so called label using the feature descriptor, and describing the edge by a so called weight using the spatial relation measure, e.g. length and orientation of the link. Thus the feature correspondence problem can by assigning the vertices in two labeled weighted graphs, abbreviated by graph below, which problem is known as graph matching.

Mathematically, the collection of the weights in a graph $\mathcal{G}$ can be represented by weighted adjacency matrices $\mathbf{G}^i, i = 1 \cdots d$. The number of weighted adjacency matrices d depends on the weight dimension. For instance, when using the distance between feature points, i.e. the link length, as the edge weight, only one adjacency matrix $\mathbf{G}^1$ is enough for a graph, where each non-diagonal entry $\mathbf{G}^1_{ij}$ denotes the distance between the ith and j vertices in $\mathcal{G}$. The pre-calculated differences between vertex labels are stored in a label cost matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$ where $\mathbf{L}_{ia}$ denotes the distance between the label of the ith vertex in $\mathcal{G}$ and that of the ath vertex in $\mathcal{H}$. Given two graphs $\mathcal{G}$ and $\mathcal{H}$ of sizes M and N respectively, their matching can be represented by an assignment matrix $\mathbf{X} \in \{0, 1\}^{M \times N}$, where $\mathbf{X}_{ia} = 1$ means that the ith vertex in $\mathcal{G}$ is assigned to the ath vertex in $\mathcal{H}$. If the one-to-one matching assumption is adopted, then the assignment matrix becomes a so called partial permutation matrix, defined by

$$\mathbf{X} \in \mathcal{D} := \left\{ \mathbf{X} | \sum_i \mathbf{X}_{ia} \leq 1, \sum_a \mathbf{X}_{ia} = 1, \mathbf{X}_{ia} = \{0, 1\} \right\}. \tag{1}$$

Without loss of generality, it is assumed that $M \leq N$ hereafter. Based on the above the mathematical representations, the correspondence result can be obtained by minimizing the following graph matching objective function:

$$\mathbf{X}^* = \alpha \min_{\mathbf{X}} \sum_{i=1}^{d} \|\mathbf{G}^i - \mathbf{X}\mathbf{H}^i\mathbf{X}^T\|_F + (1-\alpha)\mathrm{tr}(\mathbf{L}^T\mathbf{X}), \tag{2}$$
$$\text{s.t } \mathbf{X} \in \mathcal{D}.$$

The optimization problem is an NP-hard high order combinatorial optimization problem with factorial computational complexity, for which the approximate method are necessary. We use the graduated nonconvexity and graduated concavity (GNCCP) [14], a continuous method based combinatorial optimization framework, to approximately solve the problem. The utilize the GNCCP, the discrete domain $\mathcal{D}$ should be relaxed to its domain $\mathcal{C}$, defined by

$$\mathcal{C} := \left\{ \mathbf{X} | \sum_i \mathbf{X}_{ia} \leq 1, \sum_a \mathbf{X}_{ia} = 1, \mathbf{X}_{ia} \in [0, 1] \right\}. \tag{3}$$

And the GNCCP also makes use of the property that DD is exactly the extreme point set of $\mathcal{C}$. Specifically, the GNCCP first approximates the original optimization problem (2) by a relatively simple convex optimization problem over the continuous domain $\mathcal{C}$, and step by step implicitly transforms it to be a concave optimization problem over $\mathcal{C}$. Note by a clever design both the above convex optimization problem and concave optimization problem have exactly the same global optimum as (2) over the discrete domain $\mathcal{D}$. And the optimum point of the concave optimization problem over a convex set lies in its extreme point set, i.e. $\mathcal{D}$ by the property mentioned above. Therefore a discrete assignment matrix could be automatically obtained when the GNCCP terminates at the concave optimization problem, which usually exhibit superior performance.

In each step of the GNCCP process, the optimization problem is solved by the conditional gradient descent method [15,16], also known as the Frank-Wolfe algorithm, where the gradient of the original function (2) is needed, which is

$$\nabla = \alpha \sum_{i=1}^{d} (2\mathbf{X}(\mathbf{H}^{iT}\mathbf{X}^T\mathbf{X}\mathbf{H}^i + \mathbf{H}^i\mathbf{X}^T\mathbf{X}\mathbf{H}^{iT}) - 2(\mathbf{G}^i\mathbf{X}\mathbf{H}^{iT} + \mathbf{G}^{iT}\mathbf{X}\mathbf{H}^i)) + (1-\alpha)\mathbf{L}. \tag{4}$$

The solution $\mathbf{X}^*$ indicate the assignments between feature points in two underwater images. As mentioned in Sects. 1 and 2, once these assignments obtained, the transformation between the images ca be estimated. Before the transformation estimation, we first employ the maximum likelyhood estimation sample consensus (MLESAC) [17] to refine the assignments, or say to remove the outlier assignments. The MLESAC is a variant of the famous RANSAC [18]. Different from RANSAC, it aims at the solution which maximizes the likelihood instead of the number of inliers, and is particularly appropriate for the estimation of complex surfaces or more general manifolds from points [17]. Then the projective matrix P between two images are estimated based on the refined inlier assignments. If a frame sequence sampled from for example a video clip are provided, the projective matrices are estimated sequentially following a similar way in [19].

Once the cascaded estimations of the projective transformation between underwater images are obtained, the final step is to warp all the images according to the transformation estimation and blending them together [6]. In order to reduce the visual influence of the seam, the stitched images are blended by rendering the overlapped area by the average intensities from both images.

4 Simulations

The proposed scheme is first evaluated on an underwater image sequence shot at the Valldemossa harbour seabed (Mallorca, Spain) [13]. This dataset contains

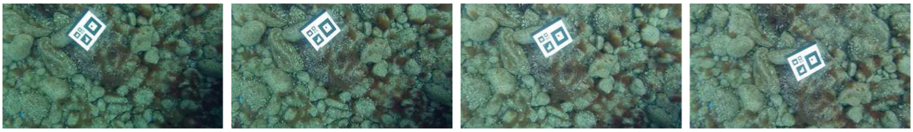

Fig. 1. Preceding 4 samples of underwater image sequence shot at Valldemossa harbour seabed.

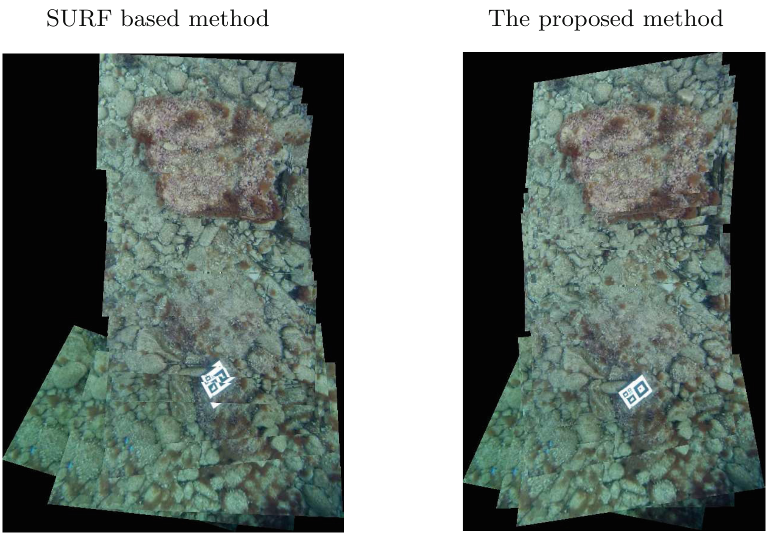

Fig. 2. Stitching result on underwater image sequence shot at Valldemossa harbour seabed.

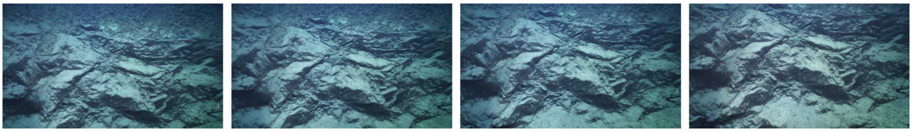

Fig. 3. Preceding 4 samples of underwater images shot by IFREMER.

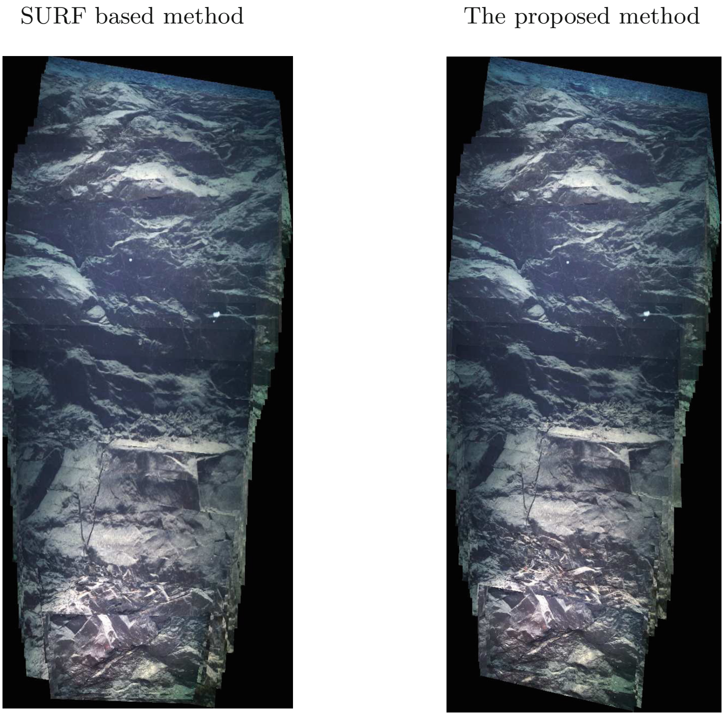

Fig. 4. Stitching result on the underwater images shot by IFREMER.

201 images with 320×180 pixels, which forms a loop around a central point. Some image samples of the dataset are shown in Fig. 1. In each image the key points are extracted by SURF [5] together with the descriptors. The graphs are constructed as described in Sect. 3. Specifically, the key points are represented by the graph vertices, with the SURF descriptor as the vertex labels. The graph structure, i.e. the set of the graph edges, are built by the Delaunay triangulation technique [20]. The length and orientation of each link in the graph structure are used as the two-dimensional edge weight. The proposed method is compared with the SURF based stitched method which does not consider the structural relation between key points. Because of the memory limit of the our computer, the preceding 40 images of the total 201 images are stitched, as illustrated in Fig. 2. It can be observed that the proposed method illustrates more smooth transition across

the images. It is directly attributed to the accurate image registration, which is essentially resulted from the robust feature correspondence by incorporating structural cues.

The proposed method is also applied to another underwater dataset[1] sampled from the video released by French research institute for exploitation of the sea (IFREMER), which is shot along the Mid-Atlantic Ridge in the North Atlantic Ocean. Some samples are illustrated in Fig. 3. The experimental setting are the same with the above experiment. The stitching result on the preceding 40 images of the total 64 images are given in Fig. 4, which validate the effectiveness of the proposed method.

5 Conclusion

This paper aims at the robust image stitching in the underwater environment, proposes to introduce the structural information to tackle the appearance ambiguity problem, a specific problem for the underwater images. Simulations witness the effectiveness of our idea. However, there may be seams or moving objects caused ghosting areas by the current version of proposed method. Therefore, we intend to make more investigations in the blending step in our future work.

Acknowledgment. This work is supported partly by the National Natural Science Foundation (NSFC) of China (grants 61503383, 61633009, U1613213, 61375005, 61210009, and 61773047), partly by the National Key Research and Development Plan of China (grant 2016YFC0300801), partly by the Beijing Municipal Science and Technology (grants D16110400140000 and D161100001416001), and Guangdong Science and Technology Department (grant 2016B090910001).

References

1. Huang, H., Tang, Q., Li, Y., Wan, L., Pang, Y.: Dynamic control and disturbance estimation of 3D path following for the observation class underwater remotely operated vehicle. Adv. Mech. Eng. **5**, 604393 (2013)
2. Huang, H., Tang, Q., Li, H., Liang, L., Li, W., Pang, Y.: Vehicle-manipulator system dynamic modeling and control for underwater autonomous manipulation. Multibody Syst. Dyn. (2017). doi:10.1007/s11044-0169538-3
3. Tang, Y., Li, S., Zhang, A.: Research on optimization design of a new type of underwater vehicle for arctic expedition. In: Proceedings of the International Symposium Underwater Technology, pp. 96–100 (2009)
4. Lowe, D.: Object recognition from local scale-invariant features. In: Proceedings of IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157 (1999)
5. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Proceedigs of the European Conference on Computer vision, pp. 404–417 (2006)
6. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vis. **74**(1), 59–73 (2007)

[1] The dataset is named by ODEMAR, which is available at https://github.com/emiliofidalgo/bimos.

7. Yang, X., Liu, Z., Qiao, H., Song, Y., Ren, S., Zheng, S.: Underwater image matching by incorporating structural constraints. Int. J. Adv. Robot. Syst. Under Revision
8. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010 Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1_56
9. Belongie, S., Malik, J.: Matching with shape contexts. In: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 20–26 (2000)
10. Leone, A., Distante, C., Mastrolia, A., Indiveri, G.: A fully automated approach for underwater mosaicking. In: OCEANS, pp. 1–6 (2006)
11. Elibol, A., Garcia, R., Gracias, N.: A new global alignment approach for underwater optical mapping. Ocean Eng. **38**(10), 1207–1219 (2011)
12. Ferreira, F., Veruggio, G., Caccia, M., Zereik, E., Bruzzone, G.: A real-time mosaicking algorithm using binary features for rovs. In: Proceedings of the Mediterranean Conference on Control and Automation, pp. 1267–1273 (2013)
13. Garcia-Fidalgo, E., Ortiz, A., Bonnin-Pascual, F., Company, J.: Fast image mosaicing using incremental bags of binary words. In: Proceedings of the IEEE International Conference Robotics and Automation, pp. 1174–1180 (2016)
14. Liu, Z., Qiao, H.: GNCCP graduated nonconvexity and concavity procedure. IEEE Trans. Pattern Anal. Mach. Intell. **36**(6), 1258–1267 (2014)
15. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Nav. Res. Logist. Q. **3**(1–2), 95–110 (1956)
16. Jaggi, M.: Revisiting Frank-Wolfe: projection-free sparse convex optimization. In: Proceedings of the International Conference on Machine Learning, pp. 427–435 (2013)
17. Torr, P., Zisserman, A.: MLESAC: a new robust estimator with application to estimating image geometry. Comput. Vis. Image Underst. **78**, 138–156 (2000)
18. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
19. Li, C., Liu, Z., Yang, X., Qiao, H., Su, J.: Stitching contaminated images. Neurocomputing **214**, 829–836 (2016)
20. Loera, J., Rambau, J., Santos, F.: Triangulations: Structures for Algorithms and Applications. Springer, Heidelberg (2010)