Title	Algorithmically Improved Framework for Image-only Robotic Mapping
Author(s)	ELIBOL, Armagan; Nak-Young, Chong
Citation	Proceedings of the 2021 18th International Conference on Ubiquitous Robots (UR): 515-520
Issue Date	2021-07
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/17571
Rights	This is the author's version of the work. Copyright (C) 2021 IEEE. Proceedings of the 2021 18th International Conference on Ubiquitous Robots (UR), 2021, pp.515-520. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	

Algorithmically Improved Framework for Image-only Robotic Mapping

Armagan Elibol and Nak Young Chong

Abstract— **Over the past two decades, technological developments in the robotics field have made it possible to gather data from areas that are hostile and not-reachable environments for humans. In line with these advancements in data collecting tools and procedures, the need and demand for computationally efficient methods for processing the gathered data have been increased from different science and engineering fields. Among the others, optical data is one of the main data sources that low-cost robotic vehicles can obtain easily nowadays. Due to the different limitations, obtained optical data usually cannot cover a large area in a single image. Therefore, optical mapping methods (image mosaicing) are needed to create higher resolution maps by combining comparatively smaller resolution images. These methods rely on pairwise image registration and one of the main bottlenecks in the case of image-only information available is that the quadratic growth of image matching attempts with respect to the total number of images. In this paper, we propose an algorithmically improved end-to-end framework for creating 2D optical maps from a set of randomly ordered images with the aim of reducing computational efforts needed via lowering the total number of image matching attempts. We present extensive and comparative experimental results with its counterpart approach using four real datasets obtained from the underwater environment using Unmanned Underwater Vehicles (UUVs).**

I. INTRODUCTION

Optical data is one of the most crucial data resources for several different science disciplines. Thanks to the rapid developments in sensor and robotic technological studies, mobile platforms that are low-cost and capable of collecting optical data have become easily and widely accessible, and usable. This has drawn attention to the developments of methods that are scalable and generic for processing obtained data. The area of interest usually cannot be viewed in a single image mainly due to the environmental conditions, dimensions of the area of interest, a necessity for the higher resolution/details, and similar other reasons. All these yielded that most of the time, creating optical maps is in great-demand and in-use by different scientists and engineers since maps provide a global view of the area being surveyed and/or inspected. Over three decades, image mosaicing methods have been developed drastically for creating optical maps in several different areas (*e.g.,* underwater mapping [1], aerial mapping [2] and many different others). Image mosaicing can be considered as composing several lower resolution images into a single higher resolution image referred to as map and/or mosaic. Image mosaicing mainly consists of successive iterations of image matching and trajectory estimation steps. Trajectory estimation is usually carried out by minimizing a pre-defined cost function on feature point positions detected in overlapping image pairs as a result of image matching steps. The cost functions are mostly non-linear and their minimization fundamentally works in iterative manner of solving a linear equation system formed through an approximation of Hessian matrix via Jacobian matrix. This procedure is widely regarded as *Bundle Adjustment* [3] and it is still computational costly, although some improvements were proposed in [4]. On the other hand, the image matching pipeline is much less demanding than trajectory estimation in terms of computational power. Nevertheless, it has to be repeated several times, which results in a considerable amount of time, especially considering the total amount of images to be composed in a dataset.

Identifying overlapping image pairs is a must in order to obtain seamless mosaic images. Overlapping image pairs and the trajectory define a topology and can be represented as a graph in which images are vertices (nodes), and having an overlap between images is denoted by an edge (a link). For a long while, Image mosaicing methods had relied on some additional navigational sensor information [1], and assumptions (*e.g.*, time-consecutive images have an overlap [5]). Later, some methods (*e.g.*, [6]) have been developed for creating image mosaics from completely randomly ordered image sets without any other sensor information. When there is no navigational data available on the mobile robot trajectory, obtaining initial similarity information from images plays a key role in recovering the initial topology and the trajectory. Initial similarity information is obtained by matching a small number of randomly selected feature descriptors without outlier rejection step from images in an all-against-all manner, which is still computationally costly. Lately, some deep-learning-based methods have been proposed for feature detection, and matching (*e.g.*, [7]) and comparative benchmarking was presented in [8]. More recently, the D2-Net framework for jointly detection and description of local features was proposed in [9]. Its performance in localization tasks outperforms the other methods while it has some limitations in image matching.

In this paper, we present an improved framework for obtaining image mosaics from image-only information with as minimum computational efforts as possible for cases where initial similarity information is noisy. The main advantages and properties of the proposed framework can be summarized as below:

- Its iteration procedure does not include the computational costly trajectory estimation step. Instead, iterations are carried out on generating Minimum Spanning Trees (MSTs). This algorithmic enhancement reduces

Authors are with the School of Information Science of Japan Advanced Institute of Science and Technology, Ishikawa, Japan {aelibol,nakyoung}@jaist.ac.jp

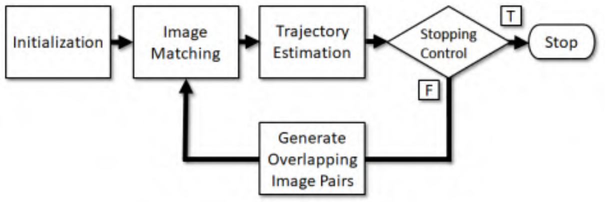


Fig. 1: The general overview of iterative topology estimation frameworks.

the overall computational cost greatly.

- It is capable of obtaining the topology with lesser total number of image matching attempts.
- It can handle cases with the noisy initial information more efficiently.

II. Algorithmically Enhanced Framework for Topology Estimation

Our proposal in this paper is built upon the existing framework proposed in [6]. Topology estimation fundamentally has been carried out as image matching and trajectory estimation steps iteratively. The general overview of the process can be seen in Fig. 1. Obtaining the initial trajectory estimate is a vital part of the estimation frameworks. To do so, initially, time-consecutive images are assumed to have an overlap [5]. This assumption is later on released using initial similarity information [6] and MST. MST was used to establish the image pairs with the highest similarity information possible to be matched in order to obtain an initial trajectory estimate. Due to the coarseness of initial similarity information, some of the image pairs in the MST are usually not matched successfully. Such unsuccessful pairs cause breaks in the connectivity of the topology. In order to keep the tree connected, adding virtual links (or observations) were proposed in [6]. Virtual links were established as identity mapping with high uncertainty between image pairs suggested by the MST. This way of processing allows for obtaining the initial trajectory estimate and keeping the images connected. However, the trajectory estimate with virtual links generates several non-overlapping image pairs as overlapping due to the presence of virtual links. The more virtual links, the more iterations the framework does or requires, thus, the more time. This yields a conclusion that the quality of initial similarity information has an indispensable role throughout the whole process since the inverse of this similarity information is used as weights during MST generation in order to maximize the similarity. The quality of the initial similarity information is directly correlated with the number and the quality of features (feature descriptors) used. Since this information is obtained in an all-images-against-all manner, using a bigger sample size would require substantial computational time, and this would be infeasible.

At this point, we present a novel and efficient initialization step, *MST Initialization Step* namely, to be used within the existing topology estimation frameworks in order to work efficiently with relatively low-quality initial similarity information. MST initialization step relies on iterating on spanning-tree generation instead of image matching and trajectory estimation steps. It is generating a minimum spanning tree and tries to match image pairs in the tree iteratively. If the successfully matched image pairs are not enough to establish a tree (so that every image is involved in at least one link), the new MST is generated with the updated initial similarity information, unlike adding virtual links with identity mapping and high uncertainty as in [6]. For the image pairs that are already attempted to be matched (either successfully matched or not matched), the initial similarity information is set to zero so that they will not appear in the MST(s) during the next iterations. We also delete the similarity information among images that are in the largest strongly connected component of the graph constructed by the successfully matched image pairs. This prevents having image pairs that are already connected in the MST(s) of the next iterations and help to focus on and select image pairs towards the ones that are not connected yet to the graph formed by successfully matched image pairs. This way of iterations is continued until the successfully matched image pairs are enough to form a tree under the assumption of a connected dataset. The step-by-step overview of the enhanced framework is given in Algorithm 1. Since the iterations are carried out over generating MST and image matching steps, its computational cost is more advantageous than its counterparts in which the trajectory estimation through non-linear minimization step is also included in iterations. This initialization step is also free of a threshold and/or user-defined parameter.

Algorithm 1: Algorithm for Enhanced Topology Estimation Framework with *MST Initialization* step

```
Input: Set of Images
Output: Optical Map and Trajectory
Compute Initial Similarity Information
while Successfully Matched Image Pairs not enough
  to form a tree do
    Compute MST using the reciprocal of similarity
      information as edge weights
    Attempt image matching for the image pairs in
      the MST
    Delete similarity information for the image pairs
      in the MST
    Delete similarity information between all image
      pairs that are in the largest strongly connected
      components of the graph constructed by
      successfully matched image pairs.
Estimate a trajectory using matched image pairs
Generate a list of overlapping image pairs using the
  estimated trajectory
Attempt to match new image pairs in the generated
  list
Re-estimate trajectory
Create the optical map using the estimated trajectory
```

III. Implementation Details

We have applied widely-accepted methods for the common steps used in the framework.

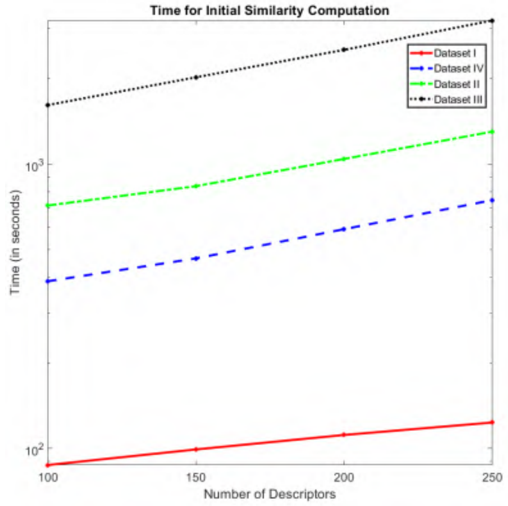
Generation of Initial Similarity Information: To obtain initial similarity information, we followed the same approach presented in [6] using Scale Invariant Feature Transform (SIFT)[10] feature descriptors. Since the initial similarity information is obtained in all-against-all matching, its computational cost increases drastically with the sample size used and the total number of images in the dataset. For descriptor matching, we used multi-threaded C implementation [6] through the mex file interface in the MATLAB© environment. We also used the implementation in VLFeat [11], and obtained times are reported in Fig. 2 to demonstrate the computational cost and time growth. Computational time grows linearly in log-scale with the total number of descriptors used. Our implementation and test platform were not fully optimized; therefore, the time reported is intended to show the computational cost and time growth as they would be similar within different implementations. There are recent studies to make image matching faster [12], [13], [14]; however, our target is to reduce the total number of image matching attempts so that the time spent would reduce regardless of any implementation used. Experiments are carried out on a desktop PC with 64GB RAM and Intel i7-6700K CPU.

Trajectory Estimation Method: For the trajectory estimation step, non-linear minimization of the error metric in Eq. 1 was employed. This error metric is independent of the (selected) global frame and it measures the error distance on the individual image frames therefore, it does not cause any shrinking effect seen in error metrics operating on map (global) frame.

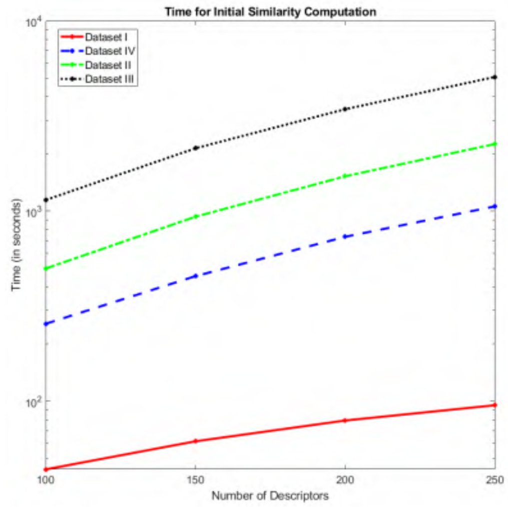
$$\min_{{}^m\mathbf{H}_2, {}^m\mathbf{H}_3, \ldots, {}^m\mathbf{H}_{N_{im}}} \sum_t \sum_r \sum_{j=1}^{\alpha_{tr}} \Big(\| {}^t\mathbf{p}_j - {}^m\mathbf{H}_t^{-1} \cdot {}^m\mathbf{H}_r \cdot {}^r\mathbf{p}_j \|_2 + \| {}^r\mathbf{p}_j - {}^m\mathbf{H}_r^{-1} \cdot {}^m\mathbf{H}_t \cdot {}^k\mathbf{p}_j \|_2 \Big) \tag{1}$$

where t and r are the successfully matched image indexes, α_{tr} is the total number of correspondences between the overlapping image pairs (t and r), $\mathbf{p} = (u, v, 1)^\top$ represents the coordinates of the points in a given image frame, and ${}^m\mathbf{H}_t$ represents the similarity type (4-Degrees of Freedom (DOFs), scale, rotation, translation in x and y) 2D planar transformation from image t to the global frame m. The first image frame is selected as a global frame. Therefore, m is equal to 1. Since $m = 1$, ${}^m\mathbf{H}_1$ becomes a constant identity mapping with $s = 1$, $\theta = 0$, $t_x = 0$, and $t_y = 0$.

Overlapping Image Pairs Generation: This step aims to predict the potentially overlapping image pairs using the initial trajectory estimate and its covariance estimate. The result of the initial trajectory estimate is a set of absolute (global) transformations, $H_i \quad i = 2, 3, \cdots, n$ (for the sake of simplicity, we omit the global frame indicator from the notation). A relative transformation



(a) Total time spent for descriptor matching with different sample size using the implementation in[6]



(b) Total time spent for descriptor matching with different sample size using the implementation in[11]

Fig. 2: Computational Time for generating Initial Similarity Information

between any pair of images can be obtained from the absolute transformations as in Eq. 2.

$$\begin{aligned} {}^rH_t &= (H_r)^{-1} \cdot (H_t) \\ {}^tH_r &= (H_t)^{-1} \cdot (H_r) \end{aligned} \tag{2}$$

By using the relative transformations, we calculate the overlapping areas between image pairs and if the calculated overlapping area is greater than the selected threshold (*e.g.*, 20%) then it is considered as potentially overlapping image pair.

Image Matching: We applied a feature-based image matching pipeline taking profit of geometric invariants [15] and makes use of SIFT [10] for distinctive point detection, description, and matching and Random Sample Consensus (RANSAC) for outlier rejection and transformation computation as an image matching step. An image pair is considered as successfully matched if it

TABLE I: Summary of Dataset Characteristics

Dataset	Total Number of Images (n)	Total Number of Image Pairs	Total Number of Overlapping Pairs	Image Resolution
Dataset I	430	92,235	5,412 ($\sim$5.87%)	384x288
Dataset II	1,650	1,360,425	16,574 ($\sim$1.22%)	512x384
Dataset III	2,403	2,886,003	18,462 ($\sim$0.64%)	512x384
Dataset IV	1,136	644,680	3,895 ($\sim$0.60%)	512x384

has at least 20 inliers.

IV. EXPERIMENTAL RESULTS

We tested our proposal using real datasets obtained by underwater robots. We assume that input images are connected, and they are randomly ordered. The motivation behind the random order is that depending on the application and area of interest, data collection procedures may span more than one deployment, and also some sensor failures may occur. The characteristics of the dataset are summarized in Table I. The total number of overlapping image pairs column represents the total number of successfully matched image pairs and their approximate percentage with respect to the total number of image pairs. Dataset I covers a large area of the seafloor (approximately $400m^2$ and was acquired by the ICTINEU underwater robot [16] during experiments in Colera on the Mediterranean coast of Spain. Datasets II, III, and IV were obtained during different surveys for monitoring coral reef communities located in the Florida Reef Tract near Key Largo in the U.S. [17].

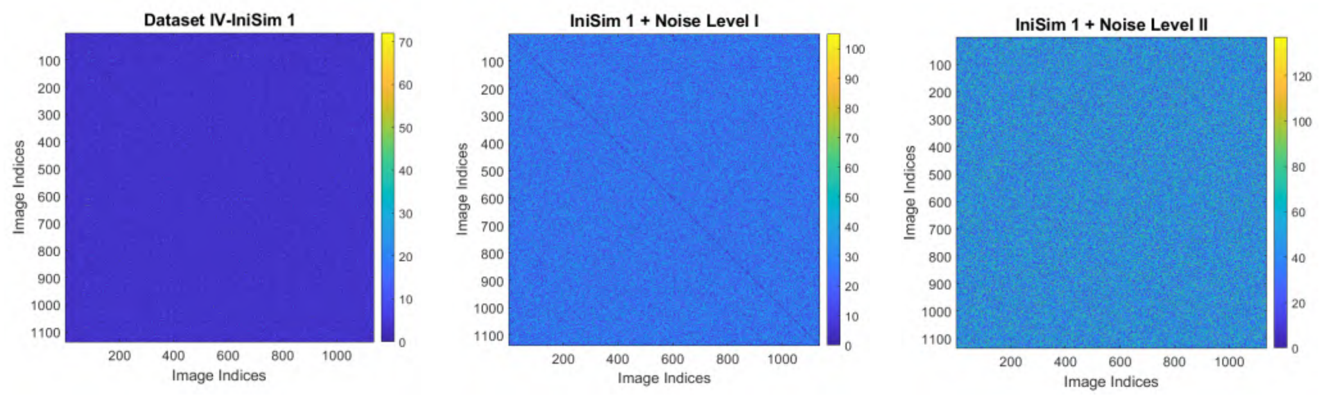
We evaluated the performance of the proposed method and its counterpart in [6] using different initial similarity information matrices that were computed using a set of a different number of sub-sample sizes (*e.g.*, 100 and 150) used in descriptor matching. An example of an initial similarity matrix is given in Fig. 3a. Initial similarity information quality is judged with the unsuccessful image matching attempts during the first iteration, thus, in the first MST, since that is the total number of virtual links needed in order to keep trajectory/topology connected [6]. Obtained results are summarized in Tables II and III. While Table II presents the obtained results of the enhanced framework in detail, Table III presents the comparison with the method in [6]. In both tables, columns *successful* and *unsuccessful* denote the number of image matching attempts that resulted correspondingly. In Table II the number of matching attempts for the first iteration is also presented to evaluate the initial similarity information quality. The presented results have achieved similar trajectory accuracy except for Dataset II. The method in [6] was not able to obtain the correct trajectory, which was obtained using all-against-all exhaustive image matching, whereas the presented method was able to achieve similar accuracy. The final mosaic images obtained are given in Fig. 4. From the tables, the proposed framework performs well in the case of noisy initial similarity information. It can be seen from Table III that the framework in [6] performed better for the cases of rather less noisy initial similarity information of Dataset IV when the total number of image matching attempts are compared. However, It should be noted that the proposed framework also has the advantage of not repeating overlapping image pairs generation step and trajectory estimation step (via non-linear minimization), which has higher computational cost than computing a MST. Therefore, even in the cases of small discrepancy between the total number of image matching attempts (*e.g.*, Dataset IV-IniSim 2 and IniSim used in [6]), we observed that the enhanced framework required less computational time during our experiments. In order to see the performance of the proposed framework in the existence of very low-quality initial similarity information, we added uniformly distributed random integers to the *IniSim 1* of the Dataset IV and ran the proposed framework. We tested two different noise levels, and for each level, we generated 10 different noise-added initial similarity matrices. For the *Noise Level 1*, noise values were drawn from the interval $[0, 50]$ while for the *Noise Level 2*, the interval was $[0, 100]$. An examples of noise added initial similarity matrices are given in Fig. 3b and Fig. 3c. Obtained results (in the form of mean, minimum, and maximum of 10 trials) are presented in Table IV. The enhanced framework was able to obtain the trajectory with similar accuracy to the one obtained using an all-against-all strategy from the cases where the first MST had very few (as low as $30.30\%\,(\frac{344}{1135})$ for Noise Level I and $21.76\%\,(\frac{247}{1135})$ for Noise Level II) successfully matched image pairs. Furthermore, if the total number of image matching attempts compared with the ones in Table III, it can be observed that it has increased by approximately $2-4$ times. Even with such noisy initial similarity information, the total number of image matching attempts is still lower than 30% of the all-against-all matching attempts. Taking into account the computational times needed for obtaining initial similarity information (referring to Fig. 2) and the time needed for image matching attempts, using a bigger number of descriptors would be more advantageous since better initial similarity information provide less the total number of image matching attempts. On the other hand, using a larger sample size of descriptors may not necessarily provide better quality since descriptor matching is applied solely without outlier rejection procedure. One way to obtain better quality initial similarity information would be using a bi-directional descriptor matching strategy. This would increase the computational time needed, presumably two times (there are also some improved methods proposed recently [18]). However, taking into account the difference would make in the total number of image matching attempts, It would reduce the overall total time spent.

V. CONCLUSIONS

Creating optical maps for a detailed and global view of the area surveyed through data obtained by mobile platforms is one of the essential steps for different scientific and engineering studies. One of the crucial steps in creating optical maps from relatively lower resolution images is pairwise image registration (image matching). In the case of no other data available except images, matching images all-against-all manner in order to obtain the topology graph is still prohibitively computational costly, although several advancements in pairwise image matching have been achieved lately.

TABLE II: Total number of image matching attempts obtained by the proposed framework with *MST Initialization* using different initial similarity information of tested datasets

Dataset	Ini. Similarity Information	First MST/Iteration		Iteration Part				Refinement Part		
		Successful	Unsuccessful	Iteration	Successful	Unsuccessful	Total	Successful	Unsuccessful	Total
Dataset I	IniSim 1	424	5	3	556	731	1,287	4,767	1,476	6,243
	IniSim 2	426	3	3	527	760	1,287	4,799	1,492	6,291
	IniSim used in [6]	428	1	2	445	413	858	4,874	1,449	6,323
Dataset II	IniSim 1	985	664	48	3,391	75,761	79,152	13,031	50,446	63,477
	IniSim 2	1,414	235	21	3,353	31,174	34,527	13,102	48,470	61,572
Dataset III	IniSim 1	2,188	214	8	5,752	13,464	19,216	12,685	132,285	144,970
	IniSim 2	2,220	182	8	5,794	13,422	19,216	12,658	110,758	123,416
Dataset IV	IniSim 1	698	437	7	1,620	6,318	7,938	2,118	30,358	32,476
	IniSim 2	786	349	7	1,665	5,630	7,295	2,076	30,925	33,001
	IniSim used in [6]	934	201	4	1,769	2,771	4,540	1,987	32,358	34,345

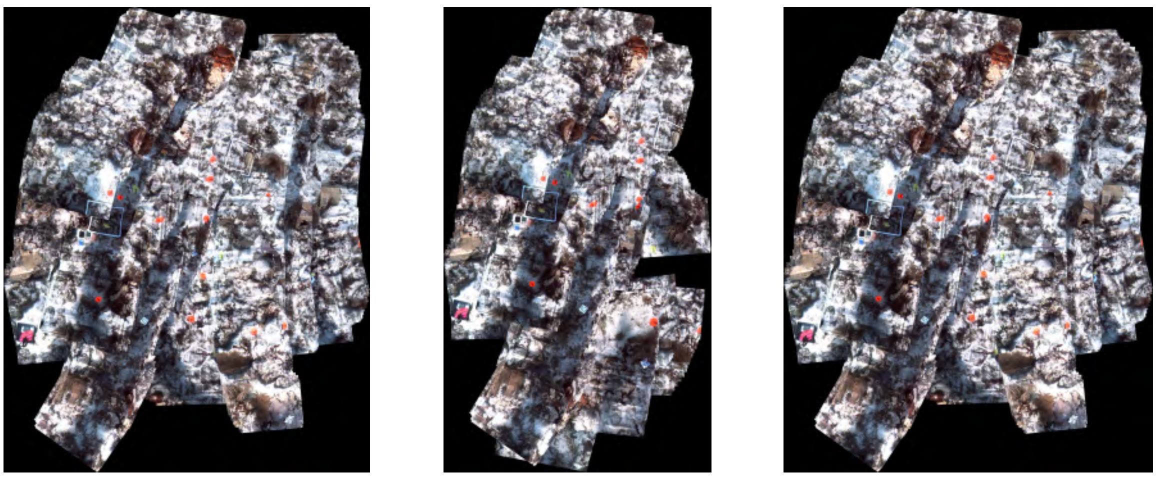


(a) Initial Similarity Matrix for Dataset IV showing the total number of matched features without rejecting outliers

(b) An example of Noise (Level I) added Initial Similarity Matrix

(c) An example of Noise (Level II) added Initial Similarity Matrix

Fig. 3: Samples of Initial Similarity Information Matrices



(a) Obtained with the proposed framework. $2,167 \times 2,753$ pixels

(b) Obtained with the approach in [6] with *IniSim 1*. $1,523 \times 2,646$ pixels

(c) Obtained with all-against-all image matching. $2,210 \times 2,781$ pixels

Fig. 4: Final Mosaic Images for Dataset II rendered with *Last-on-Top* strategy

TABLE III: Comparison of total number of image matching attempts

Dataset	Ini. Similarity Information	Framework with *MST Initialization*			Framework in [6]		
		Successful	Unsuccessful	Total	Successful	Unsuccessful	Total
Dataset I	IniSim 1	5,323	2,207	7,530	5,388	2,344	7,732
	IniSim 2	5,326	2,252	7,578	5,388	2,540	7,928
	IniSim used in [6]	5,319	1,862	7,181	5,386	1,961	7,347
Dataset II	IniSim 1	16,422	126,207	142,629	16,311	405,306	421,617
	IniSim 2	16,455	79,644	96,099	16,150	179,984	196,134
Dataset III	IniSim 1	18,437	145,749	164,186	18,440	175,161	193,601
	IniSim 2	18,452	124,180	142,632	18,448	162,467	180,915
Dataset IV	IniSim 1	3,738	36,676	40,414	3,750	43,850	47,600
	IniSim 2	3,741	36,555	40,296	3,752	36,127	39,879
	IniSim used in [6]	3,756	35,129	38,885	3,844	33,970	37,814

TABLE IV: Experimental results of the framework with *MST Initialization* using noise corrupted initial similarity information

Noise	Framework Parts		Mean	Min	Max
Level I	First Iteration/MST	Successful	364.1	344	378
		Unsuccessful	770.9	757	791
		Iteration	74.2	67	78
	Iteration Part	Successful	1,292.2	1,250	1,335
		Unsuccessful	82,718	74,733	87,068
		Total	84,010	76,009	88,364
	Refinement Part	Successful	2,425.9	2,354	2,463
		Unsuccessful	26,346.4	25,422	27,018
		Total	28,772.3	27,776	29,467
Level II	First Iteration/MST	Successful	260.6	247	278
		Unsuccessful	874.4	857	888
		Iteration	125.6	121	138
	Iteration Part	Successful	1,235	1,197	1,281
		Unsuccessful	140,893	136,096	154,585
		Total	142,128	137,327	155,811
	Refinement Part	Successful	2,498	2,467	2,536
		Unsuccessful	24,967.3	23,929	26,862
		Total	27,465.3	26,425	29,335

In this paper, we present a different initialization step, which algorithmically improves the iterative topology estimation frameworks by reducing the total number of pairwise image matching attempts in order to obtain the topology efficiently in the presence of low-quality initial similarity information. The iteration steps are changed from image matching and trajectory estimation steps to mainly obtaining MSTs and image matching. Its efficiency has been presented with extensive and comparative experimental results on both real underwater image datasets and simulations.

ACKNOWLEDGMENT

The authors would like to thank Computer Vision and Robotics Group of the University of Girona for sharing datasets used in experiments. This work was funded by the Air Force Office of Scientific Research under AFOSR/AOARD FA2386-20-1-4019 grant.

REFERENCES

[1] J. Ferrer, A. Elibol, O. Delaunoy, N. Gracias, and R. Garcia, “Large-area photo-mosaics using global alignment and navigation data,” in *MTS/IEEE OCEANS Conference*, Vancouver, Canada, November 2007, pp. 1–9.

[2] M. Xia, M. Yao, L. Li, and X. Lu, “Globally consistent alignment for mosaicking aerial images,” in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3039–3043.

[3] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, “Bundle adjustment – A modern synthesis,” in *Vision Algorithms: Theory and Practice*, ser. LNCS. Springer Verlag, 2000, pp. 298–375.

[4] S. Agarwal, K. Mierle, and Others, “Ceres solver,” http://ceres-solver.org.

[5] N. Gracias, S. Zwaan, A. Bernardino, and J. Santos-Victor, “Mosaic based navigation for autonomous underwater vehicles,” *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 609–624, Oct. 2003.

[6] A. Elibol, N. Gracias, and R. Garcia, “Fast topology estimation for image mosaicing using adaptive information thresholding,” *Robotics and Autonomous Systems*, vol. 61, no. 2, pp. 125–136, 2013.

[7] X.Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg, “Matchnet: Unifying feature and metric learning for patch-based matching,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3279–3286.

[8] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, “HPatches: A benchmark and evaluation of handcrafted and learned local descriptors,” in *CVPR*, 2017.

[9] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler, “D2-Net: A Trainable CNN for Joint Detection and Description of Local Features,” in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

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

[11] A. Vedaldi and B. Fulkerson, “VLFeat: An open and portable library of computer vision algorithms,” http://www.vlfeat.org/, 2008.

[12] C. Wu, “Towards linear-time incremental structure from motion,” in *2013 International Conference on 3D Vision - 3DV 2013*, 2013, pp. 127–134.

[13] J. Cheng, C. Leng, J. Wu, H. Cui, and H. Lu, “Fast and accurate image matching with cascade hashing for 3D reconstruction,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1–8.

[14] T. Xu, K. Sun, and W. Tao, “GPU accelerated cascade hashing image matching for large scale 3D reconstruction,” https://arxiv.org/abs/1805.08995v1, 2018.

[15] A. Elibol and N. Y. Chong, “Efficient image registration for underwater optical mapping using geometric invariants,” *Journal of Marine Science and Engineering*, vol. 7, no. 6, p. 178, 2019.

[16] D. Ribas, N. Palomeras, P. Ridao, M. Carreras, and E. Hernandez, “ICTINEU AUV wins the first SAUC-E competition,” in *IEEE International Conference on Robotics and Automation*, Roma, Italy, April 2007.

[17] D. Lirman, N. Gracias, B. Gintert, A. Gleason, R. P. Reid, S. Negahdaripour, 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.

[18] X. Liu, S. Zhou, H. Li, and K. Li, “Bidirectional scale-invariant feature transform feature matching algorithms based on priority k-d tree search,” *International Journal of Advanced Robotic Systems*, vol. 14, no. 1, p. 1729881416682700, 2017.