Pattern Recognition Letters 46 (2014) 20–26

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Efficient image mosaicing for multi-robot visual underwater mapping ☆

Armagan Elibol [a], Jinwhan Kim [a,*], Nuno Gracias [b], Rafael Garcia [b]

[a] Division of Ocean Systems Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[b] Computer Vision and Robotics Institute, University of Girona, Girona, Spain

ARTICLE INFO

Article history:
Received 8 August 2013
Available online 15 May 2014

Keywords:
Image mosaicing
Visual mapping
Unmanned underwater vehicles

ABSTRACT

Robotic platforms have advanced greatly in terms of their remote sensing capabilities, including obtaining optical information using cameras. Alongside these advances, visual mapping has become a very active research area, which facilitates the mapping of areas inaccessible to humans. This requires the efficient processing of data to increase the final mosaic quality and computational efficiency. In this paper, we propose an efficient image mosaicing algorithm for large area visual mapping in underwater environments using multiple underwater robots. Our method identifies overlapping image pairs in the trajectories carried out by the different robots during the topology estimation process, being this a cornerstone for efficiently mapping large areas of the seafloor. We present comparative results based on challenging real underwater datasets, which simulated multi-robot mapping.

1. Introduction

One of the major application fields for mobile robots is the exploration of unknown and/or hazardous environments, especially areas that have little or no access for humans, *e.g.*, space and underwater. In particular, underwater environments have been increasingly explored and mapped thanks to the rapid development of underwater robotic platforms during the last decades. Platforms such as Remotely Operated Vehicle (ROV) and Autonomous Underwater Vehicle (AUV) have demonstrated their utility and have gained significant importance in various marine application areas, including scientific, industrial, and military. Recently, low-cost robots with a very limited sensor suite (*e.g.*, just a camera and an acoustic altimeter) have attracted much attention, especially for different surveying purposes since they are relatively easy to deploy, operate, and maintain with little expertise. Although they are limited in terms of coverage and autonomy, they can be used as economical and scalable vehicle platforms. These applications are promoting the use of multiple vehicles to survey larger areas to obtain optical data, which is especially relevant in underwater applications since robots need to fly very close to the seafloor to gather images, making data collection extremely time-consuming. Automatic optical mapping of the underwater terrain (*i.e.*, building maps using digital images) with Unmanned Underwater Vehicle (UUV) is a major requirement for researchers working in the aforementioned areas. Optical underwater mapping addresses the problem of building maps of the seabed using images recorded by optical imaging systems as still images or videos. Optical imaging provides higher resolution compared to acoustic imaging [1] but the use of optical information obtained by cameras in underwater environments has some limitations because of light absorption, strong attenuation, and forward and back scattering. These limitations do not allow the area of interest to be captured in a single image. This demands methods that combine several overlapping images into a single image that encompasses the global perspective of the area surveyed. These methods are referred to as image mosaicing and the final image is referred to as a mosaic. Mosaics have been used extensively in various fields such as geological [2] and archaeological surveys [3], biology [4–6], environmental monitoring and damage assessment [7], and temporal change detection [8].

The use of multiple robotic platforms produces large volumes of raw optical imagery data, which are not suitable for direct human interpretation. Therefore, new methods are required to process the data efficiently but without sacrificing the final quality of the mosaic produced. In this paper, we propose an efficient method for aligning partially overlapping datasets obtained by using multiple robotic platforms to survey the same area of interest in order to generate a single global visual map (mosaic). When using a set of low-cost underwater robots, without expensive navigation sensors such as Ultra Short Base Line (USBL) or even Doppler Velocity Log (DVL), it is not realistic to assume that their trajectories are accurate enough to cover the right zones of a pre-determined grid.

☆ This paper has been recommended for acceptance by S. Sarkar.
* Corresponding author. Tel.: +82 42 350 1519.
E-mail addresses: aelibol@kaist.ac.kr (A. Elibol), jinwhan@kaist.ac.kr (J. Kim), ngracias@eia.udg.edu (N. Gracias), rafael.garcia@udg.edu (R. Garcia).

http://dx.doi.org/10.1016/j.patrec.2014.04.020
0167-8655/

On the contrary, their trajectories will drift, leaving gaps in their own grid, and partly covering their neighboring grids, fact that is exploited by our proposal. We assume that a set of image sequences is used as the input, which does not necessarily include additional sensor information. It is expected that time-consecutive images have overlapping areas, which is a common practice during surveys using low-cost robots equipped with very limited sensor suite (mainly a camera and an altimeter). While determining the topology of the single dataset, we actively seek similar regions in the other datasets. After a certain amount of similarity has been detected, we try to match images from different datasets. Successfully matched pairs provide transformations, which are later used to represent and update the second trajectory with respect to the first. After completing the topology estimation for the first dataset, possible overlapping image pairs in the second dataset and pairs between datasets are generated by computing the distances between the image centres. Finally, Bundle Adjustment (BA) [9] is used to obtain a globally coherent combined map. We present experimental results based on challenging real underwater image sequences.

The rest of the paper is organised as follows: the following section briefly reviews some related work, mainly in the context of multi-robot Simultaneous Localization and Mapping (SLAM). Section 3 provides the details of the proposed method. Section 4 discusses the possible advantages of the proposed method compared with existing techniques in terms of the computational resources required. Some results are presented in Section 5 and our conclusions are given in the last section.

2. Related work

Multi-robotic platforms have been widely used for mapping purposes, especially in a Simultaneous Localization and Mapping (SLAM) context. One of the main problems is to find common information in maps (or trajectories) generated by different robots so they can be jointly represented in a common global frame. This is known as the *map merging* problem. One of the basic approaches to this problem assumes that the initial positions of the robots are known [10]. This assumption does not hold in real experiments in an environment such as the ocean if the robots are near seafloor and have a limited sensor suite. Another common assumption is that the information required in (local) maps is provided by robots that observe themselves and/or communicate during the mission (referred to as the *rendezvous* problem), which allows them to estimate suitable transformations among their positions and the maps [11–13]. There are various approaches for topological map merging [14] and occupancy grid mapping [15]. Topological merging represents the surveyed area as a graph so the problem of merging maps can be modelled as finding the maximal matching subgraph, which is a (NP-hard) combinatorial optimisation problem. Similarly, the use of visual similarities was proposed in [16], which makes use of clustering methods to build visual words. Such techniques based on visual vocabularies are behind the recent widespread of panorama creation and 3D reconstruction algorithms [17]. However, some words may contain few informative features, which might occur often in several places, especially in unmanned environments as in our application domain. Similar scene elements might be repeated over several images although the images are not spatially neighbouring. Also most approaches based on visual words require a prior training (vocabulary building) phase. Another point is that the approach in [16] computed the similarity of all image pairs in sequences, unlike the method we propose in this paper. Their alignment procedure relied on finding the best matching pair of image subsequences using dynamic programming, which is known to be a highly computationally expensive technique.

All of the methods described above have been proposed in a Simultaneous Localization and Mapping (SLAM) framework where the platforms were equipped with different sensors that provided navigation feedback (*e.g.*, the angle, range, and odometry). Our application scenario involves low-cost underwater robots surveying an area of interest. In general, such vehicles are equipped only with a video camera and a pressure sensor [18,19]. We aim to obtain a map of the surveyed area using the minimum computational effort, such as a reduced number of image matching attempts. Therefore, we consider all of the possible matching pairs among all images in the different datasets obtained by each robot, which is a different problem from performing matches of the most recent image with all previous images as the robot moves like in the Simultaneous Localization and Mapping (SLAM) approach.

The process of obtaining the overlapping image pairs and the trajectory is referred to as topology estimation [20,21]. The method we propose in this paper is based on the recent method proposed in [22]. Briefly, the approach in [22] uses a graph-based topology representation, where images are nodes, and the overlap between two images is denoted by an edge or a link. The approach makes use of the initial similarity information in order to have initial trajectory estimate. This similarity information is based on comparing feature descriptors in images. It is computed in a fast and approximate way using on Scale Invariant Feature Transform (SIFT) [23]. A small subset of feature descriptors (between 100 and 200) are randomly selected from each image, and compared against the subsets of all other images. This comparison is performed using the Euclidean distance between feature descriptors [23].

In order to establish the initial links between images, a Minimum Spanning Tree (MST) where weights of the edges are the inverted initial similarity values is used. Minimum Spanning Tree (MST) provides a connected tree composed of the most similar image pairs according to the similarity information. Image pairs are attempted to be matched and then an initial trajectory estimate is obtained. Having a trajectory estimate enables to find potentially overlapping image pairs. A selected subset of the potentially overlapping image pair list is attempted to be matched and the trajectory estimate is updated by minimizing the symmetric transfer error given in Eq. (8) and including successfully matched new image pairs. MATLAB™ *lsqnonlin* function for large-scale methods was used to minimise the cost function in Eq. (8). As a final step, the uncertainty of the resulting homography estimations is propagated using Haralick's method [24]. The approach operates in an iterative manner by refining the trajectory estimation and image matching until no new image pair needs to be matched.

3. Multi-robot image mosaicing

First, time-consecutive image pairs are matched in all datasets. The associated uncertainty is obtained using first order propagation [24] under the standard assumption that the uncertainty in the matching can be modelled as additive zero-mean Gaussian noise that corrupts one of the lists of point coordinates. Next, the topology estimation framework is applied to the dataset with the largest amount of images, which is the one where the effects of the error accumulation are likely to be higher. For the other datasets, keyframe images are extracted by selecting images that have between 10% and 20% overlap, and discarding others. These extracted keyframe images are later used to compute the similarity level among images in different datasets. The aim of using keyframes is to reduce the computational cost of the similarity search between datasets. During the process of estimating the topology of the largest dataset (referred to as primary dataset hereafter), we compute similarity between images in the primary dataset and keyframes in other datasets (hereafter, referred to as the secondary