Robotics and Autonomous Systems 62 (2014) 1259–1270

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Vision-based sparse topological mapping

Hemanth Korrapati [a,*], Youcef Mezouar [a,b]

[a] Institut Pascal (UMR 6602 UBP/CNRS), Aubière-63170, France
[b] Institut Français de Mécanique Avancée (IFMA), Aubière-63170, France

HIGHLIGHTS

- A hierarchical topological mapping algorithm with sparse node representation.
- Hierarchical Inverted Files are proposed for efficient two-level map storage.
- Various filters to process the similarity vectors at node and image levels.
- A relative motion model to correlate odometry data of present and previous visits.
- Comparison with state of the art techniques, and accuracy and sparsity analysis.

ARTICLE INFO

Article history:
Available online 12 April 2014

Keywords:
Loop closure
Visual mapping
Topological maps

ABSTRACT

Most of the existing appearance-based topological mapping algorithms produce dense topological maps in which each image stands as a node in the topological graph. Sparser maps can be built by representing groups of visually similar images of a sequence as nodes of a topological graph. In this paper, we present a sparse/hierarchical topological mapping framework which uses Image Sequence Partitioning (ISP) to group visually similar images of a sequence as nodes which are then connected on the occurrence of loop closures to form a topological graph. An indexing data structure called Hierarchical Inverted File (HIF) is proposed to store the sparse maps so as to perform loop closure at the two different resolutions of the map namely the node level and image level. TFIDF weighting is combined with spatial and frequency constraints on the detected features for improved loop closure robustness. Our approach is compared with two other existing sparse mapping approaches which use ISP. Sparsity, efficiency and accuracy of the resulting maps are evaluated and compared to that of the other two techniques on publicly available outdoor omni-directional image sequences.

1. Introduction

Mapping is one of the fundamental problems of Autonomous Mobile robotics. Mapping problem can be widely categorized as Topological and Metrical [1]. Metrical mapping involves accurate position estimates of robots and landmarks of the environment. Topological mapping on the other hand represents an environment as a graph in which nodes correspond to places and the edges between them indicate some sort of connectivity. Recently, a third category called Topo-Metric Mapping [2,3] is gaining popularity. Topo-Metric mapping is a hybrid approach which uses both metrical and topological information in map building. Building an accurate map either metrical or topological depends on loop closure accuracy. Such maps are difficult to build using metrical information which is prone to gross errors in position estimation of robot and landmarks. On the other hand topological maps simply rely on the detection of topological connectivity of locations rather than demanding precise metrical accuracy of the environment. This connectivity is signaled by loop closure making it the crux of any topological mapping algorithm.

Many powerful vision-based topological mapping techniques that heavily rely on loop closure have been proposed over the past decade [4–8]. Most of them produce dense topological maps, in which every acquired image stands as a node in the topological graph. Similarly, sparse topological maps can be built in which each node represents a group of images rather than representing individual images. The images belonging to a node are sequentially contiguous and hence, spatially close, visually similar and collectively represent a place of the environment. Each node in a sparse

* Corresponding author.
E-mail addresses: hemanthkorrapati@gmail.com, Korrapati.Hemanth@univ-bpclermont.fr (H. Korrapati), youcef.mezouar@ifma.fr (Y. Mezouar).

http://dx.doi.org/10.1016/j.robot.2014.03.015

topological map can be understood as a place — region of an environment throughout which visual appearance remains more or less constant. Sparse topological maps contain far less number of nodes compared to the total number of images and hence the name sparse. Sparse maps can also be interpreted as two level hierarchical maps since the maps consist of nodes which in turn consist images. Following are some of the important advantages of sparse maps:

- For a loop closure all the images of a node/place are used rather than individual image comparison and hence accuracy can be increased.
- Hierarchical loop closure can be performed which quickly shrinks the search space to a few images which are thoroughly matched still retaining real-time operation.
- Place representation in maps can aid in accurate semantic labeling of topological map nodes [9,10] that can be used for lifelong operation and navigation of robots.
- Accurate and efficient map merging [11].

Breaking a sequence of images acquired by the robot into nodes/places is called Image Sequence Partitioning (ISP) and we achieve this using a local feature matching based similarity evaluation.

To facilitate hierarchical map representation and feature indexing, a data structure called Hierarchical Inverted File (HIF) is proposed. As opposed to the traditional inverted files [12,13], HIFs store features hierarchically in two levels – node level and image level. HIFs enable loop closure at two resolutions – a coarse node level loop closure which finds the most similar node and a finer image level loop closure which pin-points the most similar image inside a node. Constraints that exploit spatial locations and occurrence frequency of features in images are used to strengthen the image level loop closure and thereby eliminate false positives. A filter taking advantage of the temporal consistency of loop closure which also depends upon the vehicle velocity is used in validating loop closures. This paper is an extension of our previous work [14,15].

Experimentation was performed using omni-directional images from two of our own outdoor datasets and panoramic images from the popular NewCollege dataset [16], all of which are publicly available. We compare our approach to two other mapping approaches that use ISP: the first is based on GIST [17,18] and the second one uses Optical Flow [19]. Sparsity and accuracy of maps constructed using different ISP techniques are evaluated and the power of HIF representation in time efficient loop closure is demonstrated.

2. Related work

Scene Change Detection and Key Frame Selection for video segmentation and abstraction [20,21] have similar goals as that of ISP. They try to represent a video with fewer images called key frames whenever there is a sufficient change in the scene and most of them focus on the video compression domain. The major difference between these video abstraction problems and mapping is that mapping demands the localization of a query image which is obtained at a previously visited place, but with varied illumination, viewpoint, and a possible occlusion. Hence, video segmentation techniques using pixel-wise intensity measures and global image features like histograms, motion-based segmentation cannot be applied to our problem.

Quite a few loop closure techniques for topological maps have been proposed recently for both indoor [22,8,4,5,23] and outdoor environments [6,7,18,19]. However, only a few of them concentrate on generating sparse maps.

Angeli et al. [4] introduces a loop closure algorithm that incrementally adds images to a vocabulary tree. Image similarity is evaluated probabilistically using a Bayesian filter which combines temporal information to detect loop closures. [5] uses the same framework for topo-metric mapping by encoding translation and rotation values across edges of the topological graph. In [6] visual word pairs with maximum co occurrence probability are learned using chow-liu trees and used in a generative model embedded in a Bayesian filter to evaluate loop closures. This framework is extended in [7] to use inverted files and speed up the approach. [24,25] embed the motion model in the Bayesian filter to improve loop closure accuracy. All these approaches construct perform just loop closure and the resulting maps are dense where each image is represented as a node in the topological graph.

In [22,26] topological maps are built for indoor environments. They segment the topological graph of the environment using the normalized graph-cuts algorithm resulting in subgraphs corresponding to convex areas in the environment. In [23] SIFT features were used to perform matching over a sequence of images. They detect transitions between individual indoor locations depending on the number of SIFT features which can be successfully matched between the successive frames. In [27] fingerprint of an acquired image is generated using the omni-directional image and laser readings, and these fingerprints are used in loop closure. If the similarity is above a threshold the image is added to the existing node and if not a new node is formed. Change point detection has been used for indoor place segmentation in [9]. All the above works experimented on indoor environments which contain convex spaces (like rooms) and are relatively easier to be partitioned when compared to outdoor environments which in general do not have clear boundaries that separate places.

A sparse topological mapping framework using incremental spectral clustering has been presented in [28] and tested in outdoor environments. Nodes are constructed using incremental spectral clustering over the affinity matrix of the images, producing a sparse topological graph. Another ISP technique was presented in [19] which used mean optical flow vector length to signal changes in scene and nodes are formed only at those locations. [18] performs sparse mapping through ISP using rotation invariant GIST [29] features called omni-gist.

3. Framework overview

Fig. 1 depicts a global overview of our framework. Given a query image or a newly acquired image, local image features are extracted and quantized into visual words. Any one out of the rich variety of existing local feature detectors like SIFT, SURF, etc., can be used for local image feature extraction. Quantization of local image features into visual words is performed using a visual vocabulary tree learned on a training dataset. The features extracted from the query image are then used for a node level loop closure check. Basically, node level loop closure evaluates the similarity of the query image to all the reference nodes (can also be called node-image similarity) of the topological map. On the occurrence of a node level loop closure, which means that the query image is probably acquired at a previously visited place(s)/node(s) in the map, the member images of those nodes are tabbed as reference images for a more thorough image level loop closure. Image level loop closure involves an image-to-image similarity evaluation of the query image with respect to all the reference images, and filtering the similarity for valid loop closures. Howbeit, if an image level loop closure does not occur (if the query image is not similar to any reference image), the image is considered for comparison with the current place node. Likewise, if a node level loop closure does not occur, then also the query image is sent for comparison with the current place.

As can be seen in the figure, similarity evaluation of the query image to the current place node mainly constitutes ISP. At this