



High performance loop closure detection using bag of word pairs



Nishant Kejriwal^a, Swagat Kumar^{a,*}, Tomohiro Shibata^b

^a Innovation Labs, Tata Consultancy Services, New Delhi, India

^b Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan

HIGHLIGHTS

- We propose a new method for loop closure detection for topological mapping.
- It uses relative spatial co-occurrence information to improve the performance.
- We augment BoW method with a dictionary of spatially co-occurring word pairs.
- A memory map data structure is used for storing and indexing word pairs.
- We incorporate best of the existing methods to provide state-of-the-art performance.

ARTICLE INFO

Article history:

Received 29 August 2015

Accepted 9 December 2015

Available online 24 December 2015

Keywords:

Topological mapping

SLAM

BoW

BoWP

Relative spatial co-occurrence

RANSAC

Loop closure detection

Bayesian filtering

ABSTRACT

In this paper, we look into the problem of loop closure detection in topological mapping. The bag of words (BoW) is a popular approach which is fast and easy to implement, but suffers from perceptual aliasing, primarily due to vector quantization. We propose to overcome this limitation by incorporating the spatial co-occurrence information directly into the dictionary itself. This is done by creating an additional dictionary comprising of word pairs, which are formed by using a spatial neighborhood defined based on the scale size of each point feature. Since the word pairs are defined relative to the spatial location of each point feature, they exhibit a directional attribute which is a new finding made in this paper. The proposed approach, called bag of word pairs (BoWP), uses relative spatial co-occurrence of words to overcome the limitations of the conventional BoW methods. Unlike previous methods that use spatial arrangement only as a verification step, the proposed method incorporates spatial information directly into the detection level and thus, influences all stages of decision making. The proposed BoWP method is implemented in an on-line fashion by incorporating some of the popular concepts such as, K-D tree for storing and searching features, Bayesian probabilistic framework for making decisions on loop closures, incremental creation of dictionary and using RANSAC for confirming loop closure for the top candidate. Unlike previous methods, an incremental version of K-D tree implementation is used which prevents rebuilding of tree for every incoming image, thereby reducing the per image computation time considerably. Through experiments on standard datasets it is shown that the proposed methods provide better recall performance than most of the existing methods. This improvement is achieved without making use any geometric information obtained from range sensors or robot odometry. The computational requirements for the algorithm is comparable to that of BoW methods and is shown to be less than the latest state-of-the-art method in this category.

© 2015 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Simultaneous localization and mapping (SLAM) is an important problem in mobile robotics which needs to be solved in

order to achieve autonomous navigation. There exist two types of approaches to solve this problem—metric and topological. Metric SLAM aims to build a geometric map of the environment and hence requires accurate robot pose estimation. On the other hand, topological SLAM aims at building a graphical model of the environment comprising of *key locations* and their *connectivity* without explicitly making use of geometric or odometric information. Most of the topological SLAM research makes use of visual sensors which have become a common and inexpensive accessory in robotic

* Corresponding author.

E-mail addresses: nishant.kejriwal@tcs.com (N. Kejriwal), swagat.kumar@tcs.com (S. Kumar), tom@brain.kyutech.ac.jp (T. Shibata).

<http://dx.doi.org/10.1016/j.robot.2015.12.003>

0921-8890/© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

applications. There is a third type of method called topo-metric methods [1–3] that combine metric and topological informations to obtain better performance.

One of the key aspects of a SLAM system is the loop closure detection [4,5] which requires a robot to recognize previously visited places accurately and correctly when they are revisited. The challenge lies in solving the *perceptual aliasing* problem because of which two physically distinct locations may appear similar to robot sensors.

This paper focuses on topological mapping that uses appearance (image)-based methods for loop closure detection. These methods use image similarity to identify previously visited places and hence, the success of such SLAM systems rely on having a robust place recognition algorithm. The Bag-of-Words (BoW) approach [6,4,7,8] is one of the most popular methods in this category. In this method, an image is represented as a histogram of words present in a dictionary. Usually an off-line dictionary is created by clustering similar features extracted from a large set of images. Histogram comparison is used to find the similarity between a query image acquired recently with the existing images in the map. Although BoW gives good results with very less computation time, it suffers from perceptual aliasing due to vector quantization. The problem due to quantization could be solved by using *direct feature matching* approaches [9–11] where the raw features are used directly for computing image similarity instead of their quantized representation obtained through clustering. Even though these methods are shown to provide better recall performance, they have a higher computational requirement which increases with growing map size and thus, makes it prohibitive for larger maps.

In this paper, we aim to improve the recall performance of BoW methods without sacrificing its simplicity and speed of execution. This is done by incorporating spatial co-occurrence information directly into the dictionary itself. In other words, we create a dictionary of word pairs in addition to the dictionary of individual words. A word pair is formed by observing that the spatial neighborhood of a word location may include other nearby words which, together with the former, may provide better discrimination in identifying loop closures. Since the point features (like SURF [12]) are used as words in the dictionary, the extent of spatial neighborhood for each word is defined by the scale size of the point feature. This scale size depends on the scale at which it is detected in the feature extraction algorithm and thus, is not a user-defined parameter. Since the word pairs are defined relative to the spatial neighborhood of each feature, these word pairs exhibit a directional attribute which has not been exploited so far in the literature. It is possible to extend this approach to word triplets and quadruplets with increasing cost of computation and memory storage requirement. Our consideration in this paper is limited to word pairs so as to keep the computational requirement closer to that of BoW method.

The usefulness of co-occurring words in addressing the perceptual aliasing problem is well known. For instance, Cummins and Newman [6] use a Chow–Liu tree to capture the co-occurrence information into the observation likelihood. While doing so, they do not take into account their spatial proximity to each other in the image. A pair of images will be called similar as long as they contain the same set of words in them irrespective of their spatial arrangement. Secondly, creation of Chow–Liu tree is a computationally expensive process which is usually built off-line to meet the real-time requirements. In another work, Stumm et al. [13] used co-visibility maps to incorporate the co-occurrence information into the decision making process. In this method, a graph of various landmarks is maintained by linking landmarks that are visible together and then a search is performed to find a cluster of landmarks as a clique found in the query image.

This approach not only requires searching for cliques in a graph, but also requires tracking of individual landmarks over multiple frames which makes it computationally intensive compared to BoW approaches that use histogram matching for computing image similarity. In other cases, the spatial arrangement of features is used for providing better discrimination while detecting loop closures as in [4,10,7,14]. This is usually done by using RANSAC or multi-view geometry (MVG) constraints to discard outliers. These methods are known to be computationally expensive and hence, used as a second stage of verification. Some other authors have attempted to incorporate spatial information into bag of words methods as in [14,15]. In [15], spatial neighborhood is created by dividing the image into regular grids. On the other hand, the authors in [14] use fixed radial distance to decide the spatial neighborhood. Both of these methods suffer from two limitations—first, the spatial neighborhood requires user-defined parameters and second, they are not invariant to scale variations.

The proposed method differs from the above methods in two ways. First, we use relative spatial co-occurrence information that combines spatial proximity with co-occurrence information to provide better loop closure detection. This spatial occurrence has an associated directional attribute which is unique to our approach. Secondly, spatial information is used at every level of decision making unlike previous methods [4,10,7,14] that use it only as a second stage of verification. Finally, the extent of spatial neighborhood is decided automatically by using scale size of point features thereby making the algorithm scale invariant unlike methods [14,15] which use an user-defined spatial neighborhood to group features. Through experiments on standard datasets, we show that the proposed method provides significant improvement in recall performance compared to most of the existing state-of-the-art methods such as, FAB-MAP [6,16], incremental BoW [4], direct feature matching based methods [11,10] and the methods that use binary features [7,17] etc. The merit of the approach is further corroborated by making the observation that the improvement in the recall performance is more significant when the dictionary size is small (or quantization error is more). This improvement, however, is accompanied by a slight increase in the computational and memory requirement as one needs to create an additional dictionary of word pairs. We also present a completely online version of the algorithm that incorporates the best of the existing methods. This includes creating dictionary incrementally in an online fashion, using K-D tree to search for matching features, carrying out tree update at regular intervals, using Bayesian filtering to reduce transient errors in making loop closure decisions and using RANSAC as a second stage of verification.

The main contributions made in this paper are as follows: (1) We demonstrate that the recall performance of BoW approaches could be improved significantly by incorporating spatial co-occurrence information directly into the dictionary. This is done by creating an additional dictionary comprising of word pairs, which are formed by using a directional spatial neighborhood defined based on scale parameter of each point feature. This concept of relative spatial co-occurrence is a new finding which has not been exploited earlier in the literature. (2) An online version of the algorithm implementation is presented which incorporates various popular concepts like K-D tree based nearest-neighbor search for identifying potential loop closure candidates [10], tf-idf as a similarity measure [4], incrementally building dictionary [4], belief propagation based on Bayesian network to suppress transient errors in decision making [6] and RANSAC-based geometric verification stage for confirming loop closures [10,4]. This implementation which combines all of these concepts together is itself a new contribution in this field. (3) An incremental version of K-D tree implementation, available with the latest FLANN library, is used for the first time in the context of topological mapping in this paper. The features can be added to the tree incrementally with each

Download English Version:

<https://daneshyari.com/en/article/6867468>

Download Persian Version:

<https://daneshyari.com/article/6867468>

[Daneshyari.com](https://daneshyari.com)