Robotics and Autonomous Systems 107 (2018) 246–261

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

An adaptive localization system for image storage and localization latency requirements

Jun Mao [a,b,*], Xiaoping Hu [a], Michael Milford [b]

[a] *National University of Defense Technology, China*
[b] *Queensland University of Technology, Australia*

HIGHLIGHTS

- A novel particle filter combining an odometry-based place recognition with a visual place recognition.
- A particle filter-enhanced visual place recognition adaptive to sparse image databases.
- A vision reliability estimation method that calibrates the contribution of vision to particle weight.

ARTICLE INFO

Article history:
Received 26 January 2018
Received in revised form 23 April 2018
Accepted 12 June 2018
Available online 3 July 2018

Keywords:
Visual place recognition
Map-based localization
Odometry
Road map
Autonomous vehicles

ABSTRACT

Fast and efficient global localization is a critical problem for autonomous systems. Existing sequence-based visual place recognition requires a storage-intensive image database for robust localization, while more storage-efficient odometry-based place recognition approaches can require a long travel distance to obtain an accurate localization. In this paper, we present a novel particle filter-based localization system that adapts to varying degrees of map image densities, road layout ambiguity and visual appearance change. The base system combines a geometric place recognition capability utilizing odometry and roadmaps with a visual place recognition system. When using a sparse image database, particles could exist at visually unknown places, which introduces difficulties in performing sequential visual place recognition. To address this challenge, we propose to make use of *effective visual observations* to enable the system to accumulate visual belief sequentially, even when reference images are very sparse. Furthermore, we develop a vision reliability estimation method, which analyses the relationship between the visual component and the particle filter convergence, to calibrate the optimal contribution of vision to particle weighting in different visual environments and conditions. To evaluate our approach, we perform extensive experiments using four benchmark localization datasets, and control the reference image density by subsampling these datasets. Results show that the proposed technique is able to consistently and correctly localize the vehicle over a range of reference image densities, and to consistently outperform a particle filter-enhanced version of an existing state-of-the-art SeqSLAM system, which fails when image spacing exceeds 30 m. In particular, for a 600% increase in database image sparsity (from 10 m to 70 m), we show that the proposed method is able to maintain localization performance with only a 40% increase in localization latency (from 250 m to 350 m). We also provide an analysis of the results and a characterization of the system's computational requirements.

1. Introduction

Self-localization is an important capability for most mobile autonomous systems and forms the basis for navigation, path planning and other core tasks. As numerous autonomous systems such as self-driving cars, home service robots and unmanned aerial vehicles are deployed [1–4], roboticists have an increasing need to tailor localization capabilities to specific requirements, such as compact storage, low *localization latency* (which is the travel distance before obtaining a localization), efficient computation and affordable map maintenance.

Visual place recognition is becoming an increasingly viable component in self-localization. Recent developments in sequence-based visual place recognition have demonstrated condition-invariant capability [5–8]. However, existing sequence-based visual place recognition techniques utilize dense image databases,

* Correspondence to: College of Mechatronic Engineering and Automation, National University of Defense Technology, China.
E-mail address: maojun12@nudt.edu.cn (J. Mao).

https://doi.org/10.1016/j.robot.2018.06.007
0921-8890/

which contain visual information covering all places in the environment. Such dense image databases can occupy significant storage space in large-scale applications and require significant computational resources to perform image comparisons. For compute, bandwidth (in the case of cloud-based systems) or storage-sensitive applications, one proposed solution is to make use of sparse image databases, which can decrease storage and computational costs by large factor. Requiring sparse image databases also reduces the cost of image database maintenance; image database can be collected and updated by using public road service cameras through services such as Mapillary [9]. However, when map image density reduces, the effectiveness of current sequence-based localization techniques decreases, with long localization latencies due to not only the sparse visual feedback, but also the inability of current systems to scale to sparse imagery.

In recent years, researchers have shown how odometry alone combined with pre-built roadmaps can robustly localize vehicles [10–13]. Compared to image databases, roadmaps are more robust, as most road layouts remain relatively constant for years. Moreover, roadmaps only require small storage space and are easy to maintain as they consist of vectorized lines. Newly built roads can be easily updated by government traffic department or online community-driven users, as exemplified by the OpenStreetMap (OSM) community. However, odometry place recognition techniques' effectiveness depends on road geometry; vehicles need to travel for a "sufficient distance" to correctly determine their position. In commonly encountered scenarios such as long straight roads or Manhattan-like city blocks, correct localization may only occur after a long travel distance; performance is highly dependent on the geometry of the environment.

Due to the variety of challenges in self-localization, using a single technique is typically problematic. Odometry place recognition is robust and affordable but can require a long travel distance to successfully localize. Sequence-based visual place recognition is able to determine locations even in environments with significant visual appearance variation by using many sequential images [8] but requires a dense image database. This research is driven by the desire for a localization system that can scale to varying reference map image densities while still making good use of visual imagery (rather than discounting it). Our proposed base system combines odometry place recognition and visual place recognition in a novel particle filter and then adds two enhancements. The first enhancement enables the system to adapt to sparse image databases. State-of-the-art particle filter-based visual place recognition techniques [6,14] compute the visual belief at a particle's hypothesis position by comparing the current camera image to the nearby database images. However, it is problematic to apply these nearby comparison methods when the image database is sparse as *particles can be located at places where no nearby visual scenes are saved*. One solution to deal with the visually unknown places is to learn a probabilistic model on a sample image dataset [15,16], but such probabilistic models are built by utilizing single images, which is not robust to visual appearance variation. We address the challenge of performing sequential visual place recognition with sparsely available reference scenes in the first enhancement by using *effective visual observations*, which are the best visual matches remembered by the particles when passing across the reference scenes. The second enhancement calibrates the visual measurement model for particle weighting in different visual environments and conditions. We address this by analysing how sequential effective visual observations can lead to correct particle filter convergence. The main contributions of our work are listed as follow:

1. a novel particle filter combining a geometric place recognition capability utilizing odometry and roadmaps with a visual place recognition system,
2. a particle filter-enhanced sequence-based visual place recognition, which is adaptive to varying degrees of map image densities, and
3. a vision reliability estimation method, which calibrates the contribution of vision to particle weighting in different visual environments and conditions.

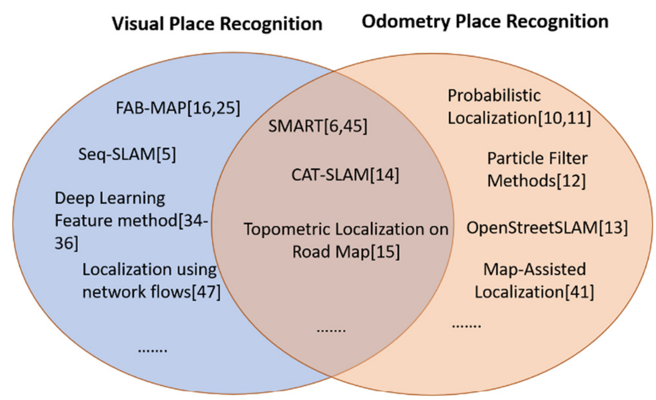

Fig. 1. Related techniques. Left: visual place recognition based on appearance only. Right: techniques that only use odometry and a reference roadmap. Middle: techniques that use both visual appearance, odometry and a roadmap.

We conducted experiments on four benchmark localization datasets with various types of road layouts and visual variations (different weather and seasonal change) over a range of map image densities. Results show that the proposed technique is able to consistently and correctly localize the vehicle with only a slight increase in localization latency under a significant increase in database image spacing, e.g. a 600% increase in image spacing (from 10 to 70 m) only leads to a 40% increase (from 250 to 350 m) in localization latency in the St Lucia dataset. We also show that consistently superior performance to an existing state-of-the-art particle filter-enhanced SeqSLAM system, especially as reference imagery becomes sparse (the reference method fails when database image spacing exceeds 30 m).

This paper is organized as follow. Section 2 briefly covers related research. Section 3 presents the details of the proposed method. The experimental setup and results are shown in Sections 4 and 5, with Section 6 studies the storage requirements and computational costs. Discussion and future work are presented in Section 7.

2. Related work

Self-Localization processes typically incorporate some form of map to represent the environment and then attempt to associate and match onboard measurements with the map to find the global position. Most existing self-localization techniques use visual appearance, odometry and prior maps, or some combination of these. Fig. 1 provides a graphical overview of some related techniques.

Vision-based or visual place recognition techniques require a geo-referenced image database described by visual features [17]. During the localization process, query image features are compared to reference images to obtain a location estimate. Visual processing in such systems can be divided into two broad categories depending on feature types; local features and global features. Local features, such as SIFT [18], SURF [19] and BRISK [20], describe image regions of interest and are widely used in visual place recognition [21–24]. One of the most well-known place recognition algorithms that use local features is FAB-MAP [16,25], which employs a bag-of-words (BOW) technique and has been successfully applied in large outdoor experiments. However, baseline local feature based visual place recognition techniques lack condition invariance [26,27];