Robotics and Autonomous Systems 75 (2016) 649–660

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Scene structure registration for localization and mapping

Eduardo Fernández-Moral [a,*], Patrick Rives [a], Vicente Arévalo [b], Javier González-Jiménez [b]

[a] INRIA Sophia Antipolis - Méditerranée, Lagadic team. 2004 route des Lucioles - BP 93, 06902 Sophia Antipolis, France

[b] Universidad de Malaga, Mapir Group. E.T.S. Ingeniería de Informática-Telecomunicación. Campus de Teatinos, 29071, Málaga, Spain

HIGHLIGHTS

- A geometric representation based on planar patches is proposed for scene registration.
- Photometric information is combined with the geometric description for robustness.
- A novel formulation encoding uncertainty information is proposed to improve accuracy.
- An implementation of this work is freely available (MRPT and PCL).

ARTICLE INFO

Article history:
Available online 30 October 2015

Keywords:
Scene registration
Scene recognition
Localization
Mapping
Planar segmentation

ABSTRACT

Image registration, and more generally scene registration, needs to be solved in mobile robotics for a number of tasks including localization, mapping, object recognition, visual odometry and loop-closure. This paper presents a flexible strategy to register scenes based on its planar structure, which can be used with different sensors that acquire 3D data like LIDAR, time-of-flight cameras, RGB-D sensors and stereo vision. The proposed strategy is based on the segmentation of the planar surfaces from the scene, and its representation using a graph which stores the geometric relationships between neighbouring planar patches. Uncertainty information from the planar patches is exploited in a hierarchical fashion to improve both the robustness and the efficiency of registration. Quick registration is achieved in indoor structured scenarios, offering advantages like a compact representation, and flexibility to adapt to different environments and sensors. Our method is validated with different sensors: a hand-held RGB-D camera and an omnidirectional RGB-D sensor; and for different applications: from visual-range odometry to loop closure and SLAM.

1. Introduction

Image registration has been a major problem in computer vision over the past decades. It implies searching an image in a database of previously acquired images to find one (or several) that fulfil some degree of similarity, e.g. an image of the same scene from a similar viewpoint. This problem is interesting in mobile robotics for topological mapping, re-localization, loop closure and object identification. Scene registration can be seen as a generalization of the above problem where the representation to match is not necessarily defined by a single image (i.e. the information may come from different images and/or sensors), attempting to exploit all information available to pursue higher performance and flexibility.

This paper addresses the problem of scene registration from 3D data using a compact description which encodes geometric information (and photometric information if it is available) about the scene's planar surfaces. This article extends a previous work which relies on the segmentation of planar patches to build a Plane-based Map (named PbMap, see Fig. 1) [1]. Such solution is extended here using a probabilistic framework to take into account the uncertainty model of the sensor. This approach is specially interesting for depth devices like range cameras or stereo vision, which deliver data in an organized way (i.e. neighbouring pixels correspond to close-by 3D points) so that it allows us to segment efficiently planar patches, referred as planes for short.

The key idea in this article is that even a small set of neighbouring planar patches (e.g. 4–10) encode enough information to recognize and register a scene. This strategy contrasts with previous methods that make use of local or global descriptors [2]. A relevant difference is that our method exploits a connected description of the scene and thus, it is less dependent on the particular field of view of the sensors, offering a piecewise continuous description that supports multi-sensor and multi-frame observations. We present a set of experiments using RGB-D cameras (Asus Xtion Pro Live -XPL-), both with a single sensor waved by hand,

* Corresponding author.
E-mail address: eduardo.fernandez-moral@inria.fr (E. Fernández-Moral).

http://dx.doi.org/10.1016/j.robot.2015.09.009
0921-8890/

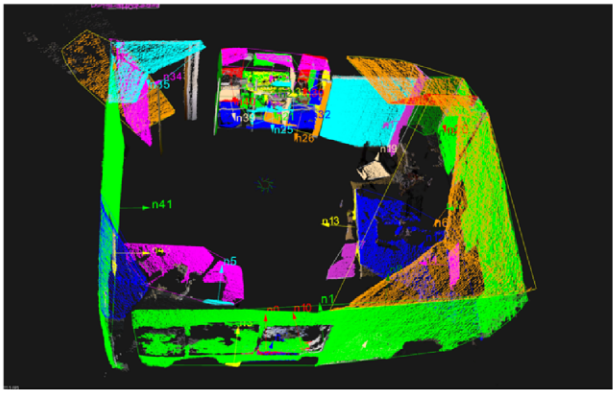

Fig. 1. Coloured planar patches of the PbMap representation of an office environment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and a rig of sensors which provide omnidirectional RGB-D images mounted on a wheeled robot [3]. These experiments demonstrate the potential of our technique for localization, loop closure, odometry and SLAM. Our registration technique has been implemented in C++ and is freely available as a module of PCL.

1.1. Related works

Most solutions for scene registration make use of intensity information from regular cameras. Such sensors have the advantage to provide detailed information at a low cost. Scene registration from images generally requires invariance to changes of illumination and viewpoint, and robustness to visual aliasing and lack of texture among others [4]. The methods proposed in the literature are traditionally classified into feature-based or dense, depending on whether they are based on local features or on global appearance. Local features like point descriptors (e.g. SIFT, SURF, ORB, etc.) are commonly applied to camera tracking (visual odometry). On the other hand, dense methods minimize the photometric difference between two images with respect to a warping function which is estimated iteratively, not requiring any previous step to extract and match salient points. However, dense methods require an initial estimate of the image warping, plus they need to make assumptions about the scene structure when 3D information is not available (monocular).

Scene recognition is a problem related to scene registration, where the former does not require to find the alignment between the scenes. Thus, scene recognition is related to topological localization while scene registration is related to metric localization. Image indexation is a particular case of scene recognition for which the data to retrieve are images [2]. In robotics, the most common approach to solve this problem is the bag-of-words (BoW) method [5,6], which creates a dictionary of features and employs a voting method to retrieve the location, is traditionally used for re-localization, loop closure or topological localization and mapping. The authors of [7] presented an alternative approach by creating a compact global descriptor which is related to the shape of the scene and the geometric relationships in it. This last is generally faster, but less suited if the relative pose between the images have to be recovered.

The use of 3D data to solve the scene registration problem is increasingly popular in mobile robotics due to several affordable range sensors that have appeared in the last years, like these of PrimeSense. Before that, registration of 3D shapes has been researched mainly for object recognition [8], where a limited amount of data, usually a point cloud of a well delimited object, is registered using Iterative-Closest-Point (ICP). This kind of solution cannot be directly applied for re-localization since first, it requires an initial estimation of the relative pose between the point clouds; and second, the point cloud description of the scene cannot be easily delimited to a set of corresponding points within a large reference map, being difficult to find a reference and a target point cloud with almost same content. Another approach which makes use of depth information (and also photometry) is direct registration [9,10], but it has the same limitations as ICP since it requires an initial estimation for the registration, and such estimation is not available in re-localization kind of problems.

The technique we propose here is an extension of our previous work in [1], which can be seen as a combined geometric–photometric global descriptor where the scene is described by a set of neighbouring planar patches which capture information about the scene's global set-up. A recent work which is based on matching a set of planar patches and line segments is described in [11], however it is restricted to match images, in a similar fashion to the SLAM solution of [12], and does not exploit the fact that planar patches can be used to build a piecewise continuous representation using a graph. In contrast to previous registration approaches, our method is not limited to image registration, and therefore it can be applied to different sensors and can use information of several images or video sequences. This advantage is crucial in situations where a single image does not provide enough information to register the scene, but whose information is useful to complete previous observations (a situation which is quite common in mobile robotics). In this sense, our work is more related to visual place recognition approaches which exploit the information of a sequence of images to boost the recognition accuracy [13,14]. The performance of such solutions is highly related to the size of the sequence describing the scene to match. This concept is similar to the size of the set of neighbour planes describing the scene. However, this last has the advantage that it is directly related to the size of the scene, and it abstracts from other aspects like sensor field of view, or proximity of the frames in the sequence. Thus, our solution frees the user from tuning sensor and robot parameters, requiring only to specify the size of the scene to be matched through the number of planes.

1.2. Contribution

The main contributions of this paper are:

- We extend our previous Plane-based Map description and registration technique to incorporate uncertainty information in a hierarchical structure to improve robustness and efficiency.
- We generalize this solution so that it can take different sources of data, from different kinds of range cameras to stereo vision. A validation is presented with different sensors and for different applications.
- An implementation of this work including a tutorial with some practical examples is made available in MRPT [15] and PCL [16].

2. PbMap: a representation of the scene's planar structure

A plane-based map (PbMap) is a representation of the scene as a set of 3D planar patches, which are described by a series of geometric and radiometric attributes [17]. It can be built from different sources of data (several range images from different sensors) if we know the relative poses of the different observations. In this case, the overlapping patches are merged together so that each planar surface in the scene is represented by a single patch.

2.1. Planar patch segmentation and parametrization

In order to obtain the planar patches the depth images are segmented with a region growing approach which is implemented