Robotics and Autonomous Systems 80 (2016) 55–68

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

Evaluation of color and texture descriptors for matching of planar surfaces in global localization scheme

Damir Filko *, Robert Cupec, Emmanuel Karlo Nyarko

Kneza Trpimira 2B, 31000 Osijek, Croatia

HIGHLIGHTS

- Indoor global localization system based on planar segments with visual descriptors.
- Applicability of 6 color and 3 texture descriptors is systematically analyzed.
- Performance increase in initial correspondence and pose hypothesis evaluation phases.
- Evaluation benchmark for visual descriptors in global localization is proposed.

ARTICLE INFO

Article history:
Received 11 June 2015
Received in revised form
10 December 2015
Accepted 18 March 2016
Available online 5 April 2016

Keywords:
3D planar features
Color
Texture
Global localization
Place recognition

ABSTRACT

This paper presents a systematic study about the applicability of color/texture descriptors in a global localization system based on planar surface segments. Two comprehensive experiments regarding matching of planar surface segments and robot pose hypothesis evaluation were conducted. The experiments show that using color/texture descriptors to prune potential surface pairs in the initial correspondence phase and to provide additional information in the hypothesis evaluation phase of a feature-based localization scheme can result in considerable speedup of the localization process and help distinguish between geometrically similar places. An experimental benchmark which enables researchers to evaluate the performance of color and texture descriptors in the context of mobile robot localization based on planar surface segments is presented.

1. Introduction

The advent of inexpensive commercially available RGB-D sensors in recent years, such as Microsoft Kinect, has caused a minor revolution in the field of robot vision and mobile robotics. It has enabled a lot more scientists to work on higher level robot vision problems instead of on lower ones such as disparity map creation from stereo vision. Until recently, researchers primarily used stereo camera systems which provided mobile robots with some localization and navigational properties. However, maps built by such systems are relatively sparse with environment information and very noisy. The recently introduced RGB-D sensors provide dense depth images at very high frame rates thus easing and speeding up the creation of very detailed 3D maps and reconstruction of the environment and objects. This advance in technology has enabled the development of extremely sophisticated SLAM systems which can be considered the pinnacle of research in mobile robotics [1–3]. Today's SLAM systems allow for the building of highly accurate metric models of large scale environments. However, for the practical application of such models in solving robotic tasks, the ability of a robot to localize itself in such a model is crucial. The problem of determining the position of an intelligent agent in a map from sensor data, where the only a priori knowledge about this position is that the robot is somewhere in the map, is referred to in the following as global localization. A problem related to global localization is the place recognition problem. The place recognition problem is usually considered in the cases where the map is represented by a discrete set of local models or by a graph whose nodes represent particular locations in the robot's environment. The task of a place recognition algorithm is to determine the local model or node which corresponds to the currently acquired image. The global localization problem can be regarded as a place recognition problem which, in addition to determining the approximate location in the map, includes estimating the robot pose relative to a local or global coordinate system. Place recognition approaches have an important application in SLAM systems, where they are

* Corresponding author.
E-mail addresses: damir.filko@etfos.hr (D. Filko), robert.cupec@etfos.hr (R. Cupec), nyarko@etfos.hr (E.K. Nyarko).

http://dx.doi.org/10.1016/j.robot.2016.03.007
0921-8890/

Fig. 1. Geometrically similar scenes which can be distinguished by considering the color and texture of the visible surfaces.

used for loop closing [2,4]. Nevertheless, while SLAM systems have been developed at astounding speed, place recognition methods have not advanced as much.

A common approach to the global localization problem is to generate a number of hypotheses about the robot's pose on the map from a set of feature matches, estimate the probability of each hypothesis and select the most probable one. If the hypothesis selection is ambiguous, several hypotheses must be considered for further evaluation using additional sensor information acquired during the robot's motion.

Current state-of-the-art place recognition approaches use a bag-of-words approach with point features and descriptors such as SIFT or SURF. These systems, such as FAB-MAP, have experimentally proven great successes in localization in outdoor areas where there are a lot of point features [1,5]. But it has also been experimentally proven that they do not have such great performance in typical indoor environments, especially in corridors lacking distinct point features and also due to the dynamic nature of human living and work environments [4]. Other place recognition systems exploit prior probability by using sequence information like those presented in [6,7]. Some indoor localization systems have opted for the use of structural elements in the environment, such as planar segments or linear features, which provide more geometrical and contextual information than point features. Planar segments are more advantageous than lines because their 3D nature and larger dimensions make them less prone to occlusion and more robust to measurement noise. Our previous system [8,9], which is the basis for this article, used planar surface segments to localize a mobile robot in 6DOF. Unfortunately, because the system used only geometrical information, it had problems discriminating places that were geometrically similar but visually different such as the ones shown in Fig. 1.

The main contribution of this paper is a systematic analysis of the application of color and texture descriptors in a global localization system based on planar features extracted from an RGB-D image. The paper investigates possible improvements of the system performance with respect to precision and computational efficiency. Furthermore, an experimental benchmark for the evaluation of color and texture surface descriptors in the context of global localization is provided.

The rest of the paper is structured as follows. In Section 2, a short survey of the related research is given. Section 3 provides an overview of the color and texture descriptors used. The mobile robot localization system based on planar surface segments is described in Section 4. Section 5 explains the experimental analysis methodology along with results. In Section 6, our evaluation benchmark is proposed. Finally, the paper is concluded with Section 7, where the obtained experimental results are discussed.

2. Related research

The problem of place recognition is similar to the image retrieval problem, where the current scene is compared to previously recorded scenes with the purpose of finding the most similar ones. In the bag-of-words (BoW) approach, place descriptors are vectors containing binary values or weighted histograms accumulating occurrences of keypoint descriptors quantized to words in a learned vocabulary. The main representative of such an approach is FAB-MAP [1,5]. In the experiments reported in [1], FAB-MAP has shown its usefulness in place recognition problem over very long outdoor trajectories. With recent advances, such systems can even operate in real-time and therefore be utilized in SLAM systems for the purpose of loop closing [2]. An alternative to using point descriptors which describe only a local area around certain keypoints, is using features such as lines and planar surface segments which give more structural information about the environment and which are especially pronounced in indoor environments. An example of a system which uses lines for place recognition is described in [4] where the authors use the mean standard-deviation line descriptor to statistically describe the area around detected lines. The authors also use the BoW approach with a vocabulary tree that is trained for line-to-line matching and apply a Bayesian filtering framework for loop closure detection. The authors report considerably better results than SIFT/SURF approaches in indoor environments especially with changes in illumination. However, since the approach proposed in [4] uses only visual lines as features, it shows reduced performance in situations where such features are either not pronounced or few.

In cases where visual features are sparse, such as bare hallways, a system that uses planar surface patches would be more appropriate. Robot localization approaches based on extracting planar surfaces from point clouds obtained by rotating laser scanner are proposed in [10–12]. Laser has many advantages over a standard RGB-D camera such as the one used in the proposed approach. Laser is more precise, it has a much longer range, it can be used outdoors in day light and it has a much wider field of view. Nevertheless, an RGB-D camera has two advantages: (i) a much lower cost and (ii) it provides color and texture information. The former is a good enough reason to investigate approaches which use RGB-D camera. Furthermore, application of color and texture information can be advantageous in situations where geometrically similar places can be distinguished by their color/texture features. A localization system which uses an RGB-D sensor is described in [13] where the authors' world representation is organized as a graph where the nodes represent planar patches and the edges connect planes that are close to one another. The authors use rigid registration test to determine candidate match between two subgraphs resulting in an estimate of the relative pose between two consecutive RGB-D observations. Because the discussed approach uses only geometrical information to match two subgraphs, significant errors can be obtained in cases of