Expert Systems With Applications 102 (2018) 273–290

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Trajectory estimation and optimization through loop closure detection, using omnidirectional imaging and global-appearance descriptors

F. Amorós, L. Payá*, J.M. Marín, O. Reinoso

Miguel Hernandez University, Department of Systems Engineering and Automation, Avda. de la Universidad, s/n. Ed. Innova. 03202 Elche (Alicante), Spain.

ARTICLE INFO

Article history:
Received 17 November 2017
Revised 6 February 2018
Accepted 28 February 2018
Available online 1 March 2018

Keywords:
Catadioptric vision sensors
Global-appearance descriptors
Fourier Signature
Histogram of Oriented Gradients
Visual odometry
Loop closure
Route estimation

ABSTRACT

Currently, the range of applications of mobile robots has extended substantially thanks to the evolution of the sensing and computing technologies. In this field, creating accurate and compact models of the environment is crucial so that the robot can estimate its position and move autonomously to the target points. Among the available alternatives, computer vision sensors have become of utmost importance to create these models, thanks to the richness of the data they can capture. However, they require the implementation of algorithms to extract relevant information from the scenes. In this work, a framework to create a model of *a priori* unknown environments is presented, which is based on the global appearance of images. The model is created on-line, as the robot explores the environment and the result is a graph whose nodes contain images and the links represent relative distance between them. The framework includes a schema that fuses the information extracted from the scenes with the angle information provided by the odometry of the robot, considering the relative reliability of each piece of information. Also, a loop closure detection algorithm is proposed, which corrects the position of the nodes and updates the map. A set of experiments has been conducted to study the influence of the most relevant parameters upon the accuracy of the model and the computational cost of the process.

1. Introduction

The navigation of autonomous mobile robots is a line of research which is in continuous development and new frameworks are continuously presented. Many possibilities are available to perform this task, depending on the kind of environment and the type of sensor to collect information and techniques to analyze this information. Among all these possibilities, visual sensors can be highlighted (Garcia-Fidalgo & Ortiz, 2015) owing to the richness of the information they provide the robot with, their relatively low cost and low power consumption, comparing to some other kinds of sensors such as laser rangefinders. There are multiple visual systems, depending on their architecture, number of sensors and field of view. Within those systems that gather a wide field of view, the most extended ones are fisheye lens cameras and catadioptric systems.

On the one hand, fisheye lenses are able to capture a field of view which is greater than 180°. Currently, some devices can be found that combine several images captured by fisheye lenses to compose an omnidirectional scene (Samsung, 2016). Also, Li, Nakano, and Chiba (2004) present a vision system composed of two fisheye lenses whose field of view is the complete sphere around the sensor. Nevertheless, as Li (2006) points out, different lighting conditions in each camera may cause some problems when creating the complete spheric image. On the other hand, catadioptric systems are formed by a camera pointing towards a reflective surface of revolution (Baker & Nayar, 1998; Gaspar, Winters, & Santos-Victor, 2000). These systems are generally easier to calibrate and cheaper than fisheye cameras, and they can be designed to present the *single effective viewpoint* property, which permits obtaining undistorted perspective images from the original omnidirectional scene (Payá, Gil, & Reinoso, 2017). In this work, a catadioptric vision system composed of a CCD color camera and a hyperbolic mirror with their axes aligned is used. It is mounted vertically on a mobile robot and captures omnidirectional images, which contain a complete field of view of 360° around the revolution axis of the mirror. This system is fully calibrated and presents the *single effective viewpoint* property.

The richness of these images permits building detailed models of the environment. However, images are highly dimensional data and storing and processing them directly would involve massive memory and computing requirements. Also, the contents of the scenes change not only when the robot moves but also under

* Corresponding author.
E-mail addresses: francisco.amoros03@graduado.umh.es (F. Amorós), lpaya@umh.es, lpaya@goumh.es (L. Payá), jmarin@umh.es (J.M. Marín), o.reinoso@umh.es (O. Reinoso).

https://doi.org/10.1016/j.eswa.2018.02.042
0957-4174/

some other circumstances, such as changes in lighting conditions, partial occlusions, changes in the position of some objects, etc. For that reasons, it is necessary to use any mechanism to extract some relevant information from the scenes, that permits modeling robustly an environment and recognizing this environment subsequently, using a limited amount of data.

These mechanisms, also named *descriptors*, can be categorized into two main groups: based on local features and based on global appearance. First, local-features descriptors entail the extraction of outstanding points, landmarks or regions from the images, either natural (Alahi, Ortiz, & Vandergheynst, 2012; Bay, Ess, Tuytelaars, & Van Gool, 2008; Lowe, 2004) or artificial (Okuyama, Kawasaki, & Kroumov, 2011). Mapping and localization are based on the extraction, description and tracking of these landmarks along a set of scenes (Valgren & Lilienthal, 2010). These tasks may result hard in real unstructured environments, where it is difficult to find stable and robust landmarks. In any case, these techniques have been used in numerous research works, and they have reached a relative maturity (Kostavelis, Charalampous, Gasteratos, & Tsotsos, 2016).

Some authors have proposed the use of local-features descriptors in visual odometry frameworks, to improve vehicle positioning in the context of Advanced Driver Assistance Systems (ADAS) and autonomous vehicles. Among them, de la Escalera et al. (2016) propose a visual odometry method which detects and tracks SIFT (Scale-Invariant Feature Transform) features on the surface of the ground, extracted from a virtual bird-view image. Alonso et al. (2012) use a stereo platform to estimate the motion trajectory of the intelligent vehicle along with a map-matching algorithm to correct the cumulative errors of the visual odometry. Also, Lee, Faundorfer, and Pollefeys (2013) present a visual egomotion estimation algorithm for a self-driving car equipped with a multi-camera system, which is treated as a generalized camera, and calculate a generalized essential matrix from visual features. Mhiri, Vasseur, Mousset, Boutteau, and Bensrhair (2014) use a set of calibrated cameras to carry out a visual odometry with metric scale estimation. The relative camera poses are estimated with a structure-from-motion approach and the extrinsic parameters are used to estimate the scale factor. Some other authors have focused on some specific issues of the visual odometry. For example Wu, Lam, and Srikanthan (2017) present a framework for ego-vehicle's motion estimation. They concentrate on the computational time of the algorithm and propose some improvements in the Kanade–Lucas–Tomasi feature detection and tracking, and in the termination condition of the RANSAC (Random Sample Consensus) algorithm. Golban, Istvan, and Nedevschi (2012) propose several improvements to cope with some challenging issues that may come out in real traffic situations, such as illumination effects, reflection and shadows. Gräter, Schwarze, and Lauer (2015) study the problem of monocular visual odometry and focus on estimating the scale for usage in ADAS. To this aim, they make use of structure from motion techniques complemented with a vanishing point estimation. All these works make use of local descriptors (landmarks) to extract the most relevant information from the images.

Over the last few years, global-appearance descriptors have emerged as an alternative to classical local descriptors. They use the information of the images as a whole, without any local segmentation of the scene. The descriptors are obtained after applying a mathematical transformation to the entire scene. Some works have proposed global appearance techniques to describe scenes, such as (Menegatti, Maeda, & Ishiguro, 2004; Oliva & Torralba, 2006). The relative simplicity of these descriptors is an advantage as they lead to models of the environment that present a good balance between compactness and computational cost of the localization process. Some researchers have made use of such techniques in the field of mapping and localization with mobile robots (Berenguer, Payá, Ballesta, & Reinoso, 2015; Rituerto, Murillo, & Guerrero, 2014; Štimec, Jogan, & Leonardis, 2008). These works show that global-appearance techniques are able to build topological maps of an environment (Bacca, Salvi, & Cufí, 2011; Payá, Reinoso, Berenguer, & Úbeda, 2016) and estimate the position of the robot within this environment (Chang, Siagian, & Itti, 2010; Murillo, Singh, Kosecka, & Guerrero, 2013).

One of the main drawbacks of the global-appearance techniques in comparison with local-features descriptors is that they do not provide any metric information about the objects that appear in the images or between scenes, so the creation of the map is usually approached from a topological point of view (unless other additional kinds of sensors are used) and the localization problem tends to be reduced to finding the nearest image in the model (Guan, Fan, Duan, & Yu, 2014; Payá, Amorós, Fernández, & Reinoso, 2014). However, more recent works have demonstrated that it is possible to estimate the relative position between two scenes using their global-appearance. In (Amorós, Payá, Reinoso, Mayol-Cuevas, & Calway, 2013) this technique is presented under the name of multi-scale analysis, where it is applied to navigation tasks in order to obtain a measurement of the displacement of the robot. This measurement is not metrical, but goes beyond the traditional connectivity concept of topology because it contains information on how long or short this distance is (i.e. it is proportional to the geometrical distance). The main feature of this technique is that, despite not having a metric map, it permits estimating relative displacements between two points using only visual information, by means of some artificial zooms of the central area of projective images.

In the present work, an algorithm is proposed to extend the multi-scale analysis to omnidirectional images. The use of omnidirectional vision permits estimating the relative position between two points independently on the relative orientation of the robot, what supposes a clear advantage. Using this algorithm, a new framework is developed to obtain a topological visual odometry system. The method that we propose makes use, first, of the global appearance of omnidirectional images to estimate the relative orientation of the robot and to detect if it goes through previously visited places (loop closing), and second, of the multi-scale analysis to determine the relative position between two images in the floor plane using a topological measurement. The result is a graph model whose nodes represent images and the links show the relative distance between nodes. This model goes beyond the classical concept of topological map because the links contain information not only about connectivity but also about relative distance. Moreover, every time that a loop closure is detected, the position of the nodes contained in the loop is corrected and the model is updated. The paper also proposes a method that fuses the orientation information provided by the global appearance descriptors of the images and the angle estimated by the odometry of the robot, through a weighting schema that considers the relative reliability of each piece of information. This way, the work focuses on the problem of knowledge discovery and management in robotics, trying to extract movement information from the scenes. This problem has received much attention in the area of expert systems (Charalampous, Kostavelis, & Gasteratos, 2016; Kostavelis & Gasteratos, 2017; Martín et al., 2014). In our work, the scenes are represented by means of a holistic descriptor, with no local features and a refinement step is included, through loop closure.

In order to test the algorithm, a set of experiments is carried out to evaluate the performance of the algorithm and to study the influence of the most relevant parameters in the accuracy and computational cost of the estimation. These experiments are performed with a set of images captured by ourselves in a specially challenging environment and under real working conditions. The data acquisition was made with a wheeled mobile robot, whose