Computers in Industry 74 (2015) 119–134

Contents lists available at ScienceDirect

Computers in Industry

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

Fusion of visual odometry and inertial navigation system on a smartphone

Simon Tomažič *, Igor Škrjanc

Faculty of Electrical Engineering, Laboratory of Autonomous Mobile Systems, University of Ljubljana, Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form 1 May 2015
Accepted 11 May 2015
Available online 27 May 2015

Keywords:
Visual odometry
Pedometer
Compass
Fusion
Kalman filter
Smartphone
Robot
Pedestrian

ABSTRACT

The paper presents the monocular visual odometry, inertial navigation system and the fusion of both these localization approaches. The visual odometry algorithm consists of four other algorithms, namely the camera calibration algorithm, KLT algorithm, algorithm for the estimation of rigid transformation and RANSAC algorithm. The inertial navigation system is based on a pedometer and digital compass. Both visual odometry and the inertial navigation system can determine the incremental movements and the positions of a robot or a pedestrian according to the world coordinate system. In order to get an even more robust and accurate localization system, the advantages of each mentioned localization approaches were combined by using the Extended Kalman Filter. The algorithms were fully implemented on a smartphone, where they were divided into several threads that could be performed simultaneously on multiple processor cores. The proposed system, which fuses information from the camera and inertial sensors, can convert the smartphone into a powerful mobile sensor unit or the so-called virtual sensor that returns relative position in relation to the starting point. This virtual sensor can be used as an advanced sensor unit on mobile robots or as part of a smartphone application which requires personal navigation system. The operation of the localization system is proved by experimental results which were obtained by attaching a smartphone on a pedestrian who walked along the reference trajectory drawn on the floor. In the experiments the described system showed big potential in many aspects since very good results were obtained.

1. Introduction

One of the most important features of all living creatures on Earth is the ability to determine their own position in the environment they live in. With the development of new localization algorithms and systems, there has been a growing tendency to enable robots, autonomous mobile systems, and particularly people who have lost this ability, e.g. due to blindness, to use this ability.

Recently, a lot of studies have focused on localization in the indoor environment, since this represents a major challenge, mainly due to the fact that GPS signals are not available there. As smartphones have become indispensable accessories of a modern man, the possibility of using their hardware for the purpose of indoor localization of persons and autonomous mobile systems has been studied. Smartphones are equipped with multiple sensors such as the accelerometer, gyroscope, magnetometer, altimeter, camera and multiple communication modules (Bluetooth, WiFi, NFC, LTE), which enable the implementation of different algorithms for indoor localization [1]. In addition, smartphones contain increasingly powerful multi-core processors, which allow for the implementation of more complex algorithms.

The development of new algorithms and systems for localization in the indoor environment is especially important due to their usefulness in many fields. Namely, an algorithm that provides accurate localization in the indoor environment can be used for different types of robots, for applications for the blind, for applications which comprise personal navigation system (PNS) for guiding [2] in large shopping malls, museums, airports, public institutions, etc. since people often spend a lot of time finding the desired location in an unfamiliar environment.

Localization of unmanned ground vehicles and robots in the indoor environment is already well developed, since they can be equipped with more powerful hardware and additional sensors, e.g. LIDAR, depth sensor, stereo camera [3]. In connection with the LIDAR sensor, especially simultaneous localization and mapping (SLAM) method [4] has been established, which can also process the information obtained by the camera. Since the camera is a low cost and relatively lightweight sensor, its use is becoming more

* Corresponding author.
E-mail addresses: simon.tomazic@fe.uni-lj.si (S. Tomažič), igor.skrjanc@fe.uni-lj.si (I. Škrjanc).

http://dx.doi.org/10.1016/j.compind.2015.05.003
0166-3615/

common for the purpose of localization of unmanned ground and aerial vehicles.

Currently the most widely used approach to indoor localization using a smartphone is based on the measurement of the WiFi, Bluetooth or GSM signal strength [1]. However, because this approach does not enable high positioning accuracy, new approaches based on the fusion of several different sensors and methods are being established. As has been shown in numerous studies, a great potential lies especially in the methods based on inertial sensors and camera, as they provide higher accuracy of localization. High accuracy and robustness of localization can be achieved through the fusion of different localization approaches which complement each other. Sirtkaya et al. [5] proposed the fusion of the information from the camera and inertial sensors by using the Kalman filter.

When camera is used for the localization purposes, there are several different methods and algorithms which can determine the movement of an agent (a vehicle, person, robot) on which the camera is attached. Among the established methods, here belong structure from motion (SFM), simultaneous localization and mapping (SLAM) [4], visual odometry (VO) [6] and image-to-map matching [7]. SFM and SLAM are considered to be compute-intensive methods and they also spend a lot of memory, because they build a 3D map of the environment besides the motion estimation. The approach with the image-to-map matching requires extensive image dataset of the indoor environment which is used for off-line reconstruction (i.e. building a map) presented with a 3D point cloud. This type of localization also requires a powerful hardware with a lot of storage [8]. Since visual odometry estimates only the motion of the camera, it can operate in real time, even on less powerful hardware. The concept of visual odometry was established by Nister et al. [9], who introduced the main concept, which is the basis for the most of the existing visual odometry algorithms. The term was chosen for its similarity to wheel odometry since both approaches incrementally estimate the motion of a vehicle.

In order to obtain an accurate localization technique which would run in real time on a smartphone, a computational optimal system based on the use of camera and inertial sensors was built in this study. The core of the system is visual odometry, which enables accurate determination of incremental movements of the smartphone, i.e. a human or a robot on which the smartphone is attached. Visual odometry is based on the assumption that a smartphone is fixed on a certain height and at an angle relative to the floor [10]. This assumption is often true when using a smartphone on a robot. For determining the movements of (the blind) persons, a smartphone needs to be fixed to the body. For the operation of visual odometry, it is necessary to know the transformation between the camera coordinate system (C.S.) and the ground C.S., which is obtained by the initial calibration. In connection with the visual odometry, RANSAC algorithm [9,11] is also often used, which enables the elimination of the traces of feature points that represent outliers in the determination of the rigid motion model.

For the purpose of testing the algorithms, a Galaxy S4 smartphone based on the operating system Android was used. In the implementation of the visual odometry, an open-source library BoofCV [12] that is written in the Java programming language and combines a multitude of useful functions in the field of machine vision was used. The visual odometry algorithm is performed in real time on the smartphone where the speed of image processing is equal to 10–15 fps. In the application the images were captured at a resolution of 320×240 pixels. The calibration was also performed on the smartphone by using the BoofCV library [12].

As a complementary system of visual odometry, the inertial navigation system based on three sensors, namely 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer, was implemented. The inertial navigation system for pedestrians is basically composed of a digital compass, which is responsible for determining the absolute heading, and a pedometer, which determines the length of the travelled path by counting the steps. Inertial navigation system can also be used on a wheel robot, whereby double integration of acceleration for determining the travelled distance is used instead of a pedometer. This approach is somewhat more susceptible to the errors accumulation but this can be reduced by considering another localization technique such as visual or wheel odometry.

Both visual odometry and inertial navigation system have advantages as well as disadvantages. Both approaches have one common weakness, namely, they belong to dead reckoning localization techniques. The phrase dead reckoning means that a previously determined position is used in the process of calculating the current position. Consequently, the newly computed positions contain cumulative errors. With the fusion of both localization approaches, the size of the accumulated errors can be reduced and also all other errors that may occur within a particular system should be eliminated. In order to get a system that would be as reliable as possible for determining the relative position, the Extended Kalman Filter (EKF), which enables optimal combining of information, was used for the fusion purposes.

The advantage of this system is that it is fully implemented on the smartphone and it does not require additional indoor infrastructure (e.g. WiFi network) for its operation. The proposed localization system, which combines information from hardware sensors (camera and inertial sensors) converts the smartphone into a powerful mobile sensor unit or the so-called virtual sensor that returns incremental movements in relation to the starting point. Therefore, the smartphone just needs to be attached to the robot and the virtual sensor already determines incremental movements of the robot. The result of the virtual sensor can be sent wirelessly, e.g. via ROS messages [13], to the central unit of the robot, where it is used in the control algorithm. The smartphone could work also in a telerobot (remote) control mode [14], wherein the whole control algorithm [15] would run on it and only commands according to the desired task would be sent to the autonomous ground vehicle. The above mentioned virtual sensor can easily be applied also in the application for guiding humans [16,17] in an indoor space.

The operation of the localization system is proved by experimental results which were obtained by attaching a smartphone on a pedestrian who walked along the reference trajectory drawn on the floor. When using a smartphone as an external sensor unit on a robot, equally good or even better results are expected than when using it on a pedestrian due to much less vibrations.

In the following sections the components of the monocular visual odometry are first presented, then visual odometry itself. Afterwards, the inertial navigation system which includes a digital compass and a pedometer is described as well as the fusion of visual odometry and inertial navigation system. Finally, the experimental results of the implemented system functionality are given.

2. Monocular visual odometry

Monocular visual odometry is the sequential estimation process of camera motions depending on the perceived movements of pixels in the image sequence. The visual odometry consists of four algorithms, namely: the camera calibration, the