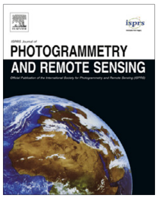
ISPRS Journal of Photogrammetry and Remote Sensing xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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



Landmark based localization in urban environment

Xiaozhi Qu *, Bahman Soheilian, Nicolas Paparoditis

Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mande, France

ARTICLE INFO

Article history:
Received 18 March 2017
Received in revised form 19 September 2017
Accepted 19 September 2017
Available online xxxx

Keywords:
Visual odometry
Landmark
Multi-camera
Local bundle adjustment
Uncertainty

ABSTRACT

A landmark based localization with uncertainty analysis based on cameras and geo-referenced landmarks is presented in this paper. The system is developed to adapt different camera configurations for six degree-of-freedom pose estimation. Local bundle adjustment is applied for optimization and the geo-referenced landmarks are integrated to reduce the drift. In particular, the uncertainty analysis is taken into account. On the one hand, we estimate the uncertainties of poses to predict the precision of localization. On the other hand, uncertainty propagation is considered for matching, tracking and landmark registering. The proposed method is evaluated on both KITTI benchmark and the data acquired by a mobile mapping system. In our experiments, decimeter level accuracy can be reached.

1. Introduction

Precise localization is desired for many applications such as MMS (Mobile Mapping System), self-driving cars, ADAS (Advanced Driving Assistance Systems) and AR (Augmented Reality). The solutions for localization can be summarized as global localization and position tracking (Thrun et al., 2005). Global localization measures absolute positions, while position tracking starts from a point and tracks the relative poses over time. The most famous global localization system is GNSS (Global Navigation Satellite System). However, the multi-path or mask of satellite signals could lead to large errors and even outage of localization in the urban environments. Moreover, GNSS provides positioning once a second and the orientations cannot be measured directly. Therefore, some direct positioning and orientation systems are developed by combining GNSS with INS (Inertial Navigation System). The gaps between GNSS points are bridged with INS measurements, which improves positioning rate and compensates the errors caused by GNSS multi-path or mask. This kind of solutions are mature enough and widely used on MMS and autonomous navigation. However, drift is an innate issue for INS and high-quality INSs confine the drift in return of high cost.

A more affordable solution is SLAM (Simultaneous Localization and Mapping), which is able to use low-cost sensors. In practice, cameras are the most widely used sensor for SLAM. Compared with some active sensors like laser, cameras are cheaper and rich information (e.g. texture, spatial) can be acquired. The solutions for SLAM using cameras are usually called Visual Odometry (VO), which estimate relative poses by tracking the correspondences across images (Nistér et al., 2004). Different techniques were proposed to improve the accuracy of VO and they can be classified as the probabilistic filter (e.g. Extended Kalman Filter (EKF), Particle Filter (PF)) and Bundle Adjustment (BA). BA achieves better accuracy than EKF or PF (Ji and Yuan, 2016), because BA provides a global least squares optimization and adjusts image poses and object points by taking into account the entire observation equations and eventual constraints. However, the size of equation system increases rapidly with the growing number of images. Then, Local Bundle Adjustment (LBA) is proposed to reduce the complexity of BA (Mouragnon et al., 2006). It employs BA on a fixed size of sliding window and propagates the uncertainties through image sequence (Eudes and Lhuillier, 2009). SLAM is a solution for position tracking, the localization is conducted on a local system and errors accumulate over time. Although the global drift can be reduced by loop closure, there is no loop in many scenarios for a moving robot or vehicle in urban areas and computation of large loops is time-consuming.

Many methods were proposed to integrate GNSS with VO, where the drift was compensated using the positions measured by GNSS (Agrawal and Konolige, 2006; Wei et al., 2011; Lhuillier, 2012). However, the accuracy depends on the precision of GNSS data and GNSS measurements are not always reliable in urban environment. The drift increases quickly when the system suffers multi-path or mask problems. Therefore, more external data needs to be considered for precise localization based on cameras.

---

* Corresponding author.
E-mail address: xiaozhi.qu@foxmail.com (X. Qu).

https://doi.org/10.1016/j.isprsjprs.2017.09.010

Please cite this article in press as: Qu, X., et al. Landmark based localization in urban environment. ISPRS J. Photogram. Remote Sensing (2017), https://doi.org/10.1016/j.isprsjprs.2017.09.010

2. Related work

Nowadays, many VO methods have been proposed and different types of maps have been designed, produced and applied for precise localization. In this section, both state-of-the-art VO approaches and map based localization methods are investigated, our landmark based localization approach is introduced briefly at the end.

2.1. Visual odometry

The visual odometry was firstly proposed by Nister (Nistér et al., 2004). Current VO approaches can be summarized as feature based and direct VO approaches. Most feature based VO methods follow the approach introduced in PTAM (Parallel Tracking and Mapping) (Klein and Murray, 2009), which has three main modules: (1) feature extraction, matching and tracking; (2) poses and map points estimation and optimization; and (3) loop closure and global optimization. Usually, interest points or lines are detected in images and matched over sequences. The optimal solutions are achieved by minimizing back projection errors with the methods such as bundle adjustment, EKF and PF. PTAM system was designed for indoor environment and the computational cost increases quickly in large scale environment. A more efficient and accurate approach on feature based VO, is proposed as ORB-SLAM (Mur-Artal and Tardós, 2016). It detects the ORB (Oriented FAST and Rotated BRIEF) features in images for matching and tracking. The local map is optimized using local bundle adjustment. The loop is recognized based on DBoW2 (Galvez-Lpez and Tardos, 2012) built on ORB features, and all the poses are optimized using global bundle adjustment once the loop is detected successfully.

Different with feature based VO methods which aim at minimizing back-projection errors, direct approaches usually minimize the photometric errors on image intensities (Forster et al., 2014). A representative work on direct VO was LSD-SLAM (Engel et al., 2014), it built semi-dense maps in large scale, based on real-time image alignment. A recent work of Engel et al. Engel et al. (2016) known as DSO (Direct Sparse Odometry), had the benefits from both direct (no feature extraction) and sparse (joint optimization of parameters), so DSO achieved better accuracy than LSD-SLAM. Nevertheless, theses direct VO methods still need features for loop closure (Mur-Artal and Tardós, 2016). Loop closure can compensate the drift of VO over time, but the absolute scale is very difficult to determine using mono based VO. Moreover, the paths of a moving agent do not always have loops in urban environment. In this case, geo-referenced maps need to be integrated into VO system for precise localization.

2.2. Map based localization methods

In this paper, we classify the map based localization methods according to the types of the map, which are *low level maps*, *conventional maps* and *semantic maps*.

2.2.1. Low level maps

The low level maps are composed of geo-referenced visual features or point clouds. They are usually used for localization in a *teach and repeat* (e.g. route following for robot) (Furgale and Barfoot, 2010). The maps are built in teaching steps with SFM (Structure From Motion) (Royer et al., 2007) or scanning point clouds by a MMS (Bodensteiner et al., 2011). In a repeating step, the robot intends to follow the same route by registering the images captured by on-board cameras with maps, according to the similarity of visual features or intensity maps for point clouds.

Low level maps are easy to produce, but they have two main drawbacks. First, the high storage volume is required due to rich features extracted from images or point clouds acquired by a laser scanner. Second, incremental updates are extremely complicated. For instance, it is difficult to know which 3D points are changed.

2.2.2. Conventional maps

Conventional maps represent generalized models of the world and they are composed of elements like segments, polygons and planes. Light storage is needed and they are easy to be updated in comparison to low level maps. Thus, more methods integrate conventional maps (e.g. OpenStreetMap, DEM (Digital Elevation Model), 3D models) with VO. For 2D maps like OpenStreetMap, curve-to-curve map matching can be applied to correct the drift for localization on road, because the road segments, building boundary, locations and attributes of objects are known. These methods usually use a cheap GPS for initial positions (Alonso et al., 2012; Brubaker et al., 2013). Recently, a graph matching based method was proposed to align the trajectory estimated by VO with 2D road network without initial positions (Gupta et al., 2016). For higher dimension conventional maps such as DEM and 3D city model, direct pose estimation is possible. For instance, the pose of an individual image is optimized by registering images with 2.5D untextured city model based on semantic image segmentation (Arth et al., 2015). A more direct strategy is to register the coarse structure generated by VO with 3D city model to correct the drift (Lothe et al., 2009) and even to estimate poses by registering images with textured 3D model based on mutual information (Caron et al., 2014). However, the precision of conventional maps is 1–5 m, which is not sufficient for precise localization.

2.2.3. Semantic maps

In urban areas, there are rich semantic features like road marks, road signs and other man-made objects which are well-defined objects. They can be detected and reconstructed precisely from ground based imagery (Soheilian et al., 2010), aerial images (Tournaire et al., 2006) and point clouds (Hervieu et al., 2015). In particular, these semantic objects are easy to be detected in images that make them be convenient to integrate in VO approaches. Different types of objects were integrated for precise localization. The lane markings extracted from aerial images, were used to correct positioning errors by aligning the lane marks detected in images with maps based on ICP (Iterative Closest Point) algorithm (Pink, 2008). A similar strategy was proposed, but the precise locations were estimated via map matching between pre-built road mark maps and road marks extracted from stereo images (Schreiber et al., 2013). These two methods need initial positions measured by a GPS. Moreover, the vehicle poses can also be optimized with the constraints generated by 2D-3D correspondences between images and geo-referenced road signs (Wei et al., 2014) or road marks (Tournaire et al., 2006). These semantic objects can be expressed using segments or polygons in the database, so they are convenient to be stored and updated.

2.3. Our strategy

A featured based VO system is developed in this paper. In our previous work, we integrated geo-referenced road signs (Qu et al., 2015) and road marks (Soheilian et al., 2016) with VO using a single camera. The road signs were reconstructed from images (Soheilian et al., 2013) and the road marks were extracted from point clouds (Hervieu et al., 2015) captured by a MMS. The localization was optimized with LBA and some Ground Control Points (GCPs) were generated by registering road signs or road marks in a pre-built database in order to reduce the drift. Besides, uncertainties were estimated to predict the uncertainty of poses. In this

Please cite this article in press as: Qu, X., et al. Landmark based localization in urban environment. ISPRS J. Photogram. Remote Sensing (2017), https://doi.org/10.1016/j.isprsjprs.2017.09.010