Wu et al. Visual Computing for Industry, Biomedicine, and Art (2018) 1:8
https://doi.org/10.1186/s42492-018-0008-z

Visual Computing for Industry, Biomedicine, and Art

REVIEW Open Access

Image-based camera localization: an overview

Yihong Wu*, Fulin Tang and Heping Li

Abstract

Virtual reality, augmented reality, robotics, and autonomous driving, have recently attracted much attention from both academic and industrial communities, in which image-based camera localization is a key task. However, there has not been a complete review on image-based camera localization. It is urgent to map this topic to enable individuals enter the field quickly. In this paper, an overview of image-based camera localization is presented. A new and complete classification of image-based camera localization approaches is provided and the related techniques are introduced. Trends for future development are also discussed. This will be useful not only to researchers, but also to engineers and other individuals interested in this field.

Keywords: PnP problem, SLAM, Camera localization, Camera pose determination

Background

Recently, virtual reality, augmented reality, robotics, autonomous driving etc., in which image-based camera localization is a key task, have attracted much attention from both academic and industrial community. It is urgent to provide an overview of image-based camera localization.

The sensors used for image-based camera localization are cameras. Many types of three-dimensional (3D) cameras have been developed recently. This study considers two-dimensional (2D) cameras. The typically used tool for outdoor localization is GPS, which cannot be used indoors. There are many indoor localization tools including Lidar, Ultra Wide Band (UWB), Wireless Fidelity (WiFi), etc.; among these, using cameras for localization is the most flexible and low cost approach. Autonomous localization and navigation is necessary for a moving robot. To augment reality in images, camera pose determination or localization is needed. To view virtual environments, the corresponding viewing angle is necessary to be computed. Furthermore, cameras are ubiquitous and people carry mobile phones that have cameras every day. Therefore, image-based camera localization has great and widespread applications.

The image features of points, lines, conics, spheres, and angles are used in image-based camera localization; of these, points are most widely used. This study focuses on points.

Image-based camera localization is a broad topic. We attempt to cover related works and give a complete classification for image-based camera localization approaches. However, it is not possible to cover all related works in this paper due to length constraints. Moreover, we cannot provide deep criticism for each cited paper due to space limit for such an extensive topic. Further deep reviews on some active important aspects of image-based camera localization will be given in the future or people interested go to read already existing surveys. There have been excellent reviews on some aspects of image-based camera localization. The most recent ones include the following. Khan and Adnan [1] gave an overview of ego motion estimation, where ego motion requires time intervals between two continuous images to be small enough. Cadena et al. [2] surveyed the current state of simultaneous localization and mapping (SLAM) and considered future directions, in which they reviewed related works including robustness and scalability in long-term mapping, metric and semantic representations for mapping, theoretical performance guarantees, active SLAM, and exploration. Younes et al. [3] specially reviewed keyframe-based monocular SLAM. Piasco et al. [4] provided a survey on visual-based localization from

* Correspondence: yhwu@nlpr.ia.ac.cn
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China, University of Chinese Academy of Sciences, Beijing, China

Springer Open

© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

heterogeneous data, where only known environment is considered.

Unlike these studies, this study is unique in that it first maps the whole image-based camera localization and provides a complete classification for the topic. "Overview" section presents an overview of image-based camera localization and is mapped as a tree structure. "Reviews on image-based camera localization" section introduces each aspect of the classification. "Discussion" section presents discussions and analyzes trends of future developments "Conclusion" section makes a conclusion of the paper.

Overview

What is image-based camera localization? Image-based camera localization is to compute camera poses under a world coordinate system from images or videos captured by the cameras. Based on whether the environment is known beforehand or not, image-based camera localization can be classified into two categories: one with known environment and the other with unknown environment.

Let n be the number of points used. The approach with known environments consists of methods with $3 \le n < 6$ and methods with $n \ge 6$. These are PnP problems. In general, the problems with $3 \le n < 6$ are nonlinear and those with $n \ge 6$ are linear.

The approach with unknown environments can be divided into methods with online and real-time environment mapping and those without online and real-time environment mapping. The former is the commonly known Simultaneous Localization and Mapping (SLAM) and the latter is an intermediate procedure of the commonly known structure from motion (SFM). According to different map generations, SLAM is divided into four parts: geometric metric SLAM, learning SLAM, topological SLAM, and marker SLAM. Learning SLAM is a new research direction recently. We think it is different from geometric metric SLAM and topological SLAM by a single category. Learning SLAM can obtain camera pose and 3D map but needs a prior dataset to train the network. The performance of learning SLAM depends on the used dataset to a great extent and it has low generalization capability. Therefore, learning SLAM is not as flexible as geometric metric SLAM and its obtained 3D map outside the used dataset is not as accurate as geometric metric SLAM most of the time. However, simultaneously, learning SLAM has a 3D map other than topology representations. Marker SLAM computes camera poses from known structured markers without knowing the complete environment. Geometric metric SLAM consists of monocular SLAM, multiocular SLAM, and multi-kind sensor SLAM. Moreover, geometric metric SLAM can be classified into filter-based SLAM and keyframe-based SLAM. Keyframe-based SLAM can be further divided into feature-based SLAM and direct SLAM. Multi-kind sensors SLAM can be divided into loosely coupled SLAM and closely coupled SLAM. These classifications of image-based camera localization methods are visualized as a logical tree structure, as shown in Fig. 1, where current active topics are indicated with bold borders. We think that these topics are camera localization from large data, learning SLAM, keyframe-based SLAM, and multi-kind sensors SLAM.

Reviews on image-based camera localization

Known environments

Camera pose determination from known 3D space points is called the perspective-n-point problem, namely, the PnP problem. When $n = 1,2$, there are no solutions for PnP problems because they are under constraints. When $n \ge 6$, PnP problems are linear. When $n = 3, 4, 5$, the original equations of PnP problems are usually nonlinear. The PnP problem dated from 1841 to 1903. Grunert [5], Finsterwalder to Scheufele [6] concluded that the P3P problem has at most four solutions and the P4P problem has a unique solution in general. The PnP problem is also the key relocalization for SLAM.

PnP problems with n = 3, 4, 5

The methods to solve PnP problems with $n = 3, 4, 5$ focus on two aspects. One aspect studies the solution numbers or multisolution geometric configuration of the nonlinear problems. The other aspect studies eliminations or other solving methods for camera poses.

The methods that focus on the first aspect are as follows. Grunert [5], Finsterwalder and Scheufele [6] pointed out that P3P has up to four solutions and P4P has a unique solution. Fischler and Bolles [7] studied P3P for RANSAC of PnP and found that four solutions of P3P are attainable. Wolfe et al. [8] showed that P3P mostly has two solutions; they determined the two solutions and provided the geometric explanations that P3P can have two, three, or four solutions. Hu and Wu [9] defined distance-based and transformation-based P4P problems. They found that the two defined P4P problems are not equivalent; they found that the transformation-based problem has up to four solutions and distance-based problem has up to five solutions. Zhang and Hu [10] provided sufficient and necessary conditions in which P3P has four solutions. Wu and Hu [11] proved that distance-based problems are equivalent to rotation-transformation-based problems for P3P and distance-based problems are equivalent to orthogonal-transformation-based problems for P4P/P5P. In addition, they showed that for any three non-collinear points, the optical center can always be found such that the P3P problem formed by these three control points and the optical center will have four solutions, which is its upper bound. Additionally, a geometric approach is provided to construct these four solutions. Vynnycky and Kanev [12]

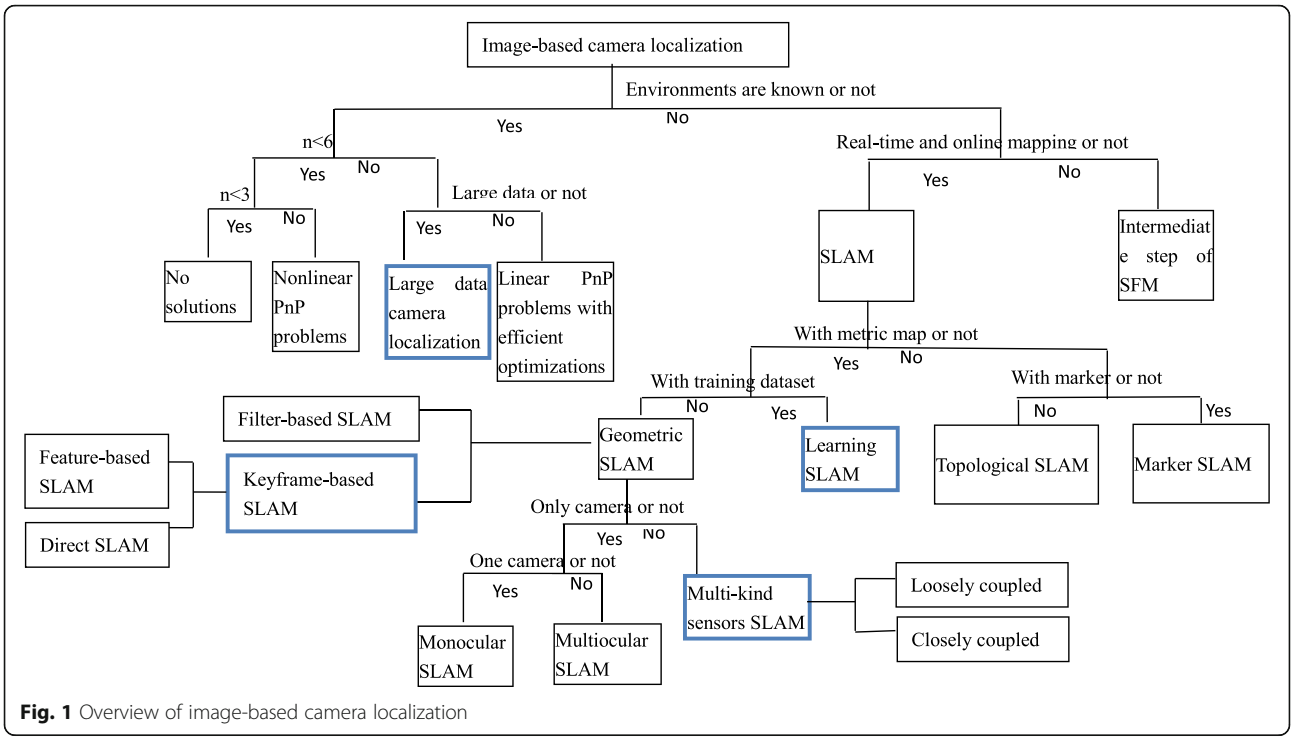

Fig. 1 Overview of image-based camera localization

studied the multisolution probabilities of the equilateral P3P problem.

The methods that focus on the second aspect of PnP problems with $n = 3, 4, 5$ are as follows. Horaud et al. [13] described an elimination method for the P4P problem to obtain a unitary quartic equation. Haralick et al. [14] reviewed six methods for the P3P problem, which are [5–7, 15–17]. Dementhon and Davis [18] presented a solution of the P3P problem by an inquiry table of quasi-perspective imaging. Quan and Lan [19] linearly solved the P4P and P5P problems. Gao et al. [20] used Wu's elimination method to obtain complete solutions of the P3P problem. Wu and Hu [10] introduced a depth-ratio-based approach to represent the solutions of the complete PnP problem. Josephson and Byrod [21] used Grobner bases method to solve the P4P problem for radial distortion of camera with unknown focal length. Hesch et al. [22] studied nonlinear square solutions of PnP with $n >= 3$. Kneip et al. [23] directly solved the rotation and translation solutions of the P3P problem. Kneip et al. [24] presented a unified PnP solution that can deal with generalized cameras and multisolutions with global optimizations and linear complexity. Kuang and Astrom [25] studied the PnP problem with unknown focal length using points and lines. Z. Kukelova et al. [26] studied the PnP problem with unknown focal length for images with radial distortion. Ventura et al. [27] presented a minimal solution to the generalized pose-and-scale problem. Zheng et al. [28] introduced an angle constraint and derived a compact bivariate polynomial equation for each P3P and then proposed a general method for the PnP problem with unknown focal length using iterations. Later, Zheng and Kneip [29] improved their work without requiring point order and iterations. Wu [30] studied PnP solutions with unknown focal length and $n = 3.5$. Albl et al. [31] studied the pose solution of a rolling shutter camera and improved the result later in 2016.

PnP problems with n ≥ 6

When $n> = 6$, PnP problems are linear and studies on them focus on two aspects. One aspect studies efficient optimizations for camera poses from smaller number of points. The other aspect studies fast camera localization from large data.

The studies on the first aspect are as follows. Lu et al. [32] gave a global convergence algorithm using collinear points. Schweighofer and Pinz [33] studied multisolutions of a planar target. Wu et al. [34] presented invariant relationships between scenes and images, and then a robust RANSAC PNP using the invariants. Lepetit et al. [35] provided an accurate O(n) solution to the PnP problem, called EPnP, which is widely used today. The pose problem of a rolling shutter camera was studied in [36] with bundle adjustments. A similar problem was also studied in [37] using B-spline covariance matrix. Zheng et al. [38] used quaternion and Grobner bases to provide a global optimized solution of the PnP problem. A very fast solution to the PnP Problem with algebraic outlier rejection was given in [39]. Svarm et al. [40] studied

accurate localization and pose estimation for large 3D models considering gravitational direction. Ozyesil et al. [41] provided robust camera location estimation by convex programming. Brachmann et al. [42] showed uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image. Feng et al. [43] proposed a hand-eye calibration-free strategy to actively relocate a camera in the same 6D pose by sequentially correcting relative 3D rotation and translation. Nakano [44] solved three kinds of PnP problems by the Grobner method: PnP problem for calibrated camera, PnPf problem for cameras with unknown focal length, PnPfr problem for cameras with unknown focal length and unknown radial distortions.

The studies on the second aspect that focus on fast camera localization from large data are as follows. Arth et al. [45, 46] presented real-time camera localizations for mobile phones. Sattler et al. [47] derived a direct matching framework based on visual vocabulary quantization and a prioritized correspondence search with known large-scale 3D models of urban scenes. Later, they improved the method by active correspondence search in [48]. Li et al. [49] devised an adaptive, prioritized algorithm for matching a representative set of SIFT features covering a large scene to a query image for efficient localization. Later Li et al. [50] provided a full 6-DOF-plus-intrinsic camera pose with respect to a large geo-registered 3D point cloud. Lei et al. [51] studied efficient camera localization from street views using PCA-based point grouping. Bansal and Daniilidis [52] proposed a purely geometric correspondence-free approach to urban geo-localization using 3D point-ray features extracted from the digital elevation map of an urban environment. Kendall et al. [53] presented a robust and real-time monocular 6-DOF relocalization system by training a convolutional neural network (CNN) to regress the 6-DOF camera pose from a single RGB image in an end-to-end manner. Wang et al. [54] proposed a novel approach to localization in very large indoor spaces that takes a single image and a floor plan of the environment as input. Zeisl et al. [55] proposed a voting-based pose estimation strategy that exhibits O(n) complexity in terms of the number of matches and thus facilitates considering more number of matches. Lu et al. [56] used a 3D model reconstructed by a short video as the query to realize 3D-to-3D localization under a multi-task point retrieval framework. Valentin et al. [57] trained a regression forest to predict mixtures of anisotropic 3D Gaussians and showed how the predicted uncertainties can be taken into account for continuous pose optimization. Straub et al. [58] proposed a relocalization system that enables real-time, 6D pose recovery for wide baselines by using binary feature descriptors and nearest-neighbor search of locality sensitive hashing. Feng et al. [59] achieved fast localization in large-scale environments by using supervised indexing of binary features, where randomized trees were constructed in a supervised training process by exploiting the label information derived from multiple features that correspond to a common 3D point. Ventura and Höllerer [60] proposed a system of arbitrary wide-area environments for real-time tracking with a handheld device. The combination of a keyframe-based monocular SLAM system and a global localization method was presented in [61]. A book of large-scale visual geo-localization was published in [62]. Liu et al. [63] showed efficient global 2D-3D matching for camera localization in a large-scale 3D map. Campbell [64] presented a method for globally optimal inlier set maximization for simultaneous camera pose and feature correspondence. Real-time SLAM relocalization with online learning of binary feature indexing was proposed by [65]. Wu et al. [66] proposed CNNs for camera relocalization. Kendall and Cipolla [67] explored a number of novel loss functions for learning camera poses, which are based on geometry and scene reprojection error. Qin et al. [68] developed a method of relocalization for monocular visual-inertial SLAM. Piasco et al. [4] presented a survey on visual-based localization from heterogeneous data. A geometry-based point cloud reduction method for mobile augmented reality system was presented in [69].

From above the studies for known environments, we see that fast camera localization from large data has attracted more and more attention. This is because there are many applications of camera localization for large data, for example, location-based services, relocalization of SLAM for all types of robots, and AR navigations.

Unknown environments

Unknown environments can be reconstructed from videos in real time and online. Simultaneously, camera poses are computed in real time and online. These are the commonly known SLAM technologies. If unknown environments are reconstructed from multiview images without requiring speed and online computation, it is the known SFM, in which solving for the camera pose is an intermediate step and not the final aim; therefore, we only mention few studies on SFM, and do not provide an in-depth overview in the following. Studies on SLAM will be introduced in detail.

SLAM

SLAM was dated from 1986 in the study [70]: "On the representation and estimation of spatial uncertainty," published in the International Journal of Robotics Research. In 1995, the acronym SLAM was then coined in the study [71]: "Localisation of automatic guided vehicles," 7th International Symposium on Robotics Research. According to different map generations, the studies on SLAM can be divided into four categories: geometric metric SLAM, learning SLAM, topological

SLAM, and marker SLAM. Due to its accurate computations, geometric metric SLAM has attracted increasing attention. Learning SLAM is a new topic gaining attention due to the development of deep learning. Studies on pure topological SLAM are decreasing. Marker SLAM is more accurate and stable. There is a study in [2] that reviews recent advances of SLAM covering a broad set of topics including robustness and scalability in long-term mapping, metric and semantic representations for mapping, theoretical performance guarantees, active SLAM, and exploration. In the following, we introduce geometric metric SLAM, learning SLAM, topological SLAM, and marker SLAM.

A. Geometric metric SLAM

Geometric metric SLAM computes 3D maps with accurate mathematical equations. Based on the different sensors used, geometric metric SLAM is divided into monocular SLAM, multiocular SLAM, and multi-kind sensors SLAM. Based on the different techniques used, geometric metric SLAM is divided into filter-based SLAM and keyframe-based SLAM and also, there is another class of SLAM: grid-based SLAM, of which minority deal with images and most deal with laser data. Recently, there was a review on keyframe-based monocular SLAM, which provided in-depth analyses [3].

A.1) Monocular SLAM

A.1.1) Filter-based SLAM

One part of monocular SLAM is the filter-based methods. The first one is the MonoSLAM proposed by Davison [72] based on extended Kalman filter (EKF). Later, the work was developed by them further in [73, 74]. Montemerlo and Thrun [75] proposed monocular SLAM based on a particle filter. Strasdat et al. [76, 77] discussed why filter-based SLAM is used by comparing filter-based and keyframe-based methods. The conference paper in ICRA 2010 by [76] received the best paper award, where they pointed out that keyframe-based SLAM can provide more accurate results. Nuchter et al. [78] used a particle filter for SLAM to map large 3D outdoor environments. Huang et al. [79] addressed two key limitations of the unscented Kalman filter (UKF) when applied to the SLAM problem: the cubic computational complexity in the number of states and the inconsistency of the state estimates. They introduced a new sampling strategy for the UKF, which has constant computational complexity, and proposed a new algorithm to ensure that the unobservable subspace of the UKF's linear-regression-based system model has the same dimension as that of the nonlinear SLAM system. Younes et al. [3] also stated that filter-based SLAM was common before 2010 and most solutions thereafter designed their systems around a non-filter, keyframe-based architecture.

A.1.2) Keyframe-based SLAM

The second part of monocular SLAM is the keyframe-based methods. Keyframe-based SLAM can be further categorized into: feature-based methods and direct methods. a) Feature-based SLAM: The first keyframe-based feature SLAM was PTAM proposed in [80]. Later the method was extended to combine edges in [81] and extended to a mobile phone platform by them in [82]. The keyframe selections were studied in [83, 84]. SLAM++ with loop detection and object recognition was proposed in [85]. Dynamic scene detection and adapting RANSAC was studied by [86]. Regarding dynamic objects, Feng et al. [87] proposed a 3D-aided optical flow SLAM. ORB SLAM [88] can deal with loop detection, dynamic scene detection, monocular, binocular, and deep images. The method of [89] can run in a large-scale environment using submap and linear program to remove outlier. b) Direct SLAM: The second part of monocular SLAM is the direct method. Newcombe et al. [90] proposed DTAM, the first direct SLAM, where detailed textured dense depth maps, at selected keyframes, are produced and meanwhile camera pose is tracked at frame rate by entire image alignment against the dense textured model. A semi-dense visual odometry (VO) was proposed in [91]. LSD SLAM by [92] provided a dense SLAM suitable for large-scale environments. Pascoe et al. [93] proposed a direct dense SLAM for road environments for LIDAR and cameras. A semi VO on a mobile phone was performed by [94].

A.2) Multiocular SLAM

Multiocular SLAM uses multiple cameras to compute camera poses and 3D maps. Most of the studies focus on binocular vision. They are also the bases of multiocular vision.

Konolige and Agrawal [95] matched visual frames with large numbers of point features using classic bundle adjustment techniques but kept only relative frame pose information. Mei et al. [96] used local estimation of motion and structure provided by a stereo pair to represent the environment in terms of a sequence of relative locations. Zou and Tan [97] studied SLAM of multiple moving cameras in which a global map is built. Engle et al. [98] proposed a

novel large-scale direct SLAM algorithm for stereo cameras. Pire et al. [99] proposed a stereo SLAM system called S-PTAM that can compute the real scale of a map and overcome the limitation of PTAM for robot navigation. Moreno et al. [100] proposed a novel approach called sparser relative bundle adjustment (SRBA) for a stereo SLAM system. Artal and Tardos [101] presented ORB-SLAM2 which is a complete SLAM system for monocular, stereo, and RGB-D cameras, with map reuse, loop closing, and relocalization capabilities. Zhang et al. [102] presented a graph-based stereo SLAM system using straight lines as features. Gomez-Ojeda et al. [103] proposed PL-SLAM, a stereo visual SLAM system that combines both points and line segments to work robustly in a wider variety of scenarios, particularly in those where point features are scarce or not well-distributed in an image. A novel direct visual-inertial odometry method for stereo cameras was proposed in [104]. Wang et al. [105] proposed stereo direct sparse odometry (Stereo DSO) for highly accurate real-time visual odometry estimation of large-scale environments from stereo cameras. Semi-direct visual odometry (SVO) for monocular and multi-camera systems was proposed in [106]. Sun et al. [107] proposed a stereo multi-state constraint Kalman filter (S-MSCKF). Compared with multi-state constraint Kalman filter (MSCKF), S-MSCKF exhibits significantly greater robustness.

Multiocular SLAM has higher reliability than monocular SLAM. In general, multiocular SLAM is preferred if hardware platforms are allowed.

A.3) Multi-kind sensors SLAM

Here, multi-kind sensors are limited to vision and inertial measurement unit (IMU); other sensors are not introduced here. This is because, recently, vision and IMU fusion has attracted more attention than others.

In robotics, there are many studies on SLAM that combine cameras and IMU. It is common for mobile devices to be equipped with a camera and an inertial unit. Cameras can provide rich information of a scene. IMU can provide self-motion information and also provide accurate short-term motion estimates at high frequency. Cameras and IMU have been thought to be complementary of each other. Because of universality and complementarity of visual-inertial sensors, visual-inertial fusion has been a very active research topic in recent years. The main research approaches on visual-inertial fusion can be divided into two categories, namely, loosely coupled and tightly coupled approaches.

A.3.1) Loosely coupled SLAM

In loosely coupled systems, all sensor states are independently estimated and optimized. Integrated IMU data are incorporated as independent measurements in stereo vision optimization in [108]. Vision-only pose estimates are used to update an EKF so that IMU propagation can be performed [109]. An evaluation of different direct methods for computing frame-to-frame motion estimates of a moving sensor rig composed of an RGB-D camera and an inertial measurement unit is given and the pose from visual odometry is added to the IMU optimization frame directly in [110].

A.3.2) Tightly coupled SLAM

In tightly coupled systems, all sensor states are jointly estimated and optimized. There are two approaches for this, namely, filter-based and keyframe nonlinear optimization-based approaches.

A.3.2.a) Filter-based approach

The filter-based approach uses EKF to propagate and update motion states of visual-inertial sensors. MSCKF in [111] uses an IMU to propagate the motion estimation of a vehicle and update this motion estimation by observing salient features from a monocular camera. Li and Mourikis [112] improved MSCKF, by proposing a real-time EKF-based VIO algorithm, MSCKF2.0. This algorithm can achieve consistent estimation by ensuring correct observability properties of its linearized system model and performing online estimation of the camera-to-inertial measurement unit calibration parameters. Li et al. [113], Li and Mourikis [114] implemented real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera. MSCKF algorithm is the core algorithm of Google's Project Tango https://get.google.com/tango/. Clement et al. [115] compared two modern approaches: MSCKF and sliding window filter (SWF). SWF is more accurate and less sensitive to tuning parameters than MSCKF. However, MSCKF is computationally cheaper, has good consistency, and improves accuracies because more features are tracked. Bloesch et al. [116] presented a monocular visual inertial odometry algorithm by directly using pixel intensity errors of image patches. In this algorithm, by directly using the intensity errors as an

innovation term, the tracking of multilevel patch features is closely coupled to the underlying EKF during the update step.

A.3.2.b) Keyframe nonlinear optimization-based approach

The nonlinear optimization-based approach uses keyframe-based nonlinear optimization, which may potentially achieve higher accuracy due to the capability to limit linearization errors through repeated linearization of the inherently nonlinear problem. Forster et al. [117] presented a preintegration theory that appropriately addresses the manifold structure of the rotation group. Moreover, it is shown that the preintegration IMU model can be seamlessly integrated into a visual-inertial pipeline under the unifying framework of factor graphs. The method is short for GTSAM. Leutenegger et al. [118] presented a novel approach, OKVIS, to tightly integrate visual measurements with IMU measurements, where a joint nonlinear cost function that integrates an IMU error term with the landmark reprojection error in a fully probabilistic manner is optimized. Moreover, to ensure real-time operation, old states are marginalized to maintain a bounded-sized optimization window. Li et al. [119] proposed tightly coupled, optimization-based, monocular visual-inertial state estimation for camera localization in complex environments. This method can run on mobile devices with a lightweight loop closure. Following ORB monocular SLAM [88], a tightly coupled visual-inertial slam system was proposed in [120].

In loosely coupled systems, it is easy to process frame and IMU data. However, in tightly coupled systems, to optimize all sensor states jointly, it is difficult to process frame and IMU data. In terms of estimation accuracy, tightly coupled methods are more accurate and robust than loosely coupled methods. Tightly coupled methods have become increasingly popular and have attracted great attention by researchers.

B. Learning SLAM

Learning SLAM is a new topic that gained attention recently due to the development of deep learning. We think it is different from geometric metric SLAM and topological SLAM by a single category. Learning SLAM can obtain camera pose and 3D map but needs a prior dataset to train the network. The performance of learning SLAM depends on the used dataset greatly and it has low generalization ability. Therefore, learning SLAM is not as flexible as geometric metric SLAM and the geometric map obtained outside the used dataset is not as accurate as geometric metric SLAM most of the time. However, simultaneously, learning SLAM has a 3D map other than 2D graph representations.

Tateno et al. [121] used CNNs to predict dense depth maps and then used keyframe-based 3D metric direct SLAM to compute camera poses. Ummenhofer et al. [122] trained multiple stacked encoder-decoder networks to compute depth and camera motion from successive, unconstrained image pairs. Vijayanarasimhan et al. [123] proposed a geometry-aware neural network for motion estimation in videos. Zhou et al. [124] presented an unsupervised learning framework for estimating monocular depth and camera motion from video sequences. Li et al. [125] proposed a monocular visual odometry system using unsupervised deep learning; they used stereo image pairs to recover the scales. Clark et al. [126] presented an on-manifold sequence-to-sequence learning approach for motion estimation using visual and inertial sensors. Detone et al. [127] presented a point tracking system powered by two deep CNNs, MagicPoint and MagicWarp. Gao and Zhang [128] presented a method for loop closure detection based on the stacked denoising auto-encoder. Araujo et al. [129] proposed a recurrent CNN-based visual odometry approach for endoscopic capsule robots.

Learning SLAM increases gradually these years. However, due to lower speed and generalization capabilities of the learning methods, using geometric methods is still centered for practical applications.

C. Topological SLAM

Topological SLAM does not need accurate computation of 3D maps and represents the environment by connectivity or topology. Kuipers and Byun [130] used a hierarchical description of the spatial environment, where a topological network description mediates between a control and metrical level; moreover, distinctive places and paths are defined by their properties at the control level and serve as nodes and arcs of the topological model. Ulrich and Nourbakhsh [131] presented an appearance-based place recognition system for topological localization. Choset and Nagatani [132] exploited the topology of a robot's free space to

localize the robot on a partially constructed map and the topology of the environment was encoded in a generalized Voronoi graph. Kuipers et al. [133] described how a local perceptual map can be analyzed to identify a local topology description and abstracted to a topological place. Chang et al. [134] presented a prediction-based SLAM algorithm to predict the structure inside an unexplored region. Blanco et al. [135] used Bayesian filtering to provide a probabilistic estimation based on the reconstruction of a robot path in a hybrid discrete-continuous state space. Blanco et al. [136] presented spectral graph partitioning techniques for the automatic generation of sub-maps. Kawewong et al. [137] proposed dictionary management to eliminate redundant search for indoor loop-closure detection based on PIRF extraction. Sünderhauf and Protzel [138] presented a back-end formulation for SLAM using switchable constraints to recognize and reject outliers during loop-closure detection by making the topology of the underlying factor graph representation. Latif et al. [139] described a consensus-based approach for robust place recognition to detect and remove past incorrect loop closures to deal with the problem of corrupt map estimates. Latif et al. [140] presented a comparative analysis for graph SLAM, where graph nodes are camera poses connected by odometry or place recognition. Vallvé et al. [141] proposed two simple algorithms for SLAM sparsification, factor descent and non-cyclic factor descent.

As shown in some above mentioned works, topological SLAM has been modified into metric SLAM as loop detection these years. Studies on pure topological SLAM are reducing.

D. Marker SLAM

We introduced studies on image-based camera localization for both known and unknown environments above. In addition, there are some studies to localize cameras using some prior environment knowledge, but not a 3D map such as markers. These works are considered to be with semi-known environments.

In 1991, Gatrell et al. [142] designed a concentric circular marker, which was modified with additional color and scale information in [143]. Ring information was considered in the marker by [144]. Kato and Billinghurst [145] presented the first augmented reality system based on fiducial markers known as the ARToolkit, where the marker used is a black enclosed rectangle with simple graphics or text. Naimark and Foxlin [146] developed a more general marker generation method, which encodes a bar code into a black circular region to produce more markers. A square marker was presented by [147]. Four circles at the corners of a square were proposed by [148]. A black rectangle enclosed with black and white blocks known as the ARTag was proposed by [149, 150]. From four marker points, Maidi et al. [151] developed a hybrid approach that combines an iterative method based on the EKF and an analytical method with direct resolution of pose parameter computation. Recently, Bergamasco et al. [152] provided a set of circular high-contrast dots arranged in concentric layers. DeGol et al. [153] introduced a fiducial marker, ChromaTag, and a detection algorithm to use opponent colors to limit and reject initial false detections and grayscale. Munoz-Salinas et al. [154] proposed to detect key points for the problem of mapping and localization from a large set of squared planar markers. Eade and Drummond [155] proposed real-time global graph SLAM for sequences with several hundreds of landmarks. Wu [156] studied a new marker for camera localization that does not need matching.

SFM

In SFM, camera pose computation is only an intermediate step. Therefore, in the following, we give a brief introduction of camera localization SFM.

In the early stages of SFM development, there were more studies on relative pose solving. One of the useful studies is the algorithm for five-point relative pose problem in [157], which has less degeneracies than other relative pose solvers. Lee et al. [158] studied relative pose estimation for a multi-camera system with known vertical direction. Kneip and Li [159] presented a novel solution to compute the relative pose of a generalized camera. Chatterjee and Govindu [160] presented efficient and robust large-scale averaging of relative 3D rotations. Ventura et al. [161] proposed an efficient method for estimating the relative motion of a multi-camera rig from a minimal set of feature correspondences. Fredriksson et al. [162] estimated the relative translation between two cameras and simultaneously maximized the number of inlier correspondences.

Global pose studies are as follows. Park et al. [163] estimated the camera direction of a geotagged image using reference images. Carlone et al. [164] surveyed techniques for 3D rotation estimation. Jiang et al. [165] presented a global linear method for camera pose registration. Later, the method was improved by [166] and [167].

Recently, hybrid incremental and global SFM have been developed. Cui et al. [168, 169], estimated rotations by a global method and translations by an incremental method and proposed community-based SFM. Zhu et al. [170] presented parallel SFM from local increment to global averaging.

A recent survey on SFM is presented in [171]. In addition, there are some studies on learning depth from

a single image. From binoculars, usually disparity maps are learnt. Please refer to the related works ranked in the website of KITTI dataset.

Discussion

From the above techniques, we can see that currently there are less and less studies on the PnP problem in a small-scale environment. Similarly, there are few studies on SFM using traditional geometric methods. However, for SLAM, both traditional geometric and learning methods are still popular.

Studies that use deep learning for image-based camera localization are increasing gradually. However, in practical applications, using geometric methods is still centered. Deep learning methods can provide efficient image features and compensate for geometric methods.

The PnP problem or relocalization of SLAM in a large-scale environment has not been solved well and deserves further research. For reliability and low cost practical applications, multi low cost sensor fusion for localization but vision sensor centered is an effective way.

In addition, some works study the pose problem of other camera sensors, such as the epipolar geometry of a rolling shutter camera in [172, 173] and radial-distorted rolling-shutter direct SLAM in [174]. Gallego et al. [175], Vidal et al. [176], Rebecq et al. [177] studied event camera SLAM.

With the increasing development of SLAM, maybe it starts the age of embedded SLAM algorithms as shown by [178]. We think integrating the merits of all kinds of techniques is a trend for a practical SLAM system, such as geometric and learning fusion, multi-sensor fusion, multi-feature fusion, feature based and direct approaches fusion. Integration of these techniques may solve the current challenging difficulties such as poorly textured scenes, large illumination changes, repetitive textures, and highly dynamic motions.

Conclusion

Image-based camera localization has important applications in fields such as virtual reality, augmented reality, robots. With the rapid development of artificial intelligence, these fields have become high-growth markets, and are attracting much attention from both academic and industrial communities.

We presented an overview of image-based camera localization, in which a complete classification is provided. Each classification is further divided into categories and the related works are presented along with some analyses. Simultaneously, the overview is described in a tree structure, as shown in Fig. 1. In the tree structure, the current popular topics are denoted with bold blue borders. These topics include large data camera localization, learning SLAM, multi-kind sensors SLAM, and keyframe-based SLAM. Future developments were also discussed in the Discussion section.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 61421004, 61572499, 61632003.

Authors' contributions

All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 3 April 2018 Accepted: 6 July 2018
Published online: 05 September 2018

References

1. Khan NH, Adnan A. Ego-motion estimation concepts, algorithms and challenges: an overview. Multimed Tools Appl. 2017;76:16581–603.
2. Cadena C, Carlone L, Carrillo H, Latif Y, Scaramuzza D, Neira J, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age. IEEE Trans Robot. 2016;32:1309–32.
3. Younes G, Asmar D, Shammas E, Zelek J. Keyframe-based monocular SLAM: design, survey, and future directions. Rob Auton Syst. 2017;98:67–88.
4. Piasco N, Sidibé D, Demonceaux C, Gouet-Brunet V. A survey on visual-based localization: on the benefit of heterogeneous data. Pattern Recogn. 2018;74:90–109.
5. Grunert JA. Das pothenotische problem in erweiterter gestalt nebst bemerkungen über seine anwendung in der Geodäsie. In: Archiv der mathematik und physik, Band 1. Greifswald; 1841. p. 238–48.
6. Finsterwalder S, Scheufele W. In: Finsterwalder zum S, editor. Das ruckwartseinschneiden im raum, vol. 75; 1937. p. 86–100.
7. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM. 1981;24:381–95.
8. Wolfe WJ, Mathis D, Sklair CW, Magee M. The perspective view of three points. IEEE Trans Pattern Anal Mach Intell. 1991;13:66–73.
9. Hu ZY, Wu FC. A note on the number of solutions of the noncoplanar P4P problem. IEEE Trans Pattern Anal Mach Intell. 2002;24:550–5.
10. Zhang CX, Hu ZY. A general sufficient condition of four positive solutions of the P3P problem. J Comput Sci Technol. 2005;20:836–42.
11. Wu YH, Hu ZY. PnP problem revisited. J Math Imaging Vis. 2006;24:131–41.
12. Vynnycky M, Kanev K. Mathematical analysis of the multisolution phenomenon in the P3P problem. J Math Imaging Vis. 2015;51:326–37.
13. Horaud R, Conio B, Leboulleux O, Lacolle B. An analytic solution for the perspective 4-point problem. Comput Vis Graph Image Process. 1989;47:33–44.
14. Haralick RM, Lee CN, Ottenburg K, Nölle M. Analysis and solutions of the three point perspective pose estimation problem. In: Proceedings of 1991 IEEE computer society conference on computer vision and pattern recognition. Maui: IEEE; 1991. p. 592–8.
15. Merritt EL. Explicity three-point resection in space. Photogramm Eng. 1949; 15:649–55.
16. Linnainmaa S, Harwood D, Davis LS. Pose determination of a three-dimensional object using triangle pairs. IEEE Trans Pattern Anal Mach Intell. 1988;10:634–47.
17. Grafarend EW, Lohse P, Schaffrin B. Dreidimensionaler ruckwartsschnitt, teil I: die projektiven Gleichungen. In: Zeitschrift für vermessungswesen. Stuttgart: Geodätisches Institut, Universität; 1989. p. 172–5.
18. DeMenthon D, Davis LS. Exact and approximate solutions of the perspective-three-point problem. IEEE Trans Pattern Anal Mach Intell. 1992; 14:1100–5.
19. Quan L, Lan ZD. Linear n-point camera pose determination. IEEE Trans Pattern Anal Mach Intell. 1999;21:774–80.
20. Gao XS, Hou XR, Tang JL, Cheng HF. Complete solution classification for the perspective-three-point problem. IEEE Trans Pattern Anal Mach Intell. 2003; 25:930–43.

21. Josephson K, Byrod M. Pose estimation with radial distortion and unknown focal length. In: Proceedings of 2009 IEEE conference on computer vision and pattern recognition. Miami: IEEE; 2009. p. 2419–26.
22. Hesch JA, Roumeliotis SI. A direct least-squares (DLS) method for PnP. In: Proceedings of 2011 international conference on computer vision. Barcelona: IEEE; 2012. p. 383–90.
23. Kneip L, Scaramuzza D, Siegwart R. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In: CVPR 2011. Providence: IEEE; 2011. p. 2969–76.
24. Kneip L, Li HD, Seo Y. UPnP: an optimal O(*n*) solution to the absolute pose problem with universal applicability. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer vision – ECCV 2014. Cham: Springer; 2014. p. 127–42.
25. Kuang YB, Åström K. Pose estimation with unknown focal length using points, directions and lines. In: Proceedings of 2013 IEEE international conference on computer vision. Sydney: IEEE; 2013. p. 529–36.
26. Kukelova Z, Bujnak M, Pajdla T. Real-time solution to the absolute pose problem with unknown radial distortion and focal length. In: Proceedings of 2013 IEEE international conference on computer vision. Sydney: IEEE; 2014. p. 2816–23.
27. Ventura J, Arth C, Reitmayr G, Schmalstieg D. A minimal solution to the generalized pose-and-scale problem. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 422–9.
28. Zheng YQ, Sugimoto S, Sato I, Okutomi M. A general and simple method for camera pose and focal length determination. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 430–7.
29. Zheng YQ, Kneip L. A direct least-squares solution to the PnP problem with unknown focal length. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE; 2016. p. 1790–8.
30. Wu CC. P3.5P: pose estimation with unknown focal length. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition. Boston: IEEE; 2015. p. 2440–8.
31. Albl C, Kukelova Z, Pajdla T. R6P - rolling shutter absolute pose problem. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition. Boston: IEEE; 2015. p. 2292–300.
32. Lu CP, Hager GD, Mjolsness E. Fast and globally convergent pose estimation from video images. IEEE Trans Pattern Anal Mach Intell. 2000;22:610–22.
33. Schweighofer G, Pinz A. Robust pose estimation from a planar target. IEEE Trans Pattern Anal Mach Intell. 2006;28:2024–30.
34. Wu YH, Li YF, Hu ZY. Detecting and handling unreliable points for camera parameter estimation. Int J Comput Vis. 2008;79:209–23.
35. Lepetit V, Moreno-Noguer F, Fua P. EP*n*P: an accurate O(*n*) solution to the P*n*P problem. Int J Comput Vis. 2009;81:155–66.
36. Hedborg J, Forssén PE, Felsberg M, Ringaby E. Rolling shutter bundle adjustment. In: Proceedings of 2012 IEEE conference on computer vision and pattern recognition. Providence: IEEE; 2012. p. 1434–41.
37. Oth L, Furgale P, Kneip L, Siegwart R. Rolling shutter camera calibration. In: Proceedings of 2013 IEEE conference on computer vision and pattern recognition. Portland: IEEE; 2013. p. 1360–7.
38. Zheng YQ, Kuang YB, Sugimoto S, Åström K, Okutomi M. Revisiting the P*n*P problem: a fast, general and optimal solution. In: Proceedings of 2013 IEEE international conference on computer vision. Sydney: IEEE; 2013. p. 2344–51.
39. Ferraz L, Binefa X, Moreno-Noguer F. Very fast solution to the PnP problem with algebraic outlier rejection. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 501–8.
40. Svärm L, Enqvist O, Oskarsson M, Kahl F. Accurate localization and pose estimation for large 3D models. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 532–9.
41. Özyesil O, Singer A. Robust camera location estimation by convex programming. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition. Boston: IEEE; 2015. p. 2674–83.
42. Brachmann E, Michel F, Krull A, Yang MY, Gumhold S, Rother C. Uncertainty-driven 6D pose estimation of objects and scenes from a single RGB image. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE; 2016. p. 3364–72.
43. Feng W, Tian FP, Zhang Q, Sun JZ. 6D dynamic camera relocalization from single reference image. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE; 2016. p. 4049–57.
44. Nakano G. A versatile approach for solving PnP, PnPf, and PnPfr problems. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer vision – ECCV 2016. Cham: Springer; 2016. p. 338–52.
45. Arth C, Wagner D, Klopschitz M, Irschara A, Schmalstieg D. Wide area localization on mobile phones. In: Proceedings of the 8th IEEE international symposium on mixed and augmented reality. Orlando: IEEE; 2009. p. 73–82.
46. Arth C, Klopschitz M, Reitmayr G, Schmalstieg D. Real-time self-localization from panoramic images on mobile devices. In: Proceedings of the 10th IEEE international symposium on mixed and augmented reality. Basel: IEEE; 2011. p. 37–46.
47. Sattler T, Leibe B, Kobbelt L. Fast image-based localization using direct 2D-to-3D matching. In: Proceedings of 2011 international conference on computer vision. Barcelona: IEEE; 2011. p. 667–74.
48. Sattler T, Leibe B, Kobbelt L. Improving image-based localization by active correspondence search. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C, editors. Computer vision – ECCV 2012. Berlin, Heidelberg: Springer; 2012. p. 752–65.
49. Li YP, Snavely N, Huttenlocher DP. Location recognition using prioritized feature matching. In: Daniilidis K, Maragos P, Paragios N, editors. Computer vision – ECCV 2010. Berlin, Heidelberg: Springer; 2010. p. 791–804.
50. Li YP, Snavely N, Huttenlocher D, Fua P. Worldwide pose estimation using 3D point clouds. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C, editors. Computer vision – ECCV 2012. Berlin, Heidelberg: Springer; 2012. p. 15–29.
51. Lei J, Wang ZH, Wu YH, Fan LX. Efficient pose tracking on mobile phones with 3D points grouping. In: Proceedings of 2014 IEEE international conference on multimedia and expo. Chengdu: IEEE; 2014. p. 1–6.
52. Bansal M, Daniilidis K. Geometric urban geo-localization. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 3978–85.
53. Kendall A, Grimes M, Cipolla R. PoseNet: a convolutional network for real-time 6-DOF camera relocalization. In: Proceedings of 2015 IEEE international conference on computer vision. Santiago: IEEE; 2015. p. 2938–46.
54. Wang SL, Fidler S, Urtasun R. Lost shopping! Monocular localization in large indoor spaces. In: Proceedings of 2015 IEEE international conference on computer vision. Santiago: IEEE; 2015. p. 2695–703.
55. Zeisl B, Sattler T, Pollefeys M. Camera pose voting for large-scale image-based localization. In: Proceedings of 2015 IEEE international conference on computer vision. Santiago: IEEE; 2015. p. 2704–12.
56. Lu GY, Yan Y, Ren L, Song JK, Sebe N, Kambhamettu C. Localize me anywhere, anytime: a multi-task point-retrieval approach. In: Proceedings of 2015 IEEE international conference on computer vision. Santiago: IEEE; 2015. p. 2434–42.
57. Valentin J, Nießner M, Shotton J, Fitzgibbon A, Izadi S, Torr P. Exploiting uncertainty in regression forests for accurate camera relocalization. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition. Boston: IEEE; 2015. p. 4400–8.
58. Straub J, Hilsenbeck S, Schroth G, Huitl R, Möller A, Steinbach E. Fast relocalization for visual odometry using binary features. In: Proceedings of 2013 IEEE international conference on image processing. Melbourne: IEEE; 2013. p. 2548–52.
59. Feng YJ, Fan LX, Wu YH. Fast localization in large-scale environments using supervised indexing of binary features. IEEE Trans Image Process. 2016;25:343–58.
60. Ventura J, Höllerer T. Wide-area scene mapping for mobile visual tracking. In: Proceedings of 2012 IEEE international symposium on mixed and augmented reality. Atlanta: IEEE; 2012. p. 3–12.
61. Ventura J, Arth C, Reitmayr G, Schmalstieg D. Global localization from monocular SLAM on a mobile phone. IEEE Trans Vis Comput Graph. 2014; 20:531–9.
62. Zamir AR, Hakeem A, Van Gool L, Shah M, Szeliski R. Large-scale visual geo-localization. Cham: Springer; 2016.
63. Liu L, Li HD, Dai YC. Efficient global 2D-3D matching for camera localization in a large-scale 3D map. In: Proceedings of 2017 IEEE international conference on computer vision. Venice: IEEE; 2017.
64. Campbell D, Petersson L, Kneip L, Li HD. Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence. In: Proceedings of 2017 IEEE international conference on computer vision. Venice: IEEE; 2017.
65. Feng YJ, Wu YH, Fan LX. Real-time SLAM relocalization with online learning of binary feature indexing. Mach Vis Appl. 2017;28:953–63.
66. Wu J, Ma LW, Hu XL. Delving deeper into convolutional neural networks for camera relocalization. In: Proceedings of 2017 IEEE international conference on robotics and automation. Singapore: IEEE; 2017. p. 5644–51.

67. Kendall A, Cipolla R. Geometric loss functions for camera pose regression with deep learning. In: Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu: IEEE; 2017.
68. Qin T, Li P, Shen S. Relocalization, global optimization and map merging for monocular visual-inertial SLAM. In: Proceedings of IEEE international conference on robotics and automation. Brisbane: HKUST; 2018.
69. Wang H, Lei J, Li A, Wu Y. A geometry-based point cloud reduction method for mobile augmented reality system. Accepted by J Compu Sci Technol. 2018.
70. Smith RC, Cheeseman P. On the representation and estimation of spatial uncertainty. Int J Robot Res. 1986;5:56–68.
71. Durrant-Whyte H, Rye D, Nebot E. Localization of autonomous guided vehicles. In: Hollerbach JM, Koditschek DE, editors. Robotics research. London: Springer; 1996. p. 613–25.
72. Davison AJ. SLAM with a single camera. In: Proceedings of workshop on concurrent mapping and localization for autonomous mobile robots in conjunction with ICRA. Washington, DC: CiNii; 2002.
73. Davison AJ. Real-time simultaneous localisation and mapping with a single camera. In: Proceedings of the 9th IEEE international conference on computer vision. Nice: IEEE; 2003. p. 1403–10.
74. Davison AJ, Reid ID, Molton ND, Stasse O. MonoSLAM: real-time single camera SLAM. IEEE Trans Pattern Anal Mach Intell. 2007;29:1052–67.
75. Montemerlo M, Thrun S. Simultaneous localization and mapping with unknown data association using FastSLAM. In: Proceedings of 2003 IEEE international conference on robotics and automation. Taipei: IEEE; 2003. p. 1985–91.
76. Strasdat H, Montiel JMM, Davison AJ. Real-time monocular SLAM: why filter? In: Proceedings of 2010 IEEE international conference on robotics and automation. Anchorage: IEEE; 2010. p. 2657–64.
77. Strasdat H, Montiel JMM, Davison AJ. Visual SLAM: why filter? Image Vis Comput. 2012;30:65–77.
78. Nüchter A, Lingemann K, Hertzberg J, Surmann H. 6D SLAM—3D mapping outdoor environments. J Field Robot. 2007;24:699–722.
79. Huang GP, Mourikis AI, Roumeliotis SI. A quadratic-complexity observability-constrained unscented Kalman filter for SLAM. IEEE Trans Robot. 2013;29:1226–43.
80. Klein G, Murray D. Parallel tracking and mapping for small AR workspaces. In: Proceedings of the 6th IEEE and ACM international symposium on mixed and augmented reality. Nara: IEEE; 2007. p. 225–34.
81. Klein G, Murray D. Improving the agility of keyframe-based SLAM. In: Forsyth D, Torr P, Zisserman A, editors. Computer vision – ECCV 2008. Berlin, Heidelberg: Springer; 2008. p. 802–15.
82. Klein G, Murray D. Parallel tracking and mapping on a camera phone. In: Proceedings of the 8th IEEE international symposium on mixed and augmented reality. Orlando: IEEE; 2009. p. 83–6.
83. Dong ZL, Zhang GF, Jia JY, Bao HJ. Keyframe-based real-time camera tracking. In: Proceedings of the 12th international conference on computer vision. Kyoto: IEEE; 2009. p. 1538–45.
84. Dong ZL, Zhang GF, Jia JY, Bao HJ. Efficient keyframe-based real-time camera tracking. Comput Vis Image Underst. 2014;118:97–110.
85. Salas-Moreno RF, Newcombe RA, Strasdat H, Kelly PHJ, Davison AJ. SLAM++: simultaneous localisation and mapping at the level of objects. In: Proceedings of 2013 IEEE conference on computer vision and pattern recognition. Portland: IEEE; 2013. p. 1352–9.
86. Tan W, Liu HM, Dong ZL, Zhang GF, Bao HJ. Robust monocular SLAM in dynamic environments. In: Proceedings of 2013 IEEE international symposium on mixed and augmented reality. Adelaide: IEEE; 2013. p. 209–18.
87. Feng YJ, Wu YH, Fan LX. On-line object reconstruction and tracking for 3D interaction. In: Proceedings of 2012 IEEE international conference on multimedia and expo. Melbourne: IEEE; 2012. p. 711–6.
88. Mur-Artal R, Montiel JMM, Tardos JD. ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans Robot. 2015;31:1147–63.
89. Bourmaud G, Mégret R. Robust large scale monocular visual SLAM. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition. Boston: IEEE; 2015. p. 1638–47.
90. Newcombe RA, Lovegrove SJ, Davison AJ. DTAM: dense tracking and mapping in real-time. In: Proceedings of 2011 international conference on computer vision. Barcelona: IEEE; 2011. p. 2320–7.
91. Engel J, Sturm J, Cremers D. Semi-dense visual odometry for a monocular camera. In: Proceedings of 2013 IEEE international conference on computer vision. Sydney: IEEE; 2013. p. 1449–56.
92. Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer vision – ECCV 2014. Cham: Springer; 2014. p. 834–49.
93. Pascoe G, Maddern W, Newman P. Direct visual localisation and calibration for road vehicles in changing city environments. In: Proceedings of 2015 IEEE international conference on computer vision workshop. Santiago: IEEE; 2015. p. 98–105.
94. Schöps T, Engel J, Cremers D. Semi-dense visual odometry for AR on a smartphone. In: Proceedings of 2014 IEEE international symposium on mixed and augmented reality. Munich: IEEE; 2014. p. 145–50.
95. Konolige K, Agrawal M. FrameSLAM: from bundle adjustment to real-time visual mapping. IEEE Trans Robot. 2008;24:1066–77.
96. Mei C, Sibley G, Cummins M, Newman P, Reid I. A constant time efficient stereo SLAM system. In: Cavallaro A, Prince S, Alexander D, editors. Proceedings of the British machine vision conference. Nottingham: BMVA; 2009. p. 54.1–54.11.
97. Zou DP, Tan P. COSLAM: collaborative visual slam in dynamic environments. IEEE Trans Pattern Anal Mach Intell. 2013;35:354–66.
98. Engel J, Stückler J, Cremers D. Large-scale direct SLAM with stereo cameras. In: Proceedings of 2015 IEEE/RSJ international conference on intelligent robots and systems. Hamburg: IEEE; 2015. p. 1935–42.
99. Pire T, Fischer T, Civera J, De Cristóforis P, Berlles JJ. Stereo parallel tracking and mapping for robot localization. In: Proceedings of 2015 IEEE/RSJ international conference on intelligent robots and systems. Hamburg: IEEE; 2015. p. 1373–8.
100. Moreno FA, Blanco JL, Gonzalez-Jimenez J. A constant-time SLAM back-end in the continuum between global mapping and submapping: application to visual stereo SLAM. Int J Robot Res. 2016;35:1036–56.
101. Mur-Artal R, Tardós JD. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Trans Robot. 2017;33:1255–62.
102. Zhang GX, Lee JH, Lim J, Suh IH. Building a 3-D line-based map using stereo SLAM. IEEE Trans Robot. 2015;31:1364–77.
103. Gomez-Ojeda R, Zuñiga-Noël D, Moreno FA, Scaramuzza D, Gonzalez-Jimenez J. PL-SLAM: a stereo SLAM system through the combination of points and line segments. arXiv: 1705.09479, 2017.
104. Usenko V, Engel J, Stückler J, Cremers D. Direct visual-inertial odometry with stereo cameras. In: Proceedings of 2016 IEEE international conference on robotics and automation. Stockholm: IEEE; 2016. p. 1885–92.
105. Wang R, Schwörer M, Cremers D. Stereo DSO: large-scale direct sparse visual odometry with stereo cameras. In: Proceedings of 2017 IEEE international conference on computer vision. Venice: IEEE; 2017.
106. Forster C, Zhang ZC, Gassner M, Werlberger M, Scaramuzza D. SVO: semidirect visual odometry for monocular and multicamera systems. IEEE Trans Robot. 2017;33:249–65.
107. Sun K, Mohta K, Pfrommer B, Watterson M, Liu SK, Mulgaonkar Y, et al. Robust stereo visual inertial odometry for fast autonomous flight. arXiv: 1712.00036, 2017.
108. Konolige K, Agrawal M, Solà J. Large-scale visual odometry for rough terrain. In: Kaneko M, Nakamura Y, editors. Robotics research. Berlin, Heidelberg: Springer; 2010. p. 201–2.
109. Weiss S, Achtelik MW, Lynen S, Chli M, Siegwart R. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments. In: Proceedings of 2012 IEEE international conference on robotics and automation. Saint Paul: IEEE; 2012. p. 957–64.
110. Falquez JM, Kasper M, Sibley G. Inertial aided dense & semi-dense methods for robust direct visual odometry. In: Proceedings of 2016 IEEE/RSJ international conference on intelligent robots and systems. Daejeon: IEEE; 2016. p. 3601–7.
111. Mourikis AI, Roumeliotis SI. A multi-state constraint Kalman filter for vision-aided inertial navigation. In: Proceedings of 2007 IEEE international conference on robotics and automation. Roma: IEEE; 2007. p. 3565–72.
112. Li MY, Mourikis AI. High-precision, consistent EKF-based visual-inertial odometry. Int J Robot Res. 2013;32:690–711.
113. Li MY, Kim BH, Mourikis AI. Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera. In: Proceedings of 2013 IEEE international conference on robotics and automation. Karlsruhe: IEEE; 2013. p. 4712–9.
114. Li MY, Mourikis AI. Vision-aided inertial navigation with rolling-shutter cameras. Int J Robot Res. 2014;33:1490–507.
115. Clement LE, Peretroukhin V, Lambert J, Kelly J. The battle for filter supremacy: a comparative study of the multi-state constraint Kalman filter and the sliding window filter. In: Proceedings of the 12th conference on computer and robot vision. Halifax: IEEE; 2015. p. 23–30.
116. Bloesch M, Omari S, Hutter M, Siegwart R. Robust visual inertial odometry using a direct EKF-based approach. In: Proceedings of 2015 IEEE/RSJ international conference on intelligent robots and systems. Hamburg: IEEE; 2015. p. 298–304.

117. Forster C, Carlone L, Dellaert F, Scaramuzza D. On-manifold preintegration for real-time visual-inertial odometry. IEEE Trans Robot. 2017;33:1–21.
118. Leutenegger S, Lynen S, Bosse M, Siegwart R, Furgale P. Keyframe-based visual–inertial odometry using nonlinear optimization. Int J Robot Res. 2015; 34:314–34.
119. Li PL, Qin T, Hu BT, Zhu FY, Shen SJ. Monocular visual-inertial state estimation for mobile augmented reality. In: Proceedings of 2017 IEEE international symposium on mixed and augmented reality. Nantes: IEEE; 2017. p. 11–21.
120. Mur-Artal R, Tardós JD. Visual-inertial monocular SLAM with map reuse. IEEE Robot Autom Lett. 2017;2:796–803.
121. Tateno K, Tombari F, Laina I, Navab N. CNN-SLAM: real-time dense monocular SLAM with learned depth prediction. In: Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu: IEEE; 2017.
122. Ummenhofer B, Zhou HZ, Uhrig J, Mayer N, Ilg E, Dosovitskiy A, et al. DeMoN: depth and motion network for learning monocular stereo. In: Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu: IEEE; 2017.
123. Vijayanarasimhan S, Ricco S, Schmid C, Sukthankar R, Fragkiadaki K. SfM-Net: learning of structure and motion from video arXiv: 1704.07804, 2017.
124. Zhou TH, Brown M, Snavely N, Lowe DG. Unsupervised learning of depth and ego-motion from video. In: Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu: IEEE; 2017.
125. Li RH, Wang S, Long ZQ, Gu DB. UnDeepVO: monocular visual odometry through unsupervised deep learning. arXiv: 1709.06841, 2017.
126. Clark R, Wang S, Wen HK, Markham A, Trigoni N. VINet: visual-inertial odometry as a sequence-to-sequence learning problem. In: Proceedings of 31st AAAI conference on artificial intelligence. San Francisco: AAAI; 2017. p. 3995–4001.
127. DeTone D, Malisiewicz T, Rabinovich A. Toward geometric deep SLAM. arXiv:1707.07410, 2017.
128. Gao X, Zhang T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system. Auton Robot. 2017;41:1–18.
129. Turan M, Almalioglu Y, Araujo H, Konukoglu E, Sitti M. Deep EndoVO: a recurrent convolutional neural network (RCNN) based visual odometry approach for endoscopic capsule robots. Neurocomputing. 2018;275:1861–70.
130. Kuipers B, Byun YT. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. Rob Auton Syst. 1991;8:47–63.
131. Ulrich I, Nourbakhsh I. Appearance-based place recognition for topological localization. In: Proceedings of 2000 ICRA. Millennium conference. IEEE international conference on robotics and automation. Symposia proceedings. San Francisco: IEEE; 2000. p. 1023–9.
132. Choset H, Nagatani K. Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization. IEEE Trans Robot Autom. 2001;17:125–37.
133. Kuipers B, Modayil J, Beeson P, MacMahon M, Savelli F. Local metrical and global topological maps in the hybrid spatial semantic hierarchy. In: Proceedings of 2004 IEEE international conference on robotics and automation. New Orleans: IEEE; 2004. p. 4845–51.
134. Chang HJ, Lee CSG, Lu YH, Hu YC. P-SLAM: simultaneous localization and mapping with environmental-structure prediction. IEEE Trans Robot. 2007; 23:281–93.
135. Blanco JL, FernÁndez-Madrigal JA, GonzÁlez J. Toward a unified Bayesian approach to hybrid metric-topological SLAM. IEEE Trans Robot. 2008;24:259–70.
136. Blanco JL, González J, Fernández-Madrigal JA. Subjective local maps for hybrid metric-topological SLAM. Rob Auton Syst. 2009;57:64–74.
137. Kawewong A, Tongprasit N, Hasegawa O. PIRF-Nav 2.0: fast and online incremental appearance-based loop-closure detection in an indoor environment. Rob Auton Syst. 2011;59:727–39.
138. Sünderhauf N, Protzel P. Switchable constraints for robust pose graph SLAM. In: Proceedings of 2012 IEEE/RSJ international conference on intelligent robots and systems. Vilamoura: IEEE; 2012. p. 1879–84.
139. Latif Y, Cadena C, Neira J. Robust loop closing over time for pose graph SLAM. Int J Robot Res. 2013;32:1611–26.
140. Latif Y, Cadena C, Neira J. Robust graph SLAM back-ends: a comparative analysis. In: Proceedings of 2014 IEEE/RSJ international conference on intelligent robots and systems. Chicago: IEEE; 2014. p. 2683–90.
141. Vallvé J, Solà J, Andrade-Cetto J. Graph SLAM sparsification with populated topologies using factor descent optimization. IEEE Robot Autom Lett. 2018; 3:1322–9.
142. Gatrell LB, Hoff WA, Sklair CW. Robust image features: concentric contrasting circles and their image extraction. In: Proceedings of SPIE volume 1612, cooperative intelligent robotics in space II, vol. 1612. Boston: SPIE; 1992. p. 235–45.
143. Cho YK, Lee J, Neumann U. A multi-ring color fiducial system and a rule-based detection method for scalable fiducial-tracking augmented reality. In: Proceedings of international workshop on augmented reality. Atlanta: International Workshop on Augmented Reality; 1998.
144. Knyaz VA, Head of Group, Sibiryakov RV. The development of new coded targets for automated point identification and non-contact surface measurements. In: 3D surface measurements, international archives of photogrammetry and remote sensing; 1998.
145. Kato H, Billinghurst M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd IEEE and ACM international workshop on augmented reality. San Francisco: IEEE; 1999. p. 85–94.
146. Naimark L, Foxlin E. Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. In: International symposium on mixed and augmented reality. Darmstadt: IEEE; 2002. p. 27–36.
147. Ababsa FE, Mallem M. Robust camera pose estimation using 2d fiducials tracking for real-time augmented reality systems. In: Proceedings of ACM SIGGRAPH international conference on virtual reality continuum and its applications in industry. Singapore: ACM; 2004. p. 431–5.
148. Claus D, Fitzgibbon AW. Reliable automatic calibration of a marker-based position tracking system. In: Proceedings of the 7th IEEE workshops on applications of computer vision. Breckenridge: IEEE; 2005. p. 300–5.
149. Fiala M. ARTag, a fiducial marker system using digital techniques. In: Proceedings of 2005 IEEE computer society conference on computer vision and pattern recognition. San Diego: IEEE; 2005. p. 590–6.
150. Fiala M. Designing highly reliable fiducial markers. IEEE Trans Pattern Anal Mach Intell. 2010;32:1317–24.
151. Maidi M, Didier JY, Ababsa F, Mallem M. A performance study for camera pose estimation using visual marker based tracking. Mach Vis Appl. 2010;21:365–76.
152. Bergamasco F, Albarelli A, Cosmo L, Rodola E, Torsello A. An accurate and robust artificial marker based on cyclic codes. IEEE Trans Pattern Anal Mach Intell. 2016;38:2359–73.
153. DeGol J, Bretl T, Hoiem D. ChromaTag: a colored marker and fast detection algorithm. In: Proceedings of 2017 IEEE international conference on computer vision. Venice: IEEE; 2017.
154. Muñoz-Salinas R, Marín-Jimenez MJ, Yeguas-Bolivar E, Medina-Carnicer R. Mapping and localization from planar markers. Pattern Recogn. 2018;73:158–71.
155. Eade E, Drummond T. Monocular SLAM as a graph of coalesced observations. In: Proceedings of the 11th international conference on computer vision. Rio de Janeiro: IEEE; 2007. p. 1–8.
156. Wu Y. Design and lightweight method of a real time and online camera localization from circles: CN, 201810118800.1. 2018.
157. Nister D. An efficient solution to the five-point relative pose problem. IEEE Trans Pattern Anal Mach Intell. 2004;26:756–70.
158. Lee GH, Pollefeys M, Fraundorfer F. Relative pose estimation for a multi-camera system with known vertical direction. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 540–7.
159. Kneip L, Li HD. Efficient computation of relative pose for multi-camera systems. In: Proceedings of 2014 IEEE conference on computer vision and pattern recognition. Columbus: IEEE; 2014. p. 446–53.
160. Chatterjee A, Govindu VM. Efficient and robust large-scale rotation averaging. In: Proceedings of 2013 IEEE international conference on computer vision. Sydney: IEEE; 2013. p. 521–8.
161. Ventura J, Arth C, Lepetit V. An efficient minimal solution for multi-camera motion. In: Proceedings of 2015 IEEE international conference on computer vision. Santiago: IEEE; 2015. p. 747–55.
162. Fredriksson J, Larsson V, Olsson C. Practical robust two-view translation estimation. In: Proceedings of 2015 IEEE conference on computer vision and pattern recognition. Boston: IEEE; 2015. p. 2684–90.
163. Park M, Luo JB, Collins RT, Liu YX. Estimating the camera direction of a geotagged image using reference images. Pattern Recogn. 2014;47:2880–93.
164. Carlone L, Tron R, Daniilidis K, Dellaert F. Initialization techniques for 3D SLAM: a survey on rotation estimation and its use in pose graph optimization. In: Proceedings of 2015 IEEE international conference on robotics and automation. Seattle: IEEE; 2015. p. 4597–604.
165. Jiang NJ, Cui ZP, Tan P. A global linear method for camera pose registration. In: Proceedings of 2013 IEEE international conference on computer vision. Sydney: IEEE; 2013. p. 481–8.

166. Cui ZP, Tan P. Global structure-from-motion by similarity averaging. In: Proceedings of 2015 IEEE international conference on computer vision. Santiago: IEEE; 2015. p. 864–72.
167. Cui ZP, Jiang NJ, Tang CZ, Tan P. Linear global translation estimation with feature tracks. In: Xie XH, Jones MW, Tam GKL, editors. Proceedings of the 26th British machine vision conference. Nottingham: BMVA; 2015. p. 46.1–46.13.
168. Cui HN, Gao X, Shen SH, Hu ZY. HSfM: hybrid structure-from-motion. In: Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu: IEEE; 2017. p. 2393–402.
169. Cui HN, Shen SH, Gao X, Hu ZY. CSFM: community-based structure from motion. In: Proceedings of 2017 IEEE international conference on image processing. Beijing: IEEE; 2017. p. 4517–21.
170. Zhu SY, Shen TW, Zhou L, Zhang RZ, Wang JL, Fang T, et al. Parallel structure from motion from local increment to global averaging. arXiv: 1702.08601, 2017.
171. Ozyesil O, Voroninski V, Basri R, Singer A. A survey on structure from motion. arXiv: 1701.08493, 2017.
172. Dai YC, Li HD, Kneip L. Rolling shutter camera relative pose: generalized epipolar geometry. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE; 2016. p. 4132–40.
173. Albl C, Kukelova Z, Pajdla T. Rolling shutter absolute pose problem with known vertical direction. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE; 2016. p. 3355–63.
174. Kim JH, Latif Y, Reid I. RRD-SLAM: radial-distorted rolling-shutter direct SLAM. In: IEEE international conference on robotics and automation. IEEE: Singapore; 2017.
175. Gallego G, Lund JEA, Mueggler E, Rebecq H, Delbruck T, Scaramuzza D. Event-based, 6-DOF camera tracking from photometric depth maps. IEEE Trans Pattern Anal Mach Intell. 2017; doi:https://doi.org/10.1109/TPAMI.2017.2769655
176. Vidal AR, Rebecq H, Horstschaefer T, Scaramuzza D. Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. IEEE Robot Autom Lett. 2018;3:994–1001.
177. Rebecq H, Horstschaefer T, Scaramuzza D. Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization. In: British machine vision conference. London: BMVA; 2017.
178. Abouzahir M, Elouardi A, Latif R, Bouaziz S, Tajer A. Embedding SLAM algorithms: has it come of age? Rob Auton Syst. 2018;100:14–26.

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com