Progress and Applications of Visual SLAM

Article · June 2018

DOI: 10.11947/j.AGCS.2018.20170652

测绘出版社
SURVEYING AND MAPPING PRESS

Progress and Applications of Visual SLAM

Kaichang DI[1], Wenhui WAN[1], Hongying ZHAO[2], Zhaoqin LIU[1], Runzhi WANG[1], Feizhou ZHANG[2]

1. *State Key Laboratory of Remote Sensing Science*, *Institute of Remote Sensing and Digital Earth*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*; 2. *Institute of Remote Sensing and Geographic Information System*, *Peking University*, *Beijing* 100871, *China*

Abstract: Visual simultaneous localization and mapping (SLAM) provides mapping and self-localization results for a robot in an unknown environment based on visual sensors, that have the advantages of small volume, low power consumption, and richness of information acquisition. Visual SLAM is essential and plays a significant role in supporting automated and intelligent applications of robots. This paper presents the key techniques of visual SLAM, summarizes the current research status, and analyses the new trends of visual SLAM research and development. Finally, specific applications of visual SLAM in restricted environments, including deep space and indoor scenarios, are discussed.

Key words: visual SLAM; feature extraction; Kalman filter; graph based optimization; loop closure detection

Citation: Kaichang DI, Wenhui WAN, Hongying ZHAO, et al. Progress and Applications of Visual SLAM[J]. Journal of Geodesy and Geoinformation Science, 2019, 2(2): 38-49.

1 Introduction

In the field of robotic automation, simultaneous localization and mapping (SLAM) is the key technology for robots to operate autonomously in unknown environments[1]. Simultaneous localization and mapping can build a map of the environment and simultaneously provide the robot with its position on the map based on data acquired from external sensors onboard the robot. The incremental construction of an environment map and the continuous positioning of a robot are the basis for robot environment perception and automation. Laser range finders are typically used as data sources for environmental perception[2]. Compared with radar, sonar, and other range-finding devices, the visual sensor has the advantages of small volume, low power consumption, and richness of information acquisition. Visual SLAM has recently received increasing attention because of the provision of abundant external environment texture information for robots[3]. Because of image degeneration caused by sensor noise, environment lighting changes, or rapid movement, visual SLAM still faces numerous challenges and, therefore, requires more advanced methods. Currently, with the development of computer vision technology, the technical merit of visual SLAM has also advanced, and it has been applied in indoor autonomous navigation, virtual reality/augmented reality (VR/AR), and other fields[4-7].

2 Key Technologies for Visual SLAM

Visual SLAM is based on sequences of images obtained by a camera. From these images the geometric relationship between the surrounding environment and the camera can be determined by combining the image information and the camera model. With the camera moving incrementally, SLAM can determine the surrounding environment map and the position of the camera in the environment map. The typical visual SLAM process can be divided into front-end processing, back-end processing, and loop closure

Received date: 2018-10-05; accepted date: 2019-03-11
Foundation support: The National Key Research and Development Program of China (2016YFB0502102); The National Natural Science Foundation of China (41471388)
First author: Kaichang DI(1967—), male, PhD, research fellow, PhD supervisor, majors in planetary mapping, visual localization and navigation.
E-mail: dikc@radi.ac.cn
Corresponding author: Feizhou ZHANG
E-mail: zhangfz@pku.edu.cn

detection[8-9], as shown in Figure 1. Front-end processing is responsible for associating sequences of images with environmental landmarks and initializing the parameters. Currently, the traditional method is to achieve the corresponding points tracking by feature extraction and matching on sequences of images. The observation of the corresponding points on the sequences of images are then associated with the environmental landmarks and the state parameters of the system are initialized. This is the required procedure for incremental map construction and independent continuous positioning. The adaptability of the front-end processing algorithm directly determines the robustness of the visual SLAM method[10]. Back-end processing is responsible for the estimation optimization for the environmental map results and positional parameters of the observation data in order to obtain high-precision positioning and mapping results[11-12]. Loop closure detection is the process of determining whether the observed environmental landmarks have been observed in the SLAM system. It is the basis of constructing loop closure constraints to eliminate error accumulation after moving for long distances[13-14]. The above three components have successively completed the data association, environmental map, positional parameters estimation, and loop closure detection in visual SLAM.

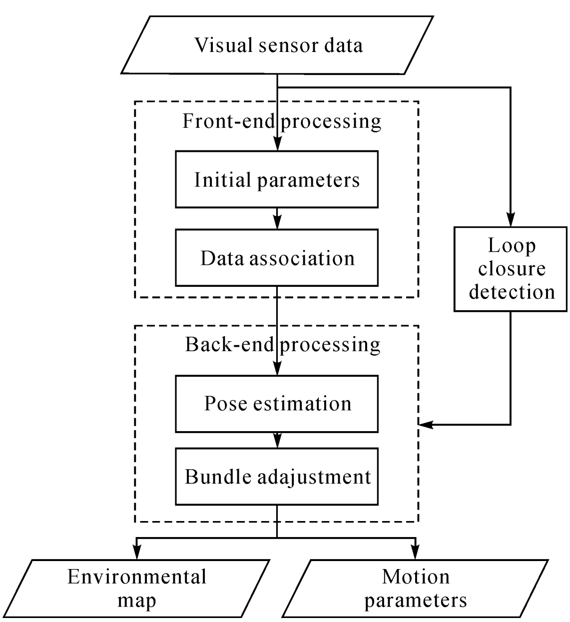

Fig.1 Flowchart of typical visual SLAM

2.1 Feature points extraction and matching

In visual SLAM, we need to associate the observation information of the image with the environment, that is, to determine the corresponding relationship between the content of sequences of images and the actual environment. Corner features are frequently used to associate sequences of images in classic visual SLAM methods. Through feature point extraction and matching between images, the relationship between space object points and the corresponding image points is formed among multiple frames of images. The appearance of the corresponding image points on the sequence images may change because of different positions and perspectives of the camera and the ambient lighting variations, that requires the expression of feature points to be independent of lighting changes and geometric changes such as rotation, zoom, and tilt. Previously, extraction of image feature points was based primarily on Harris[15], Förstner[16], and other corner feature extraction operators, while template matching[17,18-20] or optical flow tracking methods[21-22] were applied to feature tracking. The above methods were relatively effective when the perspective of adjacent images changed marginally. However, it is difficult to obtain robust tracking results, and can even lead to tracking failure, if the camera moves irregularly, causing significant perspective distortion. With the development of invariant feature description methods, such as scale-invariant feature transform (SIFT), the extraction and matching of image feature points can be performed with limited image deformation and illumination changes, that improves the applicability of visual SLAM in complex environments[23-25]. However, the significant computation requirements of SIFT limits the efficiency of positioning and mapping, and it is difficult to satisfy the real-time requirements. To improve processing efficiency, researchers have successively developed SURF[26-27], CensurE[28], BRISK[29], ORB[30], and other feature description methods. Although the performances of these operators have marginally decreased, their efficiencies have been improved by several orders of magnitude, making real-time visual SLAM a reality[31-35].

2.2 Environmental map and positional parameters estimation

There are frequently noise and errors in environmental perception data. In order to obtain high-precision results, it is necessary to estimate the entire system state and its uncertainty from noisy data. The probability theory has been used to analyze the uncertainty distribution of the environment of a robot and its positional information. Based on Bayesian rules, the state probability model of the system is constructed, and the optimal estimation of the environmental information and pose parameters are realized by combining the robot motion information and environmental observation data[36-39]. Kalman filtering (KF), a method for optimal estimation of linear systems based on the principle of minimizing the root mean square error[40], was first introduced into SLAM for the optimal estimation of the system state[41-42]. Because of the nonlinear characteristics of a SLAM system, the linear approximation of the system needs to be performed with a Taylor series expansion based on the extended KF (EKF) method[43]. In early SLAM studies, the EKF-based parameter estimation method was frequently used[44-45]. However, if the system is highly nonlinear, the linear approximation error may be great, leading to a significant decrease in EKF filtering performance. Subsequently, a number of improved filtering methods have been proposed, including unscented KF[46], cubature KF[47], and central difference KF[48]. These filtering methods can obtain better approximation precision than EKF, but still require linear calculations. Particle filtering is based on the Monte Carlo method, that approximates the probability density function through random samples propagated through the state space[49]. Particle filtering overcomes the limitations of the Gaussian distribution hypothesis and has been successfully applied in visual SLAM[50-53]. However, this type of parameter estimation method through filtering is based on the Markov hypothesis of the system, and it fails to make optimal use of all observation data. In recent years, the graph optimization method has received extensive attention in the study of SLAM. The graph optimization method estimates all the observed data and relinearizes the approximation when the system estimation changes, reducing the linearization errors. It has become the mainstream environmental mapping and positional parameters estimation method. In fact, the graph optimization method re-estimates all of the state variables in each optimization, and results in excessive computations in large scenes. Existing studies primarily decrease the computations by minimizing the number of optimization iterations[54-56] and analyzing the sparsity of the state matrix[57-59].

2.3 Loop closure detection

When the constraint of external control information is absent in the visual SLAM system, the system uncertainty increases gradually, and the positioning error accumulates after extended periods of motion. In practical applications, the camera may be moved to the previous position to form a closed loop constraint. By introducing this constraint into graph optimization, globally consistent positioning results with fewer errors can be obtained[60]. Therefore, loop closure detection is of significant importance to visual SLAM for movement over long distances. The basis of loop closure detection is to judge the correspondence between the contents of the current and previous images. Loop closure detection methods conduct feature matching between the current image and randomly selected historical images, in order to determine whether there exists an association according to the correct number of feature matches. Such methods are based on the principle that all previous images could be related to the current image. With increasing numbers of previous images, the computations increase significantly[61]. Subsequent studies have improved the detection efficiency by making preliminary judgments on the relevant possibility of previous images. Among them, one method relies on initial positioning results to determine whether the camera returns to its previous position, and is used for the possible loop closure determination[62]. However, this method is not effective with significant accumulated positioning errors over an extended period of time. Another method is to construct the bag-of-words of image descriptions based on image content. Through the combination of "words" appearing on the image, the word vector describing the whole image is used for the judgement of the possible correlation relationship by the calculation of degrees of similarity between images[63-65]. Regardless of feature distribution, the judgement of correlation relies solely on the number of corresponding features, that is more robust in practical applications.

3 New Trends in Visual SLAM Research

Currently, significant advances have been made in visual SLAM research to satisfy the application requirements of localization and mapping in a number of simple scenarios in terms of precision and efficiency. The acquisition of visual information is always affected by the richness of the environmental texture. However, changes in environmental lighting and unstable motion make visual information processing more complex. It is highly desirable to develop robust visual SLAM methods for complex environments. Recently, a number of researchers have investigated ways to obtain increased image information, and employed other types of sensors for data fusion, in order to develop more robust SLAM methods that could be applied to accurate localization and mapping in complex environments.

3.1 Extraction and matching of multiple visual features

In complex environments, such as indoor applications, there are more artificial objects and linear features in the scenes. The point feature extraction method alone is not able to make full use of the visual information of the images. In addition to the point feature, the line feature is often introduced into the visual SLAM to achieve more robust localization and mapping[66-69]. Because of the strong directional constraint of line features, they can effectively decrease the accumulation of errors caused by the reduction of adjacent image overlapping areas resulting from rapid camera movements, such as turning[70]. If the spatial plane features formed by multiple lines are taken into consideration for the optimal estimation of a system, greater localization accuracy can be achieved[71]. The addition of new types of visual features requires new multi-feature tracking and optimization models, so as to accurately estimate the uncertainty of these features.

3.2 Direct SLAM

The feature extraction and matching algorithms in visual SLAM are complex and time-consuming, and only partial image information can be used, resulting in more texture information being discarded. Methods of direct SLAM use the intensity change information of the image itself directly and skip the steps of feature extraction and matching. They take the minimized photometry error between images as the criterion to implement optimization estimation of state parameters, that can make full use of image information for localization and mapping. Improved SLAM results can also be obtained in a number of poor texture areas[72]. Unlike the feature-based visual SLAM methods, direct SLAM methods can not only generate sparse maps[73], but also semi-dense[74], and dense maps[75]. Because of the neglection of feature extraction and matching, direct SLAM methods can be used in scenes with significant real-time requirements and limited computing resources. The assumption that the grayscale of the same area between images remains unchanged is the basis of direct SLAM methods. However, the grayscale is affected by environmental lighting, camera exposure, and other factors, specifically in scenes with significant lighting changes. These factors limit the applications of direct SLAM methods in complex environments.

3.3 Multi-sensor fusion SLAM

The visual SLAM methods rely on the environmental texture information obtained by the camera, and processing results are directly affected by the environmental texture conditions. Therefore, it is difficult to perform reliably with poorly textured scenes. In order to achieve robust SLAM results in complex environments and improve algorithm performance in practical applications, multi-sensor data fusion has been employed, that compensates their respective defects to improve the robustness and accuracy of visual SLAM. Inertial measurement units (IMU) have fully independent measurement characteristics and can obtain stable localization and pose data. It can be applied to the localization and mapping in complex environments without the support of texture information, although it may accumulate significant drift errors over long distances. High-precision visual localization results could then effectively correct the rapid drift errors of the IMU and improve the SLAM accuracy[76]. Therefore, IMUs have frequently been used for fusion with cameras[77-80]. Laser rangefinders also have the advantages of small size and low power consumption[81]. They can be integrated with a single camera in order to determine the scale of the environmental map model and effectively correct the scale drift error of localization. Researchers have designed a camera, with an integrated laser rangefinder, that is portable and wearable as a navigation system to meet the autonomous localization needs of rescue workers, astronauts, and other personnel[82-84]. In recent years, the Kinect depth camera has attracted wide attention for its ability to directly obtain rich three-dimensional (3D) spatial information. Based on the measurement principles of structural light or time of flight, such sensors can actively measure spatial 3D information and generate local 3D maps directly without being affected by environmental conditions[85]. Through the integration of a depth camera and an RGB camera, the RGB-D camera is used in SLAM for the simultaneous capture of the environmental texture and dense spatial geometric information. In addition, the 3D spatial information from the depth camera improves the quality of the SLAM observation data and the robustness and accuracy of SLAM[86-89]. Although the introduction of multi-sensors enhances the richness of environmental perception data, new sensors also introduce more error sources. In order to obtain the optimal estimation results using multi-sensor fusion-based SLAM, a multi-sensor optimal estimation model needs to be built on the basis of the error analysis of each data source.

3.4 Visual SLAM based on deep learning

In the field of artificial intelligence, deep learning has recently exceeded the performance of traditional machine learning methods, and has been increasingly applied in image object recognition, voice recognition, and other aspects[90-92]. Recently, a number of researchers have attempted to introduce deep learning into visual SLAM by using the high-level features learned from deep neural networks for pose estimation between image frames[93-95] and loop closure detection[96-98]. These researchers employed the end-to-end deep learning approaches for motion estimation without feature matching and complex geometric processing. In loop closure detection, the strong recognition ability of deep learning is used to extract the robust high-level features of images. Extracted deep features can better adapt to changes of perspective and lighting, resulting in improved loop closure recognition ability. Currently, deep learning is exhibiting great potential for improving the robustness of visual SLAM algorithms. It decreases the application limitations brought by empirical feature design in traditional methods. However, deep learning requires a great number of training sample datasets, and the working scenes should be close to the sample data, otherwise the performance will decrease significantly. These problems result in the overall performance of deep learning-based visual SLAM not surpassing that of traditional methods. In the future, with the continuous development of deep learning theories and methods, it will play a critical role in visual SLAM.

4 Typical Applications of Visual SLAM

Currently, with the popularization and application of satellite positioning and navigation technology, the corresponding positioning and navigation service can be obtained through satellite receivers. However, in GNSS-denied environments, such as deep space, underground, or indoors, the localization and environmental perception of robots still face significant challenges. As an important method of autonomous localization and environmental perception, visual SLAM can provide basic data for autonomous robot operation in such environments. This section summarizes and discusses the current research on visual SLAM methods applied to deep space exploration, indoor localization, and navigation in large scenes.

4.1 Rover localization and navigation in deep space exploration

In the late 1950s, the United States and the former Soviet Union began to explore deep space. Since then, a significant number of space probes have been launched for exploration of various celestial bodies, including rovers travelling on the surface of the moon and Mars. Because of the complexity of the surface environments, planet rovers require significant automation and intelligence to overcome the challenges of long distance traversing and operating. The early exploration rovers could only perform simple observation tasks over a small range. With the expansion of the traversing distances and the increasing requirements for complex tasks, rovers require significant independent localization and environmental perception abilities to satisfy the needs of obstacle avoidance and task planning. The Mars rovers Spirit and Opportunity, launched in 2003, and the Curiosity rover, launched in 2011, were all equipped with stereo cameras for navigation. Accurate localization and mapping results were particularly critical for the rovers passing through rough areas, that played an essential role in the Mars rover exploration missions. The localization error of traversing by dead-reckoning was corrected by visual odometry and bundle adjustment techniques to provide reliable localization results to support obstacle avoidance and path planning[99-102]. The Chinese Chang'e-3 lander, launched in 2013, carried the Yutu rover to the lunar surface. Sequences of images, captured by the descent camera during landing, were used to assess the position of the landing point[103] and recreate the landing trajectory[104]. The descent images were also used to produce high-resolution (up to 5 cm) digital elevation models and digital orthophoto maps of the landing area, that were used as the base map for Chang'e-3 mission planning[105]. The Yutu rover was also equipped with a pair of stereo navigation cameras for environmental perception and accurate vision-based localization, as shown in Figure 2. The rover localization error from the guidance, navigation and control system was corrected by the visual localization method, and the error decreased from 7 % to 4 %[105-106] Figure 3 shows the Yutu rover traverse from the visual localization results. Because of the limited durations of deep space exploration missions, a rover will rarely return to the waypoints it has passed, therefore, it is difficult to form a loop closure constraint to correct the accumulated errors. Observations from high-resolution orbital and descent images can be incorporated into visual localization of the rover to reduce error accumulation over long-range traverses. In addition, it is necessary to further improve the robustness and accuracy of visual localization, so as to satisfy the demands of the new generation of deep space explorations, that will require greater automation and intelligence.

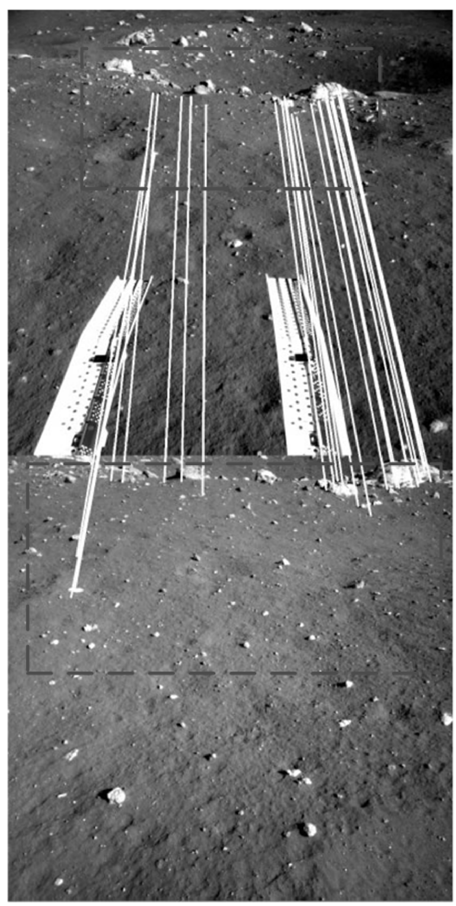

Fig.2 Matching results in cross-site navcam images of Yutu rover

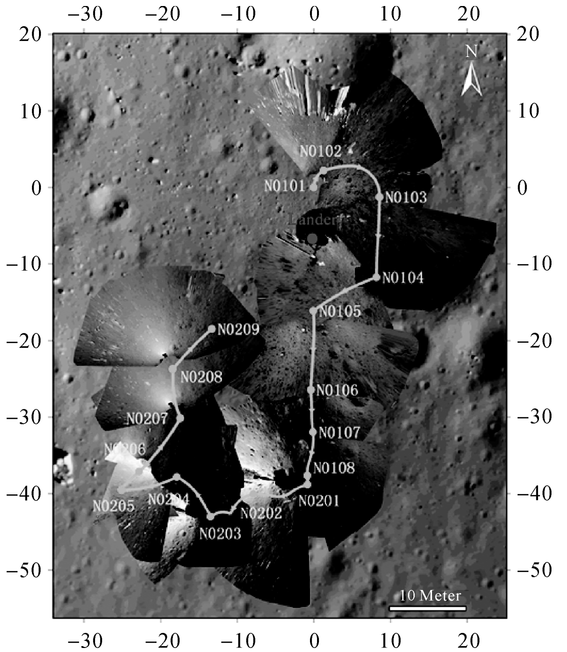

Fig.3 Traverse of Yutu rover based on visual localization

4.2 Indoor positioning and navigation

Currently, the intensity distribution estimation of wireless signals from Wi-Fi or Bluetooth base stations are commonly used in indoor locations[107-108]. However, the indoor positioning results based on wireless signals are affected by various signal noises and could not satisfy the application requirements of precise localization. Compared to wireless signals, indoor visual images contain rich and intuitive information. Visual SLAM can provide independent localization and navigation in areas not covered by wireless signals. It is not restricted by external base stations and has a wider environmental applicability. Although there have been numerous studies on indoor visual localization, only limited visual SLAM systems have been applied practically in indoor localization. Google Tango, released in 2014, was the first product for pedestrian perception in unknown indoor scenarios. By integrating an RGB-D camera and an IMU, Tango implemented dense 3D reconstruction of unknown environments and indoor localization[109], that can be applied to VR, AR, and other applications[100-112]. The HoloLens hybrid display helmet released by Microsoft adopted a similar technical route and added gesture recognition and voice recognition to realize the all-round interaction between people and the surrounding environment[113-114]. Although visual localization has the potential to satisfy the requirements of highly accurate indoor localization, current methods cannot fully cope with the high dynamics and complexity of indoor environments. In addition, large scale localization requires accurate indoor maps for the correction of localization errors. In the future, with the development of visual information processing, visual SLAM technology will play an increasingly critical role in the field of indoor positioning.

4.3 Autonomous navigation in large scenes

Recently, the popular unmanned aerial vehicles (UAV) have been increasingly applied in electric powerline inspection patrols, geological exploration, forest fire prevention, and other fields because of their good mobility, strong survivability, and long-range capabilities. These UAVs operate on preplanned mission routes, to implement efficient long-distance operations over large areas supported by GPS, BeiDou, and other GNSS systems. However, in complex unknown scenarios, the bulk of these UAVs are not capable of autonomous obstacle avoidance and path planning. The implementation of the flight mission relies on manual control by the operator, and they lack automation and intelligence. Recently, DJI-Innovations has introduced a binocular-vision sensor into the Phantom series UAVs. Based on the environmental map of visual SLAM, both the obstacle avoidance and path autonomous planning of UAVs are performed during the mission, and the ability of UAVs to survive and operate independently in a complex environment is improved[115]. Currently, with the emergence of satellite navigation signal interference and lure technology, military systems such as UAVs and missiles require full autonomous navigation and positioning capabilities in order to become independent of satellite navigation signals[116-117]. The rapid development of visual SLAM technology provides new technological means of support for autonomous navigation of aircraft over large areas. However, the rapid movement of aircraft cause an amount of degradation to the visual image, that also requires more robust real-time localization algorithms.

5 Conclusion

With the development of computer vision, digital image processing, and artificial intelligence the research and application of visual SLAM have advanced rapidly. However, visual SLAM is always affected by the condition of the environment texture and lighting, and the robustness of positioning and mapping in complex environment remains a challenge. On the one hand, present studies attempt to make full use of image information to extract as many image features as possible. On the other hand, the combination of depth cameras, IMUs, and other types of sensors is being explored to meet the demands of robust positioning mapping of visual SLAM in difficult conditions. Although the robustness of the current visual SLAM algorithm in complex scenarios needs to be improved, it has exhibited great potential in the provision of accurate and robust localization results. In future, with the improvement of SLAM technology, it will play an increasingly critical role in the field of robot automation and intelligence.

References

[1] LU Shaofang, LIU Dawei. A Survey of Research Situation on Navigation by Autonomous Mobile Robot and Its Related Techniques[J]. Transactions of the Chinese Society for Agricultural Machinery, 2002, 33(2): 112-116.

[2] ALBRECHT S. An Analysis of Visual Mono-SLAM[D]. Canada: Universität Osnabrück, 2009: 1-4. https://www.researchgate.net/publication/268031790_an_analysis_of_visual_mono-slam

[3] FUENTES-PACHECO J, RUIZ-ASCENCIO J, RENDÓN-MANCHA J M. Visual Simultaneous Localization and Mapping: A Survey[J]. Artificial Intelligence Review, 2015, 43(1): 55-81.DOI:10.1007/s10462-012-9365-8

[4] IDO J, SHIMIZU Y, MATSUMOTO Y, et al. Indoor Navigation for a Humanoid Robot Using a View Sequence[J]. The International Journal of Robotics Research, 2009, 28(2): 315-325.DOI:10.1177/0278364908095841

[5] ÇELIK K, SOMANI A K. Monocular Vision SLAM for Indoor Aerial Vehicles[J]. Journal of Electrical and Computer Engineering, 2013, 2013: 374165.

[6] COMPORT A I, MARCHAND E, PRESSIGOUT M. Real-time Markerless Tracking for Augmented Reality: The Virtual Visual Servoing Framework[J]. IEEE Transactions on Visualization and Computer Graphics, 2006, 12(4): 615-628. DOI:10.1109/TVCG.2006.78

[7] CHEKHLOV D, GEE A P, CALWAY a, et al. Ninja on a Plane: Automatic Discovery of Physical Planes for Augmented Reality Using Visual SLAM[C]// Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, Japan: IEEE, 2007: 13-16.

[8] CADENA C, CARLONE L, CARRILLO H, et al. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-perception Age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332. DOI: 10.1109/TRO.2016.2624754

[9] GAO Xiang, ZHANG Tao, LIU Yi, et al. Visual SLAM Fourteen Lectures-From Theory to Practice[M]. Bejing: China Machine Press, 2017: 17-22.

[10] KARLSSON N, DI BERNARDO E, OSTROWSKI J, et al. The vSLAM Algorithm for Robust Localization and Mapping [C]//Proceedings of 2005 IEEE International Conference on Robotics and Automation. Barcelona, Spain: IEEE, 2005.

[11] SVNDERHAUF N, PROTZEL P. Towards A Robust Back-end for Pose Graph SLAM[C]//Proceedings of 2012 IEEE International Conference on Robotics and Automation. Saint Paul, MN: IEEE, 2012.

[12] HU G, KHOSOUSSI K, HUANG Shoudong. Towards a Reliable SLAM Back-end[C]//Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE, 2013.

[13] NEWMAN P, HO K. SLAM-loop Closing with Visually Salient Features[C]//Proceedings of 2005 IEEE International Conference on Robotics and Automation. Barcelona, Spain: IEEE, 2005.

[14] HO K L, NEWMAN P. Loop Closure Detection in SLAM by Combining Visual and Spatial Appearance[J]. Robotics and Autonomous Systems, 2006, 54(9): 740-749.DOI:10.1016/j.robot.2006.04.016

[15] HARRIS C, STEPHENS M. A Combined Corner and Edge Detector [C] // Proceedings of the 4th Alvey Vision Conference. Manchester, UK: Alvety Vision Club, 1988.

[16] FÖRSTNER W, GVLCH E. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features[C]//Proceedings of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data. Interlaken, Switzerland: ISPRS, 1987.

[17] OLSON C F, MATTHIES L H, SCHOPPERS M, et al. Rover Navigation using Stereo Ego-motion[J]. Robotics and Autonomous Systems, 2003, 43(4): 215-229. DOI: 10.1016/S0921-8890(03)00004-6

[18] NISTÉR D, NARODITSKY O, BERGEN J. Visual Odometry for Ground Vehicle Applications [J]. Journal of Field Robotics, 2006, 23(1): 3-20. DOI: 10.1002/(ISSN)1556-4967.

[19] KIM J, KWEON I S. Robust Feature Matching for Loop Closing and Localization[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. San Diego, CA: IEEE, 2007.

[20] AGRAWAL M, KONOLIGE K. Real-time Localization in Outdoor Environments using Stereo Vision and Inexpensive GPS [C]//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China: IEEE, 2006.

[21] SHI Jianbo, TOMASI C. Good Features to Track[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA: IEEE, 1994.

[22] HERATH D C, KODAGODA S, DISSANAYAKE G. Simultaneous Localisation and Mapping: A Stereo Vision Based Approach[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Beijing, China: IEEE, 2006.

[23] LOWE D G. Object Recognition from Local Scale-Invariant Features[C]//Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 1150-1157.

[24] LOWE D G. Distinctive Image Features from Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[25] MOREL J M, YU Guoshen. Is SIFT Scale Invariant? [J]. Inverse Problems and Imaging, 2011, 5(1): 1-22.

[26] BAY H, TUYTELAARS T, GOOL V L. SURF: Speeded Up Robust Features[C]//Proceedings of the 9th European Conference on Compute Vision. Graz, Austria: Springer, 2006: 404-417.

[27] BAY H, ESS A, TUYTELAARS T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359. DOI: 10.1016/j.cviu.2007.09.014

[28] AGRAWAL M, KONOLIGE K, BLAS M R. CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching[C]//Proceedings of European Conference on Computer Vision. Marseille, France: INRIA Grenoble, 2008: 102-115.

[29] LEUTENEGGER S, CHLI M, SIEGWART R Y. BRISK: Binary Robust Invariant Scalable Keypoints[C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011.

[30] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An Efficient Alternative to SIFT or SURF[C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011.

[31] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System [J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.DOI:10.1109/TRO.2015.2463671

[32] GIL A, MOZOS O M, BALLESTA M, et al. A Comparative Evaluation of Interest Point Detectors and Local Descriptors for Visual SLAM[J]. Machine Vision and Applications, 2010, 21(6): 905-920. DOI:10.1007/s00138-009-0195-x

[33] HARTMANN J, KLUSSENDORFF J H, MAEHLE E. A Comparison of Feature Descriptors for Visual SLAM[C]//Proceedings of 2013 European Conference on Mobile Robots. Barcelona, Spain: IEEE, 2013.

[34] GAUGLITZ S, HÖLLERER T, TURK M. Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking [J]. International Journal of Computer Vision, 2011, 94 (3): 335-360.DOI:10.1007/s11263-011-0431-5

[35] MIKSIK O, MIKOLAJCZYK K. Evaluation of Local Detectors and Descriptors for Fast Feature Matching[C]//Proceedings of the 21st International Conference on Pattern Recognition. Tsukuba, Japan: IEEE, 2012.

[36] SMITH R C, CHEESEMAN P. On the Representation of Spatial Uncertainty [J]. International Journal of Robotics Research, 1986, 5(4): 56-68. DOI:10.1177/027836498600500404

[37] SMITH R, SELF M, CHEESEMAN P. Estimating Uncertain Spatial Relationships in Robotics[M]//COX I J, WILFONG G T. Autonomous Robot Vehicles. New York: Springer, 1990: 167-198.

[38] THRUN S, BURGARD W, FOX D. Probabilistic Robotics [M]. Massachusetts: MIT Press, 2005: 1-5.

[39] AULINAS J, PETILLOT Y R, SALVI J, et al. The SLAM Problem: A Survey[C]//Proceedings of the 2008 conference on Artificial Intelligence Research and Development: Proceedings of the 11th International Conference of the Catalan Association for Artificial Intelligence. Amsterdam, The Netherlands: IOS Press, 2008.

[40] KALMAN R E. A New Approach to Linear Filtering and Prediction Problems[J]. Journal of Basic Engineering, 1960, 82 (1): 35-45. DOI:10.1115/1.3662552

[41] CHATILA R, LAUMOND J. Position Referencing and Consistent World Modeling for Mobile Robots[C]//Proceedings of IEEE International Conference on Robotics and Automation. St. Louis, MO: IEEE, 1985.

[42] CROWLEY J L. World Modeling and Position Estimation for A Mobile Robot using Ultrasonic Ranging[C]//Proceedings of IEEE International Conference on Robotics and Automation. Scottsdale, AZ: IEEE, 1989.

[43] KALMAN R E, BUCY R S. New Results in Linear Filtering and Prediction Theory[J]. Journal of Basic Engineering, 1961, 83(1): 95-108. DOI:10.1115/1.3658902

[44] DURRANT-WHYTE H, BAILEY T. Simultaneous Localization and Mapping:Part Ⅰ[J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.

[45] THRUN S, BURGARD W, FOX D. A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots [J]. Machine Learning, 1998, 31(1-3): 29-53.

[46] WAN E A, VAN DER MERWE R. The Unscented Kalman Filter for Nonlinear Estimation[C]//Proceedings of IEEE Adaptive Systems for Signal Processing, Communications, and

Control Symposium 2000, Lake Louise, Alberta, Canada: IEEE, 2000.

[47] ARASARATNAM I, HAYKIN S. Cubature Kalman Filters[J]. IEEE Transactions on Automatic Control, 2009, 54(6): 1254-1269. DOI:10.1109/TAC.2009.2019800

[48] ITO K, XIONG K. Gaussian Filters for Nonlinear Filtering Problems[J]. IEEE Transactions on Automatic Control, 2000, 45(5): 910-927. DOI:10.1109/9.855552

[49] DOUCET A, GODSILL S, ANDRIEU C. On Sequential Monte Carlo Sampling Methods for Bayesian Filtering[J]. Statistics and Computing, 2000, 10(3): 197-208. DOI: 10.1023/A:1008935410038

[50] MONTEMERLO M, THRUN S, KOLLER D, et al. FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem[C]//Proceedings of AAAI National Conference on Artificial Intelligence. Edmonton, Canada: AAAI, 2002: 593-598.

[51] MONTEMERLO M, THRUN S, ROLLER D, et al. FastSLAM 2.0: An improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges[C]//Proceedings of the 18th International Joint Conference on Artificial Intelligence. Acapulco, Mexico: Morgan Kaufmann Publishers Inc., 2003: 1151-1156.

[52] LU F, MILIOS E. Globally Consistent Range Scan Alignment for Environment Mapping[J]. Autonomous Robots, 1997, 4(4): 333-349. DOI:10.1023/A:1008854305733

[53] GUTMANN J S, KONOLIGE K. Incremental Mapping of Large Cyclic Environments[C]//Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation. Monterey, CA: IEEE, 1999.

[54] FRESE U, LARSSON P, DUCKETT T. A Multilevel Relaxation Algorithm for Simultaneous Localization and Mapping[J]. IEEE Transactions on Robotics, 2005, 21(2): 196-207. DOI:10.1109/TRO.2004.839220

[55] OLSON E, LEONARD J, TELLER S. Fast Iterative Alignment of Pose Graphs with Poor Initial Estimates[C]//Proceedings of IEEE International Conference on Robotics and Automation. Orlando, FL: IEEE, 2006.

[56] ZHAO Liang. MonoSLAM: Theories of Parameterization, Bundle Adjustment and Subgraph Fusion[D]. Beijing: Peking University, 2012. http://www.wanfangdata.com.cn/details/detail.do?_type=degree&id=Y2498852

[57] DELLAERT F, KAESS M. Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing[J]. International Journal of Robotics Research, 2006, 25(12): 1181-1203. DOI:10.1177/0278364906072768

[58] KAESS M, RANGANATHAN A, DELLAERT F. iSAM: Fast Incremental Smoothing and Mapping with Efficient Data Association[C]//Proceedings of IEEE International Conference on Robotics and Automation. Roma, Italy: IEEE, 2007.

[59] KAESS M, JOHANNSSON H, ROBERTS R, et al. iSAM2: Incremental Smoothing and Mapping Using the Bayes Tree [J]. International Journal of Robotics Research, 2012, 31(2): 216-235. DOI:10.1177/0278364911430419

[60] BAILEY T, DURRANT-WHYTE H. Simultaneous Localization and Mapping: Part Ⅱ[J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108-117.

[61] RUBNER Y, TOMASI C, GUIBAS L J. A Metric for Distributions with Applications to Image Databases[C]//Proceedings of IEEE International Conference on Computer Vision. Bombay, India: IEEE, 1998.

[62] BOSSE M, NEWMAN P, LEONARD S J J, et al. SLAM in Large-scale Cyclic Environments Using the Atlas Framework [J]. International Journal of Robotics Research, 2004, 23: 1113-1139. DOI:10.1177/0278364904049393

[63] EADE E, DRUMMOND T. Unified Loop Closing and Recovery for Real Time Monocular SLAM[C]//Proceedings of the British Conference on Machine Vision. Leeds: BMVA Press, 2008.

[64] ANGELI A, DONCIEUX S, MEYER J A, et al. Real-time Visual Loop-closure Detection[C]//Proceedings of IEEE International Conference on Robotics and Automation. Pasadena, CA: IEEE, 2008. 4300-4305.

[65] SMITH P, REID I D, DAVISON A J. Real-time Monocular SLAM with Straight Lines[C]//Proceedings of British Conference on Machine Vision. Edinburgh, UK: BMVA Press, 2006: 17-26.

[66] LEMAIRE T, LACROIX S. Monocular-vision based SLAM Using Line Segments[C]//Proceedings of IEEE Robotics and Automation. Roma, Italy: IEEE, 2007.

[67] SOLÀ J, VIDAL-CALLEJA T, DEVY M. Undelayed Initialization of Line Segments in Monocular SLAM[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis, MO: IEEE, 2009.

[68] PERDICES E, LÓPEZ L M, CAÑAS J M. LineSLAM: Visual Real Time Localization Using Lines and UKF[M]//ARMADA A, SANFELIU A, FERRE M. ROBOT2013: First Iberian Robotics Conference. Cham: Springer, 2014.

[69] ZHOU Huazhong, ZOU Danping, PEI Ling, et al. Struct SLAM: Visual SLAM With Building Structure Lines[J]. IEEE Transactions on Vehicular Technology, 2015, 64(4): 1364-1375. DOI:10.1109/TVT.2015.2388780

[70] PUMAROLA A, VAKHITOV A, AGUDO A, et al. PL-SLAM: Real-time Monocular Visual SLAM with Points and Lines[C]//Proceedings of IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017.

[71] LI Haifeng, HU Zunhe, CHEN Xinwei. PLP-SLAM: A Visual SLAM Method Based on Point-line-plane Feature Fusion[J]. Robot, 2017, 39(2): 214-220, 229.

[72] SILVERIA G, MALIS E, RIVES P. An Efficient Direct Approach to Visual SLAM[J]. IEEE Transactions on Robotics, 2008, 24(5): 969-979. DOI:10.1109/TRO.2008.2004829

[73] ENGEL J, KOLTUN V, CREMERS D. Direct Sparse Odometry [J]. IEEE Transactions on Pattern Analysis and Machine In-

telligence, 2018, 40(3): 611-625. DOI: 10.1109/TPAMI.2017.2658577

[74] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast Semi-direct Monocular Visual Odometry[C]//Proceedings of IEEE International Conference on Robotics and Automation. Hong Kong, China: IEEE, 2014.

[75] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: Large-scale Direct Monocular SLAM[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 834-849.

[76] CORKE P, LOBO J, DIAS J. An Introduction to Inertial and Visual Sensing[J]. International Journal of Robotics Research, 2007, 26(6): 519-535. DOI:10.1177/0278364907079279

[77] LI Mingyang, MOURIKIS A I. High-precision, Consistent EKF-based Visual-inertial Odometry[J]. International Journal of Robotics Research, 2013, 33(6): 690-711.

[78] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based Visual-inertial Odometry Using Nonlinear Optimization [J]. International Journal of Robotics Research, 2014, 34 (3): 314-334.

[79] MU Xufu, CHEN Jing, ZHOU Zixiang, et al. Accurate Initial State Estimation in a Monocular Visual-inertial SLAM System [J]. Sensor, 2018, 18(2): 506. DOI:10.3390/s18020506

[80] WAN Wenhui. Theory and Methods of Stereo Vision Based Autonomous Rover Localization in Deep Space Exploration [D]. Beijng: Graduate School of Chinese Academy of Sciences, 2012.http://www.irgrid.ac.cn/handle/1471x/823847

[81] WU Kai. Monocular Vision Integrated with Laser Distance Meter for Astronaut Navigation on Lunar Surface[D]. Beijng: University of Chinese Academy of Sciences, 2013.http://www.wanfangdata. com. cn/details/detail. do? _type = degree&id =Y2609688

[82] WU Kai, DI Kaichang, SUN Xun, et al. Enhanced Monocular Visual Odometry Integrated with Laser Distance Meter for Astronaut Navigation[J]. Sensors, , 14(3): 4981-5003.DOI: 10.3390/s140304981

[83] ZHANG Xinzheng, RAD A B, WONG Y K. Sensor Fusion of Monocular Cameras and Laser Rangefinders for Line-based Simultaneous Localization and Mapping (SLAM) Tasks in Autonomous Mobile Robots[J]. Sensors, 2012, 12(1): 429-452. DOI:10.3390/s120100429

[84] DI Kaichang. Remand Analysis and Technical Proposal Discussion about Navigation for Lunar Astronauts[C]//Proceedings of Engineering Science and Technology Forum (107th) of Chinese Academy of Engineering: Manned Lunar Landing and Deep Space Exploration. Beijing: Chinese Accdemy of Engineering 2010: 213-218.

[85] KHOSHELHAM K, ELBERINK S O. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications[J]. Sensors, 2012, 12 (2): 1437-1454. DOI: 10. 3390/s120201437

[86] STURM J, ENGELHARD N, ENDRES F, et al. A Benchmark for the Evaluation of RGB-D SLAM Systems [C] // Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura, Portugal: IEEE, 2012.

[87] KERL C, STURM J, CREMERS D. Dense Visual SLAM for RGB-D Cameras [C] // Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE, 2013.

[88] WHELAN T, KAESS M, JOHANNSSON H, et al. Real-time Large-scale Dense RGB-D SLAM with Volumetric Fusion[J]. International Journal of Robotics Research, 2015, 34(4-5): 598-626.DOI:10.1177/0278364914551008

[89] DI Kaichang, ZHAO Qiang, WAN Wenhui, et al. RGB-D SLAM based on Extended Bundle Adjustment with 2D and 3D Information [J]. Sensors, 2016, 16 (8): 1285. DOI: 10. 3390/s16081285

[90] ZHU X X, TUIA D, MOU L C, et al. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 5(4): 8-36.

[91] ZHANG Zixing, GEIGER J, POHJALAINEN J, et al. Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments[J]. ACM Transactions on Intelligent Systems and Technology, 2018, 9(5): Article No.49.

[92] GARCIA-GARCIA A, ORTS-ESCOLANO S, OPERA S, et al. A Review on Deep Learning Techniques Applied to Semantic Segmentation [J]. arXiv preprint arXiv: 1704. 06857, 2017.

[93] KONDA K, MEMISEVIC R. Learning Visual Odometry with a Convolutional Network[C]//Proceedings of the 10th International Conference on Computer Vision Theory and Applications. Berlin, Germany: SCITCC Press, 2015: 486-490.

[94] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: Learning Optical Flow with Convolutional Networks[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 2758-2766.

[95] COSTANTE G, MANCINI M, VALIGI P, et al. Exploring Representation Learning with CNNs for Frame-to-frame Ego-motion Estimation [J]. IEEE Robotics and Automation Letters, 2016, 1 (1): 18-25. DOI: 10. 1109/LRA. 2015.2505717

[96] BAI Dongdong, WANG Chaoqun, ZHANG Bo, et al. Matching-range-constrained Real-time Loop Closure Detection with CNNs Features[J]. Robotics and Biomimetics, 2016, 3: 15. DOI:10.1186/s40638-016-0047-x

[97] ZHANG X W, SU Y, ZHU X H. Loop Closure Detection for Visual SLAM Systems Using Convolutional Neural Network [C]//Proceedings of IEEE International Conference on Automation and Computing. Huddersfield, UK: IEEE, 2017.

[98] GAO Xiang, ZHANG Tao. Unsupervised Learning to Detect Loops Using Deep Neural Networks for Visual SLAM System

[J]. Autonomous Robots, 2017, 41(1): 1-18. DOI: 10.1007/s10514-015-9516-2

[99] LI Rongxing, SQUYRES S W, ARVIDON R E, et al. Initial Results of Rover Localization and Topographic Mapping for the 2003 Mars Exploration Rover Mission[J]. Photogrammetric Engineering & Remote Sensing, 2005, 71(10): 1129-1142.

[100] DI Kaichang, XU Fengliang, WANG Jue, et al. Photogrammetric Processing of Rover Imagery of the 2003 Mars Exploration Rover Mission[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2008, 63(2): 181-201. DOI: 10.1016/j.isprsjprs.2007.07.007

[101] MARTIN-MUR T J, KRUIZINGA G L, BUKHART P D, et al. Mars Science Laboratory Navigation Results [C] // Proceedings of International Symposium on Space Flight Dynamics. Washington D. C.: NASA, 2012.

[102] CHENG Yang, MAIMONE M W, MATTHIES L. Visual Odometry on the Mars Exploration Rovers-A Tool to Ensure Accurate Driving and Science Imaging[J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 54-62.

[103] WAN Wenhui, LIU Zhaoqin, LIU Yiliang, et al. Descent Image Matching based Position Evaluation for Chang'e-3 Landing Point[J]. Spacecraft Engineering, 2014, 23(4): 5-12.

[104] LIU Bin, XU Bin, LIU Zhaoqin, et al. Descending and Landing Trajectory Recovery of Chang'e-3 Lander using Descent Images[J]. Journal of Remote Sensing, 2014, 18(5): 988-994. DOI: 10.11834/jrs.20144070

[105] LIU Zhaoqin, DI Kaichang, PENG Man, et al. High Precision Landing Site Mapping and Rover Localization for Chang'e-3 Mission[J]. Science China-Physics Mechanics & Astronomy, 2015, 58(1): 1-11.

[106] WAN W H, LIU Z Q, DI K C, et al. A Cross-Site Visual Localization Method for Yutu Rover[C]//Proceedings of ISPRS Technical Commission Ⅳ Symposium. Suzhou, China: ISPRS, 2014.

[107] HUANG Haosheng, GARTNER G. A Survey of Mobile Indoor Navigation Systems[M]//GARTNER G, ORTAG F. Cartography for Central and Eastern European. Berlin, Heidelberg: Springer, 2009.

[108] FALLAH N, APOSTOLOPOULOS I, BEKRIS K, et al. Indoor Human Navigation Systems: A Survey[J]. Interacting with Computers, 2013, 25(1): 21-33.

[109] FROEHLICH M, AZHAR S, VANTURE M. An Investigation of Google Tango® Tablet for Low Cost 3D scanning[C]// Proceedings of 34th International Symposium on Automation and Robotics in Construction. Hawaii: ICACR, 2017.

[110] NGUYEN K A, LUO Zhiyuan. On Assessing the Positioning Accuracy of Google Tango in Challenging Indoor Environments[C]//Proceedings of International Conference on Indoor Positioning and Indoor Navigation. Sapporo, Japan: IEEE, 2017.

[111] WINTERHALTER W, FLECKENSTEIN F, STEDER B, et al. Accurate Indoor Localization for RGB-D Smartphones and Tablets Given 2D Floor Plans[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Hambury, Germany: IEEE, 2015.

[112] LEE J. Mobile AR in Your Pocket with Google Tango[J]. Society for Information Display International Symposium Digest of Technical Papers, 2017, 48(1): 17-18. DOI: 10.1002/sdtp.11563

[113] GARON M, BOULET P O, DOIRONZ J P, et al. Real-Time High Resolution 3D Data on the HoloLens[C]//Proceedings of IEEE International Symposium on Mixed and Augmented Reality. Merida, Mexico: IEEE, 2016.

[114] EVANS G, MILLER J, PENA M I, et al. Evaluating the Microsoft HoloLens through an Augmented Reality Assembly Application[C]//Proceedings of SPIE, Volume 10197, Degraded Environments: Sensing, Processing, and Display 2017. Anaheim, CA: SPIE, 2017.

[115] ZHOU Guyue, FANG Lu, TANG Ketan, et al. Guidance: A Visual Sensing Platform for Robotic Applications[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA: IEEE, 2015: 9-14.

[116] MEJIAS L, CORREA J F, MONDRAGON I, et al. COLIBRI: A Vision-Guided UAV for Surveillance and Visual Inspection [C]//Proceedings of IEEE International Conference on Robotics and Automation. Roma, Italy: IEEE, 2015.

[117] KRAJNÍK T, NITSCHE M, PEDRE S, et al. A Simple Visual Navigation System for an UAV[C]//Proceedings of 9th International Multi-Conference on Systems, Signals and Devices. Chemnitz, Germany: IEEE, 2012.