AN INDOOR SLAM METHOD BASED ON KINECT AND MULTI-FEATURE EXTENDED INFORMATION FILTER

M. Chang[a, b], Z. Kang [a,*]

[a]School of Land Science and Technology, China University of Geosciences, 29 Xueyuan Road, Haidian District, Beijing 100083, China -zzkang@cugb.edu.cn
[b]The First Monitoring and Application Center, China Earthquake Administration,7 Naihuo Road, Hedong District,Tianjin,300180, China -charmingxz@163.com

Commission IV, WG IV/5

KEY WORDS: Indoor Positioning，RGB-D Camera，Multi-Feature Extend Information Filter Model，ICP，SLAM

ABSTRACT:

Based on the frame of ORB-SLAM in this paper the transformation parameters between adjacent Kinect image frames are computed using ORB keypoints, from which priori information matrix and information vector are calculated. The motion update of multi-feature extended information filter is then realized. According to the point cloud data formed by depth image, ICP algorithm was used to extract the point features of the point cloud data in the scene and built an observation model while calculating a-posteriori information matrix and information vector, and weakening the influences caused by the error accumulation in the positioning process. Furthermore, this paper applied ORB-SLAM frame to realize autonomous positioning in real time in interior unknown environment. In the end, Lidar was used to get data in the scene in order to estimate positioning accuracy put forward in this paper.

1. INTRODUCTION

Currently, with the continuous development of visual sensors, the technology of visual navigation and positioning has been widely used in motion platform. While algorithm and processor performance are updated continuously, how to realize robot self-localization and mapping through visual sensor in unknown environment has been a current research hotspot.

It is noteworthy that a number of RGB-D launched in recent years is widely used in the field of indoor positioning and mapping. Izadi et al.(2011) described how to get the depth image and color image of Kinect sensors in detail, and realized scene reproduction in real time with the provision of precise 3-D model. Whelan et al.(2012) extended the Kinect Fusion algorithm, after its application, the region of space mapped by Kinect Fushion algorithm could change dynamically. Furthermore, high-density point cloud data in the scene was extracted and was put into the environment. This was displayed by triangular mesh, realizing real-time processing of high-density model building for the objects in the scene. The slam method of RGB-D was firstly proposed by Newcombe et al. (2011), and ICP algorithm model was applied to work out the real time registration and mapping display of Kinect sensor. Endres et al. (2014) et al. proposed RGBD-slam system model, achieving frame registration through feature matching among frames and ICP algorithm. Raul et al. (2016) put forward ORB-SLAM system to solve the problems of monocular, stereo and the slam of RGB-D. Santos et al. (2016) proposed an adaptive registration model from coarse-to-fine with making use of RGB-D data. Xiang Gao et al. (2015) proposed feature planar features in the scene to reduce deviation accumulation.

At the present stage, slam system is often used for dealing with the problems of indoor positioning. Loop closure method included in Slam system can weaken the error calculation's influence on point cloud data. In a large-scale scene or a patency scene however, the function of loop closure is next to nothing. Because the ORB-SLAM system can quickly obtain the color image of the ORB feature points, this paper proposes multi-feature extend information filter model with using ORB-SLAM frame to weaken the influences on positioning accuracy caused by error calculation and realize real time scene location.

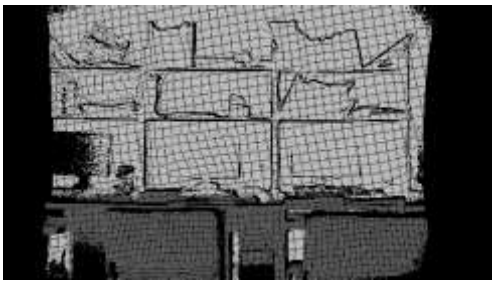
2. POSITIONING SYSTEM BASED ON MULTI-FEATURE EXTEND INFORMATION FILTER

2.1 Data Acquisition

The RGB-D sensor applied in this paper is Kinect v2.0 developed by Microsoft. This camera can acquire depth image in the range of 1-5m in the scene, combining with mutual calibration color camera, it can generate point cloud data in the scene. The experimental data in this paper has a resolution of 960*540. A frame of color and depth image respectively are as shown below:

a.Color image

* Corresponding author



b.Depth image
Figure1 Color image and depth image

2.2Data tracking

This paper completes the transaction of pair-wise points from 2-D image to 3-D point cloud data, and makes the confirmation of consecutive frame pose. Though traditional algorithm of SIFT or SURF can make good use of visual feature to realize object identification, image registration, visual image etc., it also serves as a great burden for computers.

ORB algorithm provides a new means of combining inspection method of FAST feature point with BRIEF feature descriptor while getting improvement and optimization on the primary basis as well as increasing efficiency. Tracking model applies ORB algorithm to extract the feature points of present frame of color image, registers the extracted feature point with the former frame feature points, and acquire the relative transformational correlation of present frame of data n and the previous frame of data $n-1$ *i.e.* $\Delta\mu_n = [\Delta x, \Delta y, \Delta z, \Delta\alpha, \Delta\beta, \Delta\gamma]$. Then, at present time k, the state vector of each frame of data is expressed as:

$$\mu(k) = \begin{bmatrix} \mu_1(k) \\ \vdots \\ \mu_n(k) \end{bmatrix} \quad \mu_i(k) = \begin{bmatrix} x_i(k) \\ y_i(k) \\ z_i(k) \\ \alpha_i(k) \\ \beta_i(k) \\ \gamma_i(k) \end{bmatrix} \tag{1}$$

Where n represents the pose of sequence images received up to now. $\mu_i(k)$ represents the pose of the data of i^{th} frame at the time k.Information matrix:

$$\Omega_n = \Sigma_n^{-1} = \begin{bmatrix} \Sigma_{11}(k) & \cdots & \Sigma_{1n}(k) \\ \vdots & \ddots & \vdots \\ \Sigma_{n1}(k) & \cdots & \Sigma_{nn}(k) \end{bmatrix} \tag{2}$$

Information vector: $\xi_n = \Sigma_n^{-1} * \mu_n = \Omega_n * \mu_n$
To calculate the coordinate of the present data in the global coordinate system based on X_n and ΔX_n, the resultant data is

$$\mu_{n+1}^{-}(k) = g(\mu_n(k), \Delta\mu_n) + \varpi(k) \tag{3}$$

Here, $\mu_{t+1}^{-}(k)$ represents the a-priori information vector in the global coordinate system at the time t, determined based on the present frame. $g(.)$ is a system-augmented function. $\varpi(k)$ describes a variety of uncertainties in the registration and modeling process, and it is assumed to be a Gaussian distribution expressed as white noise vector $N(0, Q)$.

The present frame pose is added into the information vector and the update of each frame of data status vector is realized.

2.3 Measurement Updates

After realizing data tracking by completing the extraction of 2-D color image features, this paper utilizes the point cloud data generated by depth data to establish multi-feature measurement model while extracting features of point and plane as well as realizing the update of information vector and information matrix.

2.3.1. Point-feature model

This paper extracts closet point of the data of the adjacent two frames by ICP algorithm, and takes these closet points as point features to establish a measurement model. It is supposed that the data of the adjacent two frames is expressed as (x_{L1}, y_{L1}, z_{L1}) in the coordinate system of i^{th} frame, and (x_{L2}, y_{L2}, z_{L2}) in the coordinate system of j^{th} frame shown as follows:

$$\begin{aligned} Z_{p1} &= h\left(X_i(k), X_j(k)\right) + v(k) \\ &= \left\{ \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + R_i \begin{bmatrix} x_{L1} \\ y_{L1} \\ z_{L1} \end{bmatrix} \right\} - \left\{ \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} + R_j \begin{bmatrix} x_{L2} \\ y_{L2} \\ z_{L2} \end{bmatrix} \right\} + v(k) \end{aligned} \tag{3}$$

Here, $h(.)$ is the systematic measurement function, and $v(k)$ denotes a variety of uncertainties in the scanning measurement and the transformation of coordinates which is supposed to comply with the Gaussian distribution expressed as white noise vector $N(0, S)$.

2.3.2. Planar feature model

This paper extracts the same planar feature of the adjacent two frames of the data to establish the measurement model of planar feature. The planar information between two frames of data only controls the rotation parameters among the transformation parameters. So, it is assumed to be the same planar unit normal vector as (a_1, b_1, c_1) and (a_2, b_2, c_2). It is expressed as (x_{l1}, y_{l1}, z_{l1}) in the i^{th} frame coordinate system, as (x_{l2}, y_{l2}, z_{l2}) in the j^{th} frame coordinate system. Based on the present pose of the survey station point of i^{th} and j^{th} respectively to estimate $X_i(k)$ and $X_j(k)$, the coordination is changed into global coordinate, which is expressed as (x_{M1}, y_{M1}, z_{M1}) and (x_{M2}, y_{M2}, z_{M2}). R is the rotation. The planar feature model is derived as

$$\begin{aligned} Z_{p2} &= h_2\left(X_i(k), X_j(k)\right) + v(k) \\ &= \left\{ R_i \begin{bmatrix} x_{M1} \\ y_{M1} \\ z_{M1} \end{bmatrix} - R_j \begin{bmatrix} x_{M2} \\ y_{M2} \\ z_{M2} \end{bmatrix} \right\} + v(k) \end{aligned} \tag{4}$$

2.3.3. Multi-feature Measurement model

This paper acquires the features of points and planes after the establishment of the measurement model of points and planes. Therefore, the measurement model derived from multiple features is as follows:

$$Z = \begin{bmatrix} Z_{p1} \\ Z_{p2} \end{bmatrix} \tag{5}$$

Z_{P1} represents feature information of points; Z_{P2} represents feature information of planes. Their accuracies are different, and the two information types are assigned different weights to enhance the final result.

$$Q = \begin{bmatrix} \lambda_1 * Q_1 & 0 \\ 0 & \lambda_2 * Q_2 \end{bmatrix} \quad (6)$$

Here, λ_1 is the weight of pair-wise points; λ_2 is the weight of planar feature; Q_1 is the covariance matrix of pair-wised points; Q_2 is the covariance matrix of multi-feature.

This paper uses measurement model, a-priori information matrix, and information vector to update the present system state as follows:

$$\left.\begin{aligned} \xi_n^+ &= \xi_n^- + \nabla h^T * Q^{-1} * \nabla h \\ \Omega_n^+ &= \Omega_n^- + \nabla h^T * Q^{-1} * \Delta z \\ \Delta z &= z_n - h(Z_n) + \nabla h|_z Z_n \end{aligned}\right\} \ldots\ldots \quad . \quad (7)$$

Where, ξ_n^+ is a-posteriori information matrix; ∇h denotes the Jacobian matrix; Ω_n^+ is a-posteriori information vector; Δz is measurement information; Ω_n^+ reflects the pose in the overall coordinate system each frame of the data.

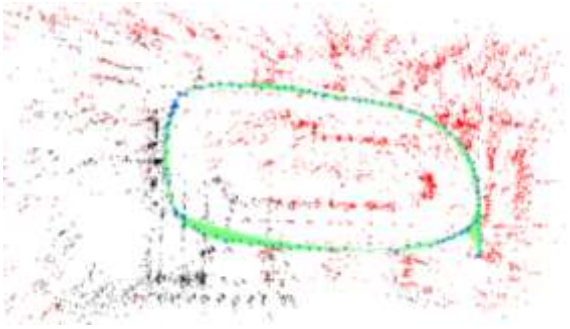
2.4. Loop Closure

In the process of indoor positioning, error accumulation is unavoidable with the continuous increase of data acquired. In order to weaken the influences of error accumulation in indoor positioning systems, the method of loop closure is provided to improve the accuracy positioning and mapping.

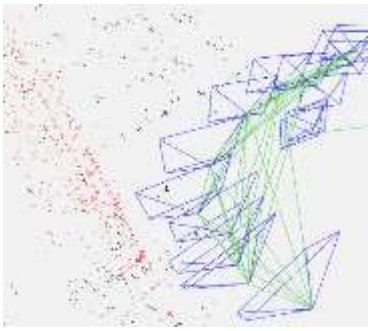
Through feature point registration, loop closure can determine whether or not to collect the data of the present scene. When the number of detected features meets a certain threshold, the data will be recognized as already acquired. Comparing the pose of present scene with the pose of repetitive scene, the value difference between the poses is the result of error accumulation of all the key frames between the two frames of the repetitive scene. Meanwhile, the information vector is modified to make the poses of two repetitive frames of data reach unanimity and complete the loop closure.

2.5. Positioning result display

With the help of Pangolin base to realize visualization user interface, the result of real-time positioning according to the pose of the present frame in global coordinate system is displayed as follows:



a Global positioning map



b Partial enlarged detail

Figure 2.Real-time positioning result display

As shown in figure 3, the process of tracing displayed points is expressed by a rectangular pyramid instead of only a point to stand for the location of the present frame. The result of positioning is expressed by the rectangular pyramid. The normal vector of the bottom plane of the rectangular pyramid is used to express the rotation magnitude of the present frame and the global coordinate system tri-axial rotation magnitude. If the green solid line of the data between every two frames is jointed, it means there are some certain identical feature points of data in the adjacent frames, hence, the loop closure detection is finished.

While displaying 3-D scene, the paper only displays feature points instead of all the generated point cloud data in order to reduce the complexity of point cloud data in the map and also to reduce the requirements of computer hardware.

3. THE ANALYSIS OF EXPERIMENT RESULT

The experimental scene chosen by this paper is the building experimental scene. Lidar is used to obtain high precision laser point cloud data in the experimental scene, and merged with Kinect sequence images by manual-registration to obtain accurate position points. At the same time, the method of ORB-SLAM is used to process the obtained sequence images, and its result is compared with the positioning result of the method proposed in this paper.

3.1. Fused data

This paper applied Lidar data to obtain color point cloud data in the experimental scene, and the result is illustrated in the following figures:

a Panoramic image of the experimental scene

b The vertical view of point cloud in the experimental scene

c The lateral view of point cloud in the experimental scene

Figure 3.The laser-point cloud data in the experimental scene

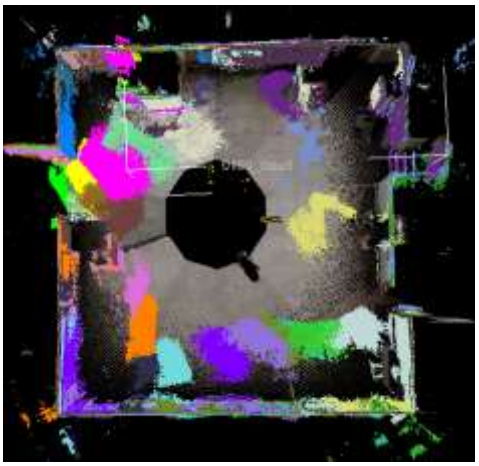
This paper uses Kinect to obtain the data of the experimental scene and 334 key frames of sequence images is obtained. The color image and the depth image of those key frames of data are shown as follows:

a.Color Image

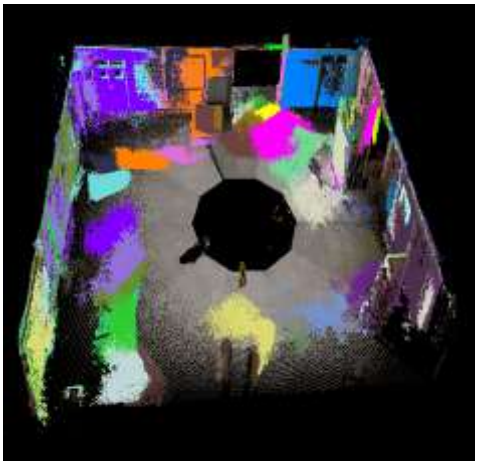
b Depth Image

Figure 4. Frame of image data obtained by Kinect

In the process of data merging in allusion to the sequence images of key frames, this paper selects corresponding color point cloud data generated by image data in every ten frames, and merges with the data obtained by Lidar. In all, 30 frames of data were totally merged. The selection of the 30 frames reduces the stress of manual-registration on the basis of not influencing the overlapping degree of point cloud data in the scene. In order to distinguish the point cloud data of Kinect and the data of Lidar. Kinect data, in the fused data is shown by single RGB values. The merging result is shown as the follow figures:



a Vertical view



b Lateral view

Figure5. Display result of fused data

In reference to the fused data, this paper calculates the track of each key frame of data and get the tracing point data of Kinect key frames of sequence image in the coordinate system of Lidar . The result is shown as follows:

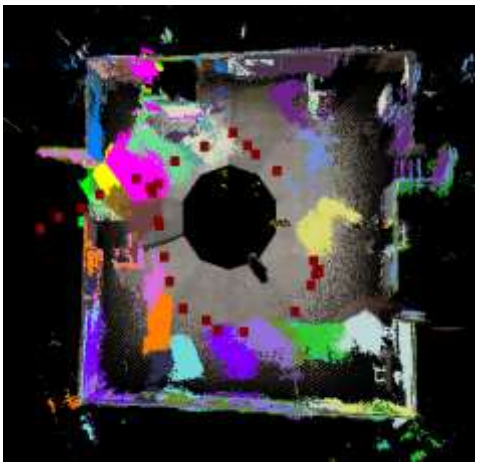
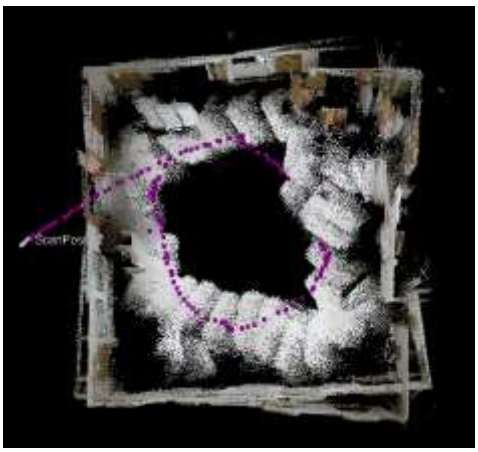


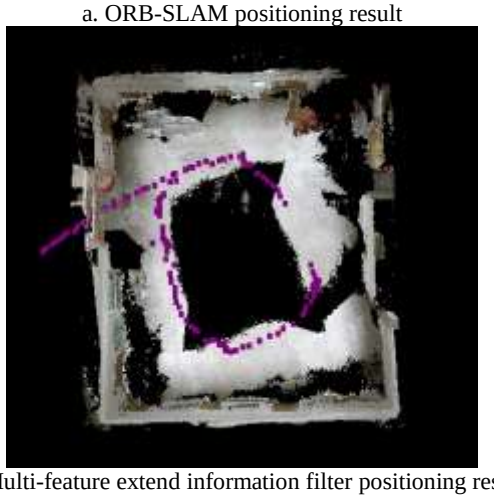
Figure6. Tracing points of fused data thinning

3.2. Positioning result

Making use of the sequence images obtained by Kinect, this paper locates sequence images with the help of ORB-SLAM system and the multi-feature extend information filter proposed in this paper. The result is shown in the following figures:



a. ORB-SLAM positioning result

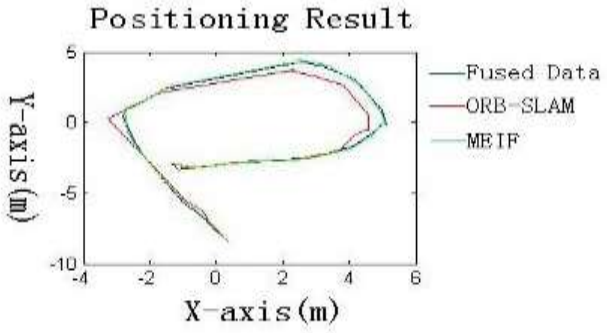


b. Multi-feature extend information filter positioning result

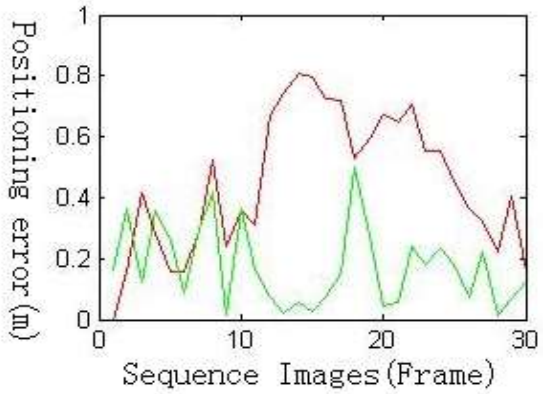
Figure7. Positioning result

As for the positioning problem of indoor scene, mapping precision and the positioning accuracy of each frame of data are correlative and coupled. While processing indoor data, high-precision mapping means high registration accuracy of sequence images – which reflects the pose accuracy of each frame of data – namely accurate positioning precision. By the comparison of the mapping accuracy of the two algorithms, there is a bigger angle deviation in wall corners with ORB-SLAM algorithm indoor data processing, which influences the mapping precision and the final positioning precision. Compared with ORB-SLAM, the method proposed in this paper is improved in the aspect of mapping accuracy.

In order to make a better comparison of the positioning point accuracy of sequence images obtained by means of fused data, ORB-SLAM and the method proposed in this paper; the tracing points obtained from these three methods are placed in the same coordinate system for comparison purposes.



a. Trajectory chart



b. Range difference of tracing point

Figure8. Comparative result of trajectory charts

Comparing the result of ORB-SLAM and the positioning result of multi-feature extend information filter with the accurate positioning of fused data as shown in figure a; the blue track is the positioning result of fused data, the green track is the positioning result of MEIF (multi-feature extend information filter), and the red track is the positioning result of ORB-SLAM. From the comparisons above, it can be seen that the positioning method proposed in this paper yielded better results compared to the results of ORB-SLAM. In order to find the positional accuracy of the two patterns precisely, a comparison of the two positioning information and the location result of fused data was made, the acquired corresponding pose of the homonymous frame was used to make a comparison, and calculate the spatial distance of the tracking points. The positioning deviation of ORB-SLAM is between 0 m to 0.8 m, and the location deviation of multi-feature extend information filter is in the range of 0 m to 0.5 m. The RMSE of ORB-SLAM and MEIF are 0.48m and 0.23m.

4. CONCLUSION

In allusion to the problem that deviation accumulation cannot be weakened effectively in large-scale scene and patency scene of

present slam, this paper extracts ORB feature points in the scene and builds an extended information filtering model which can weaken the influences of deviation accumulation effectively to realize the location of indoor scene.

ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (grant number 41471360) and the Fundamental Research Funds for the Central Universities (grant number 2652015176).

REFERENCES

Tardos J D, J, 2002.Robust mapping and localization in indoor environments using sonar data. *International Journal of Robotics Research*, 21(4), pp. 311-330.

Izadi S, Kim D, Hilliges O, et al. ,C, 2011, KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. *Proceedings of the 24th annual ACM symposium on User interface software and technology*.pp.559-568.

Whelan T, Kaess M, Fallon M, et al. J, 2012, Kintinuous: Spatially extended kinectfusion

Newcombe R A, Izadi S, Hilliges O, et al. ,C, 2011, KinectFusion: Real-time dense surface mapping and tracking. *IEEE International Symposium on Mixed and Augmented Reality. IEEE Computer Society*, pp.127-136

Besl P J, Mckay N D. ,J, 1992, Method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 14(2), pp.239-256.

Endres F, Hess J, Sturm J, et al. ,J, 2014, 3-D mapping with an RGB-D camera . *IEEE Transactions on Robotics*, 30(1), pp. 177-187.

Murartal R, Tardos J D. ,J, 2016, ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras.

Santos D R D, Basso M A, Khoshelham K, et al. ,J, 2016,Mapping Indoor Spaces by Adaptive Coarse-to-Fine Registration of RGB-D Data. *IEEE Geoscience & Remote Sensing Letters*, 13(2), pp.262-266.

Gao X, Zhang T. ,J, 2015, Robust RGB-D simultaneous localization and mapping using planar point features. *Robotics & Autonomous Systems*, 72, pp.1-14.

Rublee E, Rabaud V, Konolige K, et al. ,C, 2011,ORB: An efficient alternative to SIFT or SURF. International Conference on Computer Vision. *IEEE Computer Society*, pp.2564-2571.

Lowe D G. ,J, 2004, Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2),pp.91-110.

Bay H, Tuytelaars T, Gool L V. ,J, 2006, SURF: Speeded Up Robust Features. *Computer Vision & Image Understanding*, 110(3), pp.404-417.

Rosten E, Drummond T. ,C, 2006,Machine learning for high-speed corner detection. *European Conference on Computer Vision. Springer-Verlag*, pp.430-443.

Rosten E, Porter R, Drummond T. ,J, 2008,Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 32(1),pp.105-119.

Calonder M, Lepetit V, Strecha C, et al. ,C, 2010,BRIEF: binary robust independent elementary features. *European Conference on Computer Vision. Springer-Verlag*, pp.778-792

Labbe M, Michaud F. ,J, 2013, Appearance-based loop closure detection for online large-scale and long-term operation. *IEEE Transactions on Robotics*, 29(3) , pp.734-745.

Cummins M, Newman P. ,J, 2011,Appearance-only SLAM at large scale with FAB-MAP 2.0. *The International Journal of Robotics Research*, 30(9), pp.1100-1123.

Ho K L, Newman P. ,J, 2007,Detecting loop closure with scene sequences. *International journal of computer vision*, 74(3), pp.261-286.