IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 6, 2020, accepted January 17, 2020, date of publication January 23, 2020, date of current version February 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968958

An Improved Vision-Based Indoor Positioning Method

SONGXIANG YANG[1], LIN MA[1], (Senior Member, IEEE), SHUANG JIA[1], AND DANYANG QIN[2]

[1]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China
[2]Electronic Engineering College, Heilongjiang University, Harbin 150080, China

Corresponding author: Lin Ma (malin@hit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61971162, Grant 61771186, Grant 41861134010, and Grant 61571162, and in part by the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province under Grant UNPYSCT-2017125.

ABSTRACT Vision-based indoor positioning technology is a practical and effective method to solve the problem of indoor positioning and navigation. Compared to Bluetooth-based and WiFi-based positioning methods, vision-based positioning method can provide reliable and low-cost services using a camera without extra pre-deployed hardware. To improve the robustness and accuracy of traditional visual positioning algorithm, this paper proposes a pixel threshold based eight-point method and an improved epipolar constraint algorithm. The traditional eight-point method only uses Euclidean distance as a selection indicator for feature points. The pixel coordinates of some feature points are distorted when the positioning scene changes, which may cause mismatch. The proposed method introduces the pixel threshold constraint to improve the quality of output feature points. Further, the epipolar constraint algorithm is modified by adding a new cost function to improve the accuracy of fundamental matrix calculation, thereby improving the positioning precision. Performance simulation analysis shows that the proposed algorithm can effectively improve indoor positioning precision.

INDEX TERMS Pixel drift, pixel threshold, fundamental matrix calculation, epipolar constraint.

I. INTRODUCTION

In recent years, with the development of technologies in the field of communications and computers, the problem of positioning has become a hot issue in industry and academia. Outdoor positioning systems, represented by satellite navigation, have solved users' demand for outdoor positioning services [1], but they cannot be used for indoor positioning and navigation.

Traditional indoor positioning systems based on Bluetooth and WiFi [2], [3] need to deploy a large number of equipment in the application scene, and signal transmission is susceptible to complex indoor environments. Vision-based indoor positioning technology is based on map prior information [4], [5], without relying on additional equipment. Meanwhile, the vision-based method is less sensitive to indoor environmental changes, which is suitable for a wide range of indoor environments such as airports, parking lots, and large shopping malls [6], [7]. The online stage estimates the geographic location of pedestrian through an image retrieval algorithm and a positioning algorithm [8], [9]. The popularity of intelligent terminals and the development of image processing technology enable real-time and high-precision vision-based indoor positioning systems to meet the requirements of users [10], [11].

The vision-based positioning method relies on feature points for image matching, which can ensure strong robustness. In addition, compared with the deep learning-based method, the feature point method does not require a large amount of pre-training, and has a better adaptability to complex and variable unknown scenes [12]. In order to improve the precision of vision-based indoor positioning algorithm, an improved indoor positioning method is proposed in this paper, which focus on the accurate position estimation. The main contributions of this paper are described as follows:

(1) A pixel threshold based eight-point method is proposed to improve the quality of feature points and eliminate mismatching feature points caused by pixel drift.

(2) An improved epipolar constraint is proposed, and a new cost function is introduced to improve the accuracy of fundamental matrix calculation, which is crucial for the pose estimation of query camera. Meanwhile, the performance

The associate editor coordinating the review of this manuscript and approving it for publication was Guitao Cao.

of proposed method is evaluated on typical indoor scenes, and experimental results show that the proposed method has better improvement in positioning precision.

The rest of this paper is organized as follows. The related work is given in Section II. Section III provides the proposed pixel threshold based eight-point method. The position estimation based on improved epipolar constraint is shown in Section IV. Section V gives the simulation results. Conclusion is drawn in the last section.

II. RELATED WORK

The image-based indoor positioning technology can not only estimate the user's position, but also determine the user's orientation accurately, that is, the position and the orientation can be obtained simultaneously [13], [14]. At present, for image based indoor positioning, the commonly used methods are mostly focused on calculating the Euclidean distance between the feature points of image collected by the mobile terminal and the feature points in the database [15]. Finally, the position corresponding to the feature points with the highest matching degree is selected as the positioning result.

A positioning algorithm with high robustness is proposed in [16], which adopts a real-time positioning algorithm based on video stream. Users can determine their own position and navigate the destination by continuously obtaining image information. However, there are two problems in this method when mapping the visual position space. On the one hand, the positioning accuracy has a strong dependence on the position fingerprint. On the other hand, the fault tolerance of the system is reduced since the method matches the global feature points of the image. The authors in [17] proposes an image-based localization method for narrow corridors, which adopts the SIFT (Scale In-variant Feature Transform) feature to retrieve an image in the database closest to the input image and return the image's position. In [18], a panoramic image acquisition device is designed, and the ideal positioning precision is obtained by using panoramic photographs combined with Principle Component Analysis-SIFT (PCA-SIFT) algorithm and LSH-based nearest neighbor algorithm. The MoVIPS indoor positioning system in [19] uses the Speeded Up Robust Features (SURF) algorithm to extract feature points based on the rough positioning of WiFi, and proposes an image-based position estimation method and a video stream-based position estimation method. The authors in [20] also adopt the mobile phone to locate after the rough positioning of WiFi. The difference is that they match the building identification in the database, and finally the feedback information is directly presented on the images taken by the mobile phone. However, a mis-match result is generated when there is a pedestrian in the image captured by the camera sensor of the user's mobile phone, resulting in a larger positioning error, that is, the robustness of this method needs to be enhanced. The authors in [21] propose an Indoor Localization method via Multi-view Images and Videos (MIVIL). 2DTriPnP in [22] is proposed, which can be interpreted as the robust 2D combination of feature triangulation and PnP problems. It can be concluded that this algorithm saves time overhead to meet real-time of vision positioning, but the precision of vision positioning is seriously affected.

At present, some problems existing in the existing vision-based indoor positioning algorithms are summarized as follows:

(1) The mismatch of feature points and the single standard of selecting the eight pairs of matching feature points make the accuracy of fundamental matrix solution lower.

(2) The low precision of traditional vision-based indoor positioning methods cannot goodly meet the requirements of indoor positioning and navigation.

So, it is very necessary to propose an efficient vision-based indoor positioning algorithm to improve the precision of vision positioning.

III. PIXEL THRESHOLD BASED EIGHT-POINT METHOD

The fundamental matrix is usually solved by the traditional eight-point method. However, the relative displacement of camera and the change in angle will cause inconsistent pixel coordinates of a pair of matching feature points on matched images, namely pixel drift. Ideally, the indoor environment at the time of positioning is the same as the indoor environment when the database is collected. The pixel drift of the feature points is not serious, and the error of solving the fundamental matrix is small. However, there exist many environmental factors that affect the positioning results during the process of actual positioning. Under normal circumstances, the indoor environment when the user collects images is different from the indoor environment when the database is collected. In such a case, the pixel coordinates of the partially matching feature points are distorted, and this is not an ideal pixel shift due to the relative displacement of camera and the change in angle. Obviously, the error will be produced using such pixel coordinates to solve the fundamental matrix. In this paper, the pixel threshold based eight-point method is proposed to adapt to the complex and variable indoor environment.

The above is the theoretical analysis for the traditional eight-point method with low robustness and poor adaptability. Taking the actual indoor environment as an example, the performance of the traditional eight-point method for solving the fundamental matrix in different experimental environments is compared and the reasons for different positioning precision are analyzed. The specific situation is shown in Fig. 1. It can be seen that different positioning results are obtained when indoor positioning is performed in different positioning scenarios using the same database and positioning algorithm in Fig. 1. The positioning precision is high if the positioning environment is the same as the environment when the database is established. The positioning result is not ideal when the positioning environment changes due to human factors.

The reasons of different positioning results obtained in different experimental environments are analyzed. Fig. 2 shows the eight pairs of matching feature points selected by the traditional eight-point method in different experimental envi-

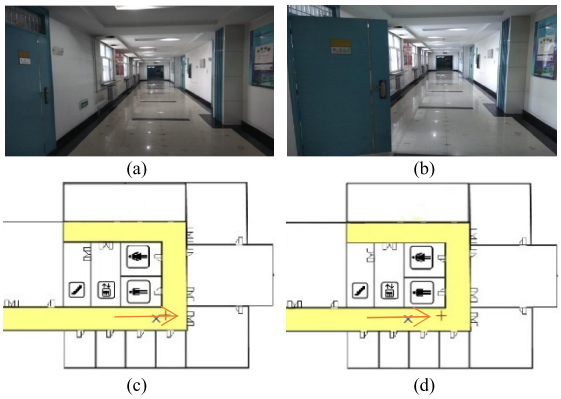

FIGURE 1. Positioning result in different scenes. (a) The image in scene 1; (b) The image in scene 2; (c) Positioning result in scene 1; (d) Positioning result in scene 2.

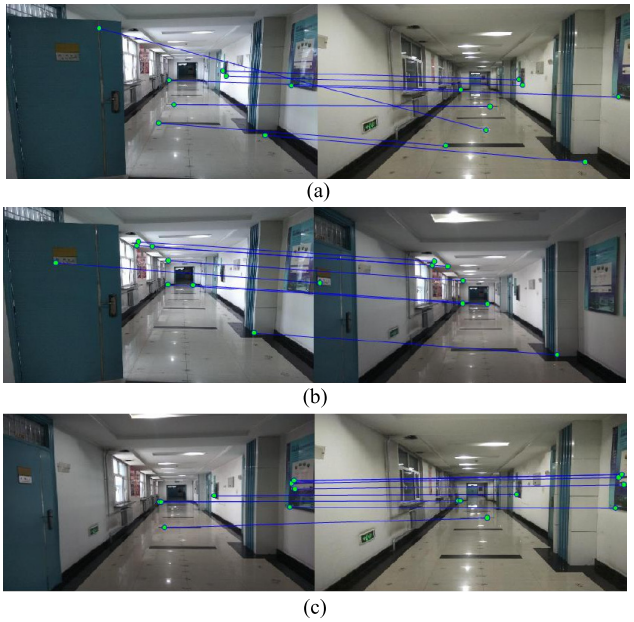

FIGURE 2. Selection of eight points method in different experimental environments. (a) Opening scene 1; (b) Opening scene 2; (c) Ideal experimental environment.

ronments. The feature points selected in the state of opening the door have significant pixel coordinate distortion compared to the feature points selected in the ideal environment. In Fig. 2(a), the feature point on the left door is matched with the feature point on the right tile, and it can be seen that the matching feature point produce a great pixel shift. The main reason is that the state of opening the door changes the environmental factors such as the intensity of light, it will cause great interference to the environment characteristics. Therefore, the feature point at the reflective spot on the metal sheet on the door is matched the feature point at the reflective point on the floor. The traditional eight-point method in Fig. 2(b) selects a pair of matching feature points at the nameplate on the door. The feature points on the door are displaced relative to the state of closing the door when the door of corridor is open. The traditional eight-point method selects such matching feature points to calculate the fundamental matrix, and there is no problem in terms of the feature point matching concept. The main reason is that the process of feature point matching is to find the feature point pairs with the smallest Euclidean distance. The traditional eight-point method also solves the fundamental matrix by selecting the first eight pairs of matching points with the smallest Euclidean distance. However, it can be seen from the Fig. 2(b) that such matching pairs have too much pixel drift relative to other good quality matching pairs.

In Fig. 2(a) and Fig. 2(b), the environmental factors causing the pixel coordinate distortion are different, but both have loopholes in the scheme of selecting eight pairs of matching points by the traditional eight-point method, and pixel distortion is introduced. The result is that the calculation accuracy of fundamental matrix is greatly affected. The ideal experimental environment is shown in Fig. 2(c), the positioning environment and the database collection environment have high similarity and less external interference in this case, so the eight pairs of matching points selected by the traditional eight-point method have less pixel drift. The error of solving the fundamental matrix by selecting such eight pairs of matching feature points is small.

It can be seen from the above analysis that the introduction of error is not the incorrect of feature point matching algorithm, but there are loopholes in the scheme of selecting eight pairs of matching points by the traditional eight-point method. The traditional eight-point method adopts the Euclidean distance as the sole criterion for selecting eight pairs of matching feature points. When the positioning environment changes, such as opening and closing of doors and windows, and object displacement, the traditional eight-point method can easily select matching feature points whose pixel coordinates are distorted. Actually, the pixel coordinate distortion of such feature points is caused by the changes in the positioning environment instead of the relative displacement or angle change between the cameras. Therefore, the selection criteria of eight pairs of matching feature points should be improved.

Aiming at the above problems, this paper proposes a pixel threshold based eight-point method, which improves the selection scheme of eight pairs of matching feature points. The proposed pixel threshold based eight-point method algorithm is shown in Table 1. The pixel threshold is added as the new selection criterion, and the quality of the matching feature point pairs is supervised. The main purpose is to avoid the presence of pixel coordinate distortion among the selected matching point pairs. In this paper, the introduction of pixel threshold can enhance the robustness of traditional eight-point method, thus making the positioning algorithm more suitable for complex and variable indoor scenes.

where φ_{query} denotes the query image, $\varphi_{\text{database}}$ denotes the image in the database, $\mathbf{P}_d$ represents the feature point in $\varphi_{\text{database}}$, $\mathbf{P}_q$ represents the feature point in φ_{query}, and θ represents the number of pairs matching feature points.

In Fig. 3, eight pairs of matching feature points selected by the pixel threshold based eight-point method are marked with

TABLE 1. Pixel threshold based eight-point method.

Algorithm I: Pixel threshold based eight-point method

1: Input: Image to be positioned φ_{query} and image in the database I_{database}
2: Output: Eight pairs of feature points $\mathbf{p}_{q_\theta}$ and $\mathbf{p}_{d_\theta}$, $1 \le \theta \le 8$ between φ_{query} and $\varphi_{\text{database}}$
3: Begin:
4: Extract image feature points of φ_{query} and $\varphi_{\text{database}}$, respectively
5: Perform feature point matching on the feature points extracted in the first step
6: Matching feature points are arranged in ascending order based on the Euclidean distance.
7: Calculate the pixel coordinates of rearranged matching feature points. If the pixel distance of matching points is greater than the pixel threshold, the pixel coordinates of matching feature point are distorted, and the matching feature points with distortion are eliminated; if it is smaller than the pixel threshold, the pixel drift is within an acceptable range, and such matching feature points are recorded.
8: Repeat step seven until the eight pairs of matching feature points with small pixel drift and minimum Euclidean distance is obtained as the output
9: end

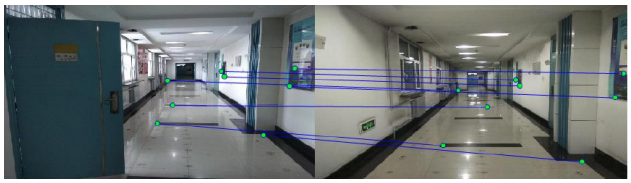

FIGURE 3. Eight pairs of matching feature points extracted by the pixel threshold based eight-point method.

green dots. It can be seen that the new eight-point selection criterion can eliminate the matching feature points with pixel coordinate distortion effectively. The selected eight pairs of matching feature points after introducing the pixel threshold are all feature points with excellent quality. Such eight pairs of feature points will have better performance for solving the fundamental matrix.

IV. QUERY CAMERA POSE ESTIMATION BY THE IMPROVED EPIPOLAR CONSTRAINT

The user's position is estimated using improved epipolar constraint after obtaining the optimal eight pairs of feature points. At first, the fundamental matrix $\mathbf{F}$ is estimated by eight pairs of high-quality points selected using the pixel threshold and the Euclidean distance. Secondly, the improved epipolar constraint is adopted, which reflects the pose relationship between the database camera and the query camera accurately. As shown in Fig. 4, $\mathbf{R}$ is rotation matrix, $\mathbf{t}$ is the translation vector, and they represent the relative position relationship between two cameras. $O_Q X_Q Y_Q Z_Q$ denotes the query database coordinate system and $O_D X_D Y_D Z_D$ denotes the database camera coordinate system.

A point in the scene is projected on the camera plane through the pinhole imaging, and a corresponding projection point is generated on the image plane. The point in the scene is represented by a vector $\mathbf{u}_{i,j} = [i, j]^{\mathrm{T}}$. A projection point

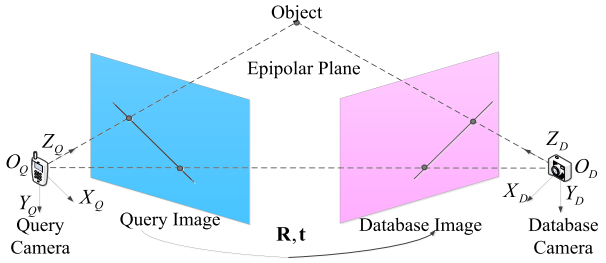

FIGURE 4. Epipolar constraint between database camera and query camera.

is generated on each of the two image planes, respectively when a point is projected onto two mutually different image planes, and there is a corresponding relationship between the two projection points. If two corresponding points or matching feature points are represented by $\left(\mathbf{u}_{i,j}, \mathbf{u}'_{i,j}\right)$, then this correspondence satisfies the epipolar constraints:

$$\mathbf{u}'^{\mathrm{T}}_{i,j} \mathbf{F} \mathbf{u}_{i,j} = 0 \tag{1}$$

where $\mathbf{F} = [f_{mn}]$ is the matrix of the 3×3 order, called the fundamental matrix, which contains the internal geometry and the relative orientation between two cameras. In addition, $\mathbf{F}$ is also constrained by (2):

$$\det \mathbf{F} = 0 \tag{2}$$

Let $\mathbf{x} = [i, j, i', j']$ denote the descriptor of the corresponding point $\left(\mathbf{u}_{i,j}, \mathbf{u}'_{i,j}\right)$, and then the solutions of (1) and (2) can be described as: a given set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of corresponding points of a pair of images and a meaningful cost function. On the one hand, the solution is used to represent any fundamental matrix $\mathbf{F}$ and each pair of corresponding points $\mathbf{x} = \mathbf{x}_m (m = 1, \ldots, n)$ deviates from the degree of formula (1). On the other hand, a fundamental matrix $\hat{\mathbf{F}} \neq 0$ is solved, and the cost function is minimized based on (2) at the same time. Since (1) and (2) have the same form when the two sides are multiplied by a non-zero scalar, $\hat{\mathbf{F}}$ is a proportionally obtained value. Excluding singular value constraints, we use a special cost function $J = J(\mathbf{F}; \mathbf{x}_1, \ldots, \mathbf{x}_n)$ to characterize the fundamental matrix $\hat{\mathbf{F}}$ that minimizes the cost function when satisfying the constraint equation:

$$\mathbf{F} = \arg\min \hat{J}(\mathbf{F}; \mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{3}$$

Assume

$$\begin{aligned} \hat{\mathbf{F}} &= \arg\min J(\mathbf{F}; \mathbf{x}_1, \ldots, \mathbf{x}_n) J(\mathbf{F}; \mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &= \|\mathbf{F}\|_F^{-2} \sum_{m=1}^{n} \left(\left(\mathbf{u}'_{i,j}\right)_m^{\mathrm{T}} \left(\mathbf{u}_{i,j}\right)_m^{\mathrm{T}} \right)^2 \end{aligned} \tag{4}$$

where $\|\mathbf{F}\|_F = \left(\sum_{m,n} f_{mn}^2\right)^{1/2}$ denotes the Frobenius norm of the fundamental matrix $\mathbf{F}$.

Define $\bar{\mathbf{u}}_{i,j}$ and $\bar{\mathbf{u}}'_{i,j}$ are centroids of $\left(\mathbf{u}_{i,j}\right)_m$ and $\left(\mathbf{u}'_{i,j}\right)_m$, respectively, which are shown as:

$$\begin{aligned} \bar{\mathbf{u}}_{i,j} &= \frac{1}{n} \sum_{m=1}^{n} \left(\mathbf{u}_{i,j}\right)_m \\ \bar{\mathbf{u}}'_{i,j} &= \frac{1}{n} \sum_{m=1}^{n} \left(\mathbf{u}'_{i,j}\right)_m \end{aligned} \tag{5}$$

Assume $\bar{\mathbf{u}}_{i,j} = \left[\bar{i}, \bar{j}\right]^{\mathrm{T}}$, $\bar{\mathbf{u}}'_{i,j} = \left[\bar{i}', \bar{j}'\right]^{\mathrm{T}}$, $(\mathbf{u}_{i,j})_m = [i_m, j_m]^{\mathrm{T}}$, $\left(\mathbf{u}'_{i,j}\right)_m = \left[i'_m, j'_m\right]^{\mathrm{T}}$, where $m = 1,\ldots,n$. The image coordinates of each corresponding point are represented by the centroid coordinates, the centroid coordinates of the m-th pairs of corresponding point can be expressed as $\left[i_m - \bar{i}, j_m - \bar{j}\right]^{\mathrm{T}}$, $\left[i_m{}' - \bar{i}', j_m{}' - \bar{j}'\right]^{\mathrm{T}}$. Then:

$$\begin{aligned} s &= \left(\frac{1}{2n}\sum_{m=1}^{n}\left\|\left(\mathbf{u}_{i,j}\right)_m - \bar{\mathbf{u}}_{i,j}\right\|^2\right)^{1/2} \\ &= \left(\frac{1}{2n}\sum_{m=1}^{n}\left(i_m - \bar{i}\right)^2 + \left(j_m - \bar{j}\right)^2\right)^{1/2} \\ s' &= \left(\frac{1}{2n}\sum_{m=1}^{n}\left\|\left(\mathbf{u}'_{i,j}\right)_m - \bar{\mathbf{u}}'_{i,j}\right\|^2\right)^{1/2} \\ &= \left(\frac{1}{2n}\sum_{m=1}^{n}\left(i_m{}' - \bar{i}'\right)^2 + \left(j_m{}' - \bar{j}'\right)^2\right)^{1/2} \end{aligned} \tag{6}$$

Then the normalized image plane coordinates can be expressed as:

$$\begin{aligned} \left(\tilde{\mathbf{u}}_{i,j}\right)_m &= \left[\left(i_m - \bar{i}\right)/s, \left(j_m - \bar{j}/s\right)\right]^{\mathrm{T}} \\ \left(\tilde{\mathbf{u}}'_{i,j}\right)_m &= \left[\left(i'_m - \bar{i}'\right)/s', \left(j'_m - \bar{j}'\right)/s'\right]^{\mathrm{T}} \end{aligned} \tag{7}$$

This definition ensures that the rms distance from $\left(\tilde{\mathbf{u}}_{i,j}\right)_m$ and $\left(\tilde{\mathbf{u}}'_{i,j}\right)_m$ to the origin of coordinate system where the corresponding point is equal to $\sqrt{2}$. Then, the distance from the plane origin can be replaced by the normalized distance $\left(\tilde{\mathbf{u}}_{i,j}\right)_m = \mathbf{T}\left(\mathbf{u}_{i,j}\right)_m$ and $\left(\tilde{\mathbf{u}}'_{i,j}\right)_m = \mathbf{T}'\left(\mathbf{u}'_{i,j}\right)_m$, where

$$\mathbf{T} = \begin{pmatrix} s^{-1} & 0 & -s^{-1}\bar{i} \\ 0 & s^{-1} & -s^{-1}\bar{j} \end{pmatrix} \quad \mathbf{T}' = \begin{pmatrix} s'^{-1} & 0 & -s'^{-1}\bar{i}' \\ 0 & s'^{-1} & -s'^{-1}\bar{j}' \end{pmatrix} \tag{8}$$

Let $\tilde{\mathbf{x}}_i = \left[\tilde{i}_m, \tilde{j}_m, \tilde{i}'_m, \tilde{j}'_m\right]^{\mathrm{T}}$, and define $\hat{\mathbf{F}}_{\mathrm{ALS}}$ as the fundamental matrix that minimizes the cost function after using the normalized value for the cost function. That is, the value of fundamental matrix when the function $\mathbf{F} \mapsto J\left(\mathbf{F}; \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n\right)$ reaches the minimum value. The expression defining the mapping relationship is:

$$\tilde{\mathbf{F}} = \left(\mathbf{T}'^{-1}\right)^{\mathrm{T}}\mathbf{F}\mathbf{T}^{-1} \tag{9}$$

Since $\tilde{\mathbf{u}}_{i,j} = \mathbf{T}\mathbf{u}_{i,j}$, $\tilde{\mathbf{u}}'_{i,j} = \mathbf{T}'\mathbf{u}'_{i,j}$, so $\left(\mathbf{u}'_{i,j}\right)^{\mathrm{T}}\mathbf{F}\mathbf{u}_{i,j} = \left(\tilde{\mathbf{u}}'_{i,j}\right)^{\mathrm{T}}\mathbf{F}\tilde{\mathbf{u}}_{i,j}$, thus (9) is derived. Then, the inverse mapping $\tilde{\mathbf{F}} \mapsto \mathbf{F}$ can be used to estimate $\mathbf{F}$ after $\hat{\mathbf{F}}_{\mathrm{ALS}}$ is obtained using the normalized data. The method for estimating the fundamental matrix $\mathbf{F}$ is expressed as $\hat{\mathbf{F}}_{\mathrm{HRT}}$, which can be obtained by:

$$\hat{\mathbf{F}}_{\mathrm{HRT}} = \mathbf{T}'^{\mathrm{T}}\hat{\mathbf{F}}_{\mathrm{ALS}}\mathbf{T} \tag{10}$$

The two smallest eigenvalues of the matrix are approximately equal when the ratio between the largest eigenvalue and the second smallest eigenvalue of a matrix is large, resulting in the instability between corresponding feature values, so $\hat{\mathbf{F}}_{\mathrm{HRT}}$ is introduced. That is, the corresponding feature vectors change greatly when the input of the matrix changes slightly. The above disadvantages can be effectively improved after the change of (10).

After the fundamental matrix is estimated by the above method, the essential matrix can be obtained:

$$\mathbf{E} = \mathbf{K_1}\mathbf{F}\mathbf{K_2} \tag{11}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ represent the camera's internal parameter matrix, respectively. The fundamental matrix $\mathbf{F}$ differs from the essential matrix $\mathbf{E}$, the essential matrix $\mathbf{E}$ contains only the relative orientation between two cameras after being multiplied by the camera's internal parameter matrix. The transfer vector $\mathbf{t}$ and the rotation matrix $\mathbf{R}$ between two cameras can be further determined according to the essential matrix $\mathbf{E}$. First of all, the sigular value decomposition of the essential matrix $\mathbf{E}$ results in $\mathbf{E}\sim\mathbf{U}\,\mathrm{diag}(1,1,0)\mathbf{V}^{\mathrm{T}}$, where $\det(\mathbf{U})>0$, $\det(\mathbf{V})>0$, then:

$$\begin{aligned} \mathbf{t} \sim \mathbf{t}_u &= [u_{13}, u_{23}, u_{33}]^T \\ \mathbf{R}_a = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}} \quad &\text{or} \quad \mathbf{R}_b = \mathbf{U}\mathbf{D}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}} \end{aligned} \tag{12}$$

There are four situations for the final solution:

$$\begin{aligned} \mathbf{P}_A &= [\mathbf{R}_a \,|\, \mathbf{t}_u], \quad \mathbf{P}_B = [\mathbf{R}_a \,|\, -\mathbf{t}_u], \\ \mathbf{P}_C &= [\mathbf{R}_b \,|\, \mathbf{t}_u], \quad \mathbf{P}_D = [\mathbf{R}_b \,|\, -\mathbf{t}_u] \end{aligned} \tag{13}$$

The point in the image must be in front of the camera according to the actual situation, so the unique solution can be obtained from the above four solutions based on this condition. In order to maintain the generality, this paper assumes that $\mathbf{P} = [\mathbf{R}_r \;\; |\mathbf{t}_r]$ is the final solution and sets $\mathbf{X}$ and $\mathbf{X}'$ as the coordinates of points in space in the database camera and query camera coordinate system, respectively. Therefore, there is:

$$\mathbf{X}' = \mathbf{R}_r\mathbf{X} + \mathbf{t}_r = \mathbf{R}_r\left(\mathbf{X} + \mathbf{R}_r^{-1}\mathbf{t}_r\right) \tag{14}$$

where $\mathbf{R}_r^{-1}\mathbf{t}_r$ denotes the transfer vector in the database camera coordinate system, denoted by $\mathbf{t}_d$. The vector passes the line between the database camera's optical center and the query camera's optical center. In other words, this vector represents the relative orientation (slope) between two cameras. Furthermore, it needs to be transformed into a world coordinate system to show the relative orientation of two cameras in the world coordinate system.

Assume that the coordinate of the reference point in the world coordinate system is $\mathbf{X}_g$, there is:

$$\mathbf{X} = \mathbf{R}\mathbf{X}_g + \mathbf{t} \tag{15}$$

where $\mathbf{R}$ is the absolute rotation matrix obtained by the database camera with reference to the world coordinate system, and $\mathbf{t}$ is the transfer vector obtained by the database camera with reference to the world coordinate system. Change (15) as follows:

$$\mathbf{X}_g = \mathbf{R}^{-1}\mathbf{X} - \mathbf{R}^{-1}\mathbf{t} \tag{16}$$

It can be seen that $-\mathbf{R}^{-1}\mathbf{t}$ denotes the conversion relationship between the transfer vector in the database camera coordinate system and the transfer vector in the world coordinate system. So, there is:

$$\mathbf{t}_{total} = -\mathbf{R}^{-1}\mathbf{t}_d = -\mathbf{R}^{-1}\mathbf{R}_r^{-1}\mathbf{t}_d \tag{17}$$

Assume that the database's absolute rotation matrix $\mathbf{R}$ is known, and then the vector $\mathbf{t}_{total}$ gives the direction relationship between the database camera and the query camera in the world coordinate system.

Assume that n images captured by the database camera at n position match with the query images, then a straight line between n database cameras' optical center and the query camera's optical center can be made by the transfer vector $\mathbf{t}_{total}$. The n straight lines intersects at one point, which is the position of the query camera's optical center, ie the user's position. However, there is an inevitable error in the estimation of the fundamental matrix $\mathbf{F}$, so that there exists an error in the estimation of the transition vector, thus causing the n straight lines not to reach a point. To solve this problem, we assume that N_i denotes the number of matching points between the i-th database image and the query image, and d_i denotes the distance from the i-th straight line. The optimal estimated position is obtained by solving the following problem:

$$\min_{x,y} \sum_i N_i d_i(x, y) \tag{18}$$

where

$$d_i(x, y) = \frac{|a_i x + b_i y + c_i|}{\sqrt{a_i^2 + b_i^2}} \tag{19}$$

where the linear equation of the i-th direction line is $a_i x + b_i y + c_i = 0$.

The above problem is essentially a convex optimization problem, so the optimal estimation point can be obtained by using the convex optimization theory. The optimal point is the final positioning result and the user's position.

V. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. EXPERIMENT ENVIRONMENT

The database adopted in this experiment is our own laboratory indoor scene, which is the 12th floor corridor of Building 2A, Science Park, Harbin Institute of Technology. There are 800 RGB images in the database, the image size is 224×224, and the floor plan is shown in Fig. 4. In addition, the length and width of this experimental corridor are 20m and 10m. For the convenience of the experiment, the lower right corner of the corridor in Fig. 5 is regarded as the coordinate origin of world coordinate system, and the direction of corridor is used as the X_w axis and Y_w axis of world coordinate system. At the same time, to evaluate the performance of proposed positioning algorithm, the experiment is performed in different indoor scenes, such as an office, a corridor, and a gymnasium.

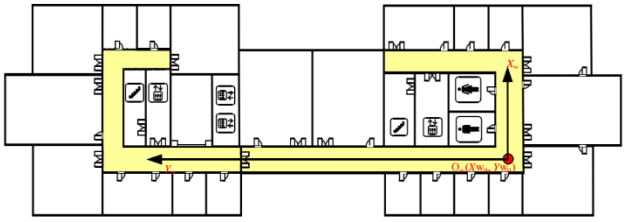

FIGURE 5. Floor plan for experiment in corridor scene.

B. PERFORMANCE EVALUATION OF PIXEL THRESHOLD BASED EIGHT-POINT METHOD

Ideally, the pixel coordinates of the matching feature points on the two camera pixel planes satisfy constraint relationship of (20). In the actual situation, the image noise is generated when the camera captures the signal, and there is no non-zero solution in (20) under the noise interference. At first, the fundamental matrix can be estimated by different methods. And then the epipolar error of fundamental matrix in (21) is adopted to calculate the solution error of the fundamental matrix:

$$\mathbf{p}_d{}^{\mathrm{T}}\mathbf{F}\mathbf{p}_q = 0 \tag{20}$$

Therefore, the epipolar error of fundamental matrix in this paper is defined as:

$$Err_\theta = \mathbf{p}_{d_\theta}{}^{\mathrm{T}}\mathbf{F}\mathbf{p}_{q_\theta} \tag{21}$$

where Err_θ represents the epipolar error of fundamental matrix solution for the θ-th pair matching feature points, $\mathbf{p}_{d_\theta}$ and $\mathbf{p}_{q_\theta}$ represent the pixel coordinates of the matching feature point pairs of images, respectively.

The traditional eight-point method, the pixel threshold based eight-point method and the traditional eight-point method + RANSAC are adopted to estimate the fundamental matrix, respectively. Then, 90 pairs of matching feature points are randomly selected to calculate the corresponding epipolar error, as shown in Fig. 6.

It can be seen from Fig. 6 that the epipolar error caused by the pixel threshold based eight-point method is significantly smaller than the epipolar error caused by the traditional eight-point method and the traditional eight-point method + RANSAC. The main reason is that the pixel threshold based eight-point method eliminate the matching feature points with severe pixel drift by using the pixel threshold, which improves the accuracy of fundamental matrix estimation.

The positioning performance analysis of pixel threshold based eight-point method is implemented in the 12th floor corridor. At the same time, the experimental scene is the same except that different fundamental matrix calculation methods. We can know that the positioning precision is greatly improved by using the pixel threshold based eight-point method in Fig. 7.

C. PERFORMANCE EVALUATION OF POSITIONING

The performance of each positioning algorithm is analyzed in the three cases. By calculating the Euclidean distances

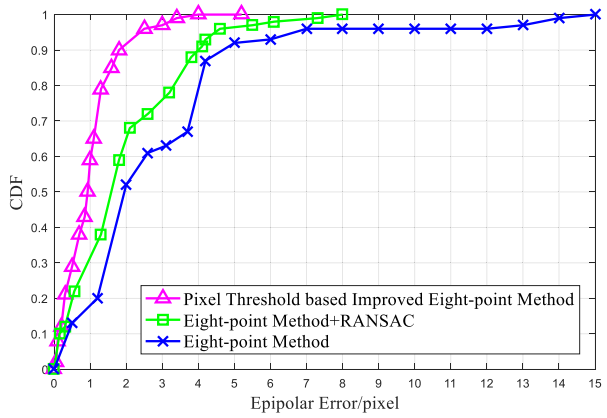

FIGURE 6. CDFs of fundamental matrix calculation epipolar error by various methods.

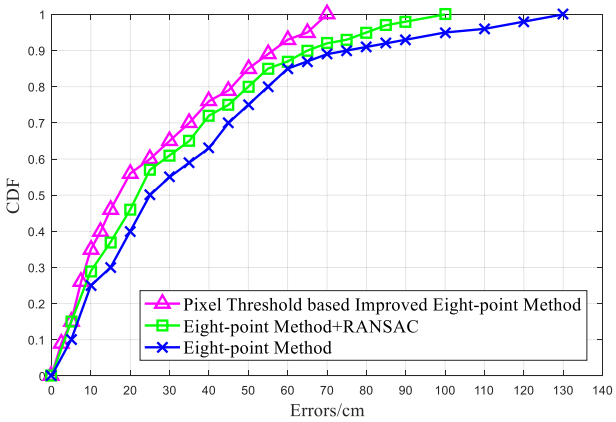

FIGURE 7. CDFs of position errors by various methods.

TABLE 2. Location errors of various positioning methods.

Scenes	Evaluation criterions	MIVIL	2DTriPnP	Proposed method
Office	Avg. (cm)	26.12	19.33	9.97
	Max. (cm)	112.77	91.55	50.02
	Impro. (%)	61.83	48.42	-
Gymnasium	Avg. (cm)	43.56	37.42	19.21
	Max. (cm)	131.46	111.86	59.42
	Impro. (%)	55.90	48.66	-
Corridor	Avg. (cm)	64.13	54.02	23.01
	Max. (cm)	149.23	131.26	70.31
	Impro. (%)	64.12	57.40	-

between the estimated and true locations of query camera, the location errors by various positioning algorithms are obtained. An accuracy improvement rate i_{im} is introduced to show the performance improvement of proposed algorithm:

$$i_{im} = \left(\left|e_p - e_c\right| / e_c\right) \cdot 100\% \tag{22}$$

where e_p and e_c denote the average errors of proposed algorithm and comparative algorithm. The experiment results are shown in Table 2. The abbreviations, i.e., Avg., Max., and Impro., are adopted to represent the average errors, the maximum errors and the improvement rates in this paper.

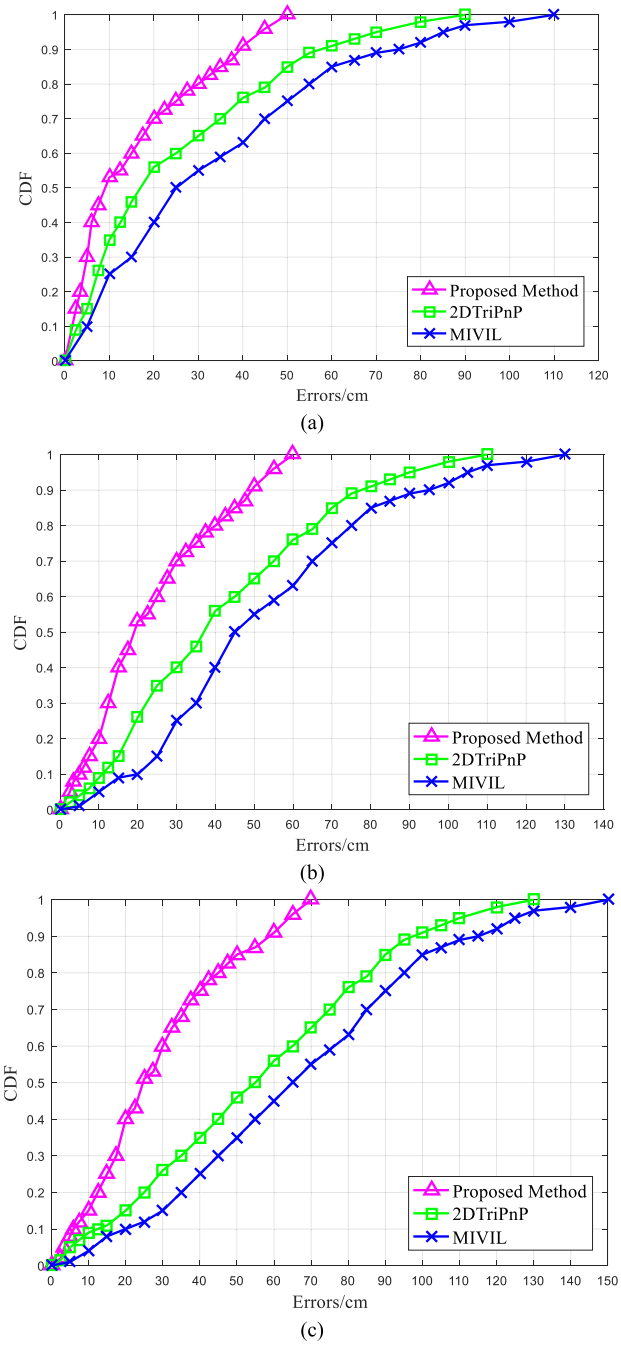

FIGURE 8. CDFs of position errors by various positioning methods. (a) CDFs of position errors in office scene; (b) CDFs of position errors in gymnasium scene; (c) CDFs of position errors in corridor scene.

As shown in Table 2, the performance of proposed algorithm evidently outperforms the other two algorithms in the three cases for the average positioning errors. The main reason is that the proposed algorithm takes advantage of both pixel threshold based eight-point method and improved epipolar constraint to estimate the position of query camera. It can be seen that the precision improvements of proposed algorithm can reach at least 48.42% in all experimental cases compared with the other algorithms.

The CDFs of the position errors by various positioning algorithms in the office scene, the gymnasium scene, and the corridor scene are shown in Fig. 8.

The maximum position errors of the proposed algorithm are limited within 50cm, 59cm, and 70cm in the three experimental scenes. Compared with the comparative algorithms, the precision of proposed positioning algorithm is improved by at least 52.88% and 45.36%, respectively. The main reason is that this paper proposes the pixel threshold based eight-point method to improve the accuracy of fundamental matrix calculation. At the same time, the improvement of traditional epipolar geometry has improved the positioning precision greatly.

VI. CONCLUSION

In this paper, a pixel threshold based eight-point method was proposed. The proposed algorithm added the pixel threshold as the new selection criterion to avoid the presence of pixel coordinate distortion among the selected matching feature point pairs, thus improving the accuracy of solving fundamental matrix. At the same time, the improved epipolar constraint was adopted to estimate the position of query camera. Performance simulation shows that the proposed algorithm has a signification improvement in terms of positioning precision.

REFERENCES

[1] C. Yang and H.-R. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.
[2] R. Faragher and R. Harle, "Location fingerprinting with Bluetooth low energy beacons," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2418–2428, Nov. 2015.
[3] J. Wu, Y. He, X. Guo, Y. Zhang, and N. Zhao, "Heterogeneous manifold ranking for image retrieval," *IEEE Access*, vol. 5, pp. 16871–16884, 2017.
[4] I. Gonzalez-Diaz, M. Birinci, F. Diaz-de-Maria, and E. J. Delp, "Neighborhood matching for image retrieval," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 544–558, Mar. 2017.
[5] K. Nagarathinam and R. S. Kathavarayan, "Moving shadow detection based on stationary wavelet transform and Zernike moments," *IET Comput. Vis.*, vol. 12, no. 6, pp. 787–795, Sep. 2018.
[6] L. M. Paz, P. PiniÉs, J. D. TardÓs, and J. Neira, "Large-scale 6-DOF SLAM with stereo-in-hand," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 946–957, Oct. 2008.
[7] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
[8] E. Deretey, M. T. Ahmed, J. A. Marshall, and M. Greenspan, "Visual indoor positioning with a single camera using PnP," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2015, pp. 1–9.
[9] H. Jegou, M. Douze, and C. Schmid, "On the burstiness of visual elements," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1169–1176.
[10] Z. Liu, L. Zhang, Q. Liu, Y. Yin, L. Cheng, and R. Zimmermann, "Fusion of magnetic and visual sensors for indoor localization: Infrastructure-free and more effective," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 874–888, Apr. 2017.
[11] H. Wei and L. Wang, "Visual navigation using projection of spatial right-angle in indoor environment," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3164–3177, Jul. 2018.
[12] X. Wang, L. Gao, and S. Mao, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 763–776, Jan. 2017.
[13] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. ICCV*, Barcelona, Spain, Nov. 2011, pp. 2564–2571.
[14] H. Sadeghi, S. Valaee, and S. Shirani, "A weighted KNN epipolar geometry-based approach for vision-based indoor localization using smartphone cameras," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2014, pp. 37–40.
[15] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based positioning in indoor environments," in *Proc. Int. Conf. Comput. Geosci. Res. Appl.*, 2016, pp. 1–5.
[16] G. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, E. Steinbach, "Mobile visual location recognition," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 77–89, Jul. 2011.
[17] Y. Zhang, L. Ma, and X. Tan, "Smart phone camera image localization method for narrow corridors based on epipolar geometry," in *Proc. IWCMC*, Paphos, Cyprus, Sep. 2016, pp. 660–664.
[18] H. Kawaji, K. Hatada, T. Yamasaki, and K. Aizawa, "Image-based indoor positioning system: Fast image matching using omnidirectional panoramic images," in *Proc. ACM Int. Workshop Multimodal Pervasive Video Anal.*, 2010, pp. 1–4.
[19] M. Werner, M. Kessel, and C. Marouane, "Indoor positioning using smartphone camera," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Sep. 2011, pp. 1–6.
[20] H. Hile and G. Borriello, "Positioning and orientation in indoor environments using camera phones," *IEEE Comput. Graph. Appl.*, vol. 28, no. 4, pp. 32–39, Jul./Aug. 2008.
[21] G. Lu, Y. Yan, N. Sebe, and C. Kambhamettu, "Indoor localization via multi-view images and videos," *Comput. Vis. Image Understand.*, vol. 161, pp. 145–160, Aug. 2017.
[22] H. Sadeghi, S. Valaee, and S. Shirani, "2DTriPnP: A robust two-dimensional method for fine visual localization using Google streetview database," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4678–4690, Jun. 2017.

SONGXIANG YANG received the bachelor's and M.Sc. degrees in communication engineering from Heilongjiang University, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, Harbin Institute of Technology. His current research interests include image processing, machine learning, and indoor positioning.

LIN MA (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 2003, 2005, and 2009, respectively, all in communication engineering. From 2013 to 2014, he has been a Visiting Scholar with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada. He is currently an Associate Professor with the School of Electronics and Information Engineering, Harbin Institute of Technology. His current research interests include location-based services, cognitive radios, and cellular networks.

SHUANG JIA received the bachelor's and M.Sc. degrees in communication engineering from Heilongjiang University, in 2014 and 2017, respectively. She is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, Harbin Institute of Technology. Her current research interests include image processing, machine learning, and indoor positioning.

DANYANG QIN received the B.Sc. degree in communication engineering, and the M.S. and Ph.D. degrees in information and communication system from the Harbin Institute of Technology, Harbin, China, in 2008 and 2011, respectively. She is currently a Professor with the Electronic Engineering College, Heilongjiang University, China. Her current research interests include wireless sensor networks, multihop routing, ubiquitous and visual sensing, machine learning, deep machine, crowdsensing, and indoor positioning.