Research Article

An Improved RANSAC Algorithm Based on Adaptive Threshold for Indoor Positioning

Jianan Bai,[1] Danyang Qin ⓘ,[1] Lin Ma ⓘ,[2] and Merhawit Berhane Teklu[3]

[1]*Heilongjiang University, Harbin 150080, China*
[2]*Harbin Institute of Technology, Harbin 150080, China*
[3]*Dire-Dawa University, Dire Dawa 1362, Ethiopia*

Correspondence should be addressed to Danyang Qin; qindanyang@hlju.edu.cn

The smart city is an important direction for the development of the highly information-based city, and indoor navigation and positioning technology is an important basis for the realization of an intelligent city. In recent years, indoor positioning technology mainly relies on WiFi, radio frequency identification (RFID), Bluetooth, and so on. Yet, the implementation of the above method requires the relevant equipment to be laid out in advance, and it is only suitable for indoor positioning with low accuracy requirements owing to interference and fading of the signal. The visual-based positioning technology can achieve high-precision positioning in enclosed, semienclosed, and multiwalled indoor environments with strong electromagnetic interference by means of epipolar geometry and image matching. The visual-based indoor positioning mostly uses the random sample consensus (RANSAC) algorithm to estimate the fundamental matrix to acquire the user's relative position. The traditional RANSAC algorithm determines the set of inliers by artificially setting a threshold to estimate the model. However, since the selection of the threshold depends on experience and prior knowledge, the reliability of the positioning results is not robust. Therefore, in order to improve the universality of the algorithm in indoor environments, this paper proposed an improved RANSAC algorithm based on the adaptive threshold and evaluated the real-time and accuracy of the algorithm by using an open-source image library. Results of the experiment show that the algorithm is more accurate than the traditional RANSAC algorithm in an enclosed and semienclosed multiwalled indoor environment, with fewer iterations.

1. Introduction

Smart city with its highly digital and intelligent features has been widely concerned with various industries; furthermore, indoor positioning and navigation technology are the fundamental part of the realization of smart city. Among several possible methods, due to the high accuracy and easy access, Global Positioning System (GPS) has been the most popular positioning method. However, GPS signal is only reliable and efficient in outdoor environments where a direct line-of-sight path can be established between the target device and the transmitting satellite [1]. In indoor environments, when a line-of-sight path cannot be established, due to the barrier or reflection of the wall, the GPS signal will be limited or even cannot be received, and the user's location information will not be provided. The corresponding alternative solution is to use other sensors, such as Bluetooth [2], WiFi [3], visible light positioning (VLC) [4], RFID [5], pedestrian dead reckoning [6], or camera [7, 8].

The Bluetooth-based indoor positioning system determines the user's location information through the received signal strength, but the complex indoor environment will cause reflection and refraction of the Bluetooth signal, which will affect the stability of transmission. In addition, the stability of the Bluetooth node itself is not strong. Although Bluetooth devices are relatively cheap and have strong spatial selectivity, they have the disadvantages of high latency [9] and limited accuracy.

Due to the continuous development of communication technology, WiFi devices have been widely deployed, and WiFi positioning has gradually become the most popular method in indoor positioning. Indoor positioning methods based on WiFi usually depend on received signal strength [10] and fingerprint technology [11]. Among the indoor positioning methods based on WiFi fingerprints, the CSI fingerprints positioning methods are easier to implement and more accurate. However, with the expansion of the fingerprint database, the training cost and processing complexity of CSI fingerprints will also greatly increase [12]. Generally, a positioning system that uses received signal strength has two components. One is a nearby anchor point whose location is known, and the other device is used for positioning. The WiFi-based indoor positioning system usually employs WiFi access points (APs) with multiple antennas as nearby anchors, while any mobile terminal with WiFi capability could be used as a positioning device. Because of the limited number of WiFi APs and the narrow system bandwidth (up to 40 MHz for 802.11n and 160 MHz for 802.11ac) [13], the electromagnetic signal itself is subject to strong interference in the indoor multiwalled environment and the WiFi positioning result depends on the accuracy of the signal strength graph. Therefore, the reliability of WiFi positioning results is not high, and it is generally only suitable for large-scale supermarkets with low-precision positioning requirements.

The vision-based indoor positioning technology relies on a priori map and feature descriptors for image retrieval and image matching [14], which can determine the location of query camera. In addition, visual positioning is an accurate and low-cost indoor positioning solution. It depends on the camera to collect the house structure information, texture differences, and static objects (doors and windows, etc.) from the environment to confirm the position, avoiding the reflection and refraction interference caused by the use of wireless electromagnetic signals when encountering obstacles. Consequently, it is feasible to adopt vision-based indoor positioning technology in an enclosed or semienclosed multiwalled indoor environment. In [15], the author proposed a continuous indoor positioning method and designed a positioning algorithm based on spatial constraint strategies. However, the method uses the traditional RANSAC algorithm to remove outliers, requiring multiple iterations and low accuracy of the manual threshold. In addition, it is necessary to constantly test and alter the threshold for changing indoor environments. In [16], the authors used an omnidirectional camera to develop an improved SLAM system for monocular vision. The ORB-SLAM framework is extended with the enhanced unified camera model as a projection function, but the traditional RANSAC algorithm is still used when calculating the fundamental matrix. It utilizes a number of iterations to estimate the best model, which will consume some time. In the literature [17], a method for precise indoor vision positioning of smartphone-based on a single image was proposed, which used PROSAC algorithm (an improved algorithm of RANSAC) to optimize the matching results of the correspondence. The PROSAC algorithm introduces the matching point evaluation function and sets the corresponding threshold to distinguish inliers or outliers and then fit the best model. Compared with the traditional RANSAC algorithm, the PROSAC algorithm has fewer iterations, but it still needs to artificially set the threshold and modify the threshold according to the changing environment. Until now, vision-based indoor positioning algorithms have typically used RANSAC and its deformations to estimate the fundamental matrix. Therefore, accelerating the convergence of RANSAC and setting a threshold to reduce the number of iterations become imperative.

To solve the problem of time consumption and unreliable results of fundamental matrix estimation caused by the RANSAC threshold setting, this paper proposed an adaptive thresholding algorithm to replace the traditional pervasive algorithm for the purpose of optimizing the results of the fundamental matrix and combined with the decomposition of the fundamental matrix to improve the localization results. Among them, the algorithm proposed by Meer et al. [18] can well adequate the shortcomings of traditional RANSAC. In this paper, we use this algorithm to calculate the fundamental matrix and then realize the localization.

The rest of this article is arranged as follows. Section 2 provides an overview of the research foundation. In Section 3, the proposed method is described in detail. Section 4 illustrates and discusses the experimental images, results, and analysis of the proposed method and other fixed-threshold methods. Finally, conclusions are presented in Section 5.

2. Epipolar Geometry and Fundamental Matrix

The vision-based indoor positioning method utilizes the epipolar constraint between the query image and the database image to determine the relative position of the query camera and the database camera. This relative relationship has nothing to do with the scene structure, but only with the internal and external parameters of the camera.

As shown in Figure 1, C is the query camera, C' is the database camera, X is a point in the scene, and the three points C, C', and X will constitute the epipolar plane. When a point in the space is projected onto two different image planes, an image point is generated on two image planes, respectively, and there will be some corresponding relationships between the two image points, which is called epipolar constraint. Under this epipolar constraint, the positional relationship between the query camera and the database camera can be represented by the rotation matrix $\mathbf{R}$ and the transfer vector $\mathbf{t}$. If the corresponding feature point is represented by (x_L, x_R), then the epipolar constraint is satisfied:

$$x_R^T \mathbf{F} x_L = 0, \tag{1}$$

where $\mathbf{F}$ is called the fundamental matrix which is a $3 * 3$ matrix with rank 2, satisfying the formula

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

The classical fundamental matrix estimation methods are the seven-point method and the eight-point method
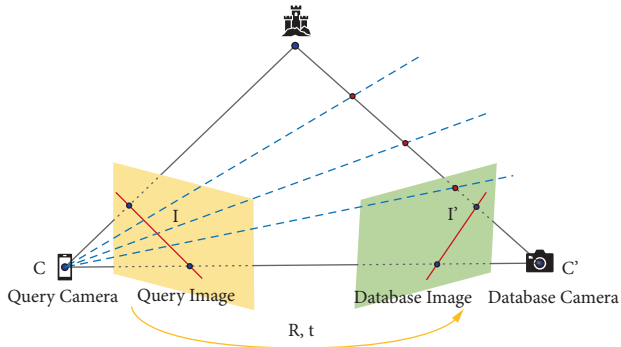
FIGURE 1: The epipolar constraint between the camera and the database camera.

proposed by Hartley, which use the corresponding points to solve the linear equation to achieve the purpose of estimating the fundamental matrix. Both the seven-point method and the eight-point method are linear methods. This method is less computationally intensive, but does not identify mismatches. Once the error matching point is used to calculate the fundamental matrix, the accuracy of the fundamental matrix will be greatly reduced. In addition, there are non-linear methods: iterative methods and robust algorithms. The iterative method relies on minimizing the ultimate geometric distance to estimate the fundamental matrix. Although it has high accuracy, the calculation is complicated and time consuming. In contrast, the robust algorithm is a method that has great anti-interference and can eliminate mismatches, and it is also the main method of estimating the fundamental matrix at present. Among them, robust algorithms are RANSAC and its improved algorithms LMedS (least median of squares), PROSAC (progressive sample consensus), NAPSAC (N-adjacent points sample consensus), etc.

LMedS uses maximum likelihood to calculate the model parameters and deviations for each subset and then selects the least deviated values among multiple sample subsets, and its corresponding model parameters are used as the estimated model parameters, also known as the fundamental matrix. LMedS does not need to set too many parameters, but all samples need to participate in the final model estimation, and the participation of outliers will make the estimation result worse. Therefore, when the outlier ratio is greater than 50%, this method will not get the ideal result [19]. PROSAC assumes that the description similarity between the corresponding relationships in the inliers is higher than that of the outliers, so the samples are sorted according to the description distance between each pair of matching points. These descriptions include SIFT (scale-invariant feature transform), Harris, SURF (speeded-up robust features), ORB (oriented FAST and rotated BRIEF), etc., and then prioritize the ahead matching points for model estimation [20]. Although this algorithm can enhance the sampling probability of correct data, thereby reducing the number of iterations of the algorithm and improving timeliness, when the relevant description of feature points is lacking, the algorithm will degenerate to ordinary RANSAC [21].

3. Indoor Positioning Method Based on Adaptive Threshold RANSAC

3.1. Adaptive Threshold RANSAC with Similar Slopes of Feature Points. The basic idea of the traditional RANSAC algorithm is as follows: (1) Randomly extract the smallest number of samples that can calculate the model parameter, namely, the fundamental matrix $\mathbf{F}$, from the sample set S with N data and an outlier ratio of v. (2) Calculate model parameters from the samples. (3) The parameters are back-substituted to all data samples and the inlier ratio is counted. If the current inlier ratio is the largest, the model is determined as the current optimal model. (4) If the inlier ratio of the current optimal model is bigger than the set threshold or the number of iterations is greater than the predetermined number, the iteration stops. Otherwise, the above steps are repeated. (5) Output the current optimal model.

The maximum number of iterations N guarantees that at least one group of sampled data is all inliers under a certain confidence probability. The calculation formula of N is deduced as follows:

$$N = \frac{\log(1-p)}{\log(1-(1-v)^m)}, \tag{3}$$

where m is the minimum number of data samples that can calculate the model parameters; p is the confidence set in advance, that is, sampling N times can guarantee the possibility of p so that the samples drawn are all inliers. The relationship between p, m, N, and v will be satisfied:

$$1 - (1-(1-v)^m)^N = p. \tag{4}$$

Equation (5) is used as a criterion for discriminating inliers, where (x_1, x_2) is a feature point, and only those with an error less than the threshold can be added to the inliers set S_i.

$$d(x_1, x_2) = \frac{(x_2 \mathbf{F} x_1)}{(\mathbf{F} x_1)_x^2 + (\mathbf{F} x_1)_y^2 + (x_2^T \mathbf{F})_x^2 + (x_2^T \mathbf{F})_y^2}. \tag{5}$$

The following results can be obtained from the traditional RANSAC algorithm flow: (1) there is no upper bound for the number of iterations to compute the parameters. The result of the algorithm depends on the number of iterations. If the number of iterations is insufficient, the result obtained may not be optimal or even wrong. Besides, RANSAC only has a certain probability of obtaining a plausible model, which is positively related to the number of iterations. (2) It is required to set thresholds related to the problem. The mathematical model does not support threshold setting and can only rely on human experience. In addition, a single threshold has no strong universality.

Instead of relying on the fixed threshold in the above-mentioned RANSAC algorithm, the method used in this paper is an adaptive threshold RANSAC. It can generate an adaptive threshold to distinguish inliers and outliers, optimize the sample selection strategy, and stop the iteration through the final sampling strategy, thereby reducing the

number of iterations. Literature [22] gives the general formula of this optimized sample selection strategy

$$\text{NFA}\left(\{\epsilon_j: i = 1, \ldots, n\}, k\right) = N_{\text{outcomes}}\left(n - N_{\text{sample}}\right)\binom{n}{k}\binom{k}{N_{\text{sample}}}\left(\epsilon_k^d \alpha_0\right)^{k - N_{\text{sample}}}, \tag{6}$$

where ϵ_j are the errors of the correspondences of the fundamental matrix, increasingly ordered, varying k represents the numbers of inliers among n point correspondences. $N_{\text{sample}} = 7$ while using the seven-point algorithm, since the seven-point algorithm will produce three fundamental matrixes and accordingly $N_{\text{outcomes}} = 3$. The error is the distance from a point to a line, which is a one-dimensional mathematical model, so $d = 1$. The best model parameter estimation, that is, the fundamental matrix can be estimated using

$$\mathbf{F} = \text{argmin}_{\mathbf{F}} \min_{k=8,\ldots,n}\left(\{\epsilon_j(\mathbf{F})\}, k\right). \tag{7}$$

Note that the distance from a point to a straight line is defined as

$$\epsilon_F(x_L, x_R) = \frac{\left|x_L^{\text{T}} \mathbf{F} x_R\right|}{\sqrt{\left(x_L^{\text{T}} \mathbf{F} \mathbf{e}_1\right)^2 + \left(x_L^{\text{T}} \mathbf{F} \mathbf{e}_2\right)^2}}, \tag{8}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the first two vectors of the standard base of 3-row vectors.

The details of the adaptive algorithm used in this paper are shown in Algorithm 1, where bestMod is the fundamental matrix, vector_inr is the inliers set, and persion is the threshold. The algorithm can be roughly summarized into three steps: first, use the RANSAC algorithm to perform several random trials to generate corresponding matches. Second, calculate the residual for each correspondence and continuously update the inliers set according to the sample selection strategy. Lastly, obtain the adaptive threshold and the best model through the final sampling strategy. Since large binomial coefficients will be generated when calculating NFA, the logarithmic form is used instead:

$$\log(\text{NFA}) = \log N_{\text{outcomes}}\left(n - N_{\text{sample}}\right) + \log\binom{n}{k}$$
$$+ \log\binom{k}{N_{\text{sample}}} + \left(d \log \varepsilon_k + \log \alpha_0\right) \tag{9}$$
$$\times \left(k - N_{\text{sample}}\right).$$

To facilitate understanding, equation (6) can also be written as the product of two terms: a number of tests,

$$\text{NFA}(S) = N(k, n) \times P(k, \tau), \tag{10}$$

which calculates all possible couplings (F, S) and an upper bound $P(k, \tau)$.

$$N(k, n) = 3(n - 7)\binom{n}{k}\binom{k}{7}. \tag{11}$$

Assuming that the image points are uniformly distributed, this upper limit can be expressed as

$$P(k, \tau) = \left(\alpha_0 \tau\right)^{k - 7}, \tag{12}$$

where α_0 is an upper bound and τ is the normalized error threshold. The model corresponding to the minimum NFA is the best model. In practice, if NFA is less than 1, it is considered that an effective fundamental matrix has been found; that is, in $\text{NFA}(S) < \varepsilon$ chooses the value $\varepsilon = 1$, which is also the final sampling strategy mentioned above. The classical RANSAC algorithm dynamically adjusts the number of iterations according to equation (3), where the outlier ratio $v = 1 - k/n$, in which k is unknown at the beginning, and the RANSAC algorithm sets the number of iterations as much as possible according to the amount of calculation that can be undertaken. When the inlier ratio is high, the ideal model can be estimated with fewer iterations, but when the inlier ratio is less than 50%, the number of iterations required will increase exponentially. Assuming that there are k inliers among n matches, the number of samples to be drawn to have a probability q of selecting the minimal matching sample of inliers is

$$N = \frac{\log(1 - q)}{\log\left(1 - \prod_{i=0}^{6} k - i/n - i\right)}. \tag{13}$$

When $q = 95\%$, $N \approx 3(n/k)^7$, which allows the outlier ratio of $1 - k/n$ to be approximately 70% to be maintained within a reasonable calculation time (10,000 trials). This number is relatively large. Moreover, it is impossible to experiment such a large number of samples in practice. In contrast, the algorithm in this paper converges faster. Once it is found that the final sampling strategy $NFA < 1$ is established, it will stop immediately. The number of tests can be as small as $N/10 + 1$ and never exceed N.

Although the above method accelerates the convergence, it can only converge the sample set to a local minimum. Therefore, to ensure an accurate result, the evaluation of the feature point location relationship was added in this paper. The orientation relationship of two matching points in the query image is always similar to the orientation relationship of the corresponding matching points in the database image. According to this characteristic, suppose (Q_i, Q_j) and (D_i, D_j) are two matching pairs of query image and database image, respectively. Then, $k(Q_i, Q_j)$ and $k(D_i, D_j)$ represent the slope. In theory, the two slopes should be the same. Moreover, this paper has used the adaptive RANSAC with similar slopes of feature points to calculate the fundamental matrix. Algorithm 2 describes the process of similar slope evaluation.

INPUT: two images and n correspondences
  //Initialization
(1) vector_inx = [1:n]
(2) bestMod = $\varnothing$
(3) vector_inr = $\varnothing$
(4) minNFA = ∞
(5) nIter=(nIterRe = nIter/10)
(6) nData = x1.ncol ()
(7) errorMax = 0
        //Main estimation loop
(8) **for** *Iters = 0* to *nIters* **do**
(9)   vector_sp = USample (size_Sample, vector_inx)
(10)   vector_mods = Fit (vector_sp)
(11) **end for**
        //Evaluate Models
(12) better = false//Whether one of the tested models improves the NFA
(13) **for** *k = 0* to *vector_mods.size* **do**
(14)   error = Error (vector_mods[k])
        //Residuals computation and ordering
(15)   **for** *i = 0* to *nData* **do**
(16)   error = Error (vector_mods[k])
(17)   vector_res = ErrorIndex (error)
(18)   sort (vector_res)
(19)   **end for**
(20) **end for**
   //Statistical detection of the best meaningful subset (inliers/outliers)
(21) ErrorIndex.best = bestNFA
        //Find a better model
(22) **if** (best.error < minNFA), **then**
(23)   minNFA = best.error
(24)   vector_inr.resize (best, inx)
(25) **end if**
(26) **for** *i = 0* to *best.index* **do**
(27)   vector_inr[i] = vector_res[i].index
(28)   errorMax = vector_res[best.index-1].error//Error threshold
(29)   bestMod = vector_mod[k]
(30)   better = true
(31)   precision = denormalizeError (errorMax)
(32) **end for**
        //Optimization
(33) **If** (*better* and *minNFA* < 0) or (*Iter* + 1 = *nIterRe*),
   then
(34)   vector_inx = vector_inr
(35)   **if** (nIterRe), **then**
(36)   nIter = Iter+1+nIterRe
(37)   nIterRe = 0
(38)   **end if**
(39) **end if**
(40) **return** bestMod, vector_inr, minNFA, precision

ALGORITHM 1: Improved algorithm based on RANSAC.

(1) **For** $(Q_i, Q_j) \in Q$, $(D_i, D_j) \in D$ **do**
(2) k_Q = slope (Q_i, Q_j);
(3) k_D = slope (D_i, D_j);
(4)   **if** S (i, j) = similarity $(k\_Q, k\_D) < 0.6$, **then**
(5)   remove
(6)   **end if**
(7) **end for**

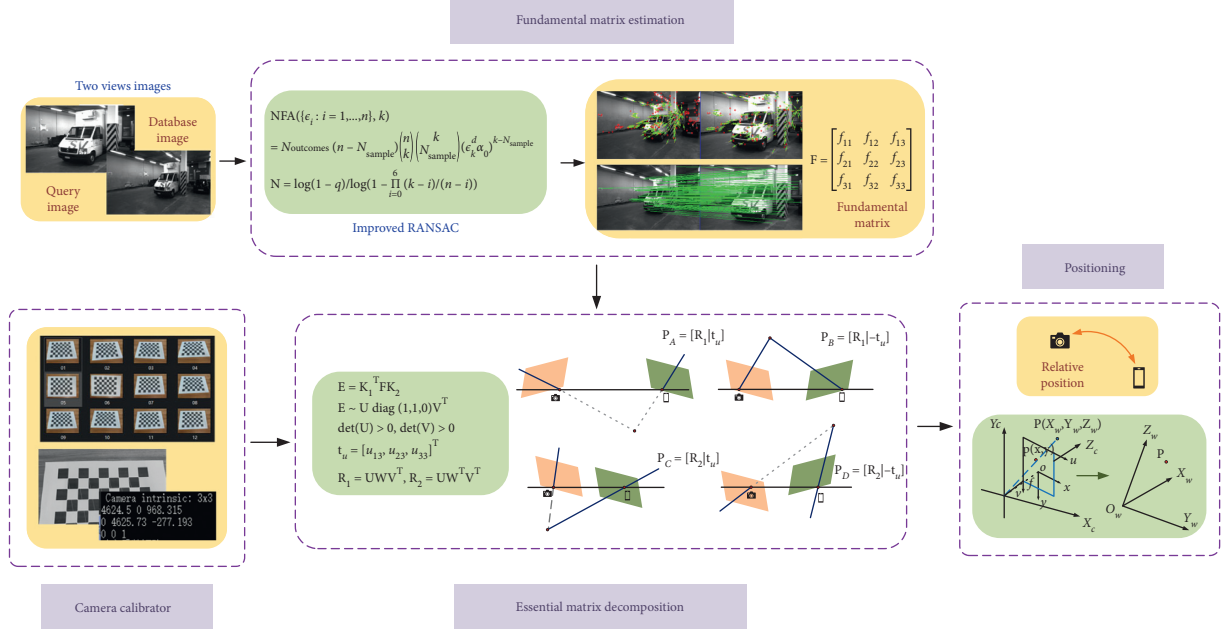ALGORITHM 2: Slope similarity evaluation.

FIGURE 2: Overall workflow of visual-based indoor positioning.

3.2. Estimate Location Based on Fundamental Matrix.

Figure 2 shows the detailed process of positioning based on the fundamental matrix. The input query image and the corresponding database image constitute an image pair. The SIFT (scale-invariant feature transform) is used to detect the feature points of the two images and to find and draw the corresponding matching points, and then, the fundamental matrix is computed according to Section 3.1.

However, due to the camera lens distortion and the degree of distortion not being the same, it is necessary to complete the camera calibration. Camera calibration can be seen as a process of constant transformation of the coordinate system: a point in space could be regarded as the point in the world coordinate system, so the position of the camera can be described in the world coordinate system. The origin of the camera coordinate system is the optical center of the camera. The transformation from the world coordinate system to the camera coordinate system is a rigid transformation, and the object will not deform during the process. The transformation from camera coordinate system to image coordinate system is a process of pinhole imaging. The pixel coordinate system reflects the arrangement of pixels in the camera chip, and it is a two-dimensional rectangular coordinate system. A point (x_w, y_w, z_w) in space can be converted to pixel coordinates in world coordinates and a point (u, v) can be represented as

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{M}_1 \mathbf{M}_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \qquad (14)$$

where s is a scale factor that is not equal to 0, $\mathbf{M}_1$ is the camera's internal parameter matrix, and $\mathbf{M}_2$ is the camera's external parameter matrix. When the camera is calibrated, it is necessary

to shoot the checkerboard from different angles, and when the number of calibrated images is greater than 10, the accuracy of the calibration result can be guaranteed. In this paper, the Redmi K20 Pro mobile phone was used to shoot the chessboard during the calibration, and a total of 12 calibration images with a size of 1600*1400 were obtained. After the calibration image shown in Figure 3 was known, the camera would be calibrated using Zhang's method to acquire the camera's internal parameter matrix and distortion parameter matrix.

When the fundamental matrix and the camera internal parameter matrix are known, the essential matrix can be calculated according to

$$\mathbf{E} = \mathbf{K}_1^{\mathrm{T}} \mathbf{F} \mathbf{K}_2, \qquad (15)$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ represent the internal parameter matrix of query camera and database camera, respectively [23]. The difference between the fundamental matrix and the essential matrix is that after multiplying the internal parameter matrix of the camera, the essential matrix $\mathbf{E}$ only contains the relative direction relationship between the two cameras. According to the essential matrix $\mathbf{E}$, the rotation matrix $\mathbf{R}$ and transfer vector $\mathbf{t}$ between the two cameras can be further obtained. First, the singular value decomposition of the essential matrix $\mathbf{E}$ results in $\mathbf{E} \sim \mathbf{U}\mathrm{diag}(1,1,0)\mathbf{V}^T$, where $\det(\mathbf{U}) > 0$, $\det(\mathbf{V}) > 0$, then

$$\mathbf{t}_u = [u_{13}, u_{23}, u_{33}]^T. \qquad (16)$$

The rotation matrix $\mathbf{R}$ has two forms, $\mathbf{R}_1 = \mathbf{UWV}^T$ and $\mathbf{R}_2 = \mathbf{UW}^T\mathbf{V}^T$, where $\mathbf{W}$ is an orthogonal matrix:

$$\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (17)$$
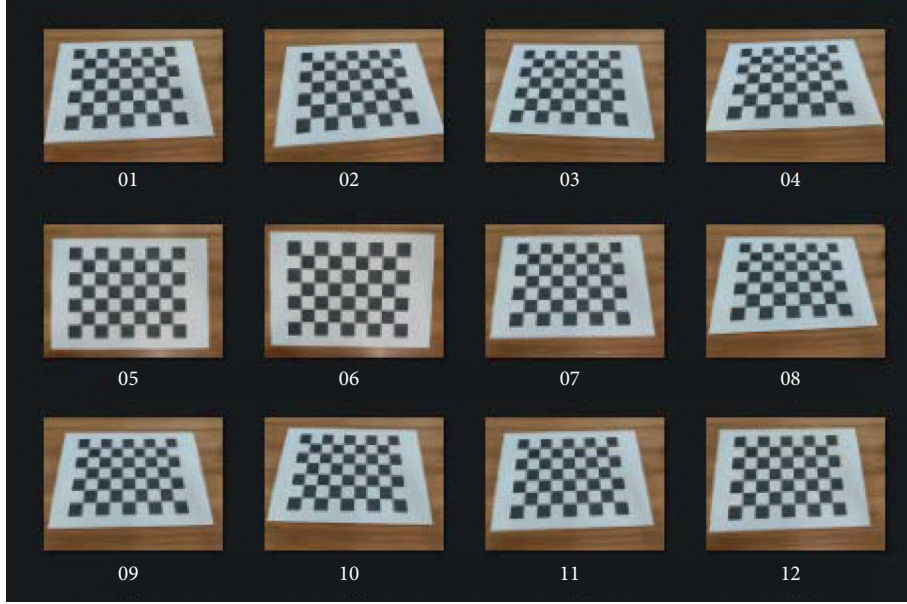
Figure 3: Calibration image collected by Redmi K20 Pro.

However, since the value of $\mathbf{t}_u$ cannot be determined, the following four final solutions will be produced:

$$\mathbf{P}_A = [\mathbf{R}_1 \mid \mathbf{t}_u], \mathbf{P}_B = [\mathbf{R}_1 \mid -\mathbf{t}_u], \mathbf{P}_C = [\mathbf{R}_2 \mid \mathbf{t}_u], \mathbf{P}_D = [\mathbf{R}_2 \mid -\mathbf{t}_u]. \tag{18}$$

According to the actual situation, the point in the image must be in front of the camera, so only one of the above four solutions is correct. Assume that $\mathbf{P} = [\mathbf{R}_c \mid \mathbf{t}_c]$ is the correct one, and that X and X' are the coordinates of the points, which are in the coordinate system of the database camera and the query camera, respectively. Then, there is

$$X' = \mathbf{R}_c X + t_c = \mathbf{R}_c \left(X + \mathbf{R}_c^{-1} t_c \right), \tag{19}$$

where $\mathbf{R}_c^{-1} \mathbf{t}_c$ represents the transfer vector in the database camera coordinate system and is represented by $\mathbf{t}_r$. The relative direction relationship between the database camera and the query camera can be represented by this vector. Assume that the world coordinate of the reference point is X_w, then there is

$$X = \mathbf{R} X_w + \mathbf{t}, \tag{20}$$

where $\mathbf{R}$ is the absolute rotation matrix, and $\mathbf{t}$ is the transfer vector. Make the following changes to equation (20):

$$X_w = \mathbf{R}^{-1} X - \mathbf{R}^{-1} \mathbf{t}, \tag{21}$$

where $-\mathbf{R}^{-1}\mathbf{t}$ is a conversion relationship, which is from database camera coordinate system to world coordinate system. Hence,

$$\mathbf{t}_{\text{total}} = -\mathbf{R}^{-1} \mathbf{t}_r = -\mathbf{R}^{-1} \mathbf{R}_c^{-1} \mathbf{t}_c. \tag{22}$$

If the absolute rotation matrix $\mathbf{R}$ is known, the vector t_{total} can represent the orientation relationship between the two cameras.

By using Algorithm 3, the position information of random points in the space can be determined.

4. Experiment and Discussion

4.1. Adaptive Threshold RANSAC to Estimate Fundamental Matrix. Due to the changing indoor environment, different image resolutions, and uncertain image quality (blurred images due to light brightness and relative motion, etc.), this article chooses ETH3D as the experimental object. The open-source ETH3D image library contains indoor scene images with dual (multiview) views, different resolutions, and even distortion. According to the proposed improved algorithm, the fundamental matrix is estimated, and the real-time performance and adaptability of the algorithm are analyzed when facing the above scenes.

Considering the effectiveness of the proposed algorithm in low-resolution images in this paper, experiments were conducted using the ETH3D low-resolution image set and the result was that the proposed algorithm could calculate the fundamental matrix of low-resolution image pairs. Three groups with different background consistency and foreground target profiles were selected as representatives, and the results of the statistical calculation method were plotted in Table 1. As shown in Table 1, Case B has the strongest background consistency, followed by Case A and finally Case C. Case A has the clearest outline of the foreground target, followed by Case C and finally Case B. The experiment also found that in images with strong background consistency, the inliers had high aggregation, which was beneficial to the selection of samples.

Observing Table 1, it can be concluded that the improved algorithm in this paper is capable of generating adaptive threshold with a threshold accuracy of 0.00001 pixels, making it more applicable in changing indoor environments. In addition, the number of iterations has been significantly reduced to better satisfy the user's requirements for real-time when positioning. For the three scenarios in Table 1, the number of iterations of adaptive threshold RANSAC was reduced by approximately 38.65% (compared

```
INPUT: imageL, imageR, F, K1, K2
(1)     Matrix_E = K1'*F*K2;
(2)     Vector_t, Matrix_R = recover (Matrix_E);
(3)     Vector_Keypoints, Mathes = feature_extract_match (imageL, imageR)
(4) keypoints1 = detectore (imageL);
(5) keypoints2 = detectore (imageR);
(6) points1 = Pointchange (keypoints1);
(7) points2 = Pointchange (keypoints2);
(8) space_points = triangulation (keypoints1, keypoints2, Matches, Matrix_R, Vector_t)
(9) for i = 0 to matches.size do
(10)      points1_cam = pixel2cam (keypoints1);
(11)      points2_cam = pixel2cam (keypoints2);
(12)      points2_trans = Matrix_R *(points[i].x, points[i].y, points[i].z) + Vector_t;
(13) end for
OUTPUT: location
```

ALGORITHM 3: Indoor positioning method based on adaptive threshold RANSAC.

TABLE 1: Results for adaptive and fixed thresholds (resolution 711 ∗ 435, SIFT ratio = 1).

Case (arranged from top to bottom corresponding to Case A, B, and C)	Threshold (pixel)	Average error (pixel)	Inliers/ outliers	Number of iterations
	Arbitrary	0.536736	92/165	307
	1	0.707051	76/184	816
	2	0.675245	91/166	702
	5	0.988990	88/169	559
	Arbitrary	0.548608	148/156	616
	1	0.693217	117/187	827
	2	0.615778	140/164	995
	5	0.938899	141/163	803
	Arbitrary	0.569602	76/177	638
	1	0.551904	71/182	887
	2	0.574652	77/176	905
	5	0.667092	73/180	863

to RANSAC with a threshold of 1 pixel), 41.29% (compared to RANSAC with a threshold of 2 pixels), and 31.48% (compared to RANSAC with a threshold of 5 pixels), respectively, compared to the other comparative methods. Although the average error of the fixed-threshold RANSAC algorithm is sometimes smaller than that of the proposed algorithm, it needs to constantly change the threshold manually, which is complicated to operate and is not universally applicable to changing indoor environments.

In the actual image acquisition process, image blur due to relative motion or poor image quality due to low light often occurs. Therefore, this paper uses the distorted image library in the ETH3D image library to test the influence of the blurred image on the algorithm proposed in this paper. Figures 4 and 5 are respectively partial distorted images in ETH3D and the epipolar diagrams after using the algorithm. The above experiments show that the resolution of the image and the distortion of the

image do not affect the method to calculate the adaptive threshold and the fundamental matrix.

4.2. Query Camera Pose Estimation by the Improved RANSAC Algorithm. In order to evaluate the use of the proposed algorithm in practice, this paper uses the 3rd floor of the laboratory building of Heilongjiang University as an experimental environment (shown in Figure 6) to assess the method proposed in this paper. Considering the texture differences of the rooms (low texture and repetitive patterns), Lab 304A and Conference Room 306 were analyzed separately as examples. For the test, the database camera and the query camera were identical, both being the Redmi K20 Pro, and the images were captured with the camera at a height of 155 cm from the ground and an elevation angle of 0. The image database was acquired with 125 images of Lab 304A and 120 images of Conference Room 306, with an

FIGURE 4: Partially distorted images from the ETH3D image library (image size 752 ∗ 480).
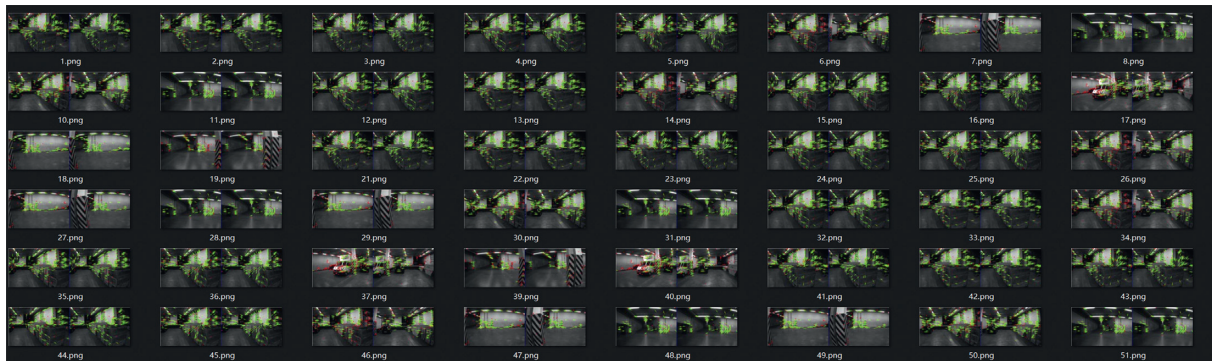

FIGURE 5: Distribution of inliers and outliers after using the algorithm.
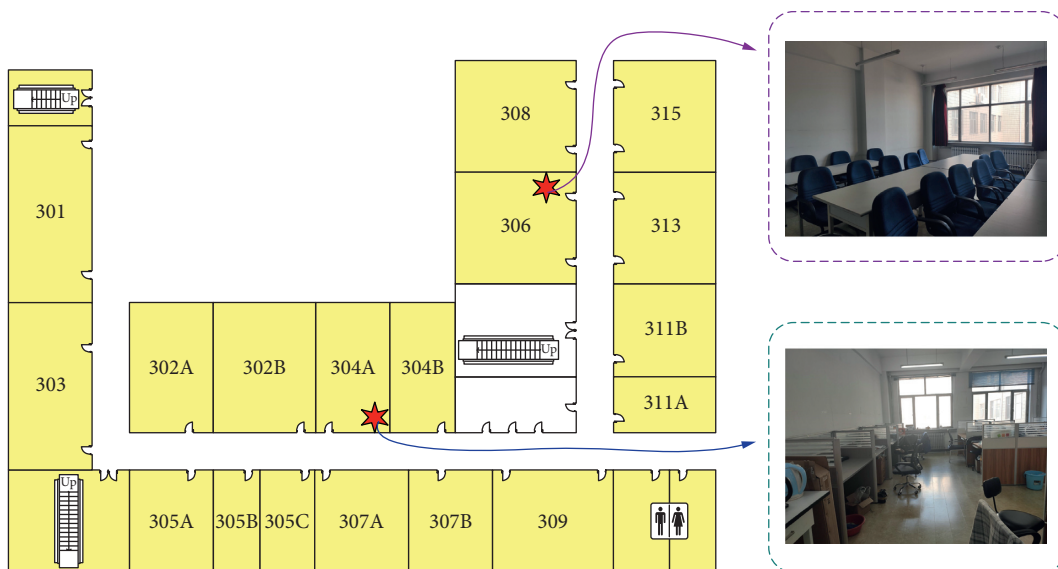
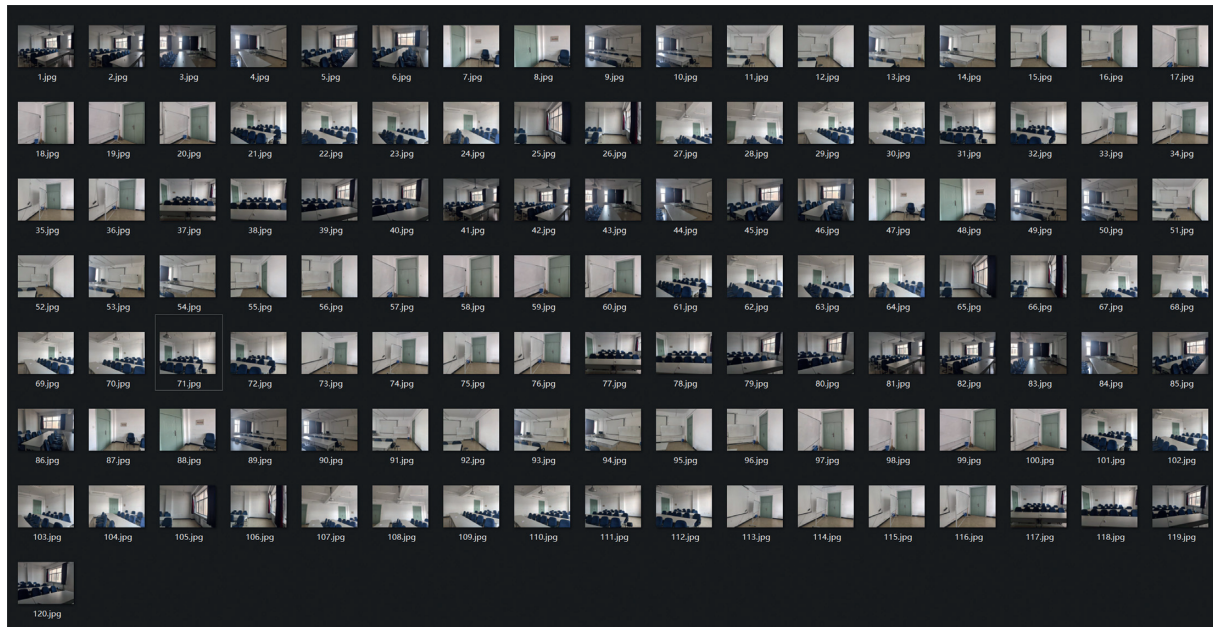

FIGURE 6: Floor plan of the experimental environment.

FIGURE 7: Database images (Room 306).

image size of $1600 * 1200$. Figure 7 shows a selection of database images.

As with the previous analysis of algorithm performance using the ETH3D image library, the proposed algorithm was used to estimate the fundamental matrix in Room 304A and Room 306, respectively. The camera external parameters (rotation matrix $\mathbf{R}$ and transfer vector $\mathbf{t}$) are deduced from Section 3.2. Knowing the position of the database camera, then the rotation matrix $\mathbf{R}$ and the transfer vector $\mathbf{t}$ can well reflect the position information of the query camera. As the root mean square error (RMSE) in experiments can be a good indicator of measurement accuracy, this paper uses the RMSE for accuracy assessment, which is mathematically expressed as

$$\text{RMSE} = \sqrt{\frac{\sum_i \left(\left(x_L^i \right)^{\text{T}} \mathbf{F} x_R^i \right)^2}{n}}, \tag{23}$$

where (x_L^i, x_R^i) indicates the key point correspondence.

In the experiments, the number of feature points will affect the speed of image matching and thus impact the real-time performance of localization; second, a smaller number of feature points will reduce the accuracy of the fundamental matrix, which will cause the localization error to increase. Considering the above reasons, the scene shown in Figure 8 is chosen as a representative. The effect of the number of feature points on the running time of the algorithm proposed in this paper is observed, and the RMSE of the algorithm in this paper and different fixed-threshold RANSAC algorithms are analyzed. Room 304 A has more feature matching points than Room 306, and in Room 304A, scene one has the most feature matching points and scene two has the least feature points; in Room 306, scene one has the least feature points and scene two has the most feature

matching points. At the same time, the experimental results are given in Figure 9.

By comparing LMedS (shown in Table 2), the traditional RANSAC, and the algorithm proposed in this paper, it can be concluded the RMSE of the LMedS algorithm is the largest for both experimental environments. The reason for this phenomenon is that the LMedS algorithm does not eliminate false matches, resulting in the accuracy of the results being closely related to the inlier ratio. The accuracy of the LMedS algorithm drops sharply when the inlier ratio is below 50%. For indoor positioning, the LMedS algorithm is not recommended because of the uncontrollable inlier ratio within the image captured using the query camera.

The fundamental matrices were estimated by using the algorithm proposed for the image sets of Room 306 and Room 304A, and some of the results are shown in Tables 3 and 4. It can be seen that the running time of the algorithm in practice generally does not exceed 3 s, and the number of iterations is positively correlated with the running time of the algorithm. Furthermore, from Tables 3 and 4, it can be concluded that the number of feature matching points does affect the running time of the algorithm. Furthermore, as the number of feature matching points decreases, the algorithm will take less time to run. The reason is that the number of feature matching points directly impacts the number of iterations of the algorithm, before the final sampling strategy is executed. However, the proposed algorithm still has a significant reduction in running time compared to the RANSAC algorithm due to its accelerating convergence.

In Conference Room 306, the location of the query camera was randomly set to analyze the positioning results of the proposed algorithm in this paper, where the positioning error results of various methods are shown in Figure 10. As the LMedS algorithm is so strongly dependent on the inlier ratio, this paper focuses on comparing the

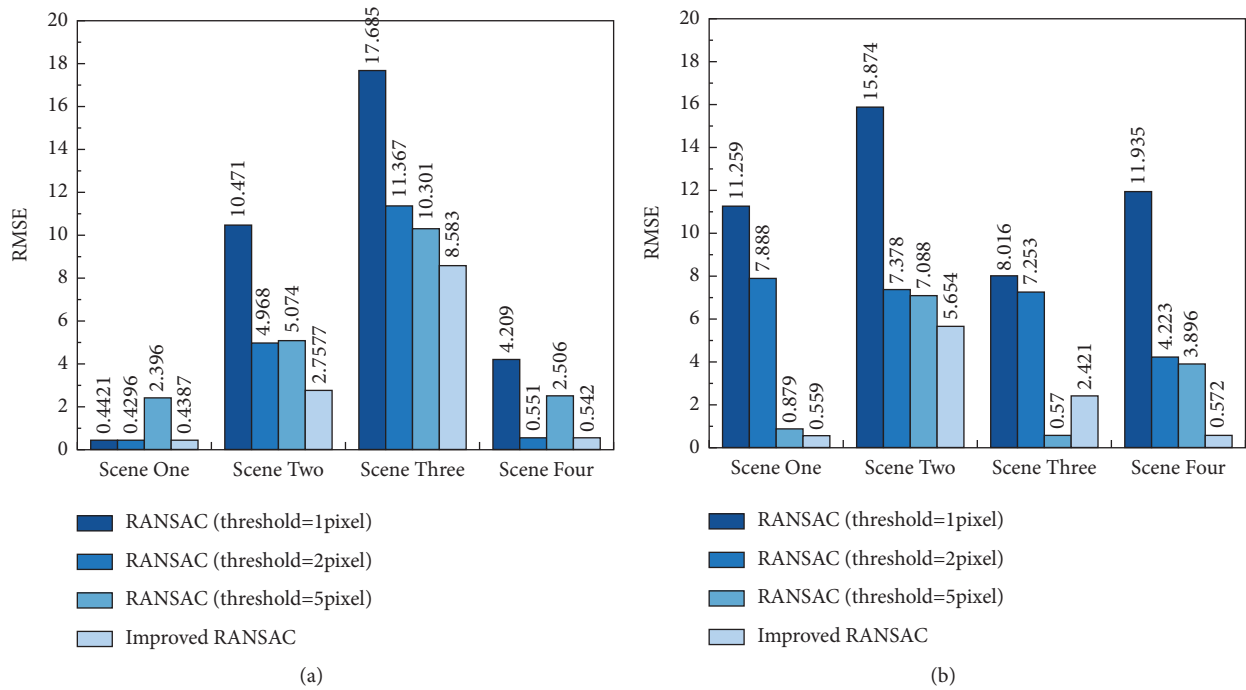FIGURE 8: Experimental scenarios. (a) Room 304A. (b) Room 306.

FIGURE 9: RMSE for various methods. (a) Room 304A. (b) Room 306.

TABLE 2: RMSE for the LMedS algorithm.

RMSE	Scene one	Scene two	Scene three	Scene four
304A	41.361	740.089	120.971	107.105
306	70.285	110.357	72.953	118.342

results with the RANSAC algorithm with the fixed thresholds. According to Figure 10, the results can be derived: the localization accuracy of the improved algorithm proposed in this paper is increased compared to the traditional RANSAC algorithm. In practice, the maximum positioning error was limited to 90 cm. Compared with other algorithms, the accuracy of the proposed localization algorithm is improved by at least 40% (compared to RANSAC with a threshold of 1 pixel), 55% (compared to RANSAC with a threshold of 2 pixels), and 70% (compared to

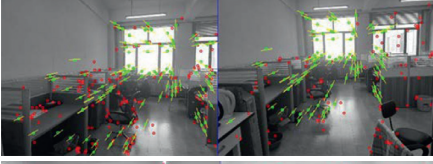TABLE 3: Results for various cases of the improved algorithm in 304A (1600 ∗ 1200 SIFT ratio = 1).
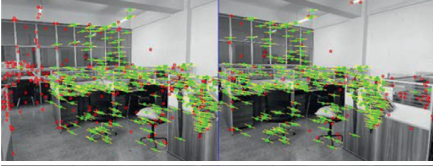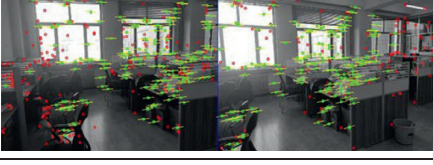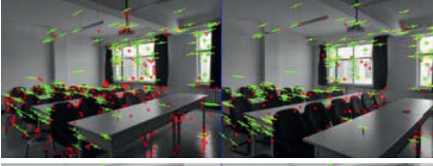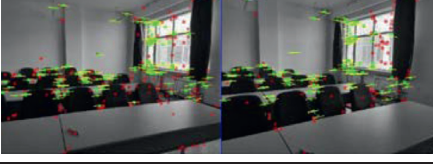
	Running time (s)	min NFA	Average error (pixel)	Threshold (pixel)
	2.87	−549.544	0.36188	0.914693
	2.66	−115.994	0.953146	2.11079
	2.80	−125.185	0.807435	1.67146
	2.77	−226.968	0.789106	1.91970

TABLE 4: Results for various cases of the improved algorithm in 306 (1600 ∗ 1200 SIFT ratio = 1).

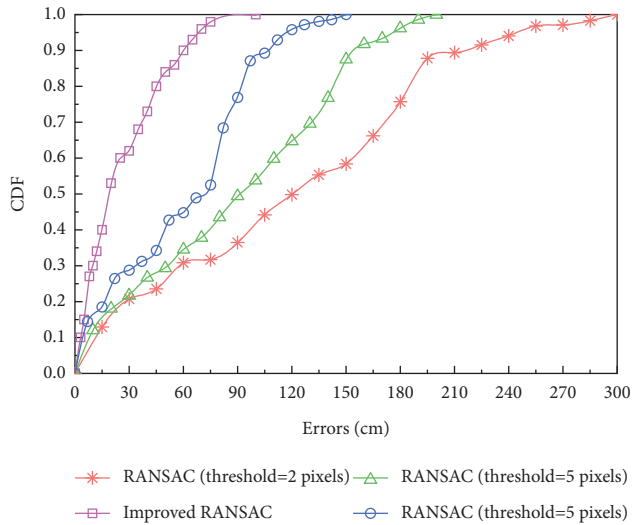|	Running time (s)	min NFA	Average error (pixel)	Threshold (pixel)
	2.16	−24.322	2.78902	6.0036
	2.65	−111.408	1.04754	2.2187
	2.15	9.49835	4.31296	9.68792
	2.46	−107.431	1.61656	3.23888

FIGURE 10: Error accumulation functions for different algorithms.

RANSAC with a threshold of 5 pixels), respectively. The main reason is that while changing indoor environments, the algorithm proposed in this paper is able to obtain thresholds adaptively to improve the accuracy of the fundamental matrix. The proposed algorithm is more specific to each situation than the traditional RANSAC algorithm, which sets a fixed threshold.

5. Conclusions

This paper presents a practical method for indoor localization. The method uses an improved algorithm based on RANSAC thresholds and the slopes of corresponding matching points to determine position. The vision-based positioning avoids the use of radio signals that are susceptible to interference; therefore, it is able to achieve precise positioning in closed or semiclosed indoor multiwalled environments. The proposed algorithm not only accelerates the convergence of the algorithm and reduces time consumption but also improves the quality of the positioning by increasing the fundamental matrix accuracy due to its targeting. Compared with traditional RANSAC, the proposed algorithm reduces the number of iterations by 30%–40%, limiting the running time to less than 3 s, and it is worth emphasizing that the resolution and blurriness of the images do not affect the accuracy of the algorithm. The results also show that the localization error of the proposed algorithm can be controlled to within 75 cm, which is a significant improvement over the traditional RANSAC algorithm.

However, the proposed method has certain limitations. When the camera captures images with fewer feature points, the accuracy of the fundamental matrix will have an impact and may lead to larger errors in the positioning results. In addition, the database image needs to be updated in time when the indoor decoration changes; otherwise, it will be difficult to match the query image with the database image. In the future, we will devote ourselves to exploring how to ensure the accuracy of indoor positioning when small changes in indoor environment decoration occur.

Data Availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] Z. Hajiakhondi-Meybodi, M. Salimibeni, A. Mohammadi, and K. N. Plataniotis, "Bluetooth low energy-based angle of arrival estimation in presence of Rayleigh fading," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 3395–3400, Toronto, ON, Canada, October 2020.

[2] R. Faragher and R. Harle, "Location fingerprinting with bluetooth low energy beacons," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, 2015.

[3] G. Huang, Z. Hu, J. Wu, H. Xiao, and F. Zhang, "WiFi and vision-integrated fingerprint for smartphone-based self-localization in public indoor scenes," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6748–6761, 2020.

[4] Z. Li, A. Yang, H. Lv, L. Feng, and W. Song, "Fusion of visible light indoor positioning and inertial navigation based on particle filter," *IEEE Photonics Journal*, vol. 9, no. 5, pp. 1–13, 2017.

[5] F. Seco and A. Jiménez, "Smartphone-based cooperative indoor localization with RFID technology," *Sensors*, vol. 18, no. 1, pp. 266–289, 2018.

[6] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1281–1293, 2013.

[7] J. Kim and H. Jun, "Vision-based location positioning using augmented reality for indoor navigation," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 954–962, 2008.

[8] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," in *Proceedings of the 2013 Fourth International Conference on Computing for Geospatial Research and Application*, pp. 70–75, San Jose, CA, USA, July 2013.

[9] S. S. Chawathe, "Low-latency indoor localization using bluetooth beacons," in *Proceedings of the 2009 12th International IEEE Conference on Intelligent Transportation Systems*, pp. 1–7, St. Louis, MO, USA, October 2009.

[10] J. Huang, D. Millman, M. Quigley, D. Stavens, S. Thrun, and A. Aggarwal, "Efficient, generalized indoor WiFi Graph-SLAM," in *Proceedings of the 2011 IEEE International*

Conference on Robotics and Automation, pp. 1038–1043, Shanghai, China, May 2011.

[11] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on wi-fi: an overview," *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, p. 135, 2017.

[12] M. Zhou, Y. Long, W. Zhang et al., "Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization," *IEEE Transactions on Evolutionary Computation*, p. 1, 2021.

[13] C. Yang and H.-r. Shao, "WiFi-based indoor positioning," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 150–157, 2015.

[14] R. Wang, W. Wan, K. Di, R. Chen, and X. Feng, "A high-accuracy indoor-positioning method with automated RGB-D image database construction," *Remote Sensing*, vol. 11, no. 21, p. 2572, 2019.

[15] X. Zhang, J. Lin, Q. Li, T. Liu, and Z. Fang, "Continuous indoor visual localization using a spatial model and constraint," *IEEE Access*, vol. 8, pp. 69800–69815, 2020.

[16] S. Liu, P. Guo, L. Feng, and A. Yang, "Accurate and robust monocular SLAM with omnidirectional cameras," *Sensors*, vol. 19, no. 20, p. 4494, 2019.

[17] M. Li, R. Chen, X. Liao, B. Guo, W. Zhang, and G. Guo, "A precise indoor visual positioning approach using a built image feature database and single user image from smartphone cameras," *Remote Sensing*, vol. 12, no. 5, p. 869, 2020.

[18] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "Robust regression methods for computer vision: a review," *International Journal of Computer Vision*, vol. 6, no. 1, pp. 59–70, 1991.

[19] O. Chum and J. Matas, "Matching with PROSAC - progressive sample consensus," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 220–226, USA, June 2005.

[20] D. R. Myatt, P. H. S. Torr, S. J. Nasuto, J. M. Bishop, and R. Craddock, "NAPSAC: high noise, high dimensional robust estimation - it's in the bag," *Proceedings of the British Machine Conference*, pp. 44.1–44.10, 2002.

[21] M. Lionel, M. Pierre, and M. Pascal, "Fundamental matlix of a stereo pair with a contrario elimination of outliers," *Image Procesing On Line*, vol. 7, no. 27, pp. 89–113, 2015.

[22] F. Espuny, P. Monasse, and L. Moisan, "A new A contrario approach for the robust determination of the fundamental matrix," in *Image and Video Technology - PSIVT 2013 Workshops*, pp. 181–192, Springer, Berlin, Germany, 2014.

[23] S. Yang, L. Ma, S. Jia, and D. Qin, "An improved vision-based indoor positioning method," *IEEE Access*, vol. 8, pp. 26941–26949, 2020.