IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 14, 2021, accepted March 9, 2021, date of publication March 15, 2021, date of current version March 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3065953

A Robust Visual Localization Method With Unknown Focal Length Camera

XILIANG YIN[1,2], LIN MA[1], (Senior Member, IEEE), XUEZHI TAN[1], (Member, IEEE), AND DANYANG QIN[3]

[1]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
[2]Department of Electronics and Information Engineering, Harbin Vocational and Technical College, Harbin 150081, China
[3]Electronic Engineering College, Heilongjiang University, Harbin 150080, China

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

This work was supported by the National Natural Science Foundation of China under Grant 41861134010 and Grant 61971162.

ABSTRACT PnP problem is well researched in many fields, such as computer vision. It is considered the fundamental method to solve the key problems of robot SLAM. However, in pedestrian visual localization, uncalibrated PnP (UPnP), specifically PnP with unknown focal length (PnPf) is more suitable for solving the problem. Recently, a few researchers proposed some methods to alleviate this problem. However, the localization accuracy of the existing methods is not satisfied when image pixel noise is larger. In other words, RANSAC should be running before solving the PnPf problem to get less noisy input, which means the localization delay increases inevitably. In this paper, we propose a more robust method for solving the PnPf problem without the help of RANSAC based on Gröbner basis and convex optimization. We build a Gröbner basis solver on the offline stage with one instance in the prime field. Then, we substitute the coefficients with real value and find multiple solutions on the online stage. Finally, we construct a convex optimization program to seek the final robust solution to PnPf problem. The other purpose of this paper is to provide a second-level localization experience for the end-user. The simulation result shows that our method can give a localization solution, which is more reliable than benchmark methods by both synthetic and real data verified.

INDEX TERMS UPnP, PnPf, visual localization, Gröbner basis, convex optimization.

I. INTRODUCTION

Indoor localization is a hot branch of localization in recent years, which compares with the concept of outdoor positioning. It is in the stage of rapid development with challenges and opportunities. Various technologies emerge with their competing and unique characters, such as WiFi [1], visible-light [2], UWB [3], WSN [37], [38], image [39], etc. For other new technologies, please refer to [40]. Visual localization gets the attention of the researchers due to its low cost and no additional device requirement. It has been applied in many well-known areas, such as robot localization and vehicle navigation. More recently, with the rise of the Internet of Video Things (IoVT) [41], visual localization will also play an important role in the system.

In addition to the applications mentioned above, visual localization also knows as a prospective solution for the last kilometer of pedestrian positioning problems in the indoor environment. To solve the problem, a visual map database firstly constructs by a site scanning method, such as [4], [5]. With the database, one can locate himself even in a particular indoor area without wireless signal coverage. The main process of localization can be summarized as follows. The user uses his/her smartphone to take an image in the interested area. Then, the image feature will extract by algorithms like SURF [6]. It proved more efficient even in some feature-barren environments [7]. With the feature in the query image and its correspondence in the visual map database, the relationship could map by algorithms, such as [8]–[10]. Finally, the correspondences will be used as a known condition to solve the PnP problem, whose solution is a fine localization result for the querying user.

The very challenging part in visual localization is mainly Perspective-n-Point (PnP), which solves the camera extrinsic matrix with several 2D-3D correspondences. It can cluster into three categories. The first one under this classification is that the camera intrinsic matrix is fully unknown. It is a rare situation in real applications due to the development

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

of the digital camera. The other one is that the camera intrinsic matrix is completely known. Typically, it obtains by camera calibration [11], this category has always been the mainstream research direction in computer vision, and many scholars have made significant research contributions, such as EPnP [12], RPnP [13], OPnP [14], ASPnP [15], etc. These algorithms have been optimally implemented in many applications, especially in Robot SLAM. One simple version of this kind is called 2DTriPnP [16], which could apply in the vehicle or pedestrian visual localization in 2D space. However, camera calibration is a requirement, which is hard to force every pedestrian to do this at the very beginning.

The last category is that the camera intrinsic matrix is partially known. The unknown element of the matrix is typically one or two of focal length, screw factor, and principal point. Since both the machine is customized and its camera has been calibrated by the researchers, this kind of algorithm is not usual in computer vision. However, in pedestrian visual localization, the camera used on the phone is most likely uncalibrated. On the other hand, the element in the intrinsic matrix could estimate by the common assumption. For instance, the principal point is in the middle of the image, the screw factor is zero, focal length could read from the EXIF tag of the JPEG file. Among these three parameters, we believe the estimation of the focal length is with a larger error than that of principal point, screw factor [17], [18]. Therefore, we call this kind of PnP problem partially known. One of the most studied branches is PnPf, which donates the focal length information is unknown in the camera intrinsic matrix. Some scholars have paid attention to this problem, and they have proposed some works, which will introduce with details in Section II.

As stated before, the assumption of the principal point and screw factor is very close to the real value. More importantly, it will minimize the cost of online computing in the image-based localization system. Above all, in consideration of the characteristics of pedestrian visual localization, we will focus on the PnPf problem in this paper.

According to the number of 2D-3D point pairs used in the PnP equations, there are two research directions. One aims to find the solutions of PnP equations with the least 2D-3D point pairs. It usually defines as a minimal solver. For instance, the PnPf problem solved with four pairs of 2D-3D correspondence is called P4P. The core of this minimal solver is the Gröbner basis method, whose more details are given in [19]. It should be noted that the method is not finding a Gröbner basis for the initial equations by Buchberger [20] or F4 [21] algorithm since it not only costs too expensive in the real field but also will lead to numerical instability. As an efficient alternative, researchers find the leading monomials of the corresponding Gröbner basis are the same as the monomials in the basis of the quotient ring. By this property, Z. Kukelova *et al.* propose an automatic generator of this minimal case solver in [22]. Many solvers of the PnP problem with coefficients uncorrelated could generate by this tool without losing accuracy performance. However, not all the PnP problems are directly suitable to this automatic tool. The researchers may sometimes build the solver by hand. The main difference between these solvers is the size of the elimination template, which is mainly due to the different simplification of PnP equations. The size of the elimination template proposed in [23] is 154×180, which is further down to 53×63 in [24]. As far as we know, the least is 20×30, which proposes in [25].

The other one tries to use all pairs to solve the problem, which will reduce the sensitivity of minimal solver to noise transition. A representative algorithm was proposed in [26]. It was well known as UPnP, whose idea borrowed from EPnP. Later, E. Kanaeva *et al.* proposed a regularized distance constraints method for finding a more accurate solution for the UPnP equation in [35]. Be different from the methods [26], [35], the PnPf problem with all point correspondences can also solve by Gröbner basis solver. Y. Zheng *et al.* proposed a 36×52 elimination template in [27] for solving the equations generated from all 2D-3D correspondences. A versatile approach was proposed by G. Nakano in [36], which provided a novel solution for the least square PnPf problem. However, its improvement in solution accuracy was limited when comparing to [27]. More recently, a 130×251 template was built in [28] to improve the accuracy. Although these algorithms could alleviate the unstable solution generated by the minimal solver. Unfortunately, the state-of-art algorithms [26], [28] still depend on RANSAC for selecting inlier correspondences to attain the accuracy requirements when more noise exists in the real image. The application of RANSAC on the online visual localization step will inevitably increase the computational cost. Thus, the main purpose of this paper is to propose a robust solution for the PnPf problem without the help of online RANSAC when larger image pixel noise exists. The main contributions of this paper are in three folds:

1). A modified minimal solver proposes in this paper. We derive the solver more rigorously and show the entire back substitution process, the solver can obtain more accurate results by the constraints of unchanged rigid body distance than P3.5P, although the single running time is slightly longer.

2). A convex optimization-based method proposes in this paper for extracting a more accurate solution from our minimal solver. It could promote the accuracy performance of the minimal solver efficiently when the noise is larger in the image pixel coordinate. Compared to the state-of-art PnPf algorithm, our method is more robust to combat the noise without RANSAC, which proves by synthetic and real data simulation results.

3). The computation complexity of our proposed method is low, which approximates to $O(n)$. In this way, the proposed overall system model can provide a *second-level* of experience for pedestrian, who uses the image-based localization. Therefore, this will be more practical in real applications.

The rest of this paper organizes as follows. In Section II, some related works will discuss. Section III describes the PnPf problem in pedestrian visual localization and presents the system model. In Section IV, we propose our method

based on a Gröbner basis and convex optimization. Section V provides the simulation results, and the conclusion draws in Section VI.

II. RALATED WORKS

Although the main purpose of this paper is to propose a robust visual localization method with an unknown focal length camera, we need to introduce the most widely used algorithm EPnP [12], which is classical for its O(n) complexity with known camera internal parameters. The algorithm utilizes the linearization and re-linearization method for solving the weight of a linear combination of a matrix eigenvector, which is derived from 3D-2D correspondences. With these weights, the camera coordinates of the 3D point can be calculated. Then, with the help of SVD for solving the matrix maximum trace problem, the rotation matrix and the translation vector can be decomposed. Further, a more accurate result will achieve by setting the closed-form solution as the initial input of the Gauss-Newton scheme. As an alternative to RANSAC [29], which is time-consuming and unsuited for online computation, the standard scheme for none-convex optimization can see as a refinement of the solutions. Most recently, a more advantageous polishing algorithm is proposed in [30], which is called HARD-PnP. These two algorithms could also embed in our framework as the final refinement.

Inspired by EPnP, UPnP proposes in [26]. The algorithm takes the focal length as the denominator and the z-axis coordinate of the virtual control point as the numerator to form a whole new variable. With this synthetic variable, the solution architecture could be nearly unchanged comparing to EPnP. The solution could solve by linearization, relinearization, and exhaustive linearization consecutively, and the process could terminate whenever the solution generated from one of these methods is within a threshold of reprojection error. When the solutions from all these three methods are over the threshold, the solution with the least reprojection error will rank as the best one. Alternatively, a regularized distance constraint method substitutes for the method mentioned above in [35]. However, the regularization coefficient needs to adjust dynamically to achieve a balance between accuracy and time. Therefore, the method proposed in [35] is not chosen as the benchmark for comparison in our paper. Be similar to EPnP, the final solution of UPnP could also polish by the Gauss-Newton scheme, which always defines as UPnP-GN.

The linearization method used in the algorithms described above is an approximate solution of the polynomial equation. Naturally, the result contains some errors. Thus, the Gröbner basis method with the more accurate result applies in the state-of-the-art minimal solver [25], which is named P3.5P. Beyond this method, an Euler angle with none-unit quaternion representation for a rotation matrix is very important for its success. Another common trick for elimination by rank deficient matrix and SVD for recovery also embeds into our minimal solver to improve the performance. It should be noted that the main purpose of the solver proposed in [25] is to find the least pairs of 2D-3D correspondence for solving the PnPf problem, while our goal is trying to reduce the influence of noise on the solution without RANSAC. The P3.5P minimal solver is not embedded in our framework, the reason is the parameterization and back substitution process are still improvable. It inspires us to propose a modified minimal solver instead of using P3.5P directly in this paper.

As the state-of-the-art none-minimal solver for the PnPf problem, the expression of the rotation matrix in [25] also uses in [28]. However, be different from [25], the basic idea is to construct a proper objective function after variable elimination. Two equations generate from the derivation of an objective function. By the Gröbner basis method, all the stationary points solve. Then, one of the solutions is ranked, which will minimize the objective function. Finally, all unknowns find by back substitution. As far as we know, this method is the latest one for solving the PnPf problem with all 2D-3D correspondences. It is partly due to the special acknowledgments and skills for the Gröbner basis method. The advantage of this algorithm over the minimal ones is balancing the contribution from each 2D-3D correspondence. However, it is better to treat each equation differently than this balanced approach when the noise distribution is unbalanced. Without loss of generality, it can assume as Gaussian distribution. Therefore, we propose an algorithm based on Gröbner basis minimal solver and convex optimization for solving the solution of the PnPf problem, which could weigh equations from combinations of 2D-3D correspondence.

III. SYSTEM MODEL

A. PROBLEM STATEMENT

A diagram shows in **Fig. 1**, which can describe the UPnP problem encountered in the pedestrian visual localization. The localization equipment is only the mobile phone holding by the querying user. Typically, it is assumed that the intrinsic matrix of the camera embedded in the phone is unknown. It means that the camera is uncalibrated mostly. The camera rotates at some angle around the x-axis, y-axis, and z-axis, respectively. After the rigid body rotation, a world point and image pixel feature could form a mapping relationship. It defines as 2D-3D correspondence. With such correspondences, the extrinsic matrix (rotation matrix and translation vector) and the intrinsic matrix will be recovered from the solution of UPnP solvers. Further, the principal point, aspect

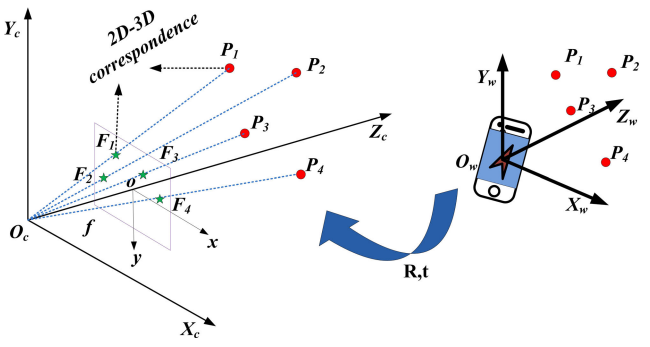


FIGURE 1. UPnP in the pedestrian visual localization.

ratio, and screw factor in the intrinsic matrix could be estimated quite approximately to the ground-truth value. It makes sense that the UPnP problem turns into a PnPf problem eventually in the image-based localization system.

However, most existing solvers including the state of the arts perform well on the prerequisites of treating inlier 2D-3D correspondences as their inputs. When the filtering algorithm is operated by the user terminal on the online localization stage, it is impossible to complete the positioning process in a very short time in consideration of time requirements from other algorithms running in the system. It is why RANSAC only applies for refining the correspondences on the offline stage in the image-based localization system, such as [16]. Without the aid of RANSAC, that means the correspondences may contain outliers. In such conditions, the accuracy of the solver reduced significantly comparing to inlier inputs. Thus, this dilemma inspires us to devise a new approach, which could provide better performance in the presence of outliers.

B. RESEARCH FRAMEWORK

To show the relationship among different parts of the system, a brief illustration of the overall indoor image-based localization provides in **Fig. 2**. The whole system can divide into two stages, which is an offline and online stage, respectively. On the offline stage, two independent parts need to be done. One is collecting visual fingerprint samples at the interested place and generating a visual map database by an automatic fingerprinting method. The time consumed by this procedure is mainly due to the area of the indoor environment. The other part is exploiting the Gröbner basis solver from particular polynomial equations with Prime field coefficients, which will take several minutes to complete the entire process.

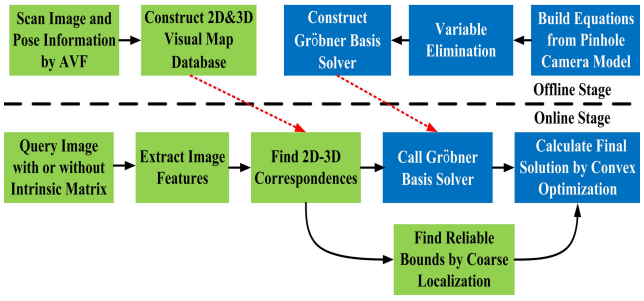


FIGURE 2. Overall framework of indoor image-based localization system.

As shown in **Fig. 2**, we expect the entire online localization process to finish within one second according to our proposed system model, which is very promising and attractive for indoor image-based localization applications. On the other hand, each step is continuous on the online stage. From the very beginning, one will take a shot anywhere in the indoor environment with a calibrated or uncalibrated camera. From the perspective of the application, we can assume most cameras are uncalibrated. Secondly, image information transforms into 2D features. With these 2D features, a matching algorithm applies for finding reliable 2D-3D or 2D-2D-3D correspondences. Moreover, a coarse localization result could calculate in this step at a lower cost. As the input parameters of the Gröbner basis minimal solver, these 2D-3D correspondences without filtering by RANSAC are used to recover the intrinsic and extrinsic parameters of the camera. In consideration of the noise, several quadruple 2D-3D correspondences test to find reliable solutions. Different solutions from the solver with reasonable bounds will form a QCQP problem, which is a convex optimization problem. It will give the final optimal solution to our PnPf problem.

The main research content of this paper is limited to the blue block diagram. We will explain the projection model and show how it parameterizes in Section IV-A. Then, the variable elimination procedure details in Section IV-B. Next, we present the complete back substitution process, which will find the solution of all variables in Section IV-C. Section IV-D describes the construction details of the Gröbner basis solver as the final step on the offline stage. Finally, on the online stage, the core block fuses different outputs of the minimal solver, a final solution closer to the real value is obtained by solving the constructed QCQP problem. It will deduce in Section IV-E. In Section IV-F, we give the computational complexity analysis.

IV. PROBLEM FORMULATION

A. PARAMETERIZATION

Be similar to most literature in computer vision and visual localization, the pinhole camera model is used here to represent relationships between the pixel and world coordinates. The famous model can express as

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K\left(R \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + t\right), \tag{1}$$

where $[X_i, Y_i, Z_i]^T$ is the world coordinate of the ith point, u_i and v_i is the corresponding coordinate in pixel system respectively, λ_i donates the depth factor of the ith point, K is the intrinsic matrix of the camera, R is the rotation matrix, t is the translation vector. Usually, R and t are defined as external parameters, which are more concerned by the image-based localization system. K defines as

$$K = \begin{bmatrix} f & \gamma & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

where f is the focal length, γ is the screw factor, u_0 and v_0 is the principal point of the image plane. And for most cameras, the assumption of $\gamma = 0$, the focal length aspect ratio is 1, and u_0, v_0 is in the middle of the image are valid.

By taking advantage of unit-norm quaternion $q = [q_w, q_x, q_y, q_z]^T$, the expression of R is consistent with [25], [28], which is

$$R = \begin{bmatrix} q_w^2+q_x^2-q_y^2-q_z^2 & 2q_xq_y-2q_wq_z & 2q_xq_z+2q_wq_y \\ 2q_xq_y+2q_wq_z & q_w^2-q_x^2+q_y^2-q_z^2 & 2q_yq_z-2q_wq_x \\ 2q_xq_z-2q_wq_y & 2q_yq_z+2q_wq_x & q_w^2-q_x^2-q_y^2+q_z^2 \end{bmatrix}. \tag{3}$$

According to Hamilton product of quaternions, R can be further uniquely decomposed into $R = R_1R_2$, where R_1 and R_2 could express by the quaternion $[q_w, 0, 0, q_z]^T$ and $[1, (q_wq_x + q_yq_z)/(q_w^2 + q_z^2), (q_wq_y - q_xq_z)/(q_w^2 + q_z^2), 0]^T$ when $q_w^2 + q_z^2 \neq 0$. Meanwhile, the other decomposition proves to be not unique when $q_w = q_z = 0$. Due to its rarity in practice, the former decomposition of R adopts. Thus, according to (3), R_1 could express as

$$R_1 = k\begin{bmatrix} Q_1 & -Q_2 & 0 \\ Q_2 & Q_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

where $k = q_w^2 + q_z^2$, $Q_1 = (q_w^2 - q_z^2)/k$, $Q_2 = 2q_wq_z/k$. Obviously, it also has a nice property between Q_1 and Q_2, which is $Q_1^2 + Q_2^2 = 1$. It should be noted that this is different from the parameterization in [25]. Also, the recovery process is different, which details in section IV-C. As for R_2, corresponding to the initial quaternion a new parameter quaternion $[1, q_1, q_2, 0]^T$ redefines as

$$R_2 = \begin{bmatrix} 1 + q_1^2 - q_2^2 & 2q_1\, q_2 & 2q_2 \\ 2q_1\, q_2 & 1 - q_1^2 + q_2^2 & -2q_1 \\ -2q_2 & 2q_1 & 1 - q_1^2 - q_2^2 \end{bmatrix}. \tag{5}$$

Before variable elimination, (4) and (5) are substituted into (1), which could deform as

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = k \begin{bmatrix} f_1 & -f_2 & u_0 \\ f_2 & f_1 & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_2^1 \\ R_2^2 \\ R_2^3 \end{bmatrix} X_i + \hat{t}, \tag{6}$$

where $f_1 = fQ_1, f_2 = fQ_2, \hat{t} = Kt$, R_2^i is the ith row vector of matrix R_2, X_i is the column vector of the ith point world coordinate. The number of unknowns is 10 in (6).

B. VARIABLES ELIMINATION

To construct more robust and efficient polynomial equations, variables are eliminated in this subsection. The depth factor λ_i is first substituted from (6). Since $k \neq 0$, two equations from one 2D-3D correspondence could represent as

$$u_iR_2^3X_i + \tilde{t}_3u_i = f_1R_2^1X_i - f_2R_2^2X_i + u_0R_2^3X_i + \tilde{t}_1 \tag{7a}$$
$$v_iR_2^3X_i + \tilde{t}_3v_i = f_2R_2^1X_i + f_1R_2^2X_i + v_0R_2^3X_i + \tilde{t}_2, \tag{7b}$$

where $\tilde{t}_1 = \hat{t}_1/k$, $\tilde{t}_2 = \hat{t}_2/k$, and $\tilde{t}_3 = \hat{t}_3/k$. Note that equations in (7) are from one 2D-3D correspondence. It can be assumed there are n correspondences, hence $2n$ equations can be generated by (7). A natural idea of eliminating $\tilde{t}_1$ and $\tilde{t}_2$ is to take the sum of each equation in (7) and get an average of the sum respectively, which is referred to in [18], [28].

Consequently, we subtract each equation in (7) from its average. Two denormalized equations will reform, from which $\tilde{t}_3$ can be obtained by

$$\tilde{t}_3 = \frac{1}{s}\sum_{i=1}^{n} g_i(f_1, f_2, q_1, q_2) + h_i(f_1, f_2, q_1, q_2), \tag{8}$$

where $s = \sum_{i=1}^{n}(\tilde{u}_i^2 + \tilde{v}_i^2)$, $\tilde{u}_i = u_i - \frac{1}{n}\sum_{i=1}^{n} u_i$, $\tilde{v}_i = v_i - \frac{1}{n}\sum_{i=1}^{n} v_i$, g_i and h_i are the polynomial function with parameters f_1, f_2, q_1, q_2. Since s must be bigger than 0, $\tilde{t}_3$ is known when f_1, f_2, q_1, q_2 are solved. Once $\tilde{t}_3$ is obtained, $\tilde{t}_1$ and $\tilde{t}_2$ can be calculated by (7). Otherwise, we can eliminate $\tilde{t}_3$ by the two denormalized equations, which are different from [28]. An equation from one 2D-3D correspondence with four unknowns is finally deformed as

$$\tilde{v}_iG_i(f_1, f_2, q_1, q_2) - \tilde{u}_iH_i(f_1, f_2, q_1, q_2) = 0, \tag{9}$$

Note that there will be n equations rather than $2n$ equations in [28]. Since there still exists 4 unknowns, 4 pairs of 2D-3D correspondences are needed for solving (9). For the sake of efficient computation on the online localization stage, a general trick is to express the equations in linear algebra form. An equivalent expression shows as

$$B(q_1, q_2)\begin{bmatrix} f_1 & f_2 & 1 \end{bmatrix}^T = 0, \tag{10}$$

where polynomial matrix B is 4×3. To find the none-trivial solutions, the rank of each 3×3 submatrices of B must be 2. Thus, we have $\det(B_{3\times3}) = 0$. It should be noted that any 2 of 4 polynomial equations from the determinant equation could use for finding the solutions, although there are some false ones. In summary, the final polynomial equations have 2 unknowns q_1, q_2 with the highest degree 6, these polynomials will support the implementation of the overall online localization algorithms proposed in this paper.

C. BACK SUBSTITUTION

In this subsection, we will find the values of all variables in (1). With the real solutions of q_1 and q_2, f_1 and f_2 can recover by SVD of matrix B in (10). Consequently, the focal length f could solve by

$$f = \sqrt{f_1^2 + f_2^2}. \tag{11}$$

It should be noted that the filtering of the final solution depends on the ground truth value of f_g in P3.5P [25], which is contradictory to the condition. Thus, we use f_r instead of f_g. It is a reference value, which could estimate from the image information. $\tilde{t}_3$ can be calculated by (5) and (8). Then, $\tilde{t}_1$ and $\tilde{t}_2$ are reached by equations in (7), respectively. Since the points are rigid, the distance of any two points from the world coordinates system to the camera coordinates system will remain unchanged. k could calculate by

$$\left\|R_1R_2(X_i - X_j)\right\|_2 = \left\|X_i - X_j\right\|_2, i \neq j, \tag{12}$$

where X_i and X_j are world coordinates of two independent points. We can have 6 solutions of k from 4 point coordinates in the solver. To reduce the impact of noise, the mean value takes. The translation vector t could obtain by

$$t = K^{-1}k\tilde{t}. \tag{13}$$

With all recovery unknowns, the focal length f, rotation matrix R, and translation vector t could solve as one tuple solution of one particular PnPf problem.

D. GRÖBNER BASIS SOLVER

In this subsection, the goal is to find the solutions of two polynomial equations with the 6 highest degrees. Several methods can be applied, such as the hidden variable method with 'polyeig' in Matlab. However, the Gröbner basis method outperforms others in terms of accuracy, which will be called GB solver for simplicity in the rest of this paper.

Thus, we also choose this standard method for computing the two polynomial equations with the 6 highest degrees, which deforms from (10). The other advantage is the solver itself can be computed on the offline stage with the coefficient chosen from finite Prime Field $\mathbb{Z}_p$, which is time-consuming compared with the online computing requirement by the solver. With the help of software, such as Maple or Macaulay2 [31], the solution number of the initial polynomial equations and the basis of the quotient ring could be obtained. Then, an Elimination Template is generated, which will record the monomial eliminated path. With this crucial elimination template, a final Action Matrix can be used to find the solution by the eigenvalue decomposition method, such as 'eig' provided in Matlab. By simulating one instance from synthetic data, integer coefficients are used to design this GB solver with a 36×70 elimination template and a 34×34 Action Matrix, respectively. For a full description of the calculation process please refers to [22].

E. CONVEX OPTIMIZATION PROGRAMMING

With the estimated intrinsic and extrinsic camera matrix, we can evaluate the accuracy by a golden standard, which is called Reprojection Error. It shows as

$$\epsilon_i^{Reproj} = \sum_{i=1}^{n} \left\| u_i - u_i' \right\|_2, \tag{14}$$

where u_i' donates the ith vector of reprojected pixel coordinates from the world coordinates, u_i is the corresponding primitive vector in the image plane. Unfortunately, the solutions given by the GB solver are unstable due to the different noise from 4 randomly chosen pairs of 2D-3D correspondence. However, the solver is fast enough to allow dozens to hundreds of repetitions on the online localization stage according to the computation platform. Typically, RANSAC is used to filter out the solutions with larger deviations from the real value, yet it is not suited for online computing with its time-consuming characteristic.

Since most literature has proved that PnPf is non-convex, we need to transform it into a convex problem for solving efficiently on the online stage. Now, we revisited equation (1) and turned it into

$$\lambda_i / f = (r_{11}X_i + r_{12}Y_i + r_{13}Z_i + t_x)/u_i' \tag{15a}$$
$$\lambda_i / f = (r_{21}X_i + r_{22}Y_i + r_{23}Z_i + t_y)/v_i' \tag{15b}$$
$$\lambda_i = r_{31}X_i + r_{32}Y_i + r_{33}Z_i + t_z, \tag{15c}$$

where r_{ij} is the ith row and jth column element of the rotation matrix, $u_i' = u_i - u_0$, $v_i' = v_i - v_0$. For each solution of GB solver, we have the numeric solution of each equation. Suppose the GB solver can provide N sets of solutions, we use equations (15) for building the optimization problem as

$$\begin{aligned} min \ & \sum_{n=1}^{N} w_i \left\| x - c_i \right\|_2^2 \\ s.t. \ & \left\| x - c_i \right\|_2^2 \le \epsilon, \\ & x_l \le x \le x_u, \end{aligned} \tag{16}$$

where c_i is a vector calculated from the ith solution of GB solver in the right side of equation (15), ϵ is the tolerance, x is the optimization variable, x_l and x_u is the reliable lower and upper bound of the final solution provided by coarse localization, w is the weight vector. Each w_i could be calculated simply by $w_i = 1/\epsilon_{Reproj}$ before normalization. Note that the optimization variable is only the element of the rotation matrix and the translation vector. The least-square solution of the focal length could deduce easily after solving the variables in (16). Equations in (15) could simplify as

$$fX_i^c = Z_i^c u_i' \tag{17a}$$
$$fY_i^c = Z_i^c v_i', \tag{17b}$$

where X_i^c, Y_i^c, Z_i^c is the coordinate of the ith point in the camera system. The problem in (16) is a standard QCQP form according to [32]. With the interior point method, the global optimal solution could be calculated efficiently by discipline convex optimization software, such as cvx [33]. Since our proposed method mainly bases on Gröbner basis solver and convex optimization, we name it as gcPnPf. The proposed method summarizes in **Algorithm 1**.

Algorithm 1 The Proposed gcPnPf Method

Input: $u, X, \epsilon_t, \epsilon_f, f_r, n_l, x_l, x_u$
Output: $t, \alpha, \beta, \gamma, f$

1: **for** $i = 1; i < n_l; i++$ **do**
2: Put the randomly chosen 4 pairs of u and X into the GB solver in section IV.D
3: Calculate $x_p = [t_p, \alpha_p, \beta_p, \gamma_p, f_p]$
4: **if** $\epsilon_{Reproj} < \epsilon_t$ && $|f_p - f_r| < \epsilon_f$ **then**
5: Save x_p within the threshold
6: Put x_p and its correlated weight w with x_u and x_l into the convex optimization solver in section IV.E
7: Calculate the final result by interior point algorithm
8: **return** $t, \alpha, \beta, \gamma, f$

F. COMPUTATIONAL COMPLEXITY

Given the online localization is more important for the application, we analyze the time complexity of our proposed algorithm on the online stage. According to Section IV.D and IV.E, the core of our algorithm is the eigenvalue decomposition of matrix and interior point method in convex optimization. The time complexity of eigenvalue decomposition is $O(n_a^3)$, where n_a is the dimension of the action matrix. The lower complexity bound and the upper bound of finding an ϵ-solution are $O(1)n_o M \ln(1/\epsilon)$ and $O(1)n_o(n_o^3 + M)ln(1/\epsilon)$,

respectively, where n_o is the dimension of the optimization problem, M is the iteration number of arithmetic operation, ϵ is the error tolerance. These prerequisites enable us to show the time complexity of the proposed algorithm clearly. Supposed that the loop number of GB solver is n_l, the lower bound and the upper bound of our proposed algorithm are $n_l O(n_a^3) + O(1)n_o M \ln(1/\epsilon)$ and $n_l O(n_a^3) + O(1)n_o(n_o^3 + M)ln(1/\epsilon)$. Therefore, the computation complexity of our proposed algorithm approximates to $O(n)$.

V. IMPLEMENTATION AND PERFORMANCE ANALYSIS

A. SYNTHETIC DATA

First of all, the experiment conducts with synthetic data for convenience. Be similar to [25]–[28], the camera coordinates of 3D points are generated randomly within a box of $[-2, 2]\times[-2, 2]\times[4, 9]$ for the general case. Then, the pixel coordinate projects by pinhole camera model with an initial point $(320, 240)$ and focal length around 1000. For the coplanar case, we set the same z-axis value of the dataset. The x-axis, y-axis, and z-axis angles randomly choose from $0° - 30°$, which defines as α, β, and γ. Meanwhile, the translation vector distributes from $0 - 3$. We will show each element of this vector in our simulation figures, specifically t_1, t_2, and t_3. With the rotation matrix and translation vector, the world coordinates of the 3D points generate. The results of each set calculate from 100 random repetition groups. Some parameters of our proposed method are listed in Table 1. It should be noted that f_r is the reference focal length, which is only used to evaluate the accuracy of the solution. In real data simulation, $f_r = (f_x + f_y)/2$.

TABLE 1. Parameters used in the simulation.

Parameter	Symbol	Value for synthetic data	Value for real data
Loop number of GB solver	n_l	300	300
Reprojection error threshold	ϵ_t	100	200
Focal length error threshold	ϵ_f	50	100
Reference value of the focal length	f_r	1000	1378.8
Tolerance in the constraint	ϵ	0.01	0.1

Since our proposed algorithm composes of GB solver and convex optimization, we define it as gcPnPf for short in the simulation figures. To compare our proposed algorithm with the benchmark thoroughly, we divided the synthetic data simulation results into four parts. The first, second, and third parts are general point cases, which are non-coplanar, while the fourth part is the coplanar case. Different parameter variation applies in the general point case.

The first part of the comparative results is with the configuration of focal length $f = 1000$, aspect ratio 1, 20 2D-3D correspondences, which shows in **Fig. 3**. Zero-mean Gaussian noise with standard deviation $\sigma_{img} = 5$ is added to the image

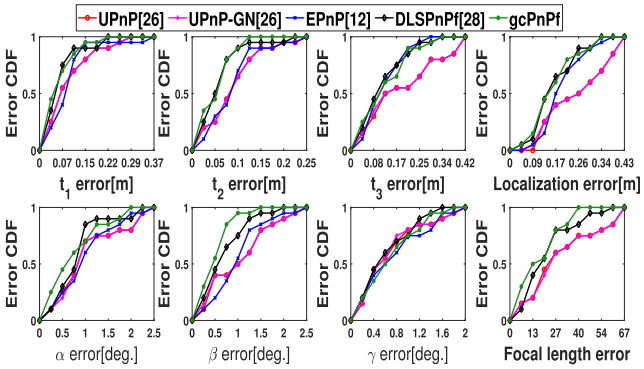


FIGURE 3. Results with image Gaussian noise $\sigma_{img} = 5$.

feature coordinates. Under this setting, the cumulative error probability density uses for evaluating the accuracy among algorithms, which shows more details than the mean-error or median-error curve. More specifically, rotation and translation of x-axis, y-axis, z-axis, focal length, and localization error evaluate separately. The localization error defines as the Euclidean distance of the translation vector between real value and estimated value. To compare with the real data, the localization error and translation error are limited to meter level, which is original dimensionless. Otherwise, we use the absolute error to evaluate the results, which is more intuitive than relative error under a particular setting for a localization system.

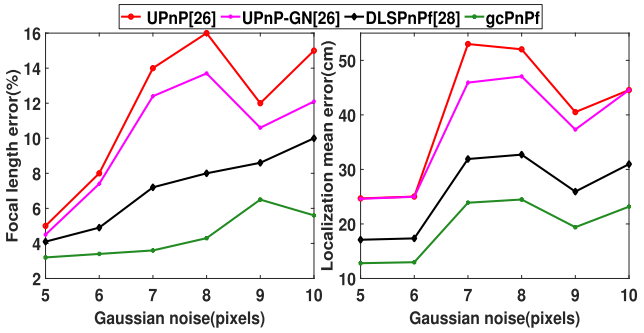


FIGURE 4. Focal length and localization results with different image Gaussian noise $\sigma_{img} = 5 - 10$.

Next, the mean relative focal length and the mean localization error show in **Fig. 4**. Different image Gaussian noise σ_{img} is varied from 5 to 10. From **Fig. 3** and **Fig. 4**, it is almost certain that DLSPnPf and our proposed gcPnPf outperforms UPnP and its enhanced version UPnP-GN in terms of accuracy. The fundamental reason is that the method is different for solving polynomial equations. Further, gcPnPf shows better results than DLSPnPf in most comparisons. This reveals that our global convex optimization solver is more robust than the direct least-square objective function in [28]. In addition, Table 2 gives a summary of different localization statistical errors under image Gaussian noise σ_{img} varied from 5 to 10.

TABLE 2. Different statistical error of localization.

Gaussian Noise	ϵ_{max}	ϵ_{min}	ϵ_{mean}	ϵ_{std}
5	0.3157	0.0047	0.1311	0.0947
6	0.3174	0.0136	0.1092	0.0914
7	0.8717	0.0016	0.103	0.1881
8	0.5720	0.0324	0.2111	0.1521
9	0.6271	0.0039	0.2699	0.2114
10	1.369	0.003	0.1938	0.2988

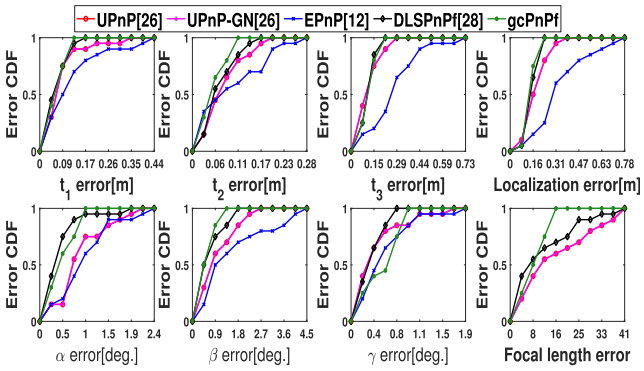


FIGURE 5. Results with 100 2D-3D correspondences.

The second part of the results are from focal length $f = 1000$ and image Gaussian noise $\sigma_{img} = 10$, which show in **Fig. 5**. The number of 2D-3D correspondences is 100. It describes the performance of different algorithms comparing to the former part, whose 2D-3D correspondences number is particularly 20. From **Fig. 5**, the performance of accuracy is promoted compared to that of the former setting.

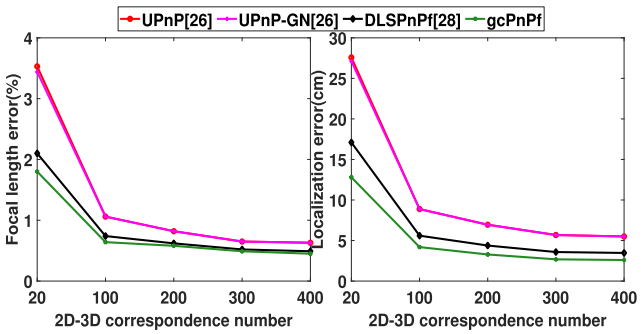


FIGURE 6. Focal length and localization results with different 2D-3D correspondences 20-400.

Fig. 6 shows the curve of the mean relative focal length error and the mean localization error w.r.t. the number of 2D-3D correspondences. Our proposed method shows better results under a different number of 2D-3D correspondences. Although the accuracy performance of UPnP approaches the other two algorithms with more 2D-3D correspondences, the computation cost dramatically increases. Numerically, the time cost of UPnP changes from 0.03s to 7.5s, while it changes slightly in DLSPnPf and our proposed method.

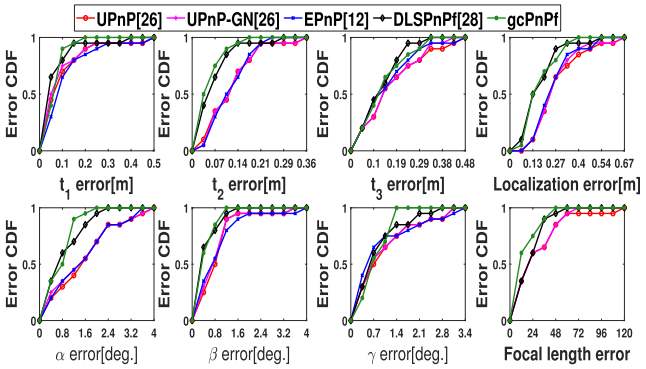


FIGURE 7. Results with $\sigma_f = 20$.

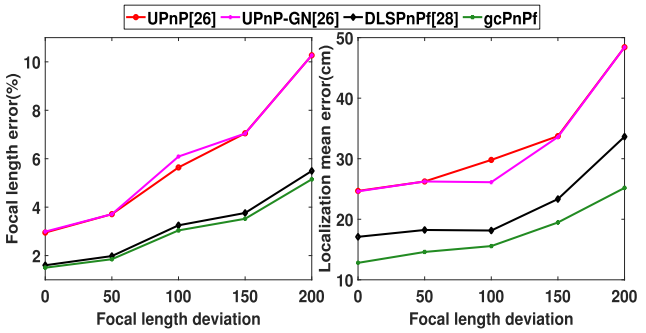


FIGURE 8. Focal length and localization results from $\sigma_f = 0$ to $\sigma_f = 200$.

The third part of the results is with image Gaussian noise $\sigma_{img} = 5$, 2D-3D correspondences number 20. The focal length f_x and f_y are chosen differently around 1000 with standard deviation σ_f, whose value is 20 in **Fig. 7**. **Fig. 8** shows the relationships between the mean error and the focal length deviation variation. From these results, the error introduced by the difference between the pinhole camera model proposed in this paper and the one in the practical application is shown. More specifically, the aspect ratio is not equal to 1. Since the model in [26], [28], and this paper treats that the focal length has a square aspect ratio, the focal length error evaluates by f from the x-axis. According to **Fig. 8**, the performance of our proposed algorithm is better than the other benchmark algorithms under this variation.

We also compare the accuracy performance of our proposed GB solver with P3.5P, which shows in Table 3. The configuration is $\sigma_{img} = 5$, $n = 20$, $f = 1000$ in non-planar case from 500 trials.

TABLE 3. Accuracy comparison(mean).

Solver	Rotation Error (degree)	Translation Error (cm)	Focal Length Error
Our proposed GB solver	4.7	0.2	214
P3.5P	6.4	0.35	259

The fourth part is from the coplanar case comparing to the former general one, which is with 20 2D-3D correspondences, image Gaussian noise $\sigma_{img} = 8$, $f = 1000$.

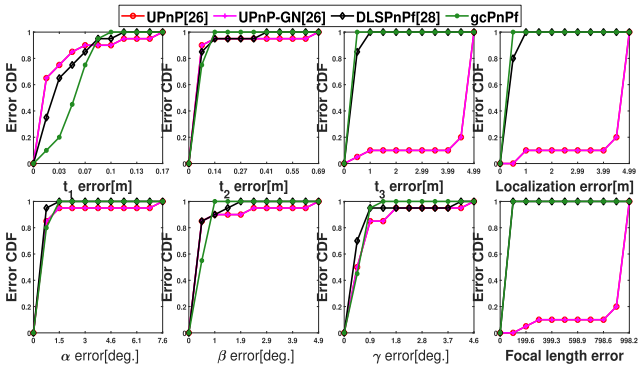


FIGURE 9. Results with coplanar case.

The simulation results show in **Fig. 9**. Note that the error of translation and rotation angle from all algorithms will significantly increase when $\sigma_{img} > 8$. UPnP degenerates more seriously than DLSPnPf and gcPnPf, partially due to its combination of translation in the z-axis and focal length into a variable. In terms of localization error, the performance of our proposed algorithm is better than the other benchmarks. The improvement can be done by reformulating the PnP equation under this coplanar case, which is beyond the scope of this paper. From the simulation results under the variation of different parameters, the accuracy of our proposed algorithm is better than the other benchmarks. The simulation results of synthetic data can summarize as follows.

(1) When σ_{img} increases, the accuracy will reduce on the overall trend.

(2) The accuracy will promote by increasing the number of 2D-3D correspondences.

(3) The accuracy will decrease by increasing the focal length deviation.

(4) The accuracy degenerates seriously in the coplanar case than the general case.

As mentioned in Section III, our goal is to provide a one-second level localization experience for a pedestrian. Since the implementation of each benchmark is different, we will not compare the computation time totally among different algorithms. Instead, we list the average latency in our proposed scheme, which shows in Table 4. Although the running time of our proposed GB solver is slightly slower than P3.5P [25], whose average computational time on our simulation platform with Intel Core CPU @2.67GHz and 8GB memory is 0.5ms. It also should be noted that the latency performance influences by the different implementation of matrix simplest form and eigenvalue decomposition algorithm.

TABLE 4. Computational time(milliseconds).

Step	Latency (mean)	Loop Number	Total Latency (mean)
GB solver	0.7	300	210
QCQP solver	306	1	306
Total			516

B. REAL DATASET

The 2D-3D correspondences calculate by algorithms proposed in [6], [9]. An open-access dataset TumIndoor [34] uses for the sake of fairness. It is used commonly in visual localization for verifying the performance of different algorithms. TumIndoor 1st Floor dataset is one of the biggest subsets, which chose as a real dataset in this paper. It provides 3146 DLSR images, a 59.7M points cloud, and 42×6 query images. The shooting positions of the querying image snap off the reference image track. In this paper, to evaluate the performance of our proposed algorithm in the real application, we only use the 42×6 query images as the query instead of choosing from the 3146 DLSR images. Some samples in TUM1 show in **Fig. 10**.

FIGURE 10. Sample images from TUM1.

Although this dataset provides a 3D points cloud with location information, it loses descriptive information of each point. We use the SURF descriptor to identify each point in our simulation and regenerate the point cloud by the common SFM technique. Then, the matching algorithm proposed in [9] is applied to calculate the 2D-3D correspondences, which finally import into the minimal solver proposed in this paper. The bound of localization is provided by KNN algorithms, which is also proposed in [9].

The localization result shows in **Fig. 11**. The outcome is from 149 images of 252 queries, which donates the success ratio is 59.13%. The reason is that the matching algorithm fails to provide enough points for the PnP localization problem. Since the dataset does not provide the rotation

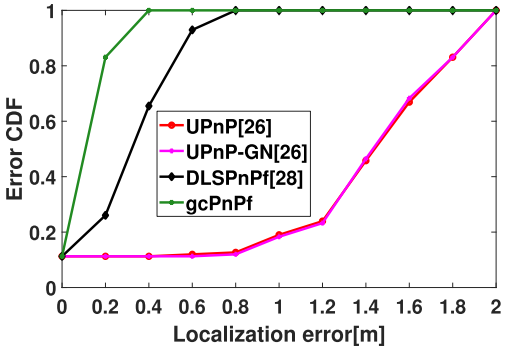


FIGURE 11. Results from 149 of 252 images in real data from TUM1.

matrix of the queries, we set the reliable bound of Eular angle at the range of $-90°$-$90°$, and the bound of translation vector is within 3 meters from the ground truth, which is the mean result of coarse localization.

The other real dataset achieves from our research center, which is in the Science Park of Harbin Institute of Technology. It covers a whole floor in building #12, whose area is close to 260m^2. The dataset is with 242 reference images, whose resolution is 1280×720. It also contains 100 test images, which are shot at the different locations with the reference images. Some samples in the dataset present in **Fig. 12**. We name it HIT-B12.

FIGURE 12. Sample images from HIT-B12.

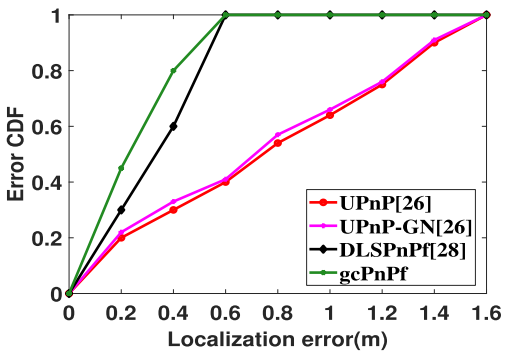


FIGURE 13. Results from 100 images in real data from HIT-B12.

The other setups of the experiment keep the same. The CDF of the localization error shows in **Fig. 13**. Since the environment of HIT-B12 is smaller than TUM1, the preorder algorithm of visual localization provides a 100% success ratio. Therefore, the results are achieved from all the test images. Our proposed algorithm still shows better accuracy than the benchmarks.

VI. CONCLUSION

In this paper, we propose an algorithm for solving the PnPf problem in an indoor image-based system based on Gröbner basis minimal solver jointed with convex optimization, which expects to provide a robust solution without running RANSAC on the online localization stage when outliers exist. Also, the solver aims at calculating the optimal results in a second-level time when one is locating himself by our proposed overall system. The proposed GB solver can obtain more accurate results by the constraints of unchanged rigid body distance comparing to the state-of-the-art solver. The entire procedure could complete within 0.6 seconds. In this way, together with the other parts of the image-based localization system, a second-level positioning experience could present. The simulation results achieve thoroughly from synthetic data by various configurations. Besides, two real experiments apply to test the performance comparing to synthetic data. In summary, our proposed algorithm can provide more reliable results than the benchmarks within the delay requirements, due to the proposed scheme in this paper. Future research will dedicate to find optimal quadruple 2D-3D correspondences, which will further reduce the repetitions of the minimal solver.

ACKNOWLEDGMENT

The authors would like to thank Z. Kukelova for sharing her code, and the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] M. Z. Chen, K. Z. Liu, J. Ma, Y. Gu, Z. Dong, and C. Liu, "SWIM: Speed-aware WiFi-based passive indoor localization for mobile ship environment," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 765–1779, Feb. 2021.

[2] Y. Yue, X. Zhao, and Z. Li, "Enhanced and facilitated indoor positioning by visible-light GraphSLAM technique," *IEEE Internet Things J.*, vol. 8, no. 2, pp. 1183–1196, Jan. 2021.

[3] Y. Xu, Y. S. Shmaliy, C. K. Ahn, T. Shen, and Y. Zhuang, "Tightly coupled integration of INS and UWB using fixed-lag extended UFIR smoothing for quadrotor localization," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1716–1727, Feb. 2021.

[4] F. Vedadi and S. Valaee, "Automatic visual fingerprinting for indoor image-based localization applications," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 1, pp. 305–317, Jan. 2020.

[5] X. Yin, L. Ma, X. Tan, and D. Qin, "A SOCP-based automatic visual fingerprinting method for indoor localization system," *IEEE Access*, vol. 7, pp. 72862–72871, 2019.

[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speed-up robust features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.

[7] M. Oelsch, D. V. Opdenbosch, and E. Steinbach, "Survey of visual feature extraction algorithm in a mars-like environment," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Taiwan, China, Dec. 2017, pp. 322–325.

[8] X. Yin, L. Ma, and X. Tan, "A PCLR-GIST algorithm for fast image retrieval in visual indoor localization system," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–5.

[9] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," in *Proc. 4th Int. Conf. Comput. Geospatial Res. Appl.*, San Jose, CA, USA, Jul. 2013, pp. 70–75.

[10] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704–1716, Sep. 2012.

[11] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

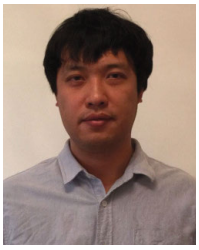
[12] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, Feb. 2009.

[13] S. Li, C. Xu, and M. Xie, "A robust O(n) solution to the perspective-n-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012.

[14] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi, "Revisiting the PnP problem: A fast, general and optimal solution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2344–2351.

[15] Y. Zheng, S. Sugimoto, and M. Okutomi, "ASPnP: An accurate and scalable solution to the perspective-n-point problem," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 7, pp. 1525–1535, 2013.

[16] 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.
[17] M. Bujňák, Z. Kukelova, and T. Pajdla, "New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion," in *Proc. Asian Conf. Comput. Vis. (ACCV)*, Queenstown, New Zealand, 2010, pp. 11–24.
[18] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[19] D. A. Cox, J. Little, and D. Oshea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. New York, NY, USA: Springer, 2007.
[20] B. Buchberger, "Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimentionalen Polynomideal," Ph.D. dissertation, Math. Inst., Univ. Innsbruck, Innsbruck, Austria, 1965.
[21] J. C. Faugére, "A new efficient algorithm for computing Gröbner bases(F_4)," *J. Pure Appl. Algebra*, vol. 139, nos. 1–3, pp. 61–88, 1999.
[22] Z. Kukelova, M. Bujňák, and T. Pajdla, "Automatic generator of minimal problem solvers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Marseille, France, 2008, pp. 302–315.
[23] M. Bujňák, Z. Kukelova, and T. Pajdla, "A general solution to the P4P problem to the absolute pose problem for camera with unknown focal length and radial distortion," in *Proc. IEEE. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Anchorage, AK, USA, Jun. 2008, pp. 24–26.
[24] M. Bujňák, "Algebraic solutions to absolute pose problems," Ph.D. dissertation, Math. Inst., Czech Tech. Univ., Prague, Czech, 2012.
[25] C. Wu, "P3.5P: Pose estimation with unknown focal length," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 2440–2448.
[26] A. Penate-Sanchez, J. Andrade-Cetto, and F. Moreno-Noguer, "Exhaustive linearization for robust camera pose and focal length estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2387–2400, Oct. 2013.
[27] Y. Zheng, S. Sugimoto, I. Sato, and M. Okutomi, "A general and simple method for camera pose and focal length determination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 430–437.
[28] Y. Zheng and L. Kneip, "A direct least-squares solution to the PnP problem with unknown focal length," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 1790–1798.
[29] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[30] S. Hadfield, K. Lebeda, and R. Bowden, "HARD-PnP: PnP optimization using a hybrid approximate representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 768–774, Mar. 2019.
[31] D. R. Grayson and E. M. Stillman, *Macaulay2, a Software System for Research in Algebraic Geometry*. Accessed: Sep. 12, 2019. [Online]. Available: http://facaulty.math.illinois.edu/Macaulay2
[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[33] M. Grant and S. Boyd. (Sep. 2013). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.0 Beta*. [Online]. Available: http://cvxr.com/cvx
[34] R. Huitl, G. Schroth, S. Hilsenbeck, F. Schweiger, and E. Steinbach, "TUMindoor: An extensive image and point cloud dataset for visual indoor localization and mapping," in *Proc. 19th IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 2012, pp. 1773–1776.
[35] E. Kanaeva, L. Gurevich, and A. Vakhitov, "Camera pose and focal length estimation using regularized distance constraints," in *Proc. 26th Brit. Mach. Vis. Conf. (BMVC)*, U.K., Sep. 2015, pp. 1–13.
[36] G. Nakano, "A versatile approach for solving PnP, PnPf, and PnPfr problems," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 338–352.
[37] L. Cheng, Y. Li, M. Xue, and Y. Wang, "An indoor localization algorithm based on modified joint probabilistic data association for wireless sensor network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 63–72, Jan. 2021.
[38] S. Kumar and S. K. Das, "Target detection and localization methods using compartmental model for Internet of Things," *IEEE Trans. Mobile Comput.*, vol. 19, no. 9, pp. 2234–2249, Sep. 2020.
[39] M. Liu, J. Du, Q. Zhou, Z. Cao, and Y. Liu, "EyeLoc: Smartphone vision enabled plug-n-play indoor localization in large shopping malls," *IEEE Internet Things J.*, early access, Oct. 15, 2020, doi: 10.1109/JIOT.2020.3031285.
[40] F. Potortì *et al.*, "The IPIN 2019 indoor localisation competition—Description and results," *IEEE Access*, vol. 8, pp. 206674–206718, 2020.
[41] C. W. Chen, "Internet of video things: Next-generation IoT with visual sensors," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6676–6685, Aug. 2020.



XILIANG YIN received the B.S. and M.S. degrees from Harbin Engineering University, Harbin, China, in 2005 and 2008, respectively. He is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, Harbin Institute of Technology. From 2008 to 2009, he was with ZTE, Shanghai, where he worked on the research and development of the 3G cellular system. From 2009 to 2014, he was with Hytera, Harbin, where he worked on the research and development of digital trunking system. Since 2015, he has been an Instructor with the Harbin Vocational and Technical College, Harbin. His research interests include location-based service, machine learning, and convex optimization.

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 was 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 service, cognitive radio, and cellular networks.

XUEZHI TAN (Member, IEEE) received the M.S. and Ph.D. degrees in communication engineering from the Harbin Institute of Technology, Harbin, China, in 1986 and 2005, respectively. He is currently a Professor with the School of Electronics and Information Engineering, Harbin Institute of Technology. His major research interests include data communication, broadband multimedia trunk communication, and cognitive radio networks. He is also a Senior Member of the China Institute of Communications and the China Institute of Electronics, the Executive Director of the Heilongjiang Province Institute of Electronics, the Vice President of the Software Association of Heilongjiang Province, and the Director of the China Radio Association.

DANYANG QIN received the B.Sc. degree in communication engineering and the M.Sc. and Ph.D. degrees in information and communication system from the Harbin Institute of Technology, in 2006, 2008, and 2011, respectively. She is currently an Associate Professor with Heilongjiang University. Her current research interests include wireless sensor networks, wireless multihop routing, and ubiquitous sensing.