Re-weighting and 1-Point RANSAC-Based PnP Solution to Handle Outliers

Haoyin Zhou, Tao Zhang, *Senior Member, IEEE*, and Jayender Jagadeesan, *Member, IEEE*

Abstract—The ability to handle outliers is essential for performing the perspective-n-point (PnP) approach in practical applications, but conventional RANSAC+P3P or P4P methods have high time complexities. We propose a fast PnP solution named R1PPnP to handle outliers by utilizing a soft re-weighting mechanism and the 1-point RANSAC scheme. We first present a PnP algorithm, which serves as the core of R1PPnP, for solving the PnP problem in outlier-free situations. The core algorithm is an optimal process minimizing an objective function conducted with a random control point. Then, to reduce the impact of outliers, we propose a reprojection error-based re-weighting method and integrate it into the core algorithm. Finally, we employ the 1-point RANSAC scheme to try different control points. Experiments with synthetic and real-world data demonstrate that R1PPnP is faster than RANSAC+P3P or P4P methods especially when the percentage of outliers is large, and is accurate. Besides, comparisons with outlier-free synthetic data show that R1PPnP is among the most accurate and fast PnP solutions, which usually serve as the final refinement step of RANSAC+P3P or P4P. Compared with REPPnP, which is the state-of-the-art PnP algorithm with an explicit outliers-handling mechanism, R1PPnP is slower but does not suffer from the percentage of outliers limitation as REPPnP.

Index Terms—Perspective-n-point, 1-point RANSAC, soft re-weighting, robustness to outliers

✦

1 INTRODUCTION

THE perspective-n-point (PnP) problem aims to determine the position and orientation of a calibrated camera from n known correspondences between three-dimensional (3D) object points and their two-dimensional (2D) image projections. PnP is a core problem in the computer vision field and has found many applications, such as robot vision navigation [1], augmented reality [2], and computer animation. In the past decades, many effective PnP approaches have been proposed with very fast computational speed [3], [4] and high accuracy [5], [6].

To date, most PnP algorithms are designed under the assumption that no outlier exists among the given 3D-2D correspondences. However, in practical applications, this outlier-free assumption is often difficult to satisfy. This is because image feature detection and matching approaches, such as SURF [7], BRISK [8] and ORB [9], do not always give perfect results due to scaling, illumination, shadow and occlusion. Outliers are often unavoidable and they have a significant impact on the PnP methods. Even a small percentage of outliers will lead to a significant decrease in accuracy. Hence, the ability to handle outliers is essential

for performing PnP algorithms in practical applications. The most common outliers handling mechanism is to combine a PnP ($n = 3$ or 4) algorithm [10], [11], [12] with the RANSAC-based scheme [13] to eliminate outliers, and then perform a more accurate PnP algorithm with the remaining inliers to refine the result. A number of very fast closed-form P3P [14] or P4P [15] algorithms have been proposed. However, their RANSAC combination scheme still needs many trials until the selected three or four 3D-2D correspondences are all inliers, which results in a high time complexity. Hence, the computational speed decreases significantly as the percentage of outliers increases. To reduce the time complexity, one natural idea is to utilize the PnP algorithm with a smaller n. However, when $n = 1$ or 2, the PnP problem has infinitely many solutions, which makes the conventional RANSAC-based scheme infeasible.

To the best of our knowledge, except for RANSAC-based methods, the only PnP method that addresses outliers is REPPnP [3] proposed by Ferraz et al., which is the state-of-the-art PnP method that is robust to outliers. REPPnP integrates an outlier rejection mechanism with camera pose estimation. It formulates the pose estimation problem as a low-rank homogeneous system in which the solution lies on its one-dimensional (1D) null space. Outlier correspondences are those rows of the linear system that perturb the null space and are progressively detected by projecting them on an iteratively estimated solution of the null space. Although REPPnP is very fast and accurate, it suffers from a severe limitation that it cannot handle more than approximately 50 percent of outliers.

In this paper, we propose a robust 1-point RANSAC-based PnP method named R1PPnP. We first present an optimal iterative process as the core PnP algorithm of R1PPnP.

- H. Zhou and J. Jagadeesan are with the Surgical Planning Laboratory, Brigham and Women's Hospital, Harvard Medical SchoolBoston, MA 02115. E-mail: {zhouhaoyin, jayender}@bwh.harvard.edu.
- T. Zhang is with the Department of Automation, School of Information Science and Technology, Tsinghua University, Beijing 100086, China. E-mail: taozhang@tsinghua.edu.cn.

The core algorithm takes a random 3D-2D correspondence as the control point. To address outliers, we propose a soft weight assignment method according to reprojection errors to distinguish inliers and outliers, and integrate it into the core algorithm. The weight factors associated with outliers decrease significantly during the iteration to reduce the impact of outliers. Finally, we employ the 1-point RANSAC scheme to try different control points for the core PnP algorithm. By using this combination of the RANSAC scheme and the soft weight assignment, the algorithm is capable of eliminating outliers when the selected control point is an inlier.

The main advantage of R1PPnP is that it has much lower time complexity and is much faster than conventional RANSAC+P3P or P4P methods, especially when the percentage of outliers is large. Compared with REPPnP, the proposed R1PPnP does not suffer from the percentage of outliers limitation.

This paper is organized as follows. In Section 2, we describe the fundamental model used in R1PPnP. The details of the core algorithm are given in Section 3, in which we also provide its proof of convergence, local minima analysis and the strategy to select control points. The outliers handling mechanism, including soft weight assignment and the 1-point RANSAC scheme, is introduced in Section 4. We also provide details of termination conditions in Section IV. Evaluation results are presented in Section 5. A discussion and description of planned future work is described in Section 6.

1.1 Related Works

The PnP problem, coined by Fischler and Bolles [13], is articulated as follows: *Given the relative spatial locations of n control points, and given the angle to every pair of control points P_i from an additional point called the center of perspective C, find the lengths of the line segments joining C to each of the control points.* The PnP problem has been studied for many years. In early studies, direct linear transformation (DLT) [16] was used as a solution in a straightforward way by solving a linear system. However, DLT ignores the intrinsic camera parameters, which are assumed to be known, and therefore generally leads to a less stable pose estimate.

In the past decade, researchers have proposed many PnP methods to improve speed, accuracy, and robustness to outliers. The PnP methods can be roughly classified into non-iterative and iterative methods. Generally speaking, non-iterative methods are more efficient but are unstable under image noise and outliers. Many non-iterative PnP methods are based on a set of small number of points ($n = 3, 4$). They are referred to as P3P [17], [11], [14] or P4P [18], [15], [19], [20] methods. P3P is the smallest subset of control points that yields a finite number of solutions [14], [21]. When the intrinsic camera parameters are known and we have $n \geq 4$ points, the solution is generally unique. Triggs proposed a PnP method with four- or five- correspondences [22]. These PnP methods based on less than four correspondences do not make use of redundant points and are very sensitive to noise and outliers. However, due to their efficiency and capability to calculate from a small point set, P3P or P4P methods are very useful for combing a RANSAC-like scheme to reject outliers. There are also many non-iterative PnP methods that are able to make use of redundant points but are quite time consuming. For example, Ansar's method is $O(n^8)$ [23] and

Fiore's is $O(n^2)$ [24]. Schweighofer proposed an $O(n)$ PnP method named SDP, but is slow [25]. In recent years, three excellent $O(n)$ effective non-iterative PnP methods, EPnP [4], RPnP [26] and UPnP[27], have been proposed, and these methods are very efficient and accurate even compared to iterative methods.

Iterative PnP methods [6], [28], [29], [30] are mostly optimization methods that decrease their energy function in the iterative process. They are generally more accurate and robust, but slower. For example, Dementhon proposed POSIT that is easy to implement [29] and further proposed SoftPOSIT to handle situations when the correspondence relationships are unknown [31]. Although SoftPOSIT has a certain ability to handle outliers, the strong assumption that all correspondences are unknown make it slow. Lu's method [6] is the most accurate iterative PnP method but may get stuck in local minima. Schweighofer discussed the local minima situation of Lu's method and proposed a method to avoid this limitation [32].

PnP algorithms are widely used in applications such as structure from motion [33] and monocular SLAM [34], which require dealing with hundreds or even thousands of noisy feature points and outliers in real-time. The fact that outliers have a much greater impact on PnP accuracy than image Gaussian white noise makes it is necessary for the PnP algorithm to handle outliers efficiently. Conventional method to handle outliers is to combine a RANSAC-like scheme with the P3P or P4P algorithms. Besides, L1-norm is also widely used to handle a certain amount of outliers [35], [36] because the L1-norm penalty is less sensitive to outliers than the L2-norm penalty. Although a L1-norm-based energy function is more robust to outliers, it cannot absolutely get rid of outliers and its computation is more complex.

Ferraz et al. proposed a very fast PnP method that can handle up to 50 percent of outliers [3]. The outlier rejection mechanism is integrated within the pose estimation pipeline with negligible computational overhead. Compared to Ferraz's method, the R1PPnP algorithm proposed in this paper demonstrates much stronger robustness, but is slower.

2 FUNDAMENTAL MODEL

In this paper we denote the camera frame as c and the world frame as w. For point i, without taking into account the distortion, the perspective projection equations are employed to describe the pinhole camera model,

$$u_i = f \frac{x_i^c}{z_i^c}, v_i = f \frac{y_i^c}{z_i^c}, \qquad (1)$$

where f is the camera focal length, $\mathbf{x}_i = [u_i, v_i, f]^T$ is the image homogeneous coordinate in pixels, and $\mathbf{X}_i^c = [x_i^c, y_i^c, z_i^c]^T$ is the real-world coordinate with respect to the camera frame.

According to Eq. (1),

$$\mathbf{X}_i^c = \lambda_i^* \mathbf{x}_i, \qquad (2)$$

where $\lambda_i^* = z_i^c/f$ is the normalized depth of point i. Eq. (2) indicates that an object point lies on the straight line of sight of the related image point.

The relationship between the camera and world frame coordinate of point i is

$$\mathbf{X}_i^c = \mathbf{R}\mathbf{X}_i^w + \mathbf{t}, \qquad (3)$$

where $\mathbf{R} \in SO(3)$ is the rotation matrix and $\mathbf{t} \in R^3$ is the translation vector. $\mathbf{R}$ and $\mathbf{t}$ are the variables that need to be estimated in the PnP problem.

Similarly to the translation elimination method used in works [37], [38], with two points i and o,

$$\mathbf{X}_i^c - \mathbf{X}_o^c = \mathbf{R}(\mathbf{X}_i^w - \mathbf{X}_o^w), \quad i \neq o. \tag{4}$$

In the proposed R1PPnP algorithm, $o \in [1, N]$ suggests the index of the control point, N is the number of 3D-2D correspondences. R1PPnP represents the shape of the point cloud by the relative positions between the control point o and other $N - 1$ points. Denoting $\mathbf{S}_i = \mathbf{X}_i^w - \mathbf{X}_o^w$, where S means "shape", then, according to Eqs. (2) and (4),

$$\lambda_i^* \mathbf{x}_i - \lambda_o^* \mathbf{x}_o = \mathbf{R}\mathbf{S}_i. \tag{5}$$

We divide both sides of Eq. (5) by the depth of the control point λ_o^*, and rewrite (5) as

$$\lambda_i \mathbf{x}_i - \mathbf{x}_o = \mu \mathbf{R}\mathbf{S}_i, \tag{6}$$

where $\lambda_i = \mu \lambda_i^*$ and $\mu = 1/\lambda_o^*$ is the scale factor. We have

$$\mathbf{t} = 1/\mu \mathbf{x}_o - \mathbf{R}\mathbf{X}_o^w. \tag{7}$$

According to Eqs. (6) and (7), the PnP problem can be solved by minimizing the objective function

$$f(\mathbf{R}, \mu, \lambda_i) = \sum_{i=1, i \neq o}^{N} \| \lambda_i \mathbf{x}_i - \mathbf{x}_o - \mu \mathbf{R}\mathbf{S}_i \|^2, \tag{8}$$

where $\|\cdot\|$ is the L2-norm.

The objective function Eq. (8) is based on euclidean distances in the 3D space. Compared with the reprojection error cost, Eq. (8) gives more weights to points with larger depths. For example, the same level of reprojection error has relatively larger effects when related to an object point with greater depth. To solve this problem, we normalize the cost function Eq. (8) with depths of points and propose the objective function of our R1PPnP algorithm, that is

$$f(\mathbf{R}, \mu, \lambda_i) = \sum_{i=1, i \neq o}^{N} \left(\frac{1}{\lambda_i} \| \lambda_i \mathbf{x}_i - \mathbf{x}_o - \mu \mathbf{R}\mathbf{S}_i \| \right)^2, \tag{9}$$

where $1/\lambda_i$ is introduced to adjust the weight of point i to eliminate the inequity among points in Eq. (8).

We estimate $\mathbf{R}$, μ and λ_i ($i = 1, \dots, N, i \neq o$) by minimizing the objective function Eq. (9), the variables of which consist of two parts: the camera pose $\{\mathbf{R}, \mu\}$ and the relative depths with respect to the control point $\{\lambda_i\}$. To describe the following algorithm intuitively, we introduce two sets of points: $\mathbf{p}_i$ and $\mathbf{q}_i$. With a randomly selected control o, points $\mathbf{p}_i$ are determined by the camera pose $\{\mathbf{R}, \mu\}$, and points $\mathbf{q}_i$ are determined by depths λ_i. We have

$$\mathbf{p}_i = \mathbf{x}_o + \mu \mathbf{R}\mathbf{S}_i, \tag{10}$$

$$\mathbf{q}_i = \lambda_i \mathbf{x}_i. \tag{11}$$

As shown in Fig. 1, points $\mathbf{p}_i$ are attached with the virtual object obtained by rotating and scaling the real object around the control point $\mathbf{p}_o = \mathbf{q}_o = \mathbf{x}_o$. $\mathbf{q}_i$ is the projection of
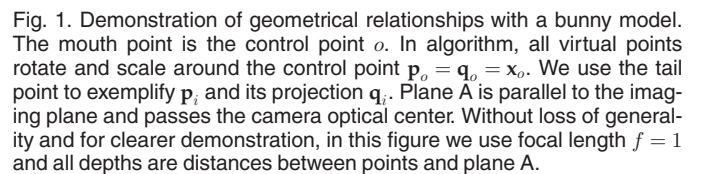

Fig. 1. Demonstration of geometrical relationships with a bunny model. The mouth point is the control point o. In algorithm, all virtual points rotate and scale around the control point $\mathbf{p}_o = \mathbf{q}_o = \mathbf{x}_o$. We use the tail point to exemplify $\mathbf{p}_i$ and its projection $\mathbf{q}_i$. Plane A is parallel to the imaging plane and passes the camera optical center. Without loss of generality and for clearer demonstration, in this figure we use focal length $f = 1$ and all depths are distances between points and plane A.

$\mathbf{p}_i$ on the corresponding line of sight. The objective function Eq. (9) is equivalent to

$$f(\mathbf{p}_i, \mathbf{q}_i) = \sum_{i=1, i \neq o}^{N} \left(\frac{1}{\lambda_i} \| \mathbf{p}_i - \mathbf{q}_i \| \right)^2. \tag{12}$$

As this objective function approaches the global optimal solution and as shown in Fig. 1, point $\mathbf{p}_i$ gets close to $\mathbf{q}_i$ and the z-component of $\mathbf{p}_i$ gets close to $f\lambda_i$. Hence, it is expected that the objective function Eq. (12) has similar optimal solutions as the conventional reprojection error cost, because

$$f(\mathbf{p}_i, \mathbf{q}_i)$$
$$= \sum_{i=1, i \neq o}^{N} \left\| \left[\frac{\mathbf{p}_{i,x}}{\lambda_i}, \frac{\mathbf{p}_{i,y}}{\lambda_i}, \frac{\mathbf{p}_{i,z}}{\lambda_i} \right]^T - \mathbf{x}_i \right\|^2$$
$$\approx \sum_{i=1, i \neq o}^{N} \left\| \left[f\frac{\mathbf{p}_{i,x}}{\mathbf{p}_{i,z}}, f\frac{\mathbf{p}_{i,y}}{\mathbf{p}_{i,z}}, f \right]^T - [u_i, v_i, f]^T \right\|^2. \tag{13}$$

3 CORE ALGORITHM DESIGN

We first introduce the core algorithm of R1PPnP, which solves the PnP problem in outlier-free situations. This section introduces the core algorithm process, proof of convergence, the local minima avoidance mechanism and the strategy to select the control point.

The core algorithm of R1PPnP is an optimal iterative process with the objective function Eq. (9) or (12). In each iteration, it estimates the points set $\mathbf{q}_i$ and $\mathbf{p}_i$ ($i = 1, \dots, N, i \neq o$) alternately by fixing one points set and updating the other one according to the objective function minimization.

(1) $\mathbf{q}_i$ estimation stage. Because each $\mathbf{q}_i$ are independent with each other, our algorithm seeks the closest $\mathbf{q}_i$ for each
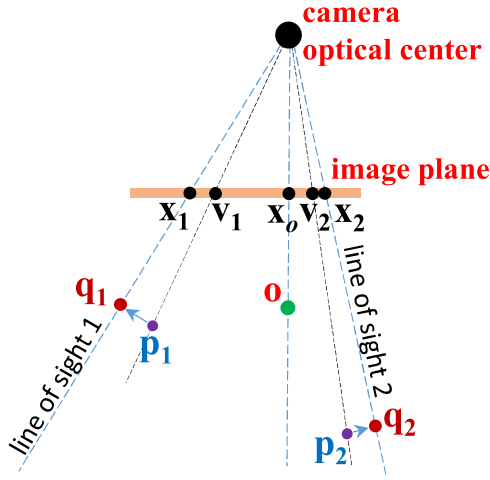
Fig. 2. Demonstration of the updating method of the scale factor μ. One possible method is to update μ according to the euclidean distances between $\mathbf{p}_i$, $\mathbf{q}_i$, and $\mathbf{o}$, which works for $\mathbf{p}_1$ and $\mathbf{q}_1$ because they have close depths as $\mathbf{o}$. However, this method may result in slow μ updating rate for $\mathbf{p}_2$ and $\mathbf{q}_2$ because $\|\mathbf{q}_2 - \mathbf{o}\| \approx \|\mathbf{p}_2 - \mathbf{o}\|$. Hence, it is more efficient to compare $\mathbf{v}_i$ and $\mathbf{x}_i$ to move points $\mathbf{p}_i$ to the related lines of sight.

$\mathbf{p}_i$. According to Eq. (11), points $\mathbf{q}_i$ are constrained to the related lines of sight. Hence, we vertically project $\mathbf{p}_i$ onto the related lines of sight to obtain the points' relative depths with respect to the control point o by

$$\lambda_i = \mathbf{x}_i^T \mathbf{p}_i / (\mathbf{x}_i^T \mathbf{x}_i), \; i = 1, \ldots, N \text{ and } i \neq o. \quad (14)$$

Then, points $\mathbf{q}_i$ are updated according to Eq. (11).

(2) $\mathbf{p}_i$ Estimation Stage. Points $\mathbf{p}_i$ are determined by $\mathbf{R}$ and μ. According to Eq. (10), the updated $\mathbf{R}$ and μ should make points $\{\mu \mathbf{R} \mathbf{S}_i\}$ have the smallest weighted sum of squared distances to points $\{\mathbf{q}_i - \mathbf{x}_o\}$, and subject to $\mathbf{R}^T \mathbf{R} = \mathbf{I}_{3 \times 3}$. According to the objective function Eq. (12), the weights used in this stage are $1/\lambda_i$ in the previous iteration.

Denoting matrices $\mathbf{A} = \left[\frac{\mathbf{q}_1 - \mathbf{x}_o}{\lambda_1}, \ldots, \frac{\mathbf{q}_N - \mathbf{x}_o}{\lambda_N}\right]_{3 \times N}$ and $\mathbf{S} = \left[\frac{\mathbf{S}_1}{\lambda_1}, \ldots, \frac{\mathbf{S}_N}{\lambda_N}\right]_{3 \times N}$, then according to Ref. [39]

$$[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}^T] = \text{svd}(\mathbf{A}\mathbf{S}^T), \mathbf{R} = \mathbf{U}\mathbf{V}^T, \quad (15)$$

Because points $\mathbf{p}_i$ are directly generated from $\mathbf{S}_i$ according to Eqs. (10), (15) suggests that $\mathbf{R}$ is updated according to the differences between points $\mathbf{p}_i$ and $\mathbf{q}_i$ in the 3D space. However, Fig. 2 demonstrates that by using this method, the updating rate of μ may be slow in situations when the range of depths is large. To achieve faster convergence rate, we update the scale factor μ by comparing the projected image coordinates of $\mathbf{p}_i$, which are denoted as $\mathbf{v}_i$, and the real image points $\mathbf{x}_i$. Denoting matrices $\mathbf{B} = [\mathbf{v}_1 - \mathbf{x}_o, \ldots, \mathbf{v}_N - \mathbf{x}_o]_{3 \times N}$ and $\mathbf{C} = [\mathbf{x}_1 - \mathbf{x}_o, \ldots, \mathbf{x}_N - \mathbf{x}_o]_{3 \times N}$. μ is updated by

$$\Delta \mu = \|\text{vector}(\mathbf{C})\| / \|\text{vector}(\mathbf{B})\| \quad (16)$$

$$\mu_{\text{new}} = \mu_{\text{old}} \Delta \mu \quad (17)$$

Finally, points $\mathbf{p}_i$ are updated according to Eq. (10).

3.1 Proof of Convergence

We first provide the mathematical proof of the convergence of R1PPnP when not using $1/\lambda_i$ as weights in the objective function Eq. (12). k denotes the number of iterations, $\mathbf{q}_i^{(k+1)}$
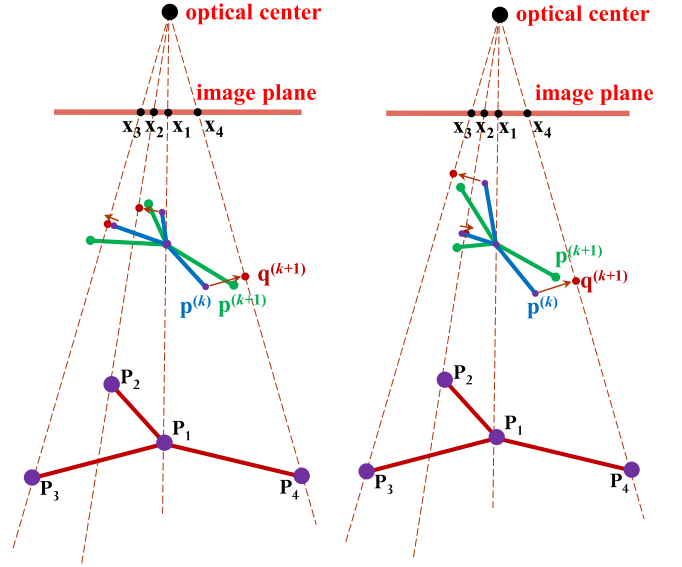

Fig. 3. Iterative process with 2D space and 1D camera imaging plane. $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, and $\mathbf{P}_4$ are the object points; $\mathbf{P}_1$ is selected as the control point ($o = 1$). (a) The process $\mathbf{p}^{(k)} \rightarrow \mathbf{q}^{(k+1)} \rightarrow \mathbf{p}^{(k+1)} \rightarrow \ldots$ makes the estimation pose approach the correct solution. (b) The rotation related to $\mathbf{p}^{(k+1)}$ is worse than that related to $\mathbf{p}^{(k)}$, which means the process is approaching a local minima, which is a mirror-image form of the true object shape.

is obtained by vertically projecting $\mathbf{p}_i^{(k)}$ to the line of sight i, and $\mathbf{q}_i^{(k+1)}$ and $\mathbf{q}_i^{(k)}$ are on the line of sight i. Hence, the three points, $\mathbf{p}_i^{(k)}$, $\mathbf{q}_i^{(k)}$, and $\mathbf{q}_i^{(k+1)}$ comprise a right-angled triangle. Therefore, for each index $i, i \neq o$,

$$\left\|\mathbf{p}_i^{(k)} - \mathbf{q}_i^{(k+1)}\right\|^2 = \left\|\mathbf{p}_i^{(k)} - \mathbf{q}_i^{(k)}\right\|^2 - \left\|\mathbf{q}_i^{(k+1)} - \mathbf{q}_i^{(k)}\right\|^2. \quad (18)$$

In the $\mathbf{p}^{(k+1)}$ updating stage, the updated $\mathbf{R}$ and μ make the objective function Eq. (12) smaller. Hence,

$$\sum_{i=1}^{N} \left\|\mathbf{p}_i^{(k+1)} - \mathbf{q}_i^{(k+1)}\right\|^2 \leq \sum_{i=1}^{N} \left\|\mathbf{p}_i^{(k)} - \mathbf{q}_i^{(k+1)}\right\|^2. \quad (19)$$

According to Eqs. (18), (19), and the objective function Eq. (12),

$$f(\mathbf{p}_i^{(k+1)}, \mathbf{q}_i^{(k+1)}) \leq f(\mathbf{p}_i^{(k)}, \mathbf{q}_i^{(k)}) - \sum_{i=1}^{N} \left\|\mathbf{q}_i^{(k+1)} - \mathbf{q}_i^{(k)}\right\|^2. \quad (20)$$

Hence, the objective function will strictly decrease until $\mathbf{q}_i^{(k+1)} = \mathbf{q}_i^{(k)}$ when not using $1/\lambda_i$ as weights. However, when $1/\lambda_i$ is applied in the objective funtion, the above convergence proof is not rigorous in mathematics because $\lambda_i^{(k+1)} \neq \lambda_i^{(k)}$. As the iteration process, the changes of λ_i become small, which makes the formula Eq. (20) hold. In addition, our experimental results in this paper also support the assumption that our algorithm is convergent.

3.2 Local Minima Avoidance

We have concluded that the iterative process of R1PPnP is convergent. However, we still need to address situations that R1PPnP may get stuck in local minima. To demonstrate the iterative process more intuitively, we introduce a 1D camera working in the 2D space, as shown in Fig. 3. In this demonstration, an object with four points $\mathbf{P}_i$, $i = 1, \ldots, 4$ are projected to the camera image plane and their image points $\mathbf{x}_i$ are obtained. $\mathbf{P}_1$ is selected as the control point, which

means $o = 1$. Different initial values may result in different convergence results.

Fig. 3a demonstrates a process that is approaching the correct global optimal results. Beginning with points $\mathbf{p}^{(k)}$, the algorithm projects $\mathbf{p}^{(k)}$ to their related lines of sight and obtains points $\mathbf{q}^{(k+1)}$. Then, according to $\mathbf{q}^{(k+1)}$, the algorithm updates the rotation $\mathbf{R}$ and scale factor μ to generate points $\mathbf{p}^{(k+1)}$. In this process, the rotation and scale factor related to $\mathbf{p}^{(k+1)}$ are closer to the truth compared to that related to $\mathbf{p}^{(k)}$, and finally the algorithm will reach the correct solution.

However, as the progress shown in Fig. 3b indicates, $\mathbf{p}^{(k+1)}$ has a larger pose error than $\mathbf{p}^{(k)}$, and the algorithm will finally get stuck in local minima. The reconstructed points $\mathbf{p}^{(k)}$ or $\mathbf{p}^{(k+1)}$ come in mirror-image forms of the real object points $\mathbf{P}$. Without loss of generality, in either 2D or 3D space, a mirror-image form suggests that the left-right-handed shape of the point cloud has been changed, which should not happen in reality. The reason the core algorithm of R1PPnP may generate points with different left-right-handed shape is that its rotation estimation Eq. (15) does not constrict $\det(\mathbf{R}) = 1$.

In practice we found it not appropriate to constrain $\det(\mathbf{R}) = 1$ from the beginning of the algorithm. Instead, we allow the iteration process to approach the mirror-image form. This is because we found that, with the constrain $\det(\mathbf{R}) = 1$ from the beginning, the algorithm has many types of local minima and they are unpredictable. However, without this constrain, the convergence direction of the core algorithm becomes predictable, with only two types of convergence. The algorithm may reach the global optimal result directly or the approximate mirror-image form. For the latter case, the estimated $\det(\mathbf{R}) = -1$.

Hence, according to the above analysis, we propose the local minima avoidance mechanism. The algorithm begins with a random initial value and control point. When the algorithm converges to a result with $\det(\mathbf{R}) = -1$, we perform a mirror flip by

$$\lambda_{i,\text{new}} = 1/\lambda_{i,\text{old}}, \; i = 1, \ldots, N \text{ and } i \neq o. \qquad (21)$$

3.3 Control Point o Selection
To select different points as the control point o may result in different convergence rates. Without taking into account noise, the correct value of rotation $\mathbf{R}$ should be the same for any control point o in a PnP task. Hence, larger rotation updating steps in the iteration process suggest that less number of iterations are required to converge to the correct value when starting from the same initial value. In R1PPnP, $\mathbf{R}$ is updated according to the differences between points $\mathbf{p}_i$ and $\mathbf{q}_i$, $i = 1, \ldots, N, i \neq o$. When point o is close to $\mathbf{p}_i$, the rotation updating steps are more likely to be large, as shown in Fig. 4a. The updating rates of μ also follow this analysis. Therefore, we are prone to select the control point o from the center of the point cloud, which has better odds of having smaller distances to the rest of the point cloud to achieve faster convergence rate, as shown in Fig. 4b.

4 OUTLIERS HANDLING MECHANISM
The robust and fast capability of handling outliers is the main contribution of the proposed R1PPnP algorithm. Our

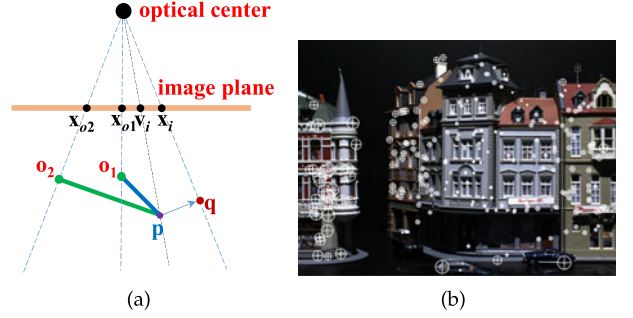(a)   (b)

Fig. 4. In R1PPnP, the selection of control point o is related to the convergence rate. (a) An example to illustrate this behavior with 2D space and 1D camera imaging plane . According to the $\mathbf{R}$ and μ updating methods in Eqs. (15) and (16). $\angle(p - o_1, q - o_1) > \angle(p - o_2, q - o_2)$ and $\|\mathbf{x}_i - \mathbf{x}_{o1}\|/\|\mathbf{v}_i - \mathbf{x}_{o1}\| > \|\mathbf{x}_i - \mathbf{x}_{o2}\|/\|\mathbf{v}_i - \mathbf{x}_{o2}\|$ suggest that the iteration process is more likely to have larger $\mathbf{R}$ and μ updating rate when o is closer to p. (b) A real-world example to illustrate this behavior, the radius of a circle represents the required number of iterations when using this feature point as the control point o.

outliers handling mechanism combines a soft weight assignment method and the 1-point RANSAC scheme.

4.1 Soft Re-Weighting
R1PPnP mainly consists of $\mathbf{q}_i$ and $\mathbf{p}_i$ estimation stages. As described in Section 3, in the $\mathbf{q}_i$ stage, calculations related to each point are independent from the others. Hence, outliers do not affect inliers in the $\mathbf{q}_i$ stage. However, in the $\mathbf{p}_i$ stage, outliers perturbs the camera pose estimation results. To reduce the impact of outliers, the basic idea of our soft re-weighting method is to assign each 3D-2D correspondence a weight factor, and to make weight factors related to outliers small when estimating the camera pose in the $\mathbf{p}_i$ stage.

One possible method to assign weights is based on least median of squares [40], however this method cannot handle more than 50 percent of outliers. We designed a soft weight assignment method embedded in the iteration process. To distinguish inliers and outliers, the weights of 3D-2D correspondences are determined by

$$w_i = \begin{cases} 1.0 & if \; e_i \leq H \\ H/e_i & if \; e_i > H, \end{cases} \qquad (22)$$

where e_i suggests the reprojection error of point i with the current $\mathbf{R}$ and μ during iteration, H is the inliers threshold that points with final reprojection errors smaller than H are considered as inliers. The reweighting rule Eq. (22) suggests that a point with a large reprojection error will have a small weight during the estimation of camera pose, which is designed under a reasonable assumption that outliers have much larger reprojection errors than inliers. Although inliers may also have larger reprojection errors than H during the iteration process, it is acceptable to assign weights that are smaller than 1 to inliers as long as outliers have much smaller weights. Hence, we simply use H as the benchmark to assign weights.

According to $\mathbf{R}$ estimation given by Eq. (15), we multiply the weight factors with each item of matrices $\mathbf{A}$ and $\mathbf{S}$,

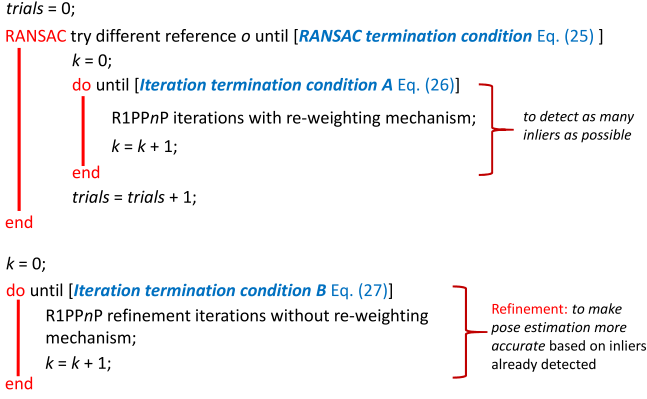$$\mathbf{A}_i = w_i \mathbf{A}_i, \mathbf{S}_i = w_i \mathbf{S}_i. \qquad (23)$$

```
trials = 0;
RANSAC try different reference o until [RANSAC termination condition Eq. (25) ]
    k = 0;
    do until [Iteration termination condition A Eq. (26)]          ] to detect as many
        R1PPnP iterations with re-weighting mechanism;             ] inliers as possible
        k = k + 1;
    end
    trials = trials + 1;
end

k = 0;
do until [Iteration termination condition B Eq. (27)]          ] Refinement: to make
    R1PPnP refinement iterations without re-weighting         ] pose estimation more
    mechanism;                                                ] accurate based on inliers
    k = k + 1;                                                ] already detected
end
```

Fig. 5. The overall flow chart of R1PPnP.

Similarity, to update μ using Eq. (16),

$$\mathbf{B}_i = w_i \mathbf{B}_i, \mathbf{C}_i = w_i \mathbf{C}_i. \tag{24}$$

Since inliers have much larger weights, $\mathbf{R}$ and μ are mainly estimated with inliers.

4.2 1-Point RANSAC Scheme

The core algorithm of R1PPnP is based on a randomly selected control point o. In outlier-free situations, our algorithm works with any control point. However, in situations with outliers, the control point o should be an inlier to make the algorithm work. Hence, we employ the 1-point RAN-SAC scheme to try different 3D-2D correspondences as the control point until the algorithm finds the correct solution. The 1-point RANSAC scheme combines the core algorithm naturally because the core algorithm can perform the computation with any control point $o \in [1, N]$. We assume that 2D-3D correspondences have the same the possibility to be an inlier, without loss of generality, we select the control point o from the center of all image points to the outside. This is because we found that R1PPnP needs less iterations to converge when o is closer to the center, the details of which has been discussed in Section 3.3.

4.3 Algorithm Flow Chart

In general, the overall flow chart of R1PPnP is shown in Fig. 5, we first detect as many inliers as possible inside the RANSAC framework, then based on the detected inliers, we perform the R1PPnP algorithm without re-weighting mechanism to get more accurate results.

Appropriate termination conditions seek balance between speed and precision for RANSAC-based or iterative algorithms. As shown in Fig. 5, two types of termination conditions need to be specified for R1PPnP.

(1) *RANSAC Termination Condition*. The standard RAN-SAC termination condition [13] was employed for R1PPnP, that is

$$trials \geq \log{(1 - p)} / \log{(1 - p_{\text{inliers}}^s)}, \tag{25}$$

where p is the certainty and we use $p = 0.99$ for all RAN-SAC-based methods in this paper, $trials$ is the number of RANSAC trials, $p_{\text{inliers}} =$ (maximum number of detected inliers) / (number of all points), s is the number of control points needed in each RANSAC trial. $s = 1$ for R1PPnP, and $s = 3, 4$ for RANSAC+P3P and P4P respectively.
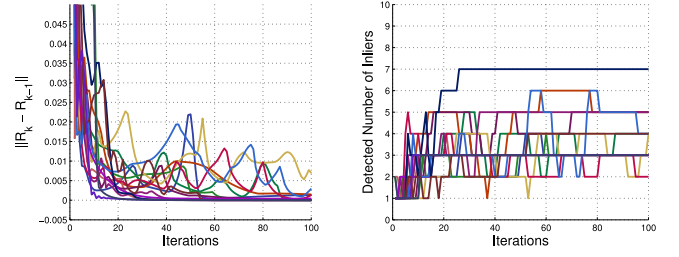

Fig. 6. Experiments with synthetic data (ordinary 3D case, 50 percent of outliers) to demonstrate the iteration process of R1PPnP when the control point o is an outlier. Randomly colored lines are results with different control points. (a) The changes of estimated camera pose between frames k and $k - 1$ are complex during the iteration process, based on which it is difficult to decide when to stop the process. (b) It is more robust and efficient to stop the process when no more inliers can be detected.

During the RANSAC process, the camera pose estimated by conventional RANSAC+P3P or P4P methods is based on very small number of points. Because of image noise, the estimated pose varies with different inliers as the control points. This is especially serious when the image noise is large. To improve accuracy, the termination condition Eq. (25) suggests that the standard RANSAC scheme may continue looking for better results after finding a large percentage of inliers. In contrast, R1PPnP takes into account all points when estimating the pose, which makes it insensitive to the selected control point o. Therefore it is a reasonable assumption that when p_{inliers} is large enough (we used the threshold of 60 percent), no improvement can be found and the RANSAC process of R1PPnP could be terminated. Accuracy evaluation results in this paper have testified the rationality of this assumption.

(2) *Termination Conditions for R1PPnP Iterations*. As shown in Fig. 5, we first detect as many inliers as possible and the related *termination condition A* is satisfied when the detected number of inliers becomes stable, that is,

$$N_{\text{inlier}}^{(k)} - N_{\text{inlier}}^{(k-20)} \leq 0 \text{ and } k > 20, \tag{26}$$

where k is the index of iterations, N_{inlier} is the number of detected inliers. According to our experience, in most cases no more inliers would be detected if N_{inlier} has not increased in 20 iterations.

A good *termination condition A* should be able to stop the iteration process as early as possible when point o is an outlier, and do not interrupt when point o is an inlier. We proposed the *termination condition A* with a window size = 20 iterations to seek balance between speed and robustness. This termination condition is not based on the comparison of parameters of adjacent iterations because in R1PPnP, the dynamically updated weights w_i may make the convergence process complex, especially when point o is an outlier. As shown in Fig. 6a, with an outlier as the control point, $\|\mathbf{R}^{(k)} - \mathbf{R}^{(k-1)}\|$ may take many iterations to converge to zero, which is slow. With the change of detected number of inliers in a larger window size, the termination decision can be more robust and efficient, as shown in Fig. 6b.

The refinement stage makes pose estimation results more accurate based on the detected inliers. Without the reweighting mechanism, the convergence process is much simple. Hence the *termination condition B* is satisfied when the estimated rotation becomes stable, that is,

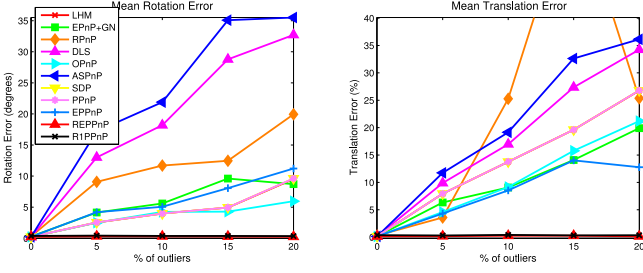$$\|\mathbf{R}^{(k)} - \mathbf{R}^{(k-1)}\| < 1e - 5. \tag{27}$$

Fig. 7. Except for REPPnP and R1PPnP, most PnP methods cannot handle outliers.

5   EXPERIMENTS

The performance of the proposed R1PPnP algorithm was evaluated by comparing against the state-of-the-art PnP methods. The source code was implemented in MATLAB scripts and executed on a computer with an Intel Core i7 2.60 GHz CPU. We used both synthetic and real-world data to conduct evaluation experiments. The initial values for R1PPnP are $\mathbf{R} = diag\{1,1,1\}$ and $\mu = 1e-4$. RANSAC +P3P or P4P methods also used the standard termination condition (25).

5.1   Synthetic Experiments

Synthetic experiments in this paper shared the following parameters. The camera focal length is 1,000 pixels with a resolution of 640×480. Two types of synthetic data were generated. (1) *Ordinary Three-Dimensional (3D) Case*: object points were randomly and uniformly distributed in a cube region $[-2,2] \times [-2,2] \times [4,8]$. (2) *Quasi-singular case*: The distribution cube is $[1,2] \times [1,2] \times [4,8]$. For each experiment result, we report the mean values of 100 trials.

For accuracy evaluation, the rotation error is measured in degrees between the truth rotation $\mathbf{R}_{\text{true}}$ and the estimated $\mathbf{R}$ as $e_{\text{rot}}(^\circ) = \|[\text{acos}(\mathbf{r}_{k,\text{true}}^T \cdot \mathbf{r}_k)_{k=1,2,3}]^T\| \times 180/\pi$, where $\mathbf{r}_{k,\text{true}}$ and $\mathbf{r}_k$ are the kth column of $\mathbf{R}_{\text{true}}$ and $\mathbf{R}$ respectively. The translation error is $e_{\text{trans}}(\%) = \|\mathbf{t}_{\text{true}} - \mathbf{t}\|/\|\mathbf{t}\| \times 100\%$.

5.1.1   Outlier-Free Synthetic Situations

Most PnP algorithms do not have the ability to handle outliers, and even a small percentage of outliers will

significantly reduce the accuracy, as shown in Fig. 7. Thus, although outlier-free situations are not the main concern of R1PPnP, we first conducted comparison experiments between the proposed R1PPnP and other PnP algorithms in outlier-free situations. The reason for this comparison is that RANSAC+P3P or P4P methods usually need other PnP methods as the final refinement step. Hence the accuracy and speed in outlier-free situations are also related to the performance in situations with outliers.

Here we only performed the core algorithm of R1PPnP without outliers handling mechanism. The termination condition for R1PPnP iterations was as Eq. (27). In this experiment, we compared our proposed R1PPnP with the following PnP methods: LHM [6], EPnP [4], RPnP [26], DLS [41], OPnP [5], ASPnP [42], SDP [25], PPnP [43], EPPnP [3] and REPPnP [3].

In our accuracy evaluation experiments, the number of points was 100 and we added different levels of Gaussian image noises from 0 to 10 pixels. As shown in Figs. 8 and 9, for both ordinary 3D and quasi-singular cases, R1PPnP gave the most accurate rotation estimation results together with OPnP and SDP. For ordinary 3D cases, R1PPnP was among the most accurate methods to estimate translation and was only sightly less accurate than OPnP. However, for quasi-singular cases, the accuracy of translation estimation of R1PPnP was not the state-of-the-art. ASPnP became unstable with large image noise hence its mean accuracy decreased significantly compared with that with small image noise. Although sometimes PPnP can provide accurate rotation estimation results in ordinary 3D cases, it also suffered from instability in some random cases, as shown by the jitter in Fig. 8. PPnP and LHM cannot handle the quasi-singular cases.

To evaluate runtime, Gaussian image noise with a standard deviation of $\sigma = 5$ pixels was added and the number of points increased from 100 to 1000. As shown in Fig. 10, the proposed R1PPnP, together with EPPnP, REPPnP and ASPnP showed superior computational speed. The runtime of R1PPnP did not grow significantly with respect to the number of points. We suspect that this results from the intrinsic parallel optimization of the matrix computation of MATLAB 2014a.
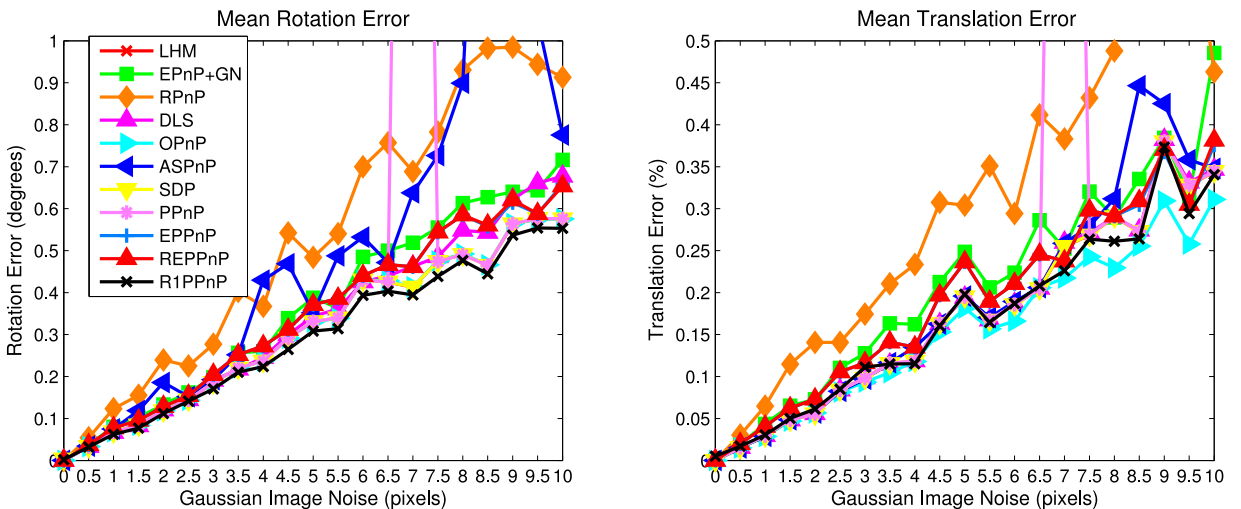

Fig. 8. Accuracy with outlier-free synthetic data (ordinary 3D cases). Number of points was 100. Different levels of image noises were added.
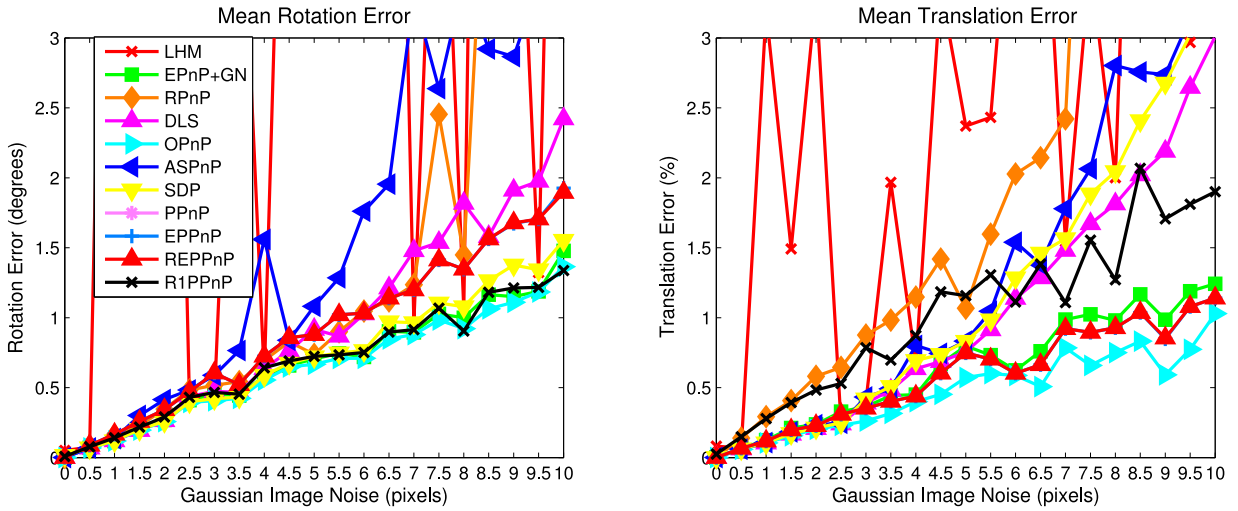
Fig. 9. Accuracy with outlier-free synthetic data (quasi-singular cases). Number of points was 100. Different levels of image noises were added. PPnP is out of range. The accuracy of all PnP methods decreased significantly compared with those in ordinary 3D cases, as shown in Fig. 8.
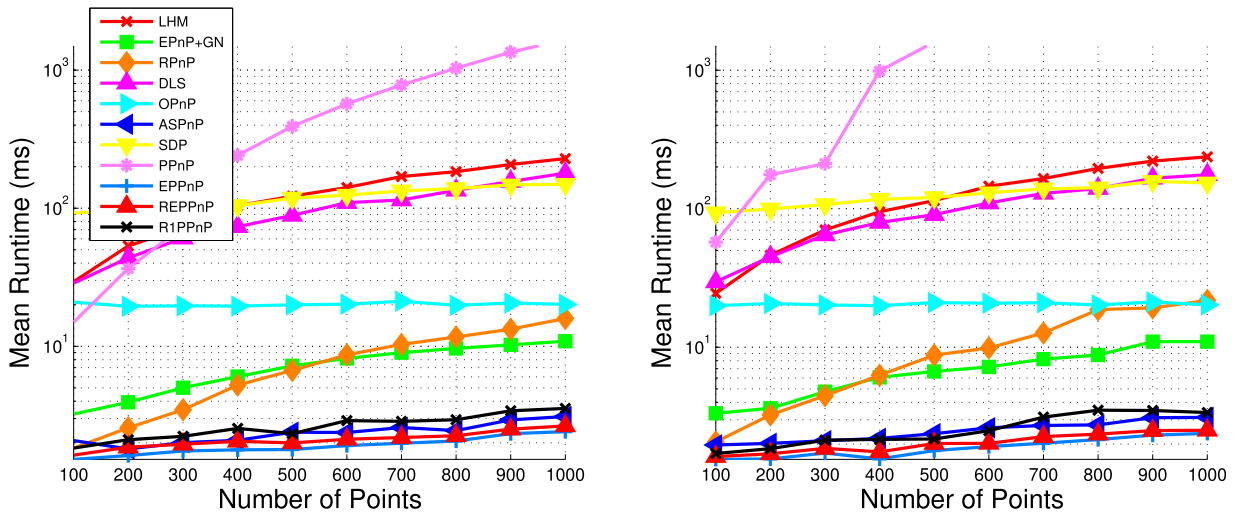

Fig. 10. Runtime results with outlier-free synthetic data. Standard deviation of image noise $\sigma = 5$ pixels. The number of points increased from 100 to 1000. (Left) Ordinary 3D cases. (Right) Quasi-singular cases.

Generally speaking, in outlier-free situations, R1PPnP was among the state-of-the-art methods in terms of both accuracy and computational speed. One drawback of R1PPnP is that the accuracy of translation estimation in quasi-singular cases was not among the best.

5.1.2 Synthetic Situations with Outliers

The main advantage of R1PPnP is that it is capable of handling a large percentage of outliers with a much faster speed than conventional methods. For demonstrating this, we introduced the following RANSAC-based PnP methods for comparison: (RANSAC+P3P [14]); (RANSAC + RP4P + RPnP [26]); (RANSAC + P3P [14] + ASPnP [42]); and (RANSAC + P3P [14] + OPnP [5]). According to evaluations in outlier-free situations, OPnP is the most accurate PnP method and ASPnP and RPnP are fast. We selected these methods as the final refinement step to fully demonstrate the performance of RANSAC+refinement-like methods. Another important method is REPPnP [3], which is the state-of-the-art PnP algorithm that addresses outliers.

The experiments were conducted as follows. $N_{\text{inlier}} = 100$ correct matches (inliers) between 3D object points and 2D image points were generated. N_{outlier} mismatches (outliers) were generated by randomly corresponding 3D and 2D points. The true percentage of outliers is $p_{\text{outlier}} = N_{\text{outlier}}/(N_{\text{inlier}} + N_{\text{outlier}})$. Gaussian image noise with a standard deviation of $\sigma = 5$ pixels was added. For R1PPnP and other RANSAC-based methods, the reprojection error threshold to distinguish inliers and outliers was $H = 10$ pixels.

As shown in Fig. 11, REPPnP began to fail when the percentage of outliers was larger than 50 percent with ordinary 3D cases, and only 5 percent with quasi-singular cases. R1PPnP and RANSAC-based methods were capable of handling situations with a large percentage of outliers. R1PPnP was more accurate than RANSAC-based methods for both rotation and translation estimation. As shown in Fig. 12. Compared to other RANSAC-based methods, R1PPnP was much faster, especially when the percentage of outliers was large.
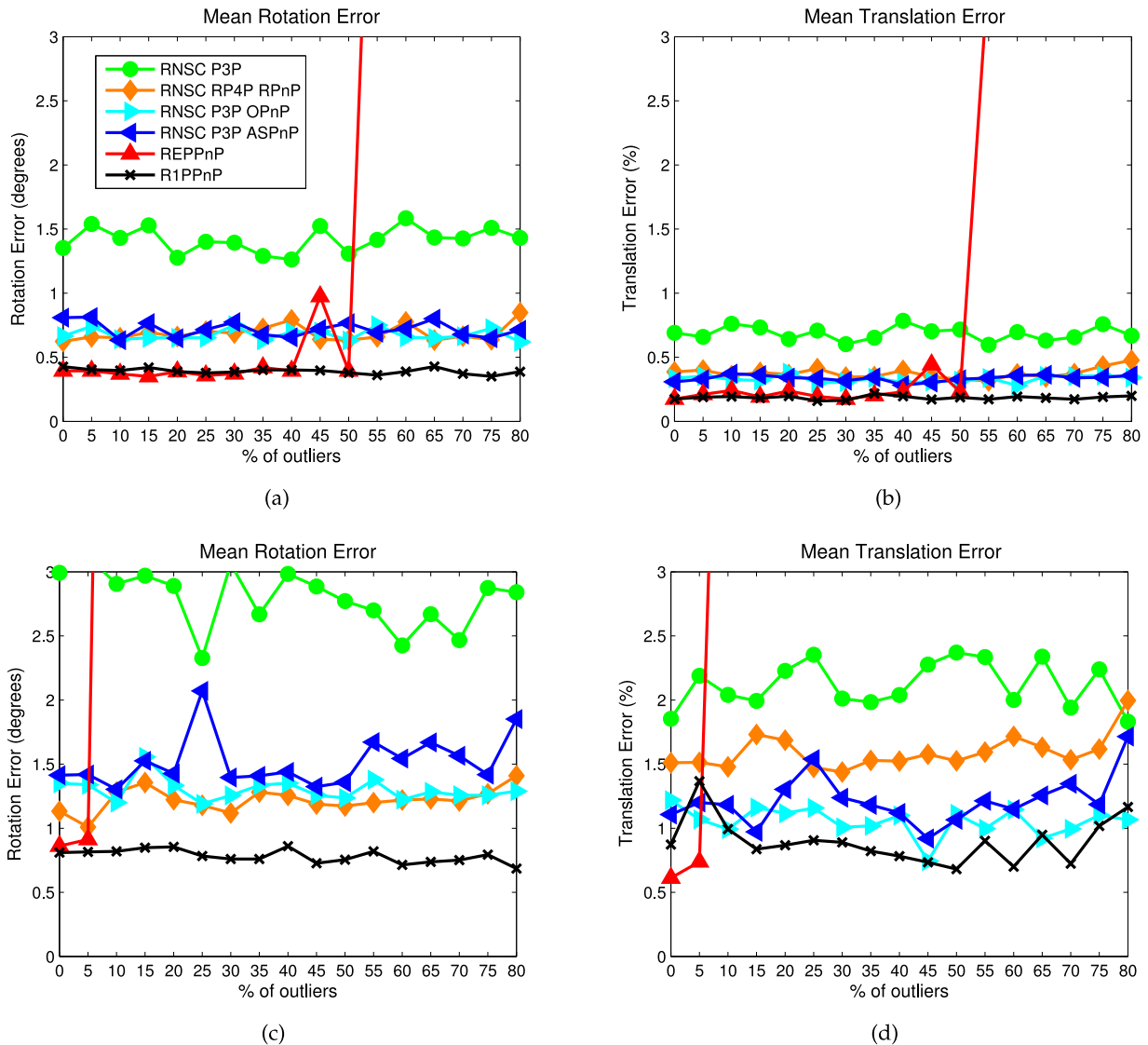
Fig. 11. Average accuracy on synthetic data with outliers. (a) and (b) Accuracy with ordinary 3D cases. (c) and (d) Accuracy with quasi-singular cases.
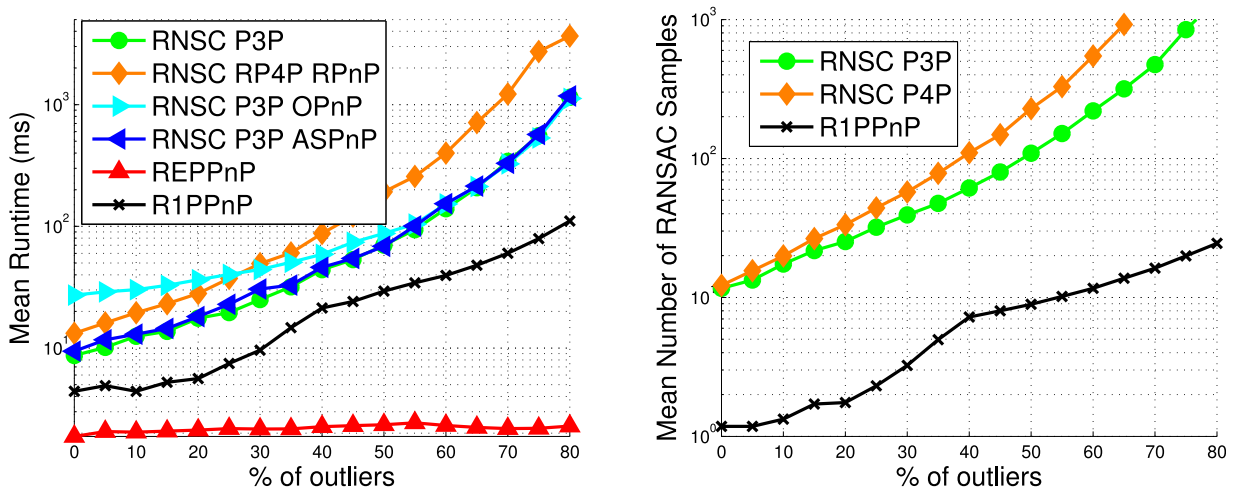

Fig. 12. Average runtime and number of required RANSAC samples with ordinary 3D cases synthetic data. We do not give the results with quasi-singular cases because they are very close to that with ordinary 3D cases. RANSAC+P3P or P4P needs more than 10 RANSAC trails when $p_{\text{outliers}} = 0$ because the large image noise ($\sigma = 5$ pixels) usually makes P3P or P4P methods unable to find the correct pose with three or four inliers to satisify the termination condition Eq. (25). In contrast, the required number of RANSAC trials of R1PPnP is not sensitive to image noise because all points are taken into account.

Fig. 13. Statistical results with real-world data. The x-axis is ranges of the percentage of outliers. (a) Rotation error. (b) Translation error. (c) Runtime. (d) The number of detected inliers compared with the maximum, zero suggests this method finds the most inliers.
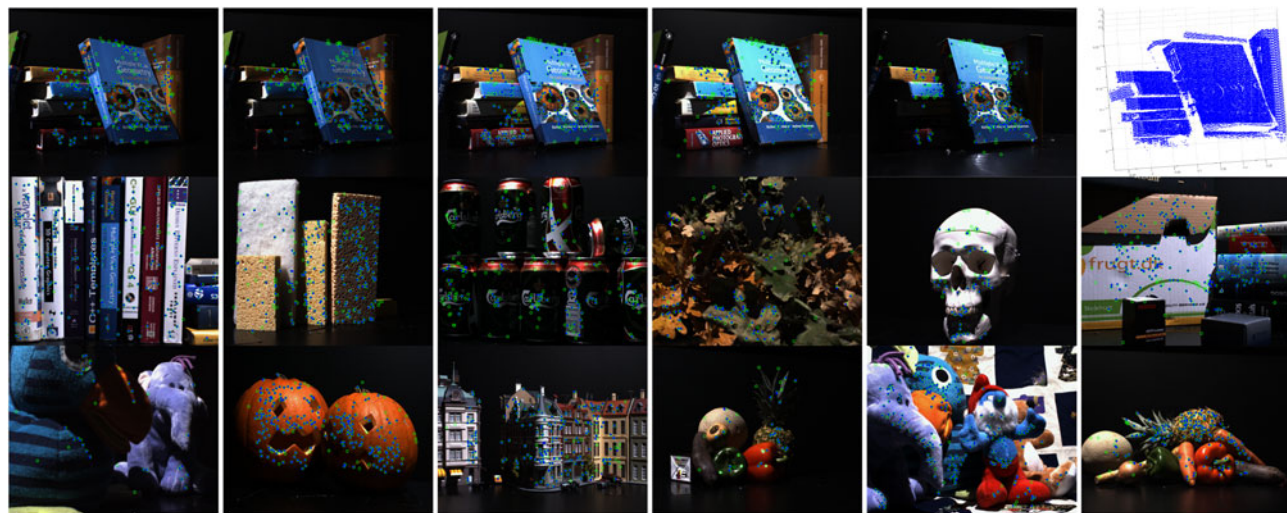

Fig. 14. Examples of images and R1PnP reprojection results. Green circles are all SURF correspondences and blue stars are the reprojected inliers detected by R1PPnP. First row: images with different illumination situations and the 3D point cloud. Second-third row: different data sets.

5.2 Real-World Image Data

Our real-world experiments were conducted on the DTU robot image data[1] [44], which provides images and the related 3D point cloud obtained by structured light scan. The true values of rotations and translations are known. Images have a resolution of 800×600. Datasets numbered 1 to 30 were used. In each dataset, images were captured under 19 different illumination situations and from 119 camera positions. We selected 10 out of 19 illumination situations. Hence, a total of $30 \times 10 \times 119 = 35700$ images were included in this evaluation. Following the instruction, for
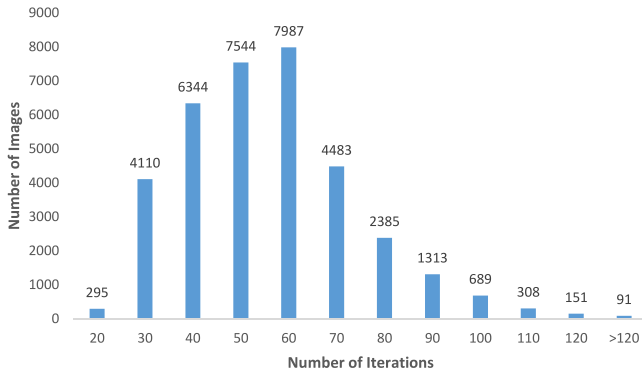
1. http://roboimagedata.imm.dtu.dk/

Fig. 15. Histogram of the number of iterations of R1PPnP in real-world experiments.

each dataset and illumination situation, we used the image numbered 25 as the reference image and performed SURF matching [7] between the reference image and other images. The inliers threshold was $H = 5$ pixels for all methods. With each image, we ran all algorithms 5 times and used the average value for the subsequent statistics.

As shown in Fig. 14, the total number of correspondences and the percentage of outliers varied with objects, illumination situations and camera poses. Although clear comparisons require that only one factor is different, this kind of variable-controlling is difficult for PnP evaluation on real-world data because SURF matching results are unpredictable. In experiments we found that the performance of PnP algorithms were mainly affected by the percentage of outliers, rather than the total number of correspondences. Therefore in this section, we report the evaluation results by comparing the statistical results of PnP methods at each percentage of outliers range. Because the true number of inliers was unknown, for each image, algorithms detected inliers and we considered the maximum number of inliers as the ground truth.

As shown in Fig. 13a, as the percentage of outliers increased, the runtime of R1PPnP did not grow significantly compared with conventional RANSAC+P3P or P4P methods. When $p_{\text{outliers}} < 30\%$, R1PP$n$P was slower than pure RANSAC+P3P, but was much more accurate as shown in Fig. 13a, 13b. To improve accuracy, RANSAC+P3P needs other PnP methods, such as OPnP or ASPnP, as the final refinement step. Compared with other refinement PnP methods, R1PPnP was slightly less accurate than OPnP, which was the most accurate PnP method according to both synthetic and real-world experiments, but much faster even when the percentage of outliers was small.

Fig. 15 shows the histogram of the number of R1PPnP iterations on all 35700 images. As shown in Fig. 5, the iteration number includes iterations with the re-weighting mechanism that obtained the best results in RANSAC trials, and the subsequent refinement iterations without re-weighting. The average number of required iterations is 51.3.

6 CONCLUSIONS

We present a fast and robust PnP solution named R1PPnP for tackling the outliers issue. We integrate a soft re-weighting method into an iterative PnP process to distinguish inliers and outliers, and employ the 1-point RANSAC scheme for selecting the control point. The number of trials is greatly reduced compared to conventional RANSAC+P3P or P4P methods; hence, it is much faster. Synthetic and real world experiments demonstrated its feasibility. Except for the good performance, another hidden advantage of R1PPnP is that its code implementation is relatively easy because all steps of R1PPnP involve only simple calculations. For example, its minima avoidance mechanism only requires to compute the determinant of the rotation matrix and to make $\lambda_{\text{new}} = 1/\lambda_{\text{old}}$. The most appropriate situations to replace conventional RANSAC+P3P methods with R1PPnP is when the percentage of outliers and/or the image white noise is large. R1PPnP is more appropriate for large point clouds because of its low time complexity and the requirement to try control points. Future works involve the development of its extension for planar cases, and applying it in the SLAM system to handle outliers when a new frame is encountered.
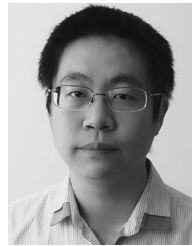
REFERENCES

[1] M.-A. Raul, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular slam system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[2] M. Müller, M.-C. Rassweiler, J. Klein, et al., "Mobile augmented reality for computer-assisted percutaneous nephrolithotomy," vol. 8, no. 4, pp. 663–675, 2013.

[3] L. Ferraz, X. Binefa, and F. Moreno-Noguer, "Very fast solution to the PnP problem with algebraic outlier rejection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 501–508.

[4] 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, 2009.

[5] Y. Zheng, Y. Kuang, et al., "Revisiting the PnP problem: A fast, general and optimal solution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2344–2351.

[6] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.

[7] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[8] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2548–2555.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2564–2571.

[10] D. Nistér, "A minimal solution to the generalised 3-point pose problem," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. 67–79.

[11] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, Aug. 2003.

[12] K. Josephson and M. Byrod, "Pose estimation with radial distortion and unknown focal length," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 2419–2426.

[13] M. A. Fischler and R. C. 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.

[14] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *Proc. IEEE CVPR*, 2011, pp. 2969–2976.

[15] M. Bujnak, Z. Kukelova, and T. Pajdla, "A general solution to the P4P problem for camera with unknown focal length," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[16] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[17] D. DeMenthon and L. S. Davis, "Exact and approximate solutions of the perspective-three-point problem," *IEEE IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 11, pp. 1100–1105, Nov. 1992.

[18] M. A. Abidi and T. Chandra, "A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 5, pp. 534–538, May 1995.

[19] R. Horaud, B. Conio, O. Leboulleux, and B. Lacolle, "An analytic solution for the perspective-4-point problem," *Comput. Vis. Graph. Image Process.*, vol. 47, no. 1, pp. 33–44, 1989.

[20] L. Quan and Z. Lan, "Linear n-point camera pose determination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 774–780, Aug. 1999.

[21] W. J. Wolfe, D. Mathis, C. W. Sklair, and M. Magee, "The perspective view of three points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 1, pp. 66–73, Jan. 1991.

[22] B. Triggs, "Camera pose and calibration from 4 or 5 known 3d points," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 278–284.

[23] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 578–589, May 2003.

[24] P. D. Fiore, "Efficient linear solution of exterior orientation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 140–148, Feb. 2001.

[25] G. Schweighofer and A. Pinz, "Globally optimal o(n) solution to the pnp problem for general camera models," in *Proc. Proc. British Mach. Vis. Conf.*, 2008, pp. 1–10.

[26] 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.

[27] L. Kneip, H. Li, and Y. Seo, "UPnP: An optimal o(n) solution to the absolute pose problem with universal applicability," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 127–142.

[28] D. G. Lowe, "Fitting parameterized three-dimensional models to images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 5, pp. 441–450, May 1991.

[29] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vis.*, vol. 15, no. 1, pp. 123–141, 1995.

[30] R. Horaud, F. Dornaika, and B. Lamiroy, "Object pose: The link between weak perspective, paraperspective, and full perspective," *Int. J. Comput. Vis.*, vol. 22, no. 2, pp. 173–189, 1997.

[31] P. David, D. Dementhon, R. Duraiswami, and H. Samet, "SoftPOSIT: Simultaneous pose and correspondence determination," *Int. J. Comput. Vis.*, vol. 59, no. 3, pp. 259–284, 2004.

[32] G. Schweighofer and A. Pinz, "Robust pose estimation from a planar target," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2024–2030, Dec. 2006.

[33] M. Havlena, A. Torii, and T. Pajdla, "Efficient structure from motion by graph optimization," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 100–113.

[34] R. Mur-Artal and J. D. Tardós, "Fast relocalisation and loop closing in keyframe-based slam," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 846–853.

[35] F. Kahl, S. Agarwal, M. K. Chandraker, D. Kriegman, and S. Belongie, "Practical global optimization for multiview geometry," *Int. J. Comput. Vis.*, vol. 79, no. 3, pp. 271–284, 2008.

[36] Q. Ke and T. Kanade, "Quasiconvex optimization for robust geometric reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1834–1847, Oct. 2007.

[37] L. Kneip, P. Furgale, and R. Siegwart, "Using multi-camera systems in robotics: Efficient solutions to the npnp problem," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2013, pp. 3770–3776.

[38] G. H. Lee, B. Li, M. Pollefeys, and F. Fraundorfer, "Minimal solutions for the multi-camera pose estimation problem," *Int. J. Robot. Res.*, vol. 34, no. 7, pp. 837–848, 2015.

[39] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.

[40] D. Simpson, "Introduction to rousseeuw (1984) least median of squares regression," *Breakthroughs in Statistics*, Berlin, Germany: Springer, 1997, pp. 433–461.

[41] J. A. Hesch and S. I. Roumeliotis, "A direct least-squares (DLS) method for PnP," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 383–390.

[42] Y. Zheng, S. Sugimoto, and M. Okutomi, "ASPnP: An accurate and scalable solution to the perspective-n-point problem," *IEICE Trans. Inform. Syst.*, vol. 96, no. 7, pp. 1525–1535, Jul. 2013.

[43] V. Garro, F. Crosilla, and A. Fusiello, "Solving the PnP problem with anisotropic orthogonal procrustes analysis," in *Proc. IEEE 2nd Int. Conf. 3D Imag. Model. Process. Vis. Transmiss.*, 2012, pp. 262–269.

[44] H. Aanæs, A. L. Dahl, and K. S. Pedersen, "Interesting interest points," *Int. J. Comput. Vis.*, vol. 97, no. 1, pp. 18–35, 2012.

Haoyin Zhou received the BS and PhD degrees from Tsinghua University, Beijing, China, in 2009 and 2014, respectively. He is a research fellow with the Surgical Planning Laboratory, Brigham and Women's Hospital, Harvard Medical School. His research interests includes PnP, structure-from-motion, 3D non-rigid dense reconstruction, and learning-based segmentation and their applications in the surgical navigation field.

Tao Zhang (M'00-SM'11) received the BS, MS, and PhD degrees from Tsinghua University, Beijing, China, in 1993, 1995, and 1999, respectively, and the second PhD degree from Saga University, Saga, Japan, in 2002. He is currently a professor and deputy head of the Department of Automation, School of Information Science and Technology, Tsinghua University, Beijing, China. He is the author or coauthor of more than 200 papers and three books. His current research includes robotics, image processing, control theory, artificial intelligent, navigation, and control of spacecraft. He is a senior member of the IEEE.

Jayender Jagadeesan received the BS degree from the Indian Institute of Technology, Kanpur, India, in 2003 and the PhD degree from the University of Western Ontario, London, ON, Canada, in 2007. He is now an assistant professor with Harvard Medical School and research associate with Brigham and Women's Hospital, Boston, Massachusetts. His research interests include surgical navigation, medical image registration and segmentation, medical robotics, and augmented reality. He is a member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.