Exterior Orientation Revisited: A Robust Method Based on l_q-norm

Jiayuan Li, Qingwu Hu, Ruofei Zhong, and Mingyao Ai

Abstract

Camera exterior orientation is essential in many photogrammetry and computer vision applications, including 3D reconstruction, digital orthophoto map (DOM) generation, and localization. In this paper, we propose a new formulation of exterior orientation that is robust against gross errors (outliers). Different from classic optimization methods whose cost function is based on the l_2-norm of residuals, we use l_q-norm ($0<q<1$) instead. We reformulate the new cost function as an augmented Lagrangian function because it is not strictly convex. In addition, we employ the alternating direction method of multipliers (ADMM) to decompose the augmented Lagrangian function into three simple sub-problems and solve them iteratively. Our work recovers the orientation and position of a camera from outliers contaminating observations without any gross error detection stage such as random sample consensus (RANSAC). Extensive experiments on both synthetic and real data demonstrate that the proposed method significantly outperforms state-of-the-art methods and can easily handle situations with up to 85 percent outliers. The source code of the proposed algorithm is made public.

Introduction

Camera exterior orientation, which is also known as perspective-n-point (PnP) problem in computer vision, refers to recovering the orientation and position of a perspective camera from n 3D-to-2D point correspondences. The PnP problem is a basic aspect of many computer vision and photogrammetric applications (e.g., camera localization, digital elevation model (DEM) production, incremental structure-from-motion, etc.), which has been studied for several decades. However, only a few studies have focused on the PnP problem with outliers in point correspondences.

The majority of PnP solutions do not directly address problems containing outliers. These solutions simply assume that all point correspondences are inliers. For example, Zheng *et al.* (2013) validated their method on simulated data that were contaminated by Gaussian noise with a zero-mean and a fixed deviation $\sigma = 2$ pixels. Traditional methods usually combine P3P algorithms with random sample consensus (RANSAC) (Fischler and Bolles, 1981) schemes to reject outliers during a preprocessing stage before the PnP methods are applied. However, RANSAC significantly reduces the efficiency, particularly for problems with numerous outliers.

Recently, Ferraz *et al.* (2014) proposed a fast, robust, and accurate PnP algorithm (REPPnP) with an algebraic outlier rejection stage. Their novel contribution was introducing a novel outlier rejection mechanism within the pose estimation framework. The PnP problem was formulated as a low-rank homogeneous system, and the solution was obtained by solving for its 1D null space. They assumed that rows of the homogeneous system which perturbed its null space were outliers. These outliers were then progressively rejected on the basis of an algebraic criterion. Their proposed REPPnP dealt well with situations with up to 50 percent outliers. However, REPPnP was ineffective when the outlier rate exceeded 50 percent. Moreover, their mechanism cannot handle planar cases and ordinary cases under a uniform framework.

In this paper, we introduce a new PnP solution that is more robust than state-of-the-art algorithms.

Compared with REPPnP (Ferraz *et al.*, 2014), our method does not require any outlier detection mechanism and handles all point configurations (planar, ordinary, and quasi-singular) in the same manner. Our work recovers the pose of a camera from correspondences corrupted by numerous outliers without any outlier detection stage such as RANSAC. The central idea of our method is to reformulate the PnP problem as an optimization function using l_q-norm ($0<q<1$) instead of l_2-norm. This idea is motivated by many other works where the l_q-norm was successfully applied, such as signal reconstruction (Marjanovic and Solo, 2012; Marjanovic and Hero, 2014; Marjanovic and Solo, 2014), compressive sensing (Candè and Wakin, 2008), point cloud processing (Bouaziz *et al.*, 2013; Li *et al.*, 2016). The l_q-norm is a sparsity-inducing norm that can minimize the number of non-zero residuals. Thus, it inherently rejects outliers. The l_q-norm cost function is then reformulated as an augmented Lagrangian function and solved by the alternating direction method of multipliers (ADMM) (Boyd *et al.*, 2011). An extensive experimental evaluation on both synthetic and real data shows that the proposed approach significantly outperforms the state-of-the-art methods and can easily handle situations with up to 85 percent outliers.

Related Work

The minimal configuration of the PnP problem is the P3P (DeMenthon and Davis 1992, Gao *et al.*, 2003; Kneip *et al.*, 2011) or P4P (Horaud *et al.*, 1989) problem. In these cases, there exist closed-form solutions finding roots of the formed fourth- or fifth-degree polynomial systems. Unfortunately, these solutions are sensitive to outliers and noise. The PnP problem, thus, focuses on over-constrained cases with larger-scale correspondence sets. Ideal PnP solutions should have the following properties: fast and unique convergence, global optimality, high accuracy, and robustness. However, finding a balance between all these features is difficult. The literature on the PnP problem can be roughly classified into two categories: non-iterative and iterative methods.

The most straightforward non-iterative PnP algorithms are the direct linear transformation (DLT) (Abdel-Aziz and Karara, 2015; Hartley and Zisserman, 2003) and its variations

Jiayuan Li, Qingwu Hu, and Mingyao Ai are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (huqw@whu.edu.cn).

Ruofei Zhong is with the Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing 100048, China.

Photogrammetric Engineering & Remote Sensing
Vol. 83, No. 1, January 2017, pp. 23–xxx.
0099-1112/17/23–xxx

doi: 10.14358/PERS.83.1.xxx

(Shan, 1996; Tsai, 1987). These methods are not very accurate because calibrated camera parameters are not considered. Several notable works have been proposed in recent years. Schweighofer and Pinz (2008) formulated the PnP problem by using an object space cost and relaxed it into a semi-definite positive program (SDP). The limitation of SDP was that the relaxation causes accuracy loss. Lepetit *et al.* (2009) proposed an efficient PnP solution (EPnP), which was the first close-form solution reducing the complexity to O(n). Their main contribution was the use of four virtual points to express the 3D object points. However, this approach applied a linearization technique, which might lose accuracy for slightly redundant cases with $n = 4$ or $n = 4$, as explained in Li *et al.* (2012).

To design more accurate non-iterative solutions, polynomial solvers without linearization were applied in the subsequent works. The direct least square (DLS) (Hesch and Roumeliotis, 2011) method, developed by Hesch and Roumeliotis, solved a fourth order polynomial system that was produced by a nonlinear least-squares cost function with the multiplication matrix technique. Unfortunately, as explained in Zheng *et al.* (2013), DLS adopted the Cayley-Gibbs-Rodriguez (CGR) to parameterize rotation, which would degenerate at 180 degrees. Li *et al.* (2012) presented a multi-stage framework called RPnP, which obtained a set of fourth order polynomials by dividing object points into several 3-point subsets and then solved a special cost function. Zheng *et al.* (2013) proposed a scalable solution (ASPnP) by minimizing algebraic error. They used unit quaternion to represent the rotation matrix and obtained a global optimum based on the Gröbner basis solver. In another study of theirs, OPnP (Zheng *et al.*, 2013), the rotation was parameterized by a non-unit quaternion. The PnP problem was then reformulated as an unconstrained optimization problem. These more recently proposed methods can usually achieve the complexity of O(n) and be initial-guess free, which are more suitable for real time applications.

On the other hand, iterative approaches formulate the PnP problem from a nonlinear least-squares perspective by iteratively minimizing geometric errors or algebraic errors. Typically, the Gauss-Newton method can be applied to the nonlinear optimization problems (Lowe, 1991). Iterative methods usually have better accuracy while being more time-consuming compared with non-iterative ones. Furthermore, whether they can correctly converge relies on having a good initial guess. POSIT (Dementhon and Davis, 1995) iteratively solved a linear system with scaled orthographic projections to obtain a better approximation of the true pose. The drawback of this work is its failing in quasi-singular cases or planar cases, which often appear in photogrammetric applications. A natural alternative to geometric error is algebraic error. Lu *et al.* (2000) developed a method to minimize the collinearity error of the object space. In the Procrustes PnP (PPnP) (Garro *et al.*, 2012), the authors defined the PnP problem as an anisotropic orthogonal Procrustes problem. More recently, Ferraz *et al.* (2014) proposed a fast, robust, and accurate PnP solution with algebraic outlier rejection. Their main contribution was introducing a novel outlier rejection mechanism into pose estimation framework.

Although most of the above-mentioned methods can achieve impressing results, the essential problem of dealing with outliers has not been well-addressed. As mentioned above, traditional methods usually combine P3P solutions with RANSAC-like schemes as a preprocessing stage. The REPPnP becomes unreliable in situations with more than 50 percent outliers.

Motivated by studies in the literature Marjanovic and Solo (2012 and 2014), we reformulate the PnP problem as a l_q optimization problem and derive its solution by using the augmented Lagrangian function and ADMM. The later experiments will show that the proposed method can yield accurate results in situations with up to 85 percent outliers.

Classical Exterior Orientation

Given n non-collinear 3D points $\boldsymbol{q}_i = [x_i, y_i, z_i]^T$, $i = 1, 2, \ldots, n$, $n \geq 3$ in an object reference system and their corresponding normalized 2D image projections $\boldsymbol{p}_i = [u_i, v_i]^T$, we can recover the pose of the image. Assume that the camera internal calibration matrix $\boldsymbol{K}$ is known, the perspective camera model which maps 3D reference points to 2D projections can be formulated as follows:

$$d_i \begin{bmatrix} \boldsymbol{p}_i \\ 1 \end{bmatrix} = \boldsymbol{K}[\boldsymbol{R}, \boldsymbol{t}] \begin{bmatrix} \boldsymbol{q}_i \\ 1 \end{bmatrix} \tag{1}$$

where,

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}_1^T \\ \boldsymbol{r}_2^T \\ \boldsymbol{r}_3^T \end{bmatrix} \in SO(3) \;\; and \;\; \boldsymbol{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{2}$$

are an orthogonal rotation matrix and a translation vector, respectively. And d_i is a depth factor. The goal of the PnP problem is to recover the pose ($\boldsymbol{R}$, $\boldsymbol{t}$) of a camera from Equation 1.

After elimination of depth factor d_i, Equation 1 is transformed to:

$$\begin{cases} u_i = \dfrac{\boldsymbol{r}_1^T \boldsymbol{q}_i + t_x}{\boldsymbol{r}_3^T \boldsymbol{q}_i + t_z} \\ v_i = \dfrac{\boldsymbol{r}_2^T \boldsymbol{q}_i + t_y}{\boldsymbol{r}_3^T \boldsymbol{q}_i + t_z} \end{cases}. \tag{3}$$

Equation 3 is called the *collinearity equation* in photogrammetry and is the mathematic basis of many other problems such as bundle adjustment and structure from motion.

Classical iterative methods usually minimize the following nonlinear least-squares cost function:

$$\arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{i=1}^{n} \left\| \boldsymbol{p}_i - \hat{\boldsymbol{p}}_i \right\|_2^2 \tag{4}$$

in which $\hat{\boldsymbol{p}}_i$ is the calculated value of the observations $\boldsymbol{p}_i$ based on Equation 3 and the estimated pose ($\boldsymbol{R}$, $\boldsymbol{t}$), and $\|\cdot\|_2$ is a l_2-norm operator.

In the next section, we will detail our l_q-norm based cost function and describe how to solve it using the augmented Lagrangian function and ADMM.

Robust PnP solution Based on l_q Optimization

When the observations are corrupted by outliers, we hope that the PnP solver is able to automatically classify the residual vector $\boldsymbol{v} = [v_1, v_2, \ldots, v_n]$ into an inlier set $I(v_i \,|\, |v_i| \approx 0)$ and an outlier set $O(v_i \,|\, |v_i| \gg 0)$. However, the classic least-squares cost is based on a fundamental assumption that observations are subjected to a normal distribution and free of outliers. The estimated pose will be biased in order to reduce the large residuals of outliers (see Figure 1), which is a typical technique called residual adjustment. In contrast, l_0-norm (Mohimani *et al.*, 2007) is designed for this problem because it measures the number of non-zero elements in a residual vector. The l_0-norm is highly non-convex; thus, the l_1-norm (Mazumder *et al.*, 2010) is usually applied as the closest convex relaxation of the l_0-norm. Recently, another sparsity-inducing norm, i.e.,the l_q-norm shows great potential and can achieve better

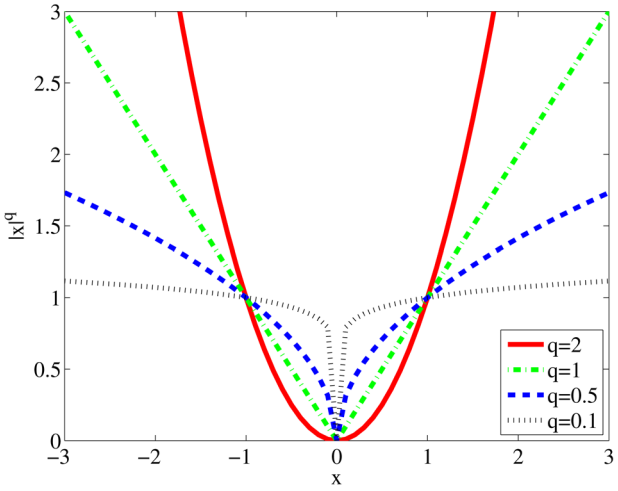


Figure 1. The penalty function curves: For large values of q, outliers or large noise are still very large after penalty function mapping; thus, the estimated pose will be biased in order to reduce the large residuals of outliers or noise.

performance than the l_1-norm, as shown in (Marjanovic and Solo, 2012 and 2014). The results of the l_q-norm are sparser than the l_1-norm and are less biased, as evidenced in (Chen *et al.*, 2010). Motivated by that, we reformulate the cost function based on the l_q-norm metric:

$$\arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{i=1}^{n} \left\| \boldsymbol{p}_i - \boldsymbol{p}_i \right\|_q^q \tag{5}$$

where $\|\cdot\|_q$ is a l_q-norm ($0<q<1$) operator.

The cost function is non-convex and non-smooth. To optimize this problem, we form it as a l_q-norm penalized least squares (l_qLS) problem which has been well studied by Marjanovic and Solo (2014). By introducing auxiliary variables $\boldsymbol{M}=[\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_n]$ into Equation 5, we rewrite the cost function as:

$$\arg\min_{\boldsymbol{R},\boldsymbol{t},\boldsymbol{M}} \sum_{i=1}^{n} \left\| \boldsymbol{m}_i \right\|_q^q \quad \text{subject to } \boldsymbol{\varepsilon}_i = \boldsymbol{p}_i - \boldsymbol{p}_i - \boldsymbol{m}_i = \boldsymbol{0}\,. \tag{6}$$

This is a constrained optimization function which is usually reformulated as an unconstrained one by using Lagrangian function. However, the strict convexity property is not satisfied in our cost function and dual ascent method may fail. Fortunately, augmented Lagrangian methods can yield convergence without assumptions of strict convexity. Thus, the augmented Lagrangian of Equation 6 is:

$$\begin{aligned} L_\rho(\boldsymbol{R},\boldsymbol{t},\boldsymbol{M},\boldsymbol{\Lambda}) &= \sum_{i=1}^{n} \left(\| \boldsymbol{m}_i \|_q^q + \boldsymbol{\lambda}_i^T \boldsymbol{\varepsilon}_i + \frac{\rho}{2} \| \boldsymbol{\varepsilon}_i \|_2^2 \right) \\ &= \sum_{i=1}^{n} \left(\| \boldsymbol{m}_i \|_q^q + \frac{\rho}{2} \| \frac{\boldsymbol{\lambda}_i}{\boldsymbol{\rho}} + \boldsymbol{\varepsilon}_i \|_2^2 - \frac{1}{2\boldsymbol{\rho}} \| \boldsymbol{\lambda}_i \|_2^2 \right) \end{aligned} \tag{7}$$

in which $\boldsymbol{\Lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]$ are dual variables or Lagrange multipliers and $\rho > 0$ is a penalty parameter.

We employ ADMM to decompose the function into three sub-problems because two sets of variables are present (pose ($\boldsymbol{R}$, $\boldsymbol{t}$) and $\boldsymbol{M}$) in Equation 7. ADMM is a method that combines the superior convergence of the basic multipliers and the decomposability property of the dual ascent method, which is why we choose ADMM instead of the basic method of multipliers. The ADMM consists of following iterations:

$$\textbf{prob 1}: \quad \boldsymbol{M}^{k+1} := \arg\min_{\boldsymbol{M}} L_\rho(\boldsymbol{R}^k, \boldsymbol{t}^k, \boldsymbol{M}, \boldsymbol{\Lambda}^k) \tag{8}$$

$$\textbf{prob 2}: \quad (\boldsymbol{R}^{k+1}, \boldsymbol{t}^{k+1}) := \arg\min_{\boldsymbol{R},\boldsymbol{t}} L_\rho(\boldsymbol{R}, \boldsymbol{t}, \boldsymbol{M}^k, \boldsymbol{\Lambda}^k) \tag{9}$$

$$\textbf{prob 3}: \quad \boldsymbol{\lambda}_i^{k+1} := \boldsymbol{\lambda}_i^k + \rho \boldsymbol{\varepsilon}_i \quad i = 1, 2, \ldots, n \tag{10}$$

where the superscript k denotes the iteration counter. Similar to the dual ascent method, sub-problem 1 and sub-problem 2 are the variable ($\boldsymbol{R}$,$\boldsymbol{t}$,$\boldsymbol{M}$) minimization step, sub-problem 3 is a dual variable update step. In sub-problem 1, only variable $\boldsymbol{M}$ is to be estimated, while others are fixed. In the same way, the pose ($\boldsymbol{R}$,$\boldsymbol{t}$) is the only variable in sub-problem 2. The ADMM alternates between these three steps until convergence. Details about augmented Lagrangian methods and ADMM can be found in literature (Boyd *et al.*, 2011).

Combining Equations 8, 9, and 7, sub-problem 1 and sub-problem 2 can be detailed as follows:

$$\begin{aligned} &\arg\min_{\boldsymbol{M}} L_\rho(\boldsymbol{R}^k, \boldsymbol{t}^k, \boldsymbol{M}, \boldsymbol{\Lambda}^k) \\ &= \arg\min_{\boldsymbol{M}} \sum_{i=1}^{n} \left(\| \boldsymbol{m}_i \|_q^q + \frac{\rho}{2} \| \frac{\boldsymbol{\lambda}_i}{\boldsymbol{\rho}} + \boldsymbol{\varepsilon}_i \|_2^2 - \frac{1}{2\boldsymbol{\rho}} \| \boldsymbol{\lambda}_i \|_2^2 \right) \\ &= \arg\min_{\boldsymbol{M}} \sum_{i=1}^{n} \left(\| \boldsymbol{m}_i \|_q^q + \frac{\rho}{2} \| \frac{\boldsymbol{\lambda}_i}{\rho} + \boldsymbol{p}_i - \boldsymbol{p}_i - \boldsymbol{m}_i \|_2^2 \right) \\ &= \arg\min_{\boldsymbol{M}} \sum_{i=1}^{n} \left(\| \boldsymbol{m}_i \|_q^q + \frac{\rho}{2} \| \boldsymbol{\delta}_i - \boldsymbol{m}_i \|_2^2 \right) \end{aligned} \tag{11}$$

$$\begin{aligned} &\arg\min_{\boldsymbol{R},\boldsymbol{t}} L_\rho(\boldsymbol{R}, \boldsymbol{t}, \boldsymbol{M}^k, \boldsymbol{\Lambda}^k) \\ &= \arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{i=1}^{n} \left(\| \boldsymbol{m}_i \|_q^q + \frac{\rho}{2} \| \frac{\boldsymbol{\lambda}_i}{\boldsymbol{\rho}} + \boldsymbol{\varepsilon}_i \|_2^2 - \frac{1}{2\rho} \| \boldsymbol{\lambda}_i \|_2^2 \right) \\ &= \arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{i=1}^{n} \left(\frac{\rho}{2} \| \frac{\boldsymbol{\lambda}_i}{\rho} + \boldsymbol{p}_i - \boldsymbol{p}_i - \boldsymbol{m}_i \|_2^2 \right) \\ &= \arg\min_{\boldsymbol{R},\boldsymbol{t}} \sum_{i=1}^{n} \left(\frac{\rho}{2} \| \boldsymbol{\gamma}_i - \hat{\boldsymbol{p}}_i \|_2^2 \right) \end{aligned} \tag{12}$$

in which $\boldsymbol{\delta}_i = \dfrac{\boldsymbol{\lambda}_i}{\rho} + \boldsymbol{p}_i - \hat{\boldsymbol{p}}_i$ and $\boldsymbol{\delta}_i = \dfrac{\boldsymbol{\lambda}_i}{\rho} + \boldsymbol{p}_i - \boldsymbol{m}_i$ are only used for compactness of notation.

Equation 11 is a non-trivial l_q-norm penalized least squares (l_qLS) problem and each variable $\boldsymbol{m}_i$ can be solved independently. We first consider the scalar version of Equation 11:

$$\arg\min_{m} \left(\| m \|_q^q + \frac{\rho}{2} \| \delta - m \|_2^2 \right) = \arg\min_{m} \left(|m|^q + \frac{\rho}{2} (\delta - m)^2 \right) \tag{13}$$

as proven in (Marjanovic and Solo, 2012 and 2014), the optimal solution $\hat{\boldsymbol{m}}$ is given by:

$$\hat{m} = \begin{cases} 0 & if\ |\delta| < \tau_a \\ \{0, \mathrm{sgn}(\delta)\beta_a\} & if\ |\delta| = \tau_a \\ \mathrm{sgn}(\delta)\beta_* & if\ |\delta| > \tau_a \end{cases} \tag{14}$$

where,

$$\beta_a = \left[\frac{2}{\rho} (1-q) \right]^{\frac{1}{2-q}}, \quad \tau_a = \beta_a + \frac{q}{\rho} \beta_a^{q-1} \tag{15}$$

sgn(·) is the signum function, and $\beta_* \in (\beta_a \mid \delta \mid)$ is the larger one in the two solutions of the following iteration:

$$\beta^{k+1} = f(\beta^k) \quad \text{where} \quad f(\beta) = |\delta| - \frac{q}{\rho}\beta^{q-1} \tag{16}$$

with initial guess $\beta^0=(\beta_a+ \mid \delta \mid)/2$. Marjanovic and Solo (Marjanovic and Solo, 2012 and 2014) pointed out that the solution would yield convergence within two iterations.

In sub-problem 1, the variable $\boldsymbol{m}_i$ is a 2D vector, and Equation 14 cannot be adopted directly. Fortunately, it can be turned into a scalar problem by simply performing a minimization operator on Equation 11 along the j^{th} ($j=1,2.$) element after completing the square. This is the coordinate-wise optimal solution for l_qLS. More details about the l_qLS problem may be seen in Marjanovic and Solo's l_q Cyclic Descent (l_qCD) algorithm (Marjanovic and Solo, 2014).

The second sub-problem is a classical least-squares function, which can be solved by many mature methods. In this paper, we adopt Gauss-Newton method and parameterize the rotation with a unit quaternion. As shown in Zheng *et al.*(2013a) and Zheng *et al.* (2013b), the unit quaternion is more powerful and less sensitive to initial guess. Using the unit quaternion to represent rotation will introduce an equality-constrained condition. Fortunately, this optimization problem can be easily transformed by a Lagrangian function.

Because our method is iterative, the final estimated pose $(\hat{\boldsymbol{R}}, \hat{\boldsymbol{T}})$ can be calculated by:

$$\begin{cases} \hat{\boldsymbol{R}} = \boldsymbol{R}^t \boldsymbol{R}^{t-1} \cdots \boldsymbol{R}^0 \\ \hat{\boldsymbol{t}} = (\boldsymbol{R}^t \boldsymbol{R}^{t-1} \cdots \boldsymbol{R}^1)\boldsymbol{t}^0 + (\boldsymbol{R}^t \boldsymbol{R}^{t-1} \cdots \boldsymbol{R}^2)\boldsymbol{t}^1 + \cdots + \boldsymbol{R}^t \boldsymbol{t}^{t-1} + \boldsymbol{t}^t \end{cases} \tag{17}$$

where the superscript t denotes the iteration count until convergence.

Experiments and Evaluation

In this section, we evaluate the proposed algorithm on both simulation data and real images. Our method (LqP*n*P in short) will be compared against state-of-the-art PnP solutions, including Lu's iterative method (LHM) (Lu *et al.*, 2000), EP*n*P+GN (Lepetit *et al.*, 2009), RPnP (Li *et al.*, 2012), the new version of DLS (Hesch and Roumeliotis, 2011), OP*n*P (Zheng *et al.*, 2013b), ASPnP (Zheng *et al.*, 2013), SDP (Schweighofer and Pinz, 2008), PPnP (Garro *et al.*, 2012), EPP*n*P without outlier rejection stage, and REPPnP with outlier detection (Ferraz *et al.*, 2014). The parameter q is set to 0.5. The implementations of all these ten methods are obtained from Ferraz's PnP toolbox in Matlab. Our approach is also implemented based on Matlab, and the source code is made public (https://sites.google.com/site/jiayuanli2016whu/home).

Simulations

A well-calibrated perspective camera capturing 2k*2k-pixel images with no principal point offset and a 1,500-pixel focal length is set in the simulation. We randomly generate n 3D points. For the ordinary configuration, the reference points are distributed in the box of [-8,8]×[-8,8]×[8,16] in the camera framework; while for the quasi-singular configuration, they are located in the box of [-8,8]×[-8,8]×[8,9] (near-planar case). Similar to OP*n*P (Zheng *et al.*, 2013b), the ground-truth translation $\boldsymbol{t}_{true}$ is set as the mean value of these 3D points, and the ground-truth rotation $\boldsymbol{R}_{true}$ is randomly generated. We obtain the three initial angles by adding a random value between [-10°, 10°] to the three angles of $\boldsymbol{R}_{true}$ and randomly generate the initial translation vector between [80%*$\boldsymbol{t}_{true}$, 120%*$\boldsymbol{t}_{true}$]. Two metric errors are used, as in (Ferraz *et al.* (2014) and Zheng *et al.* (2013b). The absolute rotation error in degrees between the estimated rotation $\boldsymbol{R}$ and the ground truth $\boldsymbol{R}_{true}$ is defined by Equation 18, and the relative translation error in percentage is defined by Equation 19.

$$e_{rot}(\text{deg}) = \max_{k=1}^{3}\{arc\cos(\boldsymbol{r}_{k,true} \cdot \boldsymbol{r}_k) \times 180 / \pi\} \tag{18}$$

$$e_{trans}(\%) = \|\boldsymbol{t}_{true} - \boldsymbol{t}\| / \|\boldsymbol{t}\| \times 100 \tag{19}$$

where $\boldsymbol{r}_{k,true}$ and $\boldsymbol{r}_k$ represent the k^{th} column of $\boldsymbol{R}_{true}$ and $\boldsymbol{R}$, respectively.

The results reported in the following plots are all based on the average rotation and translation errors of 500 independent experiments.

Number of Points and Noise Levels

In this experiment, the n generated 3D-to-2D correspondences are assumed to be only corrupted by noise. We compare the accuracy of the above-mentioned methods when varying the number of point correspondences and noise levels. We first fix the Gaussian noise with deviation σ = 2 pixels and increase the correspondence numbers n from 4 to 22. Then, we vary the level of Gaussian noise from σ = 2 to 20 pixels and fix the number of correspondences n = 10. The average rotation and translation errors are reported in Figure 2.

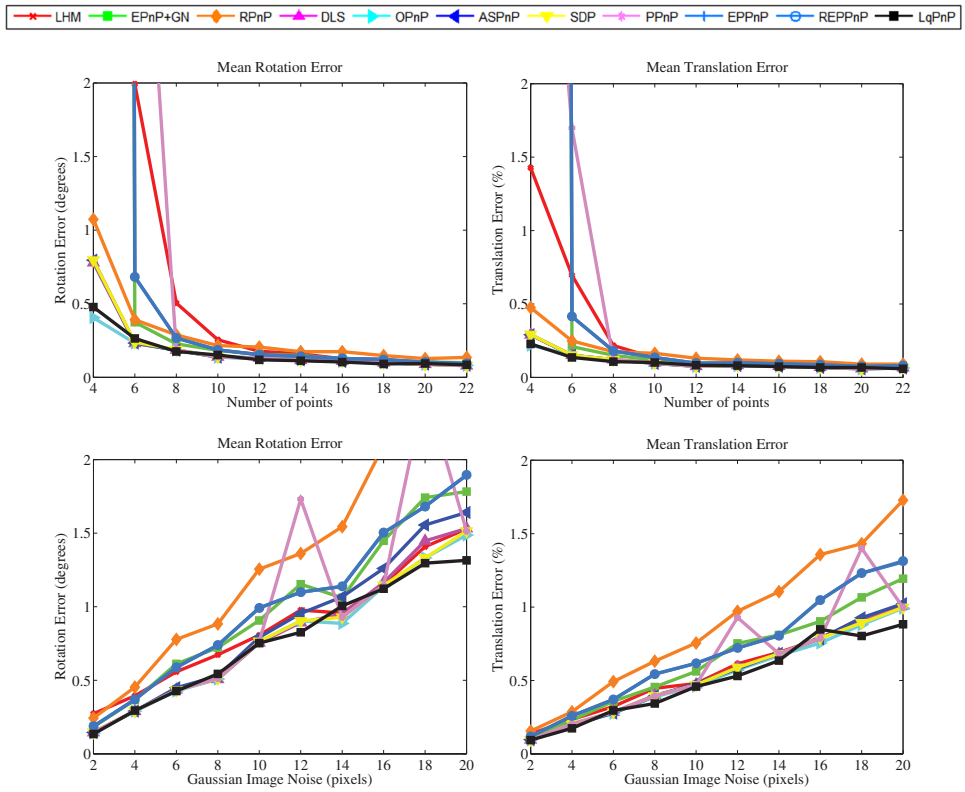
We can learn that: (1) Half of these solutions become very inaccurate when the correspondence set is small ($n \leq 8$), such as in LHM, EP*n*P+GN, PPnP, EPP*n*P, and REPPnP. The reason may be that they use approximation techniques in their works, and LHM may converge to a local optimal. (2) EP*n*P+GN, RPnP, PPnP, EPP*n*P, and REPPnP are more sensitive to Gaussian noise than others. Again, most of them apply an approximation technique. (3) LHM, EP*n*P+GN (EP*n*P in the planar configuration), RPnP, SDP, and PPnP are sensitive to the point configurations. We can observe that PPnP is less stable in quasi-singular cases than in the ordinary configurations. SDP is excluded from comparison in planar cases because it cannot handle this configuration. The performance of LHM, EP*n*P, RPnP, and PPnP drop considerably when the 3D reference points are coplanar. Although the robust version of DLS achieves good results in these plots, its failure rate is very high (almost 25 percent) in planar cases. (4) OP*n*P and ASPnP achieve significantly better accuracy than the aforementioned solutions in all synthetic experiments. This may benefit from using quaternion to parameterize the rotation matrix.

Nevertheless, our method LqP*n*P provides more impressive results than OP*n*P and ASPnP in most situations. For instance, for point number n=4 in the planar case of Figure 2, the rotation errors of OP*n*P and ASPnP reach up to 4.53 and 4.43 degrees, and their translation errors reach up to 2.06 percent and 2.05 percent, respectively. In contrast, the rotation and translation errors of the proposed algorithm LqP*n*P are 0.38 degrees and 0.41 percent, respectively. The reasons may be concluded as follows: our method LqP*n*P uses the sparse-inducing norm, l_q-norm, to formulate the cost function, which is more robust to noise and outliers. Furthermore, LqP*n*P is an iterative-based algorithm which usually has better accuracy.

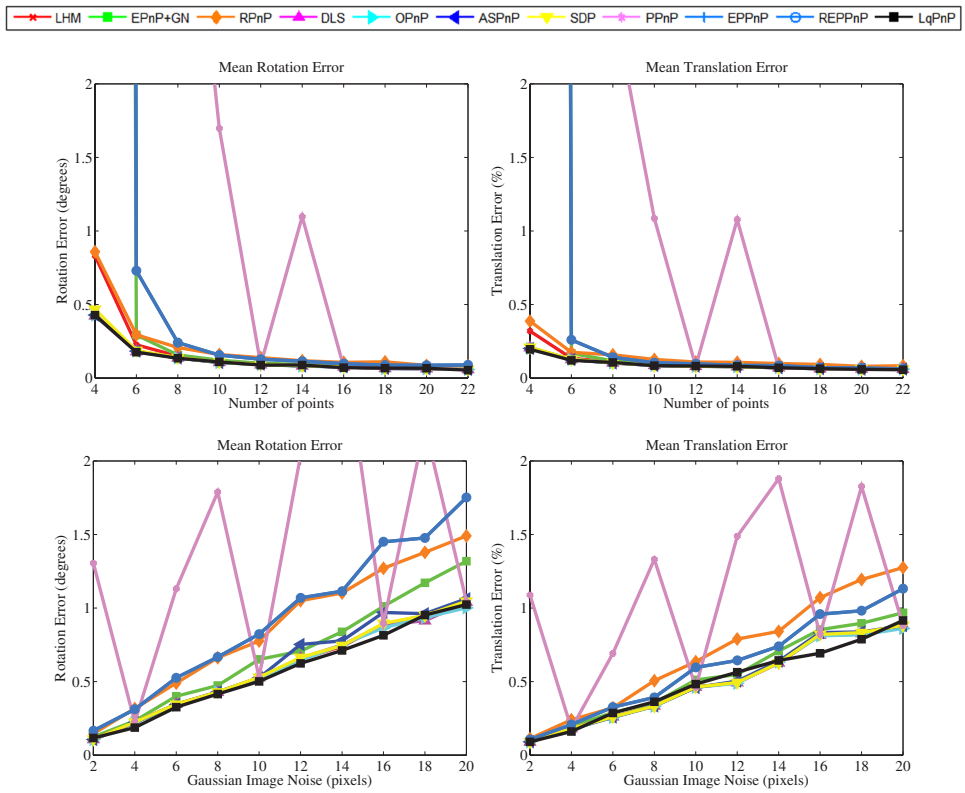
Robustness to Outliers

The main advantage of LqP*n*P compared with state-of-the-art methods is that LqP*n*P can handle situations corrupted by outliers without any outlier detection stage.

In this experiment, we fix the number of inliers to 20 and vary the outlier rates from n_{out}=5 percent to 95 percent. Thus, we generate n = 20/(1–n_{out}) 3D-to-2D correspondences in ordinary configuration and add the Gaussian noise with deviation δ=2 pixels. We randomly pick n–20 2D image projections and add errors that are randomly distributed in the range of [–300,

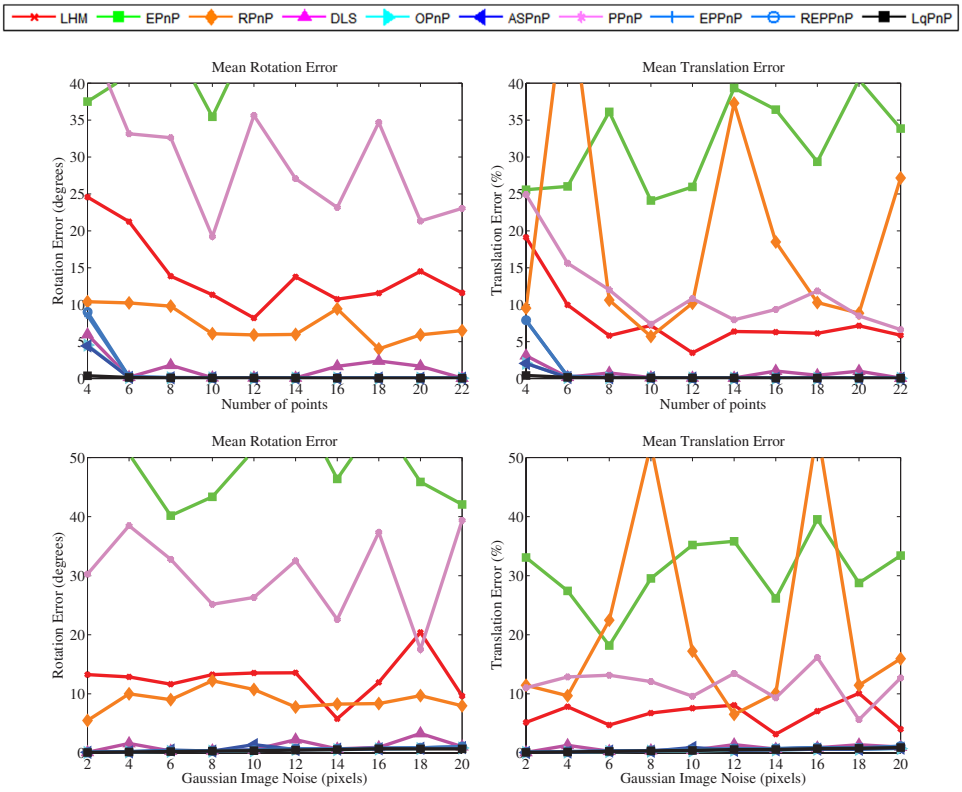


(a) Ordinary case



(b) Quasi-singular case

Figure 2. Synthetic experimental results for ordinary, quasi-singular, and planar cases, varying the correspondence numbers (first row in Figure 2a, Figure 2b, and Figure 2c) and the levels of Gaussian noise (second row in Figure 2a, Figure 2b, and Figure 2c): (a) Ordinary case, (b) Quasi-singular case, and (c) Planar case. *Continued on next page.*



(c) Planar case

Figure 2 (*continued*). Synthetic experimental results for ordinary, quasi-singular, and planar cases, varying the correspondence numbers (first row in Figure 2a, Figure 2b, and Figure 2c) and the levels of Gaussian noise (second row in Figure 2a, Figure 2b, and Figure 2c): (a) Ordinary case, (b) Quasi-singular case, and (c) Planar case.

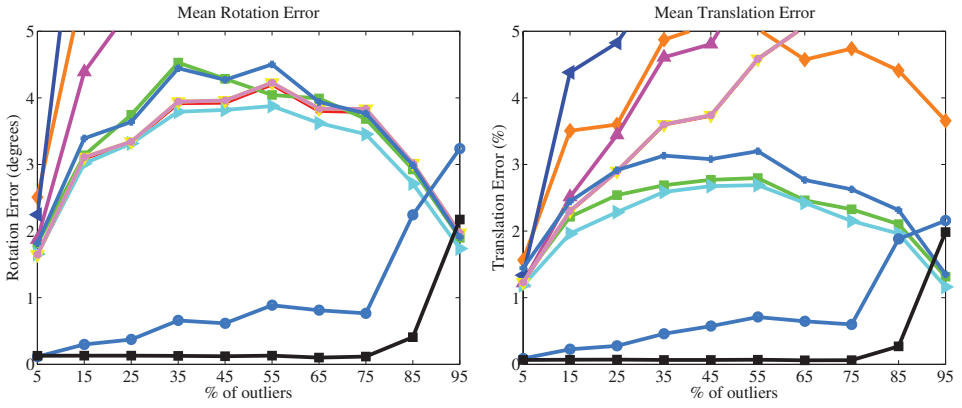


Figure 3. Experimental results with reference to increasing the levels of outlier rate.

300] pixels (the range is almost one-third of the image width). The results are shown in Figure 3.

Furthermore, we have compared the proposed solution LqP*n*P with REPPnP in terms of their sensibility to inlier numbers. We fix the outlier rate to 30 percent, increase the inlier number from 6 to 24, and add Gaussian noise with deviation δ=2 pixels. The average rotation and translation errors can be found in Figure 4.

We also compare LqP*n*P with some RANSAC-based multi-stage methods, i.e., RANSAC+P3P, RANSAC+P3P+ASPnP, RANSAC+P3P+OP*n*P, and RANSAC+RP4P+RPnP. In this experiment, we set the inlier number to 20. Gaussian noise levels are set to 2 pixels, 5 pixels, 10 pixels, and 20 pixels, respectively. We increase the outlier rate from 5 percent to 95 percent as in the above. Figure 5 plots the comparison results.

Figure 3 shows that only REPPnP and our LqP*n*P have the ability to distinguish inliers from outliers, while other nine methods all become very inaccurate when observations are corrupted by outliers. Our LqP*n*P achieves significantly better accuracy than REPPnP, especially for high outlier rates. Our LqP*n*P achieves accurate results (about 0.5 degrees for rotation error and 0.5 percent for translation error) even when the outlier rate reaches 85 percent.

Figure 4 shows that REPPnP is very sensitive to inlier numbers, while our LqP*n*P is robust. The performance of REPPnP is seriously decreased when the inlier number is less than 16,

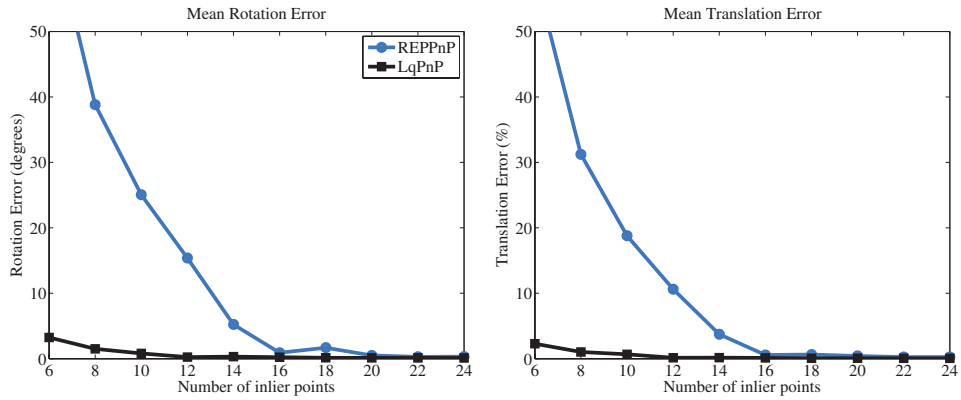
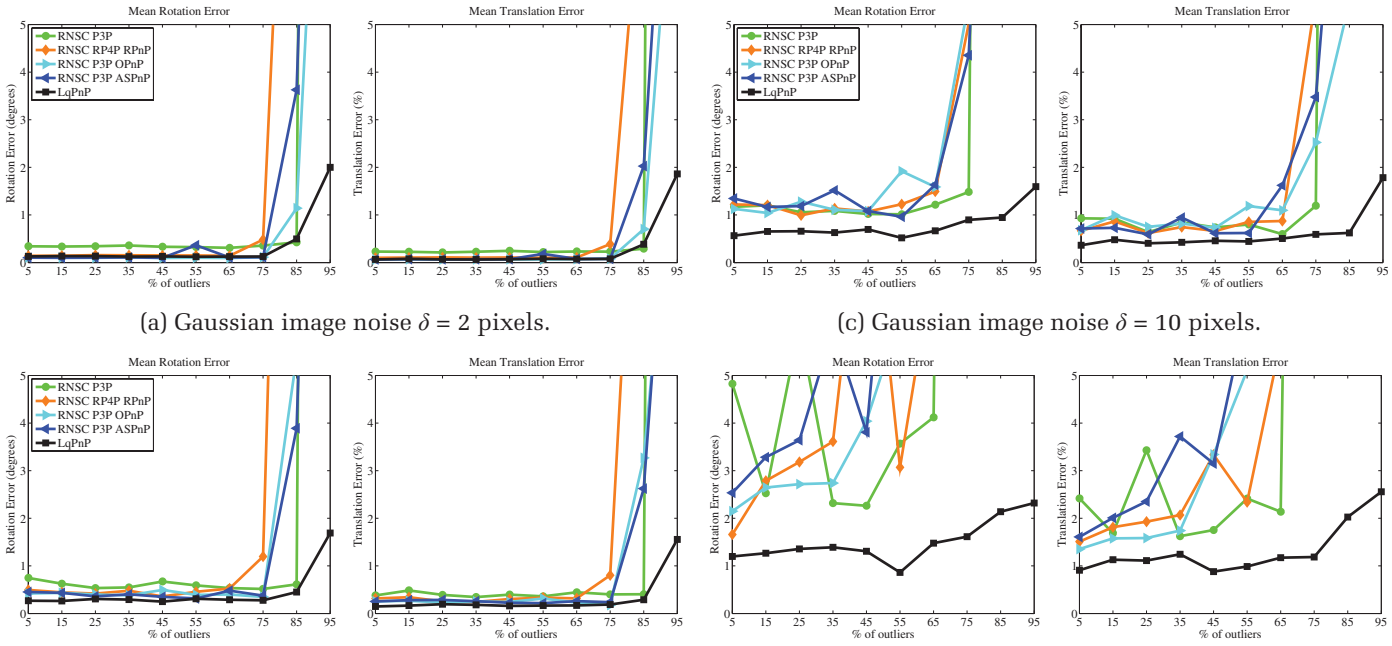


Figure 4. Comparison of LqPnP and REPPnP in terms of the sensibility to inlier numbers.



(a) Gaussian image noise δ = 2 pixels.

(c) Gaussian image noise δ = 10 pixels.

(b) Gaussian image noise δ = 5 pixels.

(d) Gaussian image noise δ = 20 pixels.

Figure 5. Performance comparison between LqPnP and RANSAC-based methods w.r.t. varying outlier rates.

while the performance of our LqPnP drops only slightly when the inlier number is less than 8.

Figure 5 indicates that our LqP*n*P achieves comparable results with RANSAC+P3P (or P4P) +PnP methods when the Gaussian noise level is small (see in Figure 5a and 5b). However, RANSAC-based methods are very sensitive to noise (see in Figure 5c and 5d). In addition, when the outlier rate is higher than 75 percent, the performance of our LqP*n*P is definitely better than RANSAC-based methods.

Real Experiment

In this experiment, we choose two large-scale aerial images captured by SWDC aerial digital camera system in Pingdingshan City, Henan, China. SWDC is composed of five digital cameras, i.e., one central camera and four oblique cameras. The first image (Figure 6a) is obtained by the central camera whose resolution is 5,406×7,160 pixels and focal length is 12,102.1 pixels. The other one (Figure 6b) is captured by an oblique camera whose resolution is 7,160×5,406 pixels and focal length of 14,671.5 pixels. There are 12 and 15 control points measured by GPS-RTK in Figure 6a and Figure 6b, respectively. The initial rotation angles (φ, ω, κ in the table) and camera position vector ([*Xs*,*Ys*,*Zs*] in the table) needed by our LqP*n*P are set as in Table 1. The comparison results are reported in Table 2 and Table 3.

Table 1. The Setting of Initial Values

Images	φ/degree	ω/degree	κ/degree	*Xs* (m)	*Ys* (m)	*Zs* (m)
1	0	0	0	0	0	100
2	0	0	-30	0	0	100

As can be seen, only 3 of 12 methods (EPnP+GN, PATB, and LqPnP) can correctly estimate the camera pose. Our LqP*n*P achieves the best RMSE of reprojection error, i.e., 0.617 and 0.678 pixels. The other nine methods converge to wrong solutions whose camera positions are under the ground. From Table 2 and Table 3, we can also infer that our LqP*n*P does not rely on good initial-guess if the observations have no gross errors. For example, the real pose of the second image is $[17.964, -44.673, -117.563, 657.875, 162.093, 647.846]^T$, while the initial values are set to $[0, 0, -30, 0, 0, 100]^T$.

We also perform an experiment with outliers. We randomly add three outliers (outlier rates of image 1 and image 2

(a) (b)

Figure 6. The aerial images used in this experiment. The red circles in the images are control points measured by GPS-RTK. (Figure 6a has been rotated for better visualization): (a) Image 1: captured by central camera of SWDC, and (b) Image 2: captured by an oblique camera of SWDC.

Table 2. The Experimental Results of the First Image

Methods	φ/degree	ω/degree	κ/degree	*Xs* (m)	*Ys* (m)	*Zs* (m)	RMSE(pixels)	True/False
LHM	-11.980	5.955	-20.396	-154.344	-39.414	-631.589	44.205	False
EPnP+GN	-3.4360	0.987	-19.807	19.2890	37.5730	650.269	1.991	True
RPnP	-16.676	3.508	-19.820	-205.987	-10.504	-617.843	46.092	False
DLS	-13.530	5.674	-20.374	-171.509	-35.707	-627.129	43.659	False
OPnP	-13.936	5.636	-20.350	-176.032	-35.127	-626.105	43.623	False
ASPnP	-12.845	6.530	-20.679	-164.067	-45.641	-628.751	44.271	False
SDP	-13.880	5.539	-20.329	-175.342	-34.055	-626.055	43.644	False
PPnP	-13.881	5.539	-20.329	-175.348	-34.050	-626.053	43.644	False
EPPnP	-11.291	6.036	-19.331	-148.154	-40.922	-640.069	62.728	False
REPPnP	-11.291	6.036	-19.331	-148.154	-40.922	-640.069	62.728	False
LqPnP	-3.2230	0.582	-19.785	16.864	32.822	650.592	0.617	True
PATB	-3.2100	0.545	-19.802	16.740	32.820	650.599	1.051	True

Note that **PATB** is a commercial software for bundle adjustment.

Table 3. The Experimental Results of the Second Image

Methods	φ/degree	ω/degree	κ/degree	*Xs* (m)	*Ys* (m)	*Zs* (m)	RMSE(pixels)	True/False
LHM	-19.033	41.645	63.137	-154.344	-39.414	-631.589	256.238	False
EPnP+GN	17.927	-44.755	-117.56	658.171	161.18	646.314	2.546	True
RPnP	-19.246	41.624	63.062	627.541	190.293	-684.325	256.902	False
DLS	-19.537	43.031	62.86	640.381	186.995	-663.044	257.031	False
OPnP	-19.716	42.911	62.702	647.656	192.066	-672.89	256.515	False
ASPnP	-19.400	42.880	62.312	639.621	186.035	-665.693	256.566	False
SDP	-20.3985	44.0064	62.27	646.752	191.336	-643.662	259.937	False
PPnP	-20.399	44.007	62.27	646.752	191.336	-643.66	259.937	False
EPPnP	5.433	3.832	59.196	49.547	-101.967	-705.835	1585.296	False
REPPnP	5.433	3.832	59.196	49.547	-101.967	-705.835	1585.296	False
LqPnP	17.964	-44.673	-117.563	657.875	162.093	647.846	0.678	True
PATB	17.99	-44.699	-117.577	657.718	162.188	646.999	1.192	True

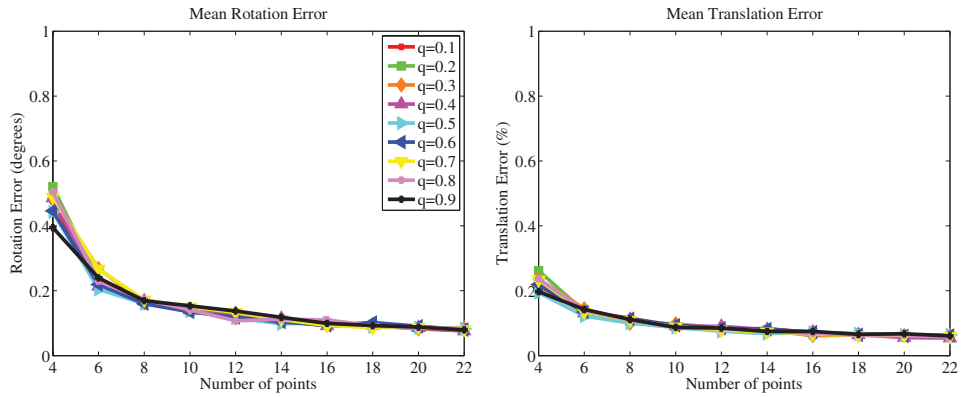
Note that **PATB** is a commercial software for bundle adjustment.

are 20 percent and 16.7 percent, respectively.) whose range is the maximum range of the all 3D-to-2D correspondences (the range of the image projections is almost as large as the image size). The average RMSEs (reprojection error) of 100 independent tests of our LqP*n*P are 1.292 and 1.571 pixels, respectively, while other methods are all failed.

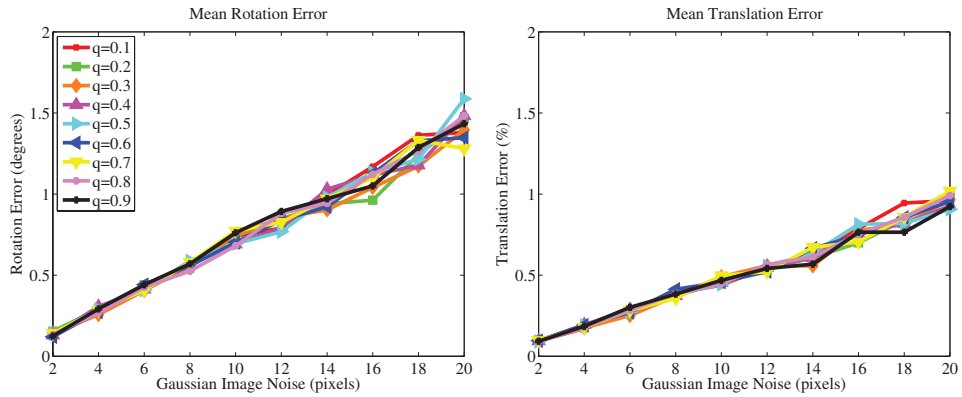
Parameter Study

In the proposed method, parameter q is very important. To quantify the sensitivity of the proposed method to parameter q, we have performed several studies whose results are reported in Figure 7. The configurations of these experiments are the same as the ordinary case in the Simulations section.

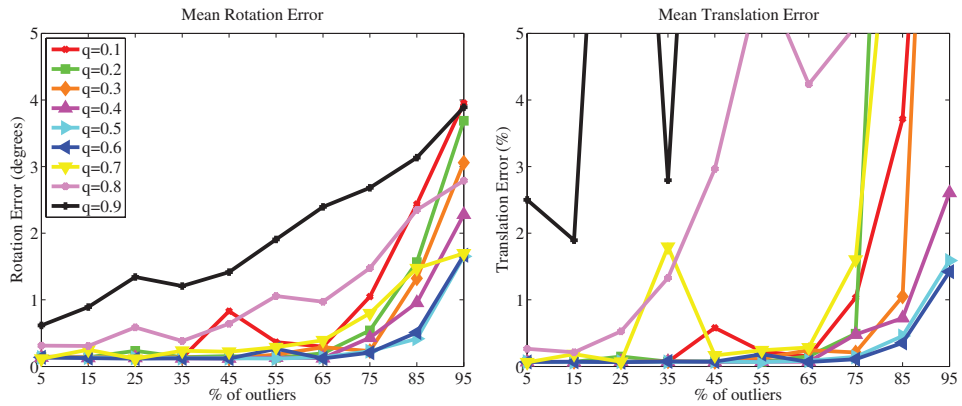
As can be seen, parameter q is not sensitive to the number of points and Gaussian noise levels (Figure 7a and Figure 7b). The performances are similar when varying the parameter q from 0.1 to 0.9. In contrast, parameter q is very sensitive to outliers. The best performances are achieved when q is set to be 0.5 and 0.6 (Figure 7c). Either small values or large values of q will decrease the accuracy and stability of the proposed



(a) Number of points.



(b) Gaussian noise levels.



(c) Outlier levels (Gaussian image noise δ = 2 pixels).

Figure 7. Results of parameter study: (a) Number of points, (b) Gaussian noise levels, and (c) Outlier levels (Gaussian image noise δ=2 pixels).

algorithm. This is the reason why q is set to be 0.5 in all our performance evaluation experiments.

Limitations

As mentioned in the above, the proposed method is iteration-based and an initial guess is needed. In other words, our LqP*n*P method has two drawbacks: (1) the convergence of LqP*n*P depends on the initial guess to some extent. Although we use unit quaternion to represent the rotation, LqP*n*P may fail when the initial guess is far from the ground truth. (2) our LqP*n*P is not suitable for real-time applications. It is more time-consuming than non-iterative ones. Fortunately, a good initial guess can be easily obtained by a set of non-iterative methods or by a position and orientation system (POS). Our LqPnP with good initial guess will converge within a few iterations.

Conclusions

In this paper, we proposed a new robust perspective-n-point solution for pose estimation problem. Our work recovers the orientation and position of a camera from observations corrupted by outliers without any outlier detection stage such as RANSAC. We use the l_q-norm metric to formulate the cost function and solve this function by adopting the augmented Lagrangian function and ADMM method. Extensive experimental evaluations on both synthetic and real data show that the proposed approach significantly outperforms the state-of-the-art methods.

Our LqPnP is more robust to noise and outliers than current state-of-the-art methods. (1) Current non-iterative PnP approaches are usually designed for small scale computer vision tasks. For large scale remote sensing images, these approaches usually become unreliable and converge to a wrong solution.

In contrast, the proposed method is suitable for both computer vision and photogrammetry applications. (2) The proposed method handles different point configurations, i.e., ordinary case, quasi- singular case and planar case, in a uniform framework. (3) The most important advantage of our method is that it removes outliers and recovers the camera pose in a single step. In contrast, RANSAC-based strategy usually rejects outliers during a preprocessing stage before the PnP methods are applied. In addition, RANSAC-based methods are less reliable and accurate than our method if the outlier rate is high or the noise level is serious.

The main limitation of LqPnP is the need of initial guess. Fortunately, the initial guess can be provided by POS or non-iterative methods. Our future studies will involve integrating LqPnP into bundle adjustment or structure-from-motion for accurate 3D reconstruction and DEM production.

Acknowledgments

The authors would like to express their gratitude to the editors and the reviewers for their constructive and helpful comments for substantial improvement of this paper. This work is supported by National Natural Science Foundation of China (No. 41271452).

References

Abdel-Aziz, Y., and H. Karara, 2015. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry, *Photogrammetric Engineering & Remote Sensing*, 81(2):103–107.

Bouaziz, S., A. Tagliasacchi, and M. Pauly, 2014. Sparse iterative closest point, *Proceedings of the Computer graphics forum, Conference Conference*, pp. 113–123.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine Learning*, 3(1):1–12.

Candè, E.J., and M.B. Wakin, 2008. An introduction to compressive sampling, *Signal Processing Magazine*, IEEE, 25(2):21–30.

Chen, X., F. Xu, and Y. Ye, 2010. Lower bound theory of nonzero entries in solutions of \ell_2-\ell_p minimization, *SIAM Journal on Scientific Computing*, 32(5):2832–2852.

DeMenthon, D., and L.S. Davis, 1992. Exact and approximate solutions of the perspective-three-point problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (11):1100–1105.

Dementhon, D.F., and L.S. Davis, 1995. Model-based object pose in 25 lines of code, *International Journal of Computer Vision*, 15(1-2):123–141.

Ferraz, L., X. Binefa, and F. Moreno-Noguer, 2014. Very fast solution to the PnP problem with algebraic outlier rejection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 501–508.

Fischler, M.A., and R.C. Bolles, 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24(6):381–395.

Gao, X.-S., X.-R. Hou, J. Tang,and H.-F. Cheng, 2003. Complete solution classification for the perspective-three-point problem, *Transactions on Pattern Analysis and Machine Intelligence*, 25(8):930–943.

Garro, V., F. Crosilla, and A. Fusiello, 2012. Solving the pnp problem with anisotropic orthogonal procrustes analysis, *Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pp. 262–269.

Hartley, R., and A. Zisserman, 2003. *Multiple View Geometry in Computer Vision*, Cambridge University Press,

Hesch, J., and S. Roumeliotis, 2011. A direct least-squares (DLS) method for PnP, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 383–390.

Horaud, R., B. Conio, O. Leboulleux, and L.B. Lacolle, 1989. An analytic solution for the perspective 4-point problem, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'89*, pp. 500–507.

Kneip, L., D. Scaramuzza and R. Siegwart, 2011. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2969–2976.

Lepetit, V., F. Moreno-Noguer, and P. Fua, 2009. Epnp: An accurate o (n) solution to the pnp problem, *International Journal of Computer Vision*, 81(2):155–166.

Li, J., R. Zhong, Q. Hu, and M. Ai, 2016. Feature-based laser scan matching and its application for indoor mapping, *Sensors*, 16(8):1265.

Li, S., C. Xu, and M. Xie, 2012. A robust O (n) solution to the perspective-n-point problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7):1444–1450.

Lowe, D.G., 1991. Fitting parameterized three-dimensional models to images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):441–450.

Lu, C.-P., G.D. Hager, and E. Mjolsness, 2000. Fast and globally convergent pose estimation from video images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):610–622.

Marjanovic, G. and V. Solo, 2012. On optimization and matrix completion. *Signal Processing, IEEE Transactions on*, 60(11):5714–5724.

Marjanovic, G., and V. Solo, 2012. l q matrix completion, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3885–3888.

Marjanovic, G., and A.O. Hero, 2014. On l q estimation of sparse inverse covariance, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3849–3853.

Marjanovic, G., and V. Solo, 2014. Sparsity *Penalized Linear Regression With Cyclic Descent*, *IEEE Transactions on Signal Processing*, 62(6):1464–1475.

Mazumder, R., T. Hastie, and R. Tibshirani, 2010. Spectral regularization algorithms for learning large incomplete matrices, *The Journal of Machine Learning Research*, 11:2287–2322.

Mohimani, G.H., M. Babaie-Zadeh, and C. Jutten, 2007. Fast sparse representation based on smoothed l0 norm, *Independent Component Analysis and Signal Separation*, Springer, pp. 389–396.

Schweighofer, G., and A. Pinz, 2008. Globally optimal O (n) solution to the PnP Problem for general camera models, *Proceedings of the BMVC*, pp. 1–10.

Shan, J., 1996. An algorithm for object reconstruction without interior orientation, *ISPRS Journal of Photogrammetry and Remote Sensing*, 51(6):299–307.

Tsai, R.Y., 1987. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Journal of Robotics and Automation*, 3(4):323–344.

Zheng, Y., S. Sugimoto, and M. Okutomi, 2013a. Aspnp: An accurate and scalable solution to the perspective-n-point problem, *IEICE Transactions on Information and Systems*, 96(7):1525–1535.

Zheng, Y., Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi, 2013b. Revisiting the pnp problem: A fast, general and optimal solution, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2344–2351.

(Received 19 May 2016; accepted 23 August 2016; final version 26 August 2016)