Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6675607, 17 pages
https://doi.org/10.1155/2021/6675607

Research Article

A Closed-Form Solution for Estimating the Accuracy of Circular Feature' Pose for Object 2D-3D Pose Estimation System

Cui Li,[1] **Derong Chen**,[1] **Jiulu Gong**,[1] **and Yangyu Wu**[1,2]

[1]*The National Laboratory for Mechatronic and Control, Beijing Institute of Technology, Beijing 100081, China*
[2]*The Key Laboratory of Intelligent Information and Big Data Processing of Ningxia Province, North Minzu University, Yinchuan 750021, China*

Correspondence should be addressed to Cui Li; licui@bit.edu.cn and Jiulu Gong; lujiugong@bit.edu.cn

Received 13 November 2020; Revised 3 February 2021; Accepted 18 February 2021; Published 17 March 2021

Academic Editor: Gaoge Hu

Copyright © 2021 Cui Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Many objects in the real world have circular feature. In general, circular feature's pose is represented by 5-DoF (degree of freedom) vector $\xi = (\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \alpha, \beta)^{\mathbf{T}}$. It is a difficult task to measure the accuracy of circular feature's pose in each direction and the correlation between each direction. This paper proposes a closed-form solution for estimating the accuracy of pose transformation of circular feature. The covariance matrix of ξ is used to measure the accuracy of the pose. The relationship between the pose of the circular feature of 3D object and the 2D points is analyzed to yield an implicit function, and then Gauss–Newton theorem is employed to compute the partial derivatives of the function with respect to such point, and after that the covariance matrix is computed from both the 2D points and the extraction error. In addition, the method utilizes the covariance matrix of 5-DoF circular feature's pose variables to optimize the pose estimator. Based on pose covariance, minimize the mean square error (Min-MSE) metric is introduced to guide good 2D imaging point selection, and the total amount of noise introduced into the pose estimator can be reduced. This work provides an accuracy method for object 2D-3D pose estimation using circular feature. At last, the effectiveness of the method for estimating the accuracy is validated based on both random data sets and synthetic images. Various synthetic image sequences are illustrated to show the performance and advantages of the proposed pose optimization method for estimating circular feature's pose.

1. Introduction

Pose estimation is an essential step in many machine vision and photogrammetric applications, and the ultimate goal of pose estimation is to identify 3D pose of an object of interest from an image or image sequence [1, 2]. The existing algorithms detect elliptic from 2D image, and the 3D pose of the circular can be extracted from single image using the inverse projection model of the calibrated camera (see Figure 1(b)) [3–5]. These methods are successfully applied to pose estimation of underwater dock [6] and pose estimation of the bait [7]. For the application of the monocular vision pose estimation system using circular feature in industrial, quite a few works on accuracy are presented; however, industrial application needs very high requirement to precision. The process of pose estimation is so complicated that it is difficult to observe the effect of the each parameter error on pose error and measure the accuracy of the circular feature's pose. To overcome this, in the paper, the covariance matrix of the pose vector is used to measure the accuracy of the circular feature's pose, and the paper proposes a closed-form solution for estimating the accuracy of circular feature's pose. However, before doing so, let us revisit several contributions related to the proposed method.

The accuracy of the monocular vision circular pose estimation system is related to many factors such as the detection error of image coordinate, the scale error of object model, and the camera calibration error. To evaluate error of elements in 3D pose estimation result, at the same time, the effects from related factors to 3D pose estimation result are analyzed and two approaches have been developed. A common approach was implemented for error analysis of 3D

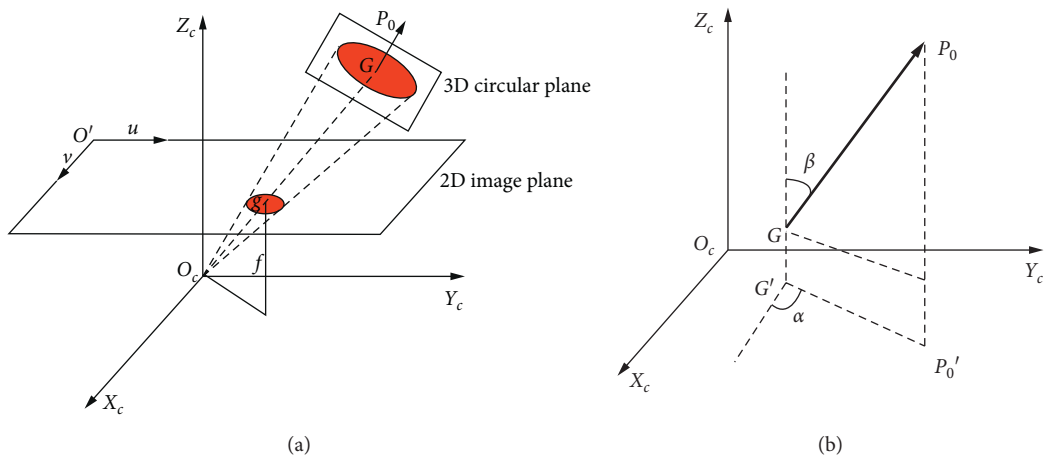


Figure 1: The projection of circular feature of 3-D object into the image plane (a) and a location and orientation of a circular feature (b).

pose estimation result with the ideal simulation model of the monocular vision pose estimation system. It does not derive the error propagation equations for the pose estimation model. An ideal monocular vision pose estimation system was adopted for experiment, and the various parameters of the monocular vision system were introduced into the pose estimation model, through which the main parameters that influence the error of pose estimation were concluded. The most effective accuracy optimization method was found in [1], [4, 8–13]. In [11], an error analysis method considering multiple error factors was proposed, and the camera internal error, lens distortion error, image point's detection error, and object error were introduced into the pose estimation model simultaneously based on actual error levels. The influence of suppressing parameter errors improving camera resolution on the pose estimation results was obtained. Above error analysis is based on the sampled model of 3D pose estimation, but this analysis just provides a rough qualitative result from the theory; for the quantitative result of error estimation, one needs to compute the covariance matrix of unknown variables.

Another approach based on a known model in the real project is that the error of the pose can be expressed by the error of the various parameters of the monocular vision system, and then one gets covariance of pose variables for the quantitative result of error estimation [14–24]. In [14–18], the simplified 3D pose estimation model is constructed instead of so complicated pose estimation model for special real project of the vision measure system, and the relation between the error of various input parameters and pose measurement error has been deduced by the first derivative of pose on each input parameter. In [21], error propagation rules in 3D reconstruction process have deduced on the base of the matrix analysis method, and a calculation method of the covariance matrix has put forward to evaluate error of elements in 3D reconstruction result. Through this work, one can conclude which parameter mainly influences the accuracy of the pose estimation so as to provide some reference for choosing different optimization technique for improving the accuracy of pose estimation.

Although the conclusion of these approaches provides important guidance for engineering application of the visual pose estimation system, the drawback of above methods is that the covariance is computed by setting multigroup actual pose parameters statistically. So, these works incur increased timing. In [23, 24], the covariance matrix of pose variables is used to measure the accuracy of relative pose, and a closed-form solution for estimating the accuracy of pose transformation is proposed. An implicit function from the three dimensional point pairs to the pose variable is defined, and then implicit theorem is employed to compute the partial derivatives of function with respect to such point pairs, and after the covariance matrix is computed from both the changing trend of implicit function and the measuring error of the camera. The proposed method provides fast numerical solutions for covariance of pose variables. Above methods rely on the direct relationship between image points and pose parameters to construct the 3D model of pose estimation for estimating the covariance matrix of pose variables. In the process of 2D-3D pose estimation using circular feature, the least square model and many matrix transformations are used and it is much complex. For measuring the accuracy of circular feature's pose variables in each direction and the correlation between each direction, it is a challenging task to have representations for the covariance matrix of 5D circular feature's pose variables with good and fast numerical solution by building the relationship directly from the image points to the 5D pose variables.

In this work, the proposed algorithm measures the accuracy of the circular feature's pose variables using the covariance matrix of pose variables. We propose a closed-form solution for estimating the covariance matrix of 5D circular feature's pose variables by exploiting 2D imaging point's coordinates and its extraction error. The main contributions of the present work can be summarized as follows:

(1) The paper proposes a closed-form solution for estimating the accuracy of 5D circular feature's pose variables using 2D imaging point's coordinate and its extraction error covariance matrix.

(2) The projection of the circular feature of 3D object yields a 2D virtual elliptic contour, and algebraic distance from 2D imaging point to the elliptic contour is employed to yield an implicit function from the image points to 5D pose variables as defined, and then Gauss–Newton theorem is employed to compute the partial derivatives of the function with respect to such point. For this, one can obtain error propagation rules in 3D pose estimation process and the relation between the extraction error of 2D imaging point's coordinate and circular feature's pose variable error. These works are used to compute the covariance matrix of circular feature's pose variables.

(3) The covariance of 2D imaging point's extraction error due to quantization can be computed by the size of quantization unit [19]. The error propagation rules are combined with the 2D imaging point's coordinate, and its quantization error covariance computes the covariance matrix of circular feature's pose variables. The proposed algorithm provides a closed-form solution for estimating the covariance matrix of circular feature's pose variables instead of Monte Carlo simulation.

(4) The covariance matrix of 5D circular feature's pose variables is employed to the object 2D-3D pose estimation system using circular feature. The proposed optimal algorithm yields high accuracy of object 2D-3D pose estimation, and low location error values are obtained when processing a sequence of 2D monocular images degraded with additive noise.

The paper is organized as follows. In Section 2, we define the representation method of the 5D circle pose parameters and then give the representation of the projection of the 3D object circular feature using the 5D circular pose variables. Section 3 explains the proposed closed-form solution for estimating the accuracy of 5D pose variables for 2D-3D object pose estimation systems using circular feature. Specifically, we discuss the definition method of an implicit function from the image points to 5D pose variables and the representations for the covariance matrix of 5D circular feature's pose variables with good and fast numerical solution by building the relationship directly from the image points to the 5D pose variables. In Section 4, we briefly explain the proposed pose optimization algorithm for object 2D-3D pose estimation using circular feature and the covariance matrix of 5D circular feature's pose for selecting good 2D imaging point for object 2D-3D pose optimization. Section 5 presents experimental results obtained with the proposed closed-form solution and the proposed pose optimization algorithm when processing synthetic image, which are discussed and compared with the obtained by Monte Carlo simulation and existing object 2D-3D pose estimation using circular feature. The conclusions of the present work are summarized in Section 6.

2. Preliminaries

2.1. The Position and Orientation of a Circular Feature in 3D. As shown in Figure 1(a), we have shown two coordinate frames. The camera frame $X_C - Y_C - Z_C$ is a 3D frame with the origin as the projection center and has its $Z_C - axis$ pointing to direction it is pointed. The image frame $u - v$ is a 2D frame with u and v axes parallel to Y_C and X_C of the camera frame, respectively. The projection of circular feature with a radius of R into the image plane is represented as elliptic g.

As shown in Figure 1(b), the position and orientation of a circular feature in 3D are completely specified by the coordinates of its center G and the direction angle of the surface normal vector $\overrightarrow{GP_0}$. We will adopt a convention that points the surfaces normal from the circular towards the direction where the circular is visible [25]. Examples are shown in Figure 1(b). The direction angle α indicates the angle between the projection of $\overrightarrow{GP_0}$ into the plane $O_C - X_C Z_C$ and the axis $O_c X_c$. The positive angle is defined as the counter clockwise rotation of $\overrightarrow{G'P_0'}$. The direction angle β indicates the angle between $\overrightarrow{GP_0}$ and $O_c Z_c$ axis. The positive angle is defined as the counter clockwise rotation of $\overrightarrow{GP_0}$. In addition, $\overrightarrow{GP_0} = 1$ and $\overrightarrow{GP_0} = (l, m, n)^T = (\sin\beta\cos\alpha, \sin\beta\sin\alpha, \cos\beta)^T$. Therefore, the position and orientation of a circular feature in 3D can be expressed as $\xi = (X, Y, Z, \alpha, \beta)^T$, where the coordinates of the center G are represented as $G = (X, Y, Z)^T$ [26].

2.2. The Representation of the Projection of the 3D Object Circular Feature Using $\xi = (X, Y, Z, \alpha, \beta)^T$. To interact with 2D imaging points, the circular feature in 3D has to be projected to the image plane and then the result yields a 2D virtual elliptic contour $\Phi(\xi, (x, y))$. In the section, we shall give the representation of the projection of the 3D object circular feature using the circular feature's pose variables $\xi = (X, Y, Z, \alpha, \beta)^T$ for constructing an implicit function from 2D imaging point to 5D pose variables.

For each pose configuration $\xi = (X, Y, Z, \alpha, \beta)^T$, one can derive $\Phi(\xi)$ as follows: let $(X, Y, Z)^T$ and $(l, m, n)^T$ denote the center coordinate and surface normal vector of the circular feature in 3D with radius R. Equation for projection of the circular feature in 3D into the image plane to yield the 2D ellipse curve $\Phi(\xi, (x, y))$ is as follows:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0\left(B^2 - 4AC = 1\right), \tag{1}$$

where the parameter $(A, B, C, D, E, F)^T$ is represented by the location $G = (X, Y, Z)^T$ and 3D orientation $(l, m, n)^T = (\sin\beta\cos\alpha, \sin\beta\sin\alpha, \cos\beta)^T$ and we need to distinguish it from reference [25]. See Appendix A for detailed derivation. Moreover, the parameters can be denoted as $\overline{\Phi}(\xi) = (A(\xi), B(\xi), C(\xi), D(\xi), E(\xi), F(\xi))^T$, and we represent this 2D virtual elliptic contour by its normalized coefficients: $\widetilde{\Phi}(\xi) = (\overline{\Phi}(\xi)/\overline{\Phi}(\xi))$.

$$\overline{\Phi}(\xi)=\begin{cases}((\sin\beta\sin\alpha)Y+\cos\beta Z)^2+(\sin\beta\cos\alpha)^2\left(Y^2+Z^2-R^2\right);\\ -2(\sin\beta\sin\alpha)X(\sin\beta\sin\alpha Y+\cos\beta Z)-2(\sin\beta\cos\alpha)Y((\sin\beta\cos\alpha)X+\cos\beta Z)\\ +2((\sin\beta\cos\alpha)\cdot(\sin\beta\sin\alpha))\left(Z^2-R^2\right);\\ (\sin\beta\sin\alpha)^2\left(X^2+Z^2-R^2\right)((\sin\beta\cos\alpha)X+\cos\beta Z)^2;\\ -2X\cos\beta f((\sin\beta\sin\alpha)Y+\cos\beta Z)+2\cos\beta(\sin\beta\cos\alpha)fY^2\\ -2f(\sin\beta\cos\alpha)\left(Z((\sin\beta\cos\alpha)X+(\sin\beta\sin\alpha)Y)-\cos\beta R^2\right);\\ 2(\sin\beta\sin\alpha)\cos\beta R^2-2\cos\beta fY((\sin\beta\cos\alpha)X+\cos\beta Z)\\ -2(\sin\beta\sin\alpha)f\left(Z((\sin\beta\cos\alpha)X+(\sin\beta\sin\alpha)Y)-\cos\beta R^2\right);\\ X^2(\cos\beta)^2+Y^2(\cos\beta)^2+((\sin\beta\cos\alpha)X+(\sin\beta\sin\alpha)Y)^2f^2-R^2(\cos\beta)^2f^2;\end{cases}\tag{2}$$

where f is the focal length of the camera.

3. Method Description

In this section, we present the closed-form solution for computing the covariance matrix of 5D circular feature's pose variable using 2D imaging points and its extraction error covariance under the implicit function. We shall discuss the definition method of implicit function by algebraic distance from 2D imaging point to the elliptic contour. Later, we shall incorporate such implicit function with the Gauss–Newton theorem to obtain error propagation rules in 3D pose estimation process and the relation between the extraction error of 2D imaging point's coordinate and circular feature's pose variable error. For this, a closed form for estimating the covariance matrix of circular feature's pose variables can be obtained.

3.1. Minimization Algebraic Distance to Yield an Implicit Function. To interact with the 2D imaging points, we give the representation of the projection of the 3D object circular feature using $\xi=(X,Y,Z,\alpha,\beta)^T$ and then these results yield an implicit function E_{circle} from all 2D imaging points to circular pose variables. Moreover, the implicit function is assumed to be constructed by the algebraic distance E_{p_ic}.

For each pose configuration $\xi=(X,Y,Z,\alpha,\beta)^T$, the algebraic distance E_{p_ic} is then defined as the points-to-contour distance $E_{p_ic}(p_i,\widetilde{\Phi}(\xi))$ between detected the ith 2D imaging point $p_i(x_i,y_i)^T$ and the 2D virtual elliptic contour $\overline{\Phi}(\xi)=(A(\xi),B(\xi),C(\xi),D(\xi),E(\xi),F(\xi))^{\text{T}}$, which can be denoted as follows:

$$E_{p_ic}\left(p_i,\widetilde{\Phi}(\boldsymbol{\xi})\right)=\widetilde{A}(\boldsymbol{\xi})\cdot x_i^2+\widetilde{B}(\boldsymbol{\xi})\cdot x_iy_i+\widetilde{C}(\boldsymbol{\xi})\cdot y_i^2,\widetilde{D}(\boldsymbol{\xi})\cdot x_i+\widetilde{E}(\boldsymbol{\xi})\cdot y_i+\widetilde{F}(\boldsymbol{\xi})=\left[x_i^2,x_iy_i,y_i^2,x_i,y_i,1\right]\cdot\widetilde{\Phi}(\boldsymbol{\xi}),\tag{3}$$

where $p_i(x_i,y_i)^T\in p$, $p=\left\{p_i(x_i,y_i)^T,\ i=1,\ldots,n\right\}$ indicates the set of all detected 2D point on the image plane that correspond to a 3D point on the circular feature in 3D. By knowing that the circular feature's pose parameters can be obtained by minimizing the square sum of all detected 2D imaging point's algebraic distance, which is defined as follows:

$$E_{\text{circle}}=\sum_{i=1}^{n}E_{p_ic}^2\left(p_i,\widetilde{\Phi}(\boldsymbol{\xi})\right)=E_{pc}\left(p,\widetilde{\Phi}(\boldsymbol{\xi})\right)_2^2=D(p)\widetilde{\Phi}(\boldsymbol{\xi})_2^2,\tag{4}$$

where $E_{pc}(p,\widetilde{\Phi}(\xi))=[E_{p_1c}(p_1,\widetilde{\Phi}(\xi)),\ldots,E_{p_nc}(p_n,\widetilde{\Phi}(\xi))]^T$ and $D(p)$ indicates the stack matrices of 2D imaging points $p=\left\{p_i(x_i,y_i)^T,\ i=1,\ldots,n\right\}$, as expressed in the following equation:

$$D(p)=\begin{bmatrix}x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ x_n^2 & x_ny_n & y_n^2 & x_n & y_n & 1\end{bmatrix}.\tag{5}$$

For convenient, we denoted $f(p,\xi)=D(p)\widetilde{\Phi}(\xi)$.

The least squares objective of circular feature's pose optimization can be written as follows:

$$\arg\min_{\boldsymbol{\xi}}E_{\text{circle}}(\boldsymbol{\xi},p)=\|f(p,\boldsymbol{\xi})\|_2^2,\tag{6}$$

where $f(p,\xi)$ is a combination of 2D imaging points and pin-hole projection. The Levenberg–Marquardt method is used for optimizing circular feature's pose variables [27].

3.2. Computing Covariance Matrix of Circular Feature's Pose Variables through Implicit Function. In this subsection, we describe the representations for the covariance matrix of 5D circular feature's pose variables with good and fast numerical solution by using 2D imaging points and its extraction error covariance under the nonlinear implicit function. Here, Gauss–Newton theorem is employed to compute the partial derivatives of the function with respect to such point, and then one can derive the covariance matrix σ as follows.

Solving the least squares objective of equation (6) needs the first-order approximation to the nonlinear function $f(p,\xi)$ linearization about initial pose ξ^s:

$$\|f(p,\boldsymbol{\xi})\|_2^2 \approx \left\|f(p,\boldsymbol{\xi}^s) + J_{\boldsymbol{\xi}}(\boldsymbol{\xi}-\boldsymbol{\xi}^s)\right\|_2^2, \tag{7}$$

where J_ξ is a block matrix of size $n \times 6$, n is the number of all detected 2D imaging points; for simplicity, it can be expressed as $J_\xi = (\partial f/\partial \xi)$.

To minimize of the first-order approximation equation (7) via Gauss–Newton, $(\mathrm{d}(f(p,\xi^s) + J_\xi(\xi-\xi^s)_2^2)/\mathrm{d}\ (\xi-\xi^s)) = 0$ is used to yield an analytic expression of pose optimization error $\Delta\xi = \xi - \xi^s$, which can be denoted as follows:

$$\Delta\boldsymbol{\xi} = J_{\boldsymbol{\xi}}^{+} \cdot f(p,\boldsymbol{\xi}^s), \tag{8}$$

where J_ξ^+ is the pseudo inverse of Jacobian matrix J_ξ. The pose estimation error is affected by 2D imaging points extraction error. Again with the first-order approximation of at the initial pose ξ^S and 2D imaging point's coordinate p^S, we can connect the pose estimation error to 2D imaging point's extraction errors:

$$\Delta\boldsymbol{\xi} = J_{\boldsymbol{\xi}}^{+} \cdot J_p(p-p^s), \tag{9}$$

where J_p is a block diagonal matrix of size $1n \times 2n$; for simplicity, it can be expressed as $J_p = (\partial f(p,\xi^s)/\partial p)$.

According to equation (9), error propagation rules from 2D imaging point's coordinate extraction error Δ_p to circular feature's pose variable error Δ_ξ can be obtained, which is denoted as follows:

$$\Delta_{\boldsymbol{\xi}} = \left(J_{\boldsymbol{\xi}}^{+} J_p\right)\Delta_p. \tag{10}$$

Consider the case that only quantization error exists and is i.i.d Gaussian with isotropic: $\Delta_p \sim N(\mu_p, \sigma_p^2)$, we can derive the pose covariance matrix:

$$\sum\nolimits_{\Delta_{\boldsymbol{\xi}}} = \left(J_{\boldsymbol{\xi}}^{+} J_p\right)\left(J_{\boldsymbol{\xi}}^{+} J_p\right)^T \sigma_p^2 = \sigma_p^2 \cdot \left(\sum_{i=1}^{n} J_{\boldsymbol{\xi}}^T(i)\left(J_p(i) J_p^T(i)\right)^{-1} J_{\boldsymbol{\xi}}(i)\right)^{-1}. \tag{11}$$

Still all 2D imaging points being detected should be utilized, and we can compute the pose covariance matrix from 2D imaging point's coordinate and its extraction error variance. The variance of 2D imaging point's quantization error can be computed by the size of quantization unit in Section 5.

4. Covariance for 2D-3D Object Pose Estimation Using Circular Feature

In 2D-3D object pose estimation system using circular feature, 2D imaging points extraction error negatively affects pose optimization result. Regarding accurate pose optimization, not all 2D imaging points being detected contribute the same. If only those valuable towards accurate pose estimation are utilized, the total amount of noise introduced into the 2D-3D object pose estimation system using circular feature can be reduced, and this work improves the accuracy of pose estimation result. Here, minimize the mean square error (Min-MSE) metric is used to select S 2D imaging points out of n candidates, and then the Max-Trace metric is introduced to guide 2D imaging points selection. 2D imaging points selection occurs prior to Levenberg–Marquardt pose optimization so that only a subset of selected features is sent to the optimizer. First, according to Min-MSE criteria, we shall give expression to evaluate a random subset of candidates to identify the current "best" 2D imaging point. Later, we shall briefly describe an algorithm for approximately solving the NP-hard Max-Trace problem.

4.1. Good 2D Imaging Points Selection Metrics. Analyzing the impact of 2D imaging points extraction error on pose optimization led to equations where the matrix trace of $(J_\xi^+ J_p)$ was connected to best case outcome. It well known that the expectation of the quadratic form is equal to the matrix trace, and it can be expressed as follows: $\mathrm{E}(\omega^{\mathrm{T}}\Psi\omega) = \mathrm{Tr}(\Sigma_\omega \Psi)$ with $\omega \sim N(0, \Sigma_\omega)$ According to this work and pose error propagation rules equation (10), $\mathrm{MSE}(\tilde{\xi})$ can be denoted as follows:

$$\begin{aligned}\mathrm{MSE}(\tilde{\boldsymbol{\xi}}) &= E\left((\tilde{\boldsymbol{\xi}}-\boldsymbol{\xi})^T(\tilde{\boldsymbol{\xi}}-\boldsymbol{\xi})\right) = E\left(\mathrm{tr}\left[(\Delta\boldsymbol{\xi})^T(\Delta\boldsymbol{\xi})\right]\right) = \mathrm{tr}\left(E\left[(\Delta\boldsymbol{\xi})^T(\Delta\boldsymbol{\xi})\right]\right)\\ &= \sigma_p^2 \cdot \mathrm{tr}\left[\left(J_{\boldsymbol{\xi}}^{+} J_p\right)^T\left(J_{\boldsymbol{\xi}}^{+} J_p\right)\right] = \sigma_p^2 \cdot \mathrm{tr}\left[\left(J_{\boldsymbol{\xi}}^{+} J_p\right)\left(J_{\boldsymbol{\xi}}^{+} J_p\right)^T\right],\end{aligned} \tag{12}$$

where equation (12) gives the relationship among the 2D imaging point's coordinate p^s, its extraction error variance, and the $\mathrm{MSE}(\tilde{\xi})$ criteria. Good 2D imaging point's selection metrics can be expressed as follows:

$$f(\cdot) = \min\left(\mathrm{tr}\left[\left(J_{\boldsymbol{\xi}}^{+} J_p\right)\left(J_{\boldsymbol{\xi}}^{+} J_p\right)^T\right]\right) = \max\left(\mathrm{tr}\left(\left(\left[\left(J_{\boldsymbol{\xi}}^{+} J_p\right)\left(J_{\boldsymbol{\xi}}^{+} J_p\right)^T\right]\right)^{-1}\right)\right). \tag{13}$$

4.2. Efficient Max – Trace Subset Selection. By knowing that subset selection with the Max – Trace metrics is equivalent to a finite combinational optimization problem, it can be denoted as follows:

$$\min_{p_s \subseteq p, |p_s|=s} \operatorname{tr}(p_s) = \operatorname{tr}\left(\left[J_\xi^+ J_p(p_s)\right]\left[J_\xi^+ J_p(p_s)\right]^T\right),$$
$$\Longrightarrow \max_{p_s \subseteq p, |p_s|=s} \operatorname{tr}(p_s) = \operatorname{tr}\left(\left(\left[J_\xi^+ J_p(p_s)\right]\left[J_\xi^+ J_p(p_s)\right]^T\right)^{-1}\right), \tag{14}$$

where s is the indices of selected row blocks from full matrix $(J_\xi^+ J_p)$. Also, it is the indices of selected 2D imaging point. Moreover, in equation (14), $([J_\xi^+ J_p(p_s)][J_\xi^+ J_p(p_s)]^T)^{-1}$ can be rewritten as follows:

$$\begin{aligned}\left(\left[J_\xi^+ J_p(p_s)\right]\left[J_\xi^+ J_p(p_s)\right]^T\right)^{-1} &= \sum_{i=1}^{s} J_\xi^T(i)\left(J_p(i)J_p^T(i)\right)^{-1} J_\xi(i) \\ &= \sum_{i=1}^{s}\left[J_p^+(i)J_\xi(i)\right]^T\left[J_p^+(i)J_\xi(i)\right].\end{aligned} \tag{15}$$

Note that the Jacobian J_p is a $1n \times 2n$ block diagonal matrix, consisting of 1×2 denoted by $J_p(i)$. Meanwhile, each row block of J_ξ can be written as $J_\xi(i)$.

To remove the need for the pseudo inverse of J_p, add one more row $[0, 1]$ to each block $J_p(i)$, where $J_p^n(i) = [J_p(i); 0\,0\,1]$. In addition, a zero row is added to each row block $J_\xi(i)$ to get new row block $J_\xi^n(i) = [J_\xi(i); 0, 0, 0, 0, 0]$. This trick does not affect the structure of the least square problem, but it does allow inversion of the new block $J_p^n(i) = [J_p(i); 0\,0\,1]$. After performing block-wise multiplication, one can obtain the combined matrix $J_c(i)$, consisting of concatenated row blocks $J_p^n(i)^{-1} J_\xi^n(i)$. Instead of working with two independent matrices J_ξ and J_p, we consider optimizing their combination J_c.

For this, good 2D imaging point's selection problem is equivalent to selecting a subset of row blocks in the matrix J_c, and then equation (15) can be rewritten as follows:

$$[S] = \max_{s \subseteq \{1,2,\ldots,n\}, |S|=k} \operatorname{tr}\left(\sum \left[J_c(S)\right]^T \left[J_c(S)\right]\right), \tag{16}$$

where the combinational optimization above can be solved with brute-force, but the exponentially growing problem space quickly becomes impractical to search, especially for real-time application. Subset selection with Max – Trace metric has been studied in fields such as feature selection [28]. Here, the stochastic-greedy-based Max – Trace 2D imaging point's selection algorithm is commonly used to approximate the original NP-hard combinational optimization problem. 2D imaging point's selection is done in three steps:

(1) Compute the full 2D imaging points Jacobian J_ξ and Jacobian J_p;

(2) Combine the two into J_c;

(3) Greedily select row blocks of J_c, based on the matrix-revealing metric Max – Trace, until reaching the target subset size. Algorithm 1 summarizes the stochastic-greedy-based Max – Trace 2D imaging point's selection algorithm [29].

5. Experiment Results and Discussion

The results obtained with the proposed algorithm are presented and discussed in this section. First, the effectiveness of the method for estimating the accuracy is validated based on random simulation and synthetic data sets. Later, integrating Max-Trace 2D imaging point's selection into a 2D-3D object pose estimation system using circular feature leads to accuracy improvements with low overhead, as demonstrated via evaluation on synthetic data sets.

5.1. A Closed-Form for Estimating Covariance vs. Monte Carlo Simulation. The effectiveness of a closed-form solution for estimating covariance is demonstrated by comparing against the Monte Carlo simulation method. The Monte Carlo simulation method is employed to compute the actual covariance value based on random data sets and synthetic data sets and then compares actual value with estimation value.

The simulation environment is depicted as follows.

Here, we produced a virtual camera based on the perspective projection model. The image size was 256×256 pixels, and the effective focal length was $f = 20$ mm, each circular feature with the radius $R = 100$ mm. After applying a small random pose transform to 3D object with circular feature, a fixed number (e.g., 180 in this simulation test) of the 2D imaging points of circular feature are generated using the 2D virtual elliptic contour model $\Phi(\xi, (x, y))$. To simulate extraction error, the 2D imaging points are perturbed with variety of measurement error: zero-mean Gaussian with standard deviation of 0.25, 0.50, and 0.75 pixel. Figure 2 illustrates examples of noisy random data for different values of σ_p^2.

Note that these random data are generated by MATLAB pseudo-random data generator. The mean and covariance of 2D imaging points coordinate extraction error can be denoted as follows:

$$\mu = [0, 0]^T,$$
$$\delta_{\text{cov}} = \begin{bmatrix} \sigma_x^2 & \\ & \sigma_y^2 \end{bmatrix}, \tag{17}$$

where the extraction error of coordinate x and y is i.i.d. Gaussian: $\Delta_x \sim N(0, \sigma_x^2)$, $\Delta_y \sim N(0, \sigma_y^2)$ To be statistically sound, 500 runs are repeated for each 2D imaging point, and one can obtain 500 numbers of 2D imaging point set. For this, 500 numbers of pose estimation value can be obtained by equation (6), and then the actual pose covariance matrix can be computed by using the Monte Carlo simulation method.

Data: $J_c = \{J_c(1), J_c(2), \ldots J_c(n)\}, k$
Result: $J_c^{\text{sub}} \subseteq J_c, |J_c^{\text{sub}}| = k$
(1) $J_c^{\text{sub}} = \varnothing$;
(2) while $|J_c^{\text{sub}}| < k$ do
(3) $J_c^R \longleftarrow$ a random subset obtained by sampling $|J_c^R| = (n/k)\log(1/\varepsilon)$ random elements from J_c;
(4) $J_c(i) \longleftarrow \text{argmax}_{J_c(i) \in J_c^R} \text{Trace}(J_c(i)^T J_c(i) + [J_c^{\text{sub}}]^T [J_c^{\text{sub}}])$;
(5) $J_c^{\text{sub}} \longleftarrow J_c^{\text{sub}} \cup J_c(i)$;
(6) $J_c \longleftarrow (J_c / J_c(i))$;

Algorithm 1: An efficient approximation algorithm for Max – Trace feature selection.

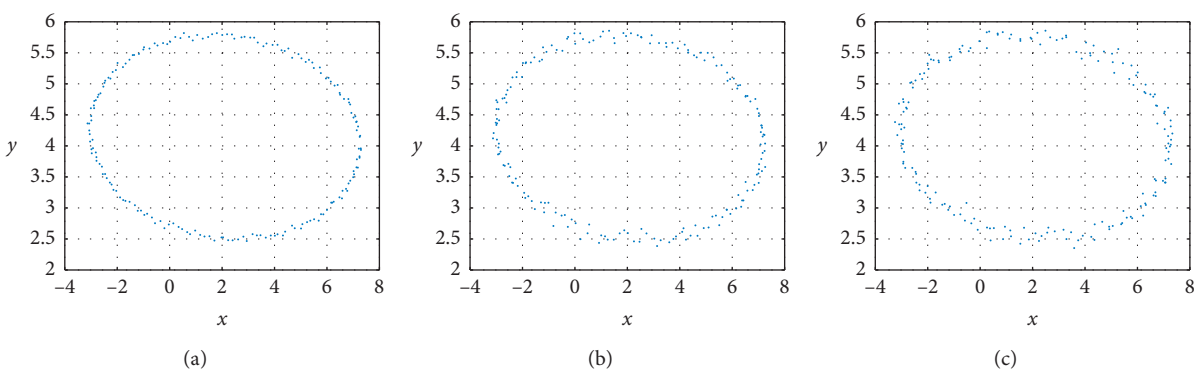


Figure 2: Random simulation data degraded with additive noise with the variance: (a) $\sigma_p = 0.25$ pixel, (b) $\sigma_p = 0.50$ pixel, and (c) $\sigma_p = 0.75$ pixel.

Another, one set is taken randomly from 500 numbers of 2D imaging point sets, and pose covariance estimation value can be obtained using the proposed a closed-form solution. Figures 3–5 present the projection of 5D ellipsoid corresponding to estimate the pose covariance matrix and actual pose covariance matrix into 2D plane, respectively, and the confidence in the ellipse is 95%. The simulation environment sets $\xi = [40, 90, 400, -(\pi/2), (\pi/18)]^T$, and it indicates the center of ellipse. Figures 3–5 present the comparisons of real and estimated covariance between translations and rotations.

In Figures 3–5, the estimated pose covariance with the proposed algorithm is indicated with 'closed'; the actual pose covariance obtained with Monte Carlo is indicated with 'Monto.' We can see that after processing random data with variety of noisy conditions, the proposed algorithm is able to estimate the pose covariance with good accuracy, and the results are closed to actual pose covariance even in highly noisy conditions. These results demonstrate that the proposed algorithm can estimate the covariance matrix of 5D circular feature's pose variables with good and fast numerical solution.

Estimating the covariance matrix of circular feature's pose variables requires the extraction error of each 2D imaging point. Here, we use the mathematical tools in [19] for the computation of the average value and the standard deviation of the extraction error due to quantization. The maximum quantization error in the detected 2D imaging points coordinates x, y being quantized is half the size of the quantization unit, and in spatial quantization, the maximum error is half a pixel, $(r/2)$. That is to say, the actual value x_a, y_a of the detected 2D imaging points coordinates x, y can be anywhere in the interval $x_a \in [x_q - (r/2), x_q + (r/2)]$ and $y_a \in [y_q - (r/2), y_q + (r/2)]$, where x_q, y_q is the value of x, y after quantization. Furthermore, the likelihood of the actual value of x, y, i.e., x_a, y_a , being at a certain place within the interval follows a uniform probability density. In other words, the probability of x_a, y_a lying inside the small interval $\mathrm{d}x, \mathrm{d}y$ around the point x_q, y_q is independent of x, y and equal to $(1/r)$. For example, the average value of the quantization error in x is $E(\triangle x) = \int_{x_q-(r/2)}^{x_q+(r/2)} (1/r)(x - x_q)\mathrm{d}x = 0$, and the average of the absolute value error in x is $E(|\triangle x|) = \int_{x_q-(r/2)}^{x_q+(r/2)} (1/r)|x - x_q|\mathrm{d}x = (r/4)$, and the standard deviation of the quantization error in x is $\sigma_x^2 = \int_{-(r/2)}^{(r/2)} (1/\ r)(\triangle x)^2 \mathrm{d}\triangle x = (r^2/12)$. So far, the extraction error of each 2D imaging point has been calculated.

To identify the estimated pose covariance that could evaluate the accuracy of circular feature's pose variables, a synthetic data simulation is carried out. The synthetic image is rendered using 3 ds Max software, each frame of an input image is composed of 3D object with circular feature that follows an unknown pose, it is embedded into a disjoint background, and the whole frame is degraded with small additive noise. Each test image is monochrome image with 256×256 pixels, and the effective focal length was

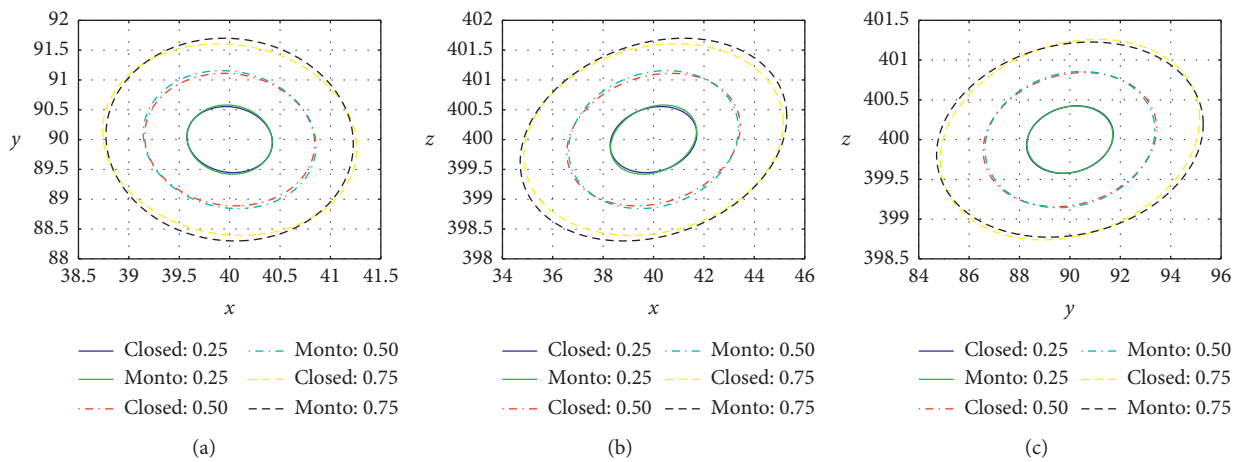


Figure 3: Comparisons of real and estimated covariance (among translations).

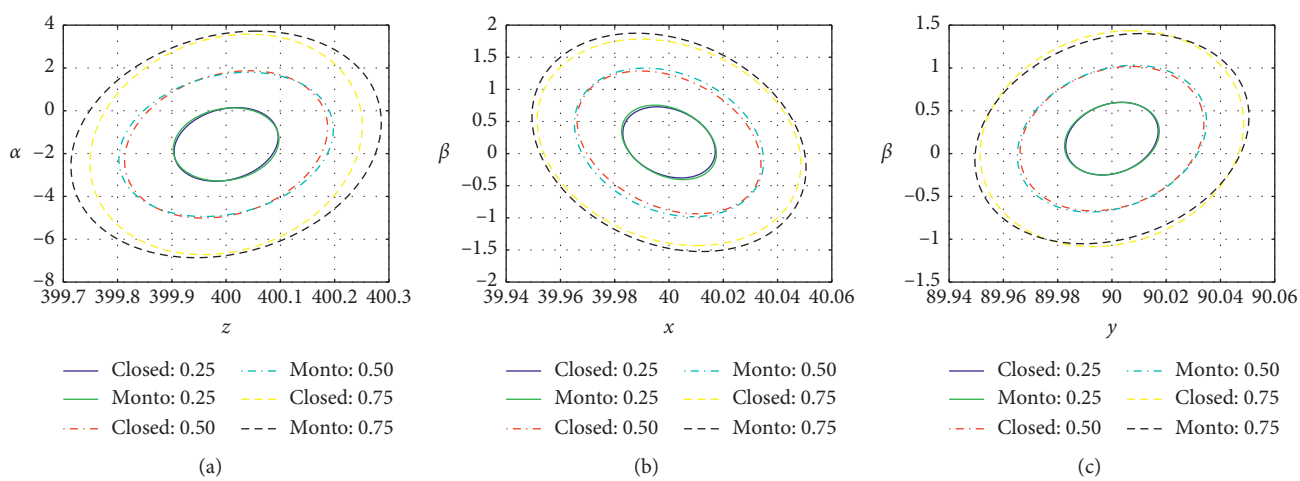


Figure 4: Comparisons of real and estimated covariance (among translations and rotations).

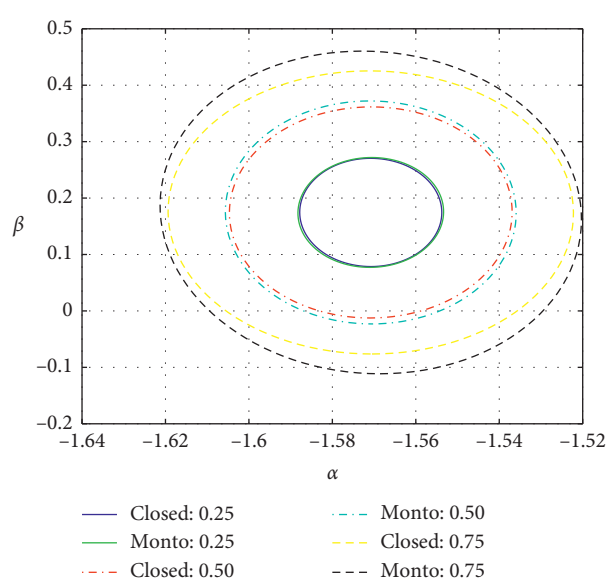


Figure 5: Comparisons of real and estimated covariance (among rotations).

$f = 20$ mm, each circular feature with the radius $R = 100$ mm, and the appearance of the object is dynamically modified by changing their orientation angles and location coordinates. By using equation (6), 500 numbers of iterations are performed for the test image, and we estimate the pose covariance matrix for each iteration by using equation (11). Figure 6 presents the projection of 5D ellipsoid corresponding to the estimation pose covariance matrix, and the confidence in the ellipse is 95%. In Figure 6, the estimated pose covariance with the proposed algorithm corresponding to each iteration is indicated with 'iteration NO.' We can see that the area of the ellipse decreases gradually as the number of iterations increases. The result indicate that the accuracy of pose estimation is improving, and estimated pose variable values are closed to actual pose $\xi = [40, 90, 400, -(\pi/2), (\pi/18)]^T$.

Figure 7 presents the estimated pose value distribution curve in each pose variable. In Figure 7, the estimated pose value distribution curve corresponding to each iteration is indicated with 'iteration NO.' We can see that the peak value of the each pose variable distribution curve is closed to the actual value gradually as the number of iterations increases, and the distribution is thin and higher. The trend is similar to the trend of Figure 6, and the result indicates the closed-form solution for estimating the covariance matrix of 5D circular feature's pose variables with good and fast numerical solution.

5.2. Good Imaging Points Selection for Circular Pose Estimation vs. Other Method. The results obtained with the proposed algorithm for 3D pose estimation of a moving object with circular feature from monocular scenes are presented and discussed in this section. These results are characterized in terms of accuracy of pose estimation when processing synthetic image sequences. All these synthetic image sequences are rendered using 3 ds Max software, each frame of an input sequence is composed of 3D object with circular feature that follows an unknown pose trajectory, it is embedded into a disjoint background, and the whole frame is degraded with additive noise. Each test sequence is composed by $\mathbf{m}$ scene frames consisting of monochrome image, and the appearance of the object during scene frames is dynamically modified by changing their orientation angles and location coordinates.

The experiments are organized as follows. First, we explore the effectiveness of the variety of matrix-revealing metrics Max-Trace, Max-logDet [30], and Max-MinEigenValue [31] for guiding good 2D imaging point selection. Second, a performance comparison of the proposed algorithm with respect to existing algorithms is presented and discussed by processing synthetic image sequences.

The accuracy of location estimation of the object is determined by the location root mean square error (rmse_L) given by

$$\text{rmse}_L = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \boldsymbol{\xi}_L^i - \widehat{\boldsymbol{\xi}}_L^{i2}}, \tag{18}$$

where ξ_L and $\widehat{\xi}_L$ are the true and estimated coordinates of the object in the scene, respectively, given in millimeters. Moreover, the orientation root mean square error (rmse_O) is given by

$$\text{rmse}_O = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \boldsymbol{\xi}_O^i - \widehat{\boldsymbol{\xi}}_O^{i2}}, \tag{19}$$

where ξ_o and $\widehat{\xi}_o$ are the true and estimated the surface normal vector of the object with respect of the observer, respectively, given in millimeters.

The performance of the tested algorithms is quantified in terms of percentages of normalized absolute errors (NAEs), between the real ξ and estimated $\widehat{\xi}$ pose parameters as follows [32]:

$$\text{NAE} = \frac{\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}}{\boldsymbol{\xi}} \times 100. \tag{20}$$

The percentages of normalized absolute errors of location estimation of object are denoted by NAE_L, and the percentages of normalized absolute errors of orientation estimation of the object are denoted by NAE_O; both are computed with equation (20).

Feature selection has been extensively studied in the fields of VSLAM [28, 31], for which several equivalent metrics exist to score the subset selection process. They are listed in Table 1. To identify the matrix-revealing metrics Max-Trace can guide good 2D imaging point selection, a simulation of least square pose optimization was carried out by processing synthetic image sequences. The evaluation scenario is depicted in Figure 8. Here, three matrix-revealing metrics in Table 1 are employed to select 2D imaging points set, and then the selected good imaging points are imputed the pose optimization system.

The test sequence is composed by 90 scene frames consisting of monochrome image with 256×256 pixels, and the effective focal length was $f = 20$ mm, each circular feature with the radius $R = 100$ mm, and the appearance of the object during scene frames is dynamically modified by a small random pose transform to 3D object. The 2D projection of 3D circular feature is detected by processing sequences of synthetic images corrupted with zero-mean additive Gaussian noise with the variance $\sigma_n^2 = 2\%$ and perfectly matched with known 3D circular feature, and the number of the detected 2D imaging points in each image exceeds 90. An optimizer equation (6) estimates the random pose from the matches. Subset size ranging from 40 to 85 is tested.

2D imaging point selection occurs prior to pose optimization so that only a subset of selected imaging point is sent to the optimizer. Each of the matrix-revealing metrics listed in Table 1 is tested. The simulation results are presented in Figure 9. For reference, we also plot the simulation results with 3σ subset selection (outlier rejection) and with all features available (All).

From Figure 9, the Max-Trace metric has the best overall performance. It more quickly approaches the baseline error

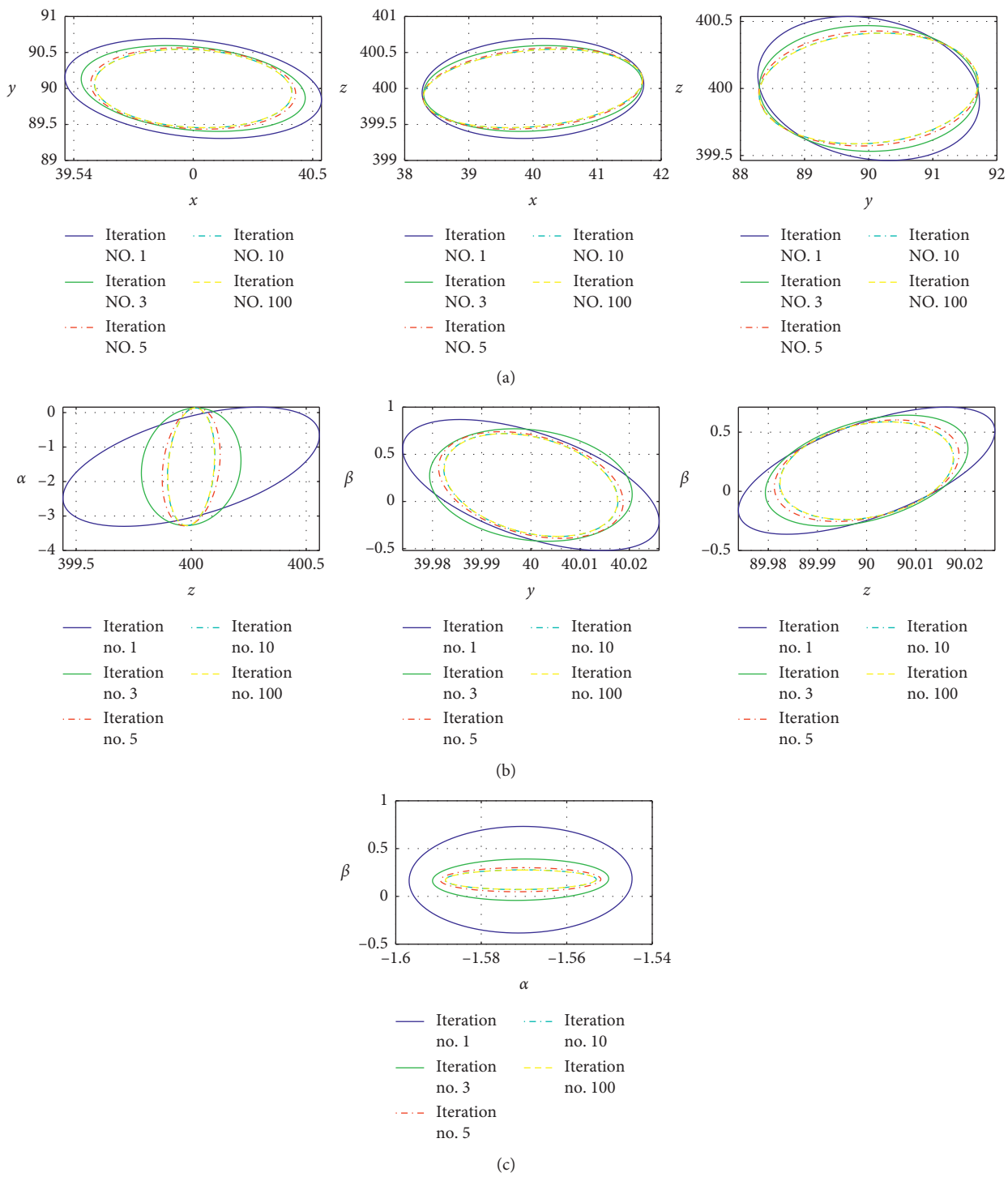


Figure 6: Estimated covariance results using the proposed closed-form solution: among translations (upper row) (a), among translations and rotations (middle row) (b), and among rotations (bottom row) (c).

(All). The orientation error goes higher the All baseline, while the locational error is significantly lower the baseline, once the subset size exceeds 55. The results point to the value of Max-Trace good features selection.

The accuracy improvement of good imaging point selection using Max-Trace is further demonstrated by comparing against other existing similar algorithms for circular feature's pose estimation using monocular images. The performance of the tested algorithms is quantified in terms of root mean square error rmse_L, between the real ξ_L and estimated $\widehat{\xi}_L$ location parameter as in equation (18). We compared the following methods:

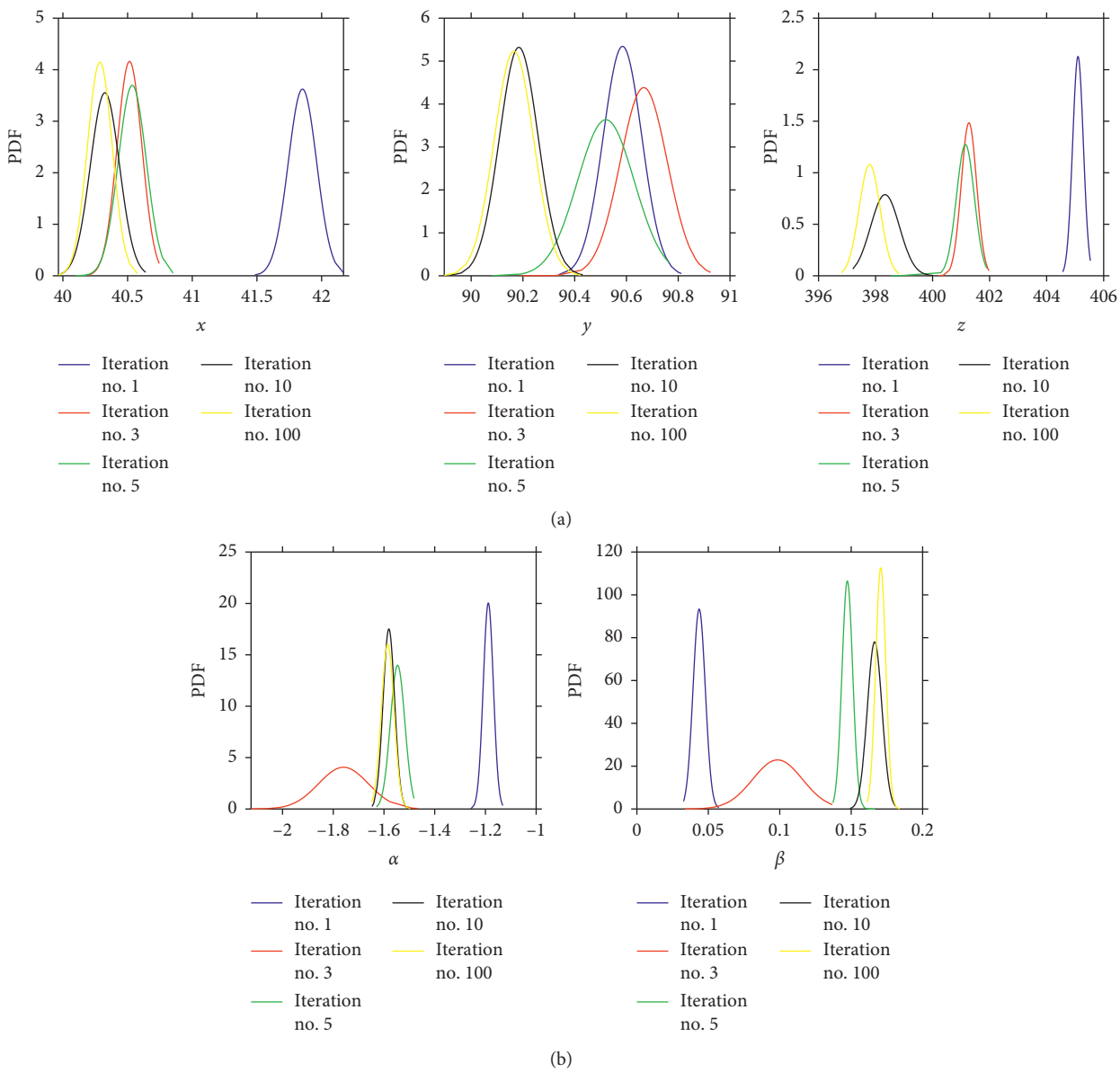


Figure 7: One dimensional probability distribution: translations (upper row) (a) and rotations (bottom row) (b).

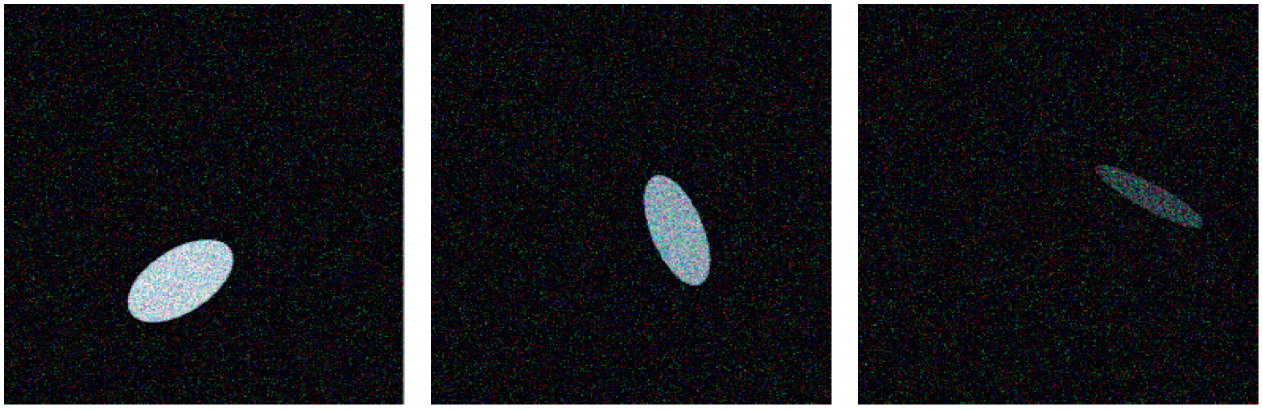
Table 1: Commonly used matrix-revealing metrics [28].

Max – Trace	Trace $\mathrm{Tr}(Q) = \sum_1^k Q_{ii}$ is max
Max – MinEigenValue	Min. eigenvalue $\lambda_k(Q)$ is max
Max – log Det	Log. of determinant $\log \mathrm{Det}(Q)$ is max

(1) The algorithm $(\cdot)^{[i]}$ in [25]: The ellipse detection [16] is also the key issue of approaches that use the 2D ellipse parameters to solve circular pose estimation problem. The 2D ellipse parameters were fitted using the least squares method.

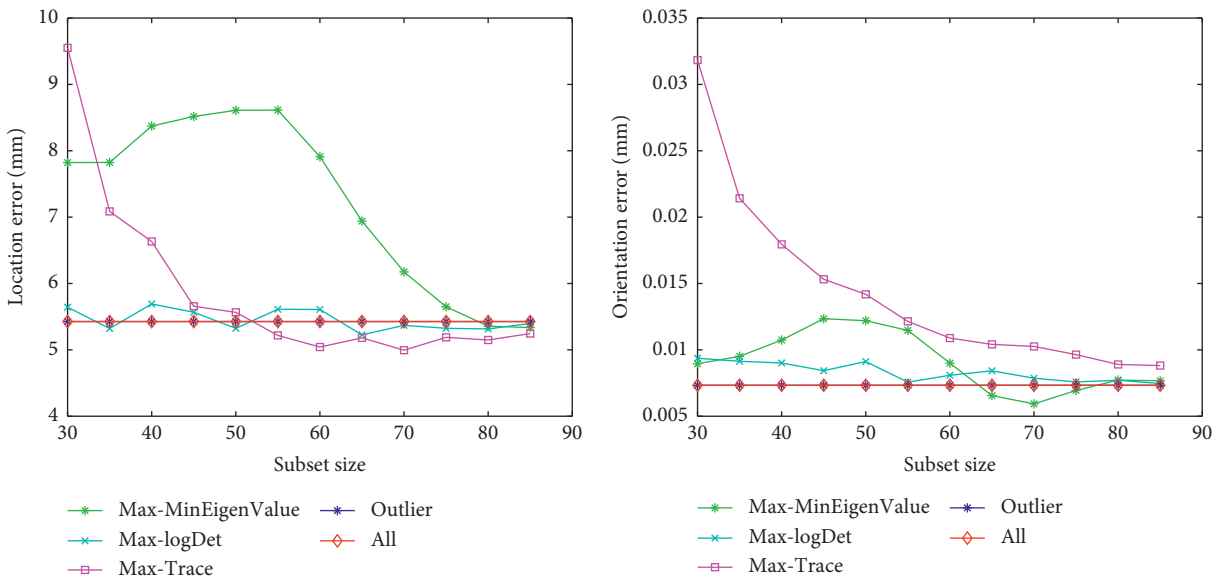
(2) The algorithm $(\cdot)^{[ii]}$ in [1]: The external feature is given, such as another circular, new points, or lines. The 2D ellipse parameters were given by the ellipse detection to solve initial estimated solution. A general frame to fuse circulars and points including all situations such as one circle one point, two or more circles, and other situations is addressed to solve the duality problem in particular cases. And then, a novel unified reprojection error for circles and points is defined to determine the optimal pose solution.

(3) The algorithm $(\cdot)^{[iii]}$: The input 2D imaging point set contains some outliers, which affect the performance of pose estimation when included. Due to the lack of explicit outlier rejection in $(\cdot)^{[i]}$, we add an outlier rejection module to the $(\cdot)^{[i]}$ pose optimization. The $\mathrm{Dis}_p \sim N(\mu_d, \sigma_d^2) \sim N(\mu, \sigma^2)$ distribution can be obtained from the distance Dis_p point-to-contour between 2D imaging point and the detected 2D ellipse. The imaging point with $\mathrm{Dis}_p \in [\mu_p - 3 \cdot \sigma_p^2, \mu + 3 \cdot \sigma_p^2]$ is preserved and is sent to pose estimation



(a) (b) (c)

Figure 8: Synthetic image sequences degraded with additive noise with the variance: $\sigma_n^2 = 2\%$.



(a) (b)

Figure 9: Simulation results of least square pose optimization: (a) first column: RMS of location error and (b) last column: RMS of orientation error.

system $(\cdot)^{[i]}$. Such an implementation of outlier rejection is far from efficient, but it will kick out most of the outliers.

(4) The algorithm $(\cdot)^{[iv]}$: The proposed Max-Trace algorithm is conducted on the whole set of input 2D imaging points directly, and then the number of 2D imaging points used is fixed (75 points per frame). The selected 2D imaging points are sent to the optimizer equation (6). Levenberg–Marquardt is used for optimizing circular feature's pose.

(5) The algorithm $(\cdot)^{[v]}$: A candidate pool of all 2D imaging points is selected using 3σ, from which the good imaging point subset is further extracted based on the proposed Max-Trace algorithm. As many as 75 imaging points can be used to optimize the pose per frame.

(6) The algorithm $(\cdot)^{[vi]}$: The algorithm $(\cdot)^{[vi]}$ presents a method for object 3D pose estimation by processing sequences of monocular images of the dynamics target. The algorithm is based on a hybrid methodology combining the good 2D imaging points selecting and unscented Kalman filter [33, 34] for dynamically adaptive processing. The good imaging point subset is extracted using 3σ and the proposed Max-Trace; these imaging points can be used to optimize the observed for object. See Appendix B for detailed of the objective function and Algorithm 2 summarizes the UKF-based 2D-3D object pose estimation algorithm.

Table 2: The mean of the location parameter absolute errors of six algorithms.

Seq.	Normalized absolute error (%)						
	$T^{[i]}_{.\text{avg}}$	$T^{[ii]}_{.\text{avg}}$	$T^{[iii]}_{.\text{avg}}$	$T^{[iv]}_{.\text{avg}}$	$T^{[v]}_{.\text{avg}}$	$T^{[vi]}_{.\text{avg}}$	Perf. gain (%)
1	0.89	0.88	0.87	0.87	**0.84**	0.39	5.60
2	0.63	0.92	0.62	0.60	**0.60**	0.47	4.80
3	0.66	1.58	0.66	**0.63**	0.66	0.54	4.50
4	0.86	0.86	0.88	0.79	**0.75**	0.62	12.8
5	1.42	1.40	0.99	0.97	**0.96**	1.16	31.40
6	**0.76**	0.76	0.77	0.85	0.82	0.77	−7.90

Initialize: the state ξ^0 is a random vector with known mean $\mu^{\mathbf{0}}_{\xi} = E[\xi^0]$ and covariance $\mathbf{P}^0_{\xi} = E[(\xi^0 - \mu^{\xi}_0)(\xi^0 - \mu^{\xi}_0)^T]$, $\mathbf{v}_k$ indicates the process noise;

Set selection of sigma points:

(1) A set of $2L+1$ sigma points: $\chi^{k-1} = \left[\xi^{k-1}\ \xi^{k-1} + \sqrt{(L+\lambda)\mathbf{P}^{k-1}_{\xi}}\ \xi^{k-1} - \sqrt{(L+\lambda)\mathbf{P}^{k-1}_{\xi}}\right]$, where L is the dimension of the state vector;

(2) The associated weights of each sigma points: $\omega^{\xi} = [(1/L+\lambda), (1/2(L+\lambda)), \ldots, (1/2(L+\lambda))]$, $\omega^{\mathbf{P}} = [(1/L+\lambda) + (1-\alpha^2_p+\beta_p), (1/2(L+\lambda)), \ldots, (1/2(L+\lambda))]$, and the weights ω^{ξ} and $\omega^{\mathbf{P}}$ are used to compute the mean and covariance, respectively. There are some suggestions for the values of some parameters: $\alpha_p \in (0,1]$, $\beta_p = 2$, $\lambda = \alpha^2_p(L+\kappa) - L, \kappa \geq 0$,

Prediction process: $\widehat{\xi}^k = \sum_{j=1}^{2L+1}\omega^{\xi}_j\widehat{\chi}^k_{[j]}$, $\widehat{\mathbf{P}}^k_{\xi} = \sum_{j=1}^{2L+1}\omega^{\mathbf{P}}_j(\widehat{\chi}^k_{[j]} - \widehat{\xi}^k)(\widehat{\chi}^k_j - \widehat{\xi}^k)^T + E(\mathbf{v}_k\mathbf{v}^T_k)$, where the index $[j]$ denotes the number of columns in the matrix $\widehat{\chi}^k$;

Resampling of sigma points: $\widehat{\chi}^k_* = \left[\widehat{\xi}^k\widehat{\xi}^k + \sqrt{(L+\lambda)\widehat{\mathbf{P}}^k_{\xi}}\ \widehat{\xi}^k - \sqrt{(L+\lambda)\widehat{\mathbf{P}}^k_{\xi}}\right]$;

Observation prediction: $\widehat{\mathbf{z}}^k = [\widehat{\Phi}(\widehat{\chi}^k_{[1]})\widehat{\Phi}(\widehat{\chi}^k_{*[2]})\ldots\widetilde{\Phi}(\widehat{\chi}^k_{*[2L+1]})]$, $\mu^k_{\mathbf{z}} = \sum_{j=1}^{2L+1}\omega^{\xi}_j\widehat{\mathbf{z}}^k_{[j]}\mathbf{P}^k_{\mathbf{z}} = \sum_{j=1}^{2L+1}\omega^{\mathbf{P}}_j(\widehat{\mathbf{z}}^k_{[j]} - \mu^k_{\mathbf{z}})(\widehat{\mathbf{z}}^k_{[j]} - \mu^k_{\mathbf{z}})^T + E(\gamma_k\gamma^T_k)$, where γ_k indicates the observation noise;

The cross covariance: $\mathbf{P}^k_{\xi\mathbf{z}} = \sum_{j=1}^{2L+1}\omega^{\mathbf{P}}_j(\widehat{\chi}^k_{[j]} - \widehat{\xi}^k)(\widehat{\mathbf{z}}^k_{[j]} - \mu^k_{\mathbf{z}})^T$;

The gain $\mathbf{K^k}$ is given by: $K^k = \mathbf{P}^k_{\xi\mathbf{z}}\mathbf{P}^{k-1}_{\mathbf{z}}$;

Update: the posterior state vector and covariance are updated after the following formula:

$\xi^k = \widehat{\xi}^k + K^k[\mathbf{z}^k - \mu^k_{\mathbf{z}}]$, $\mathbf{P}^k_{\xi} = \widehat{\mathbf{P}}^k_{\xi} - K^k\mathbf{P}^k_{\mathbf{z}}(K^k)^T$, where $\mathbf{z}^k$ is the arrived observation in the step time k.

Algorithm 2: UKF for object 2D-3D pose estimation using circular feature.

A common character of these mentioned algorithms $(\cdot)^{[i]} - (\cdot)^{[v]}$ is that the object 3D pose is obtained from a single monocular image. The means of the NAE_L of the five algorithms are shown in Table 2, which are marked as $T^{(\cdot)}_{.\text{avg}}$. For each approach type, the lowest percentages of normalized absolute errors per sequence are in bold. On almost all sequences, either $(\cdot)^{[iv]}$ or $(\cdot)^{[v]}$ has the lowest NAE_L. On synthetics sequence such as Seq.1, Seq.2, Seq.3, Seq.4, and Seq.5, the proposed Max-Trace algorithm reduces the location error significantly. The exception is Seq.6 where the proposed Max-Trace algorithm results in a higher than the lowest one (generated by $(\cdot)^{[ii]}$). Overall, the proposed Max-Trace approach reduces pose estimation error on several sequences by exploiting the structural and motion information. Integrating Max-Trace with outlier rejection (i.e., $(\cdot)^{[iii]}$) further improves performance.

Now, compare the algorithm $(\cdot)^{[v]}$ with the two algorithms $(\cdot)^{[i]}$ and $(\cdot)^{[ii]}$. On synthetics sequences, the algorithm $(\cdot)^{[v]}$ clearly leads to lower location error. Meanwhile on other sequence, the location error of the algorithm $(\cdot)^{[v]}$ is the same as baseline or slightly worse. The performance gains on the sequence which semimajor axis a and semiminor axis b of the detected ellipse are satisfied to $(a/b) \gg 1$ far outweigh the performance loss on the sequence with $(a/b) \approx 1$, as presented in the 8th column of Table 2.

Furthermore, compare the algorithm $(\cdot)^{[vi]}$ with the algorithm $(\cdot)^{[v]}$. On almost all synthetics sequences, $(\cdot)^{[vi]}$ clearly leads to lower location error. This is because the algorithm $(\cdot)^{[vi]}$ consists of two advantages. On the one hand, the algorithm takes into account the kinematics of the target in the scene, and the temporal information among the frames is used to improve the 3D pose estimation of the rigid object. On the other hand, the good imaging point subset is extracted, and these imaging points can be used to optimize the observed for object. On synthetics sequences Seq5, the location error of the algorithm $(\cdot)^{[vi]}$ is slightly worse, may be the suggestion value of the parameter is not fit the seq5. It needs to set more appropriate value of parameter for different dynamic object parameters. The filter such as CKF [35], UKF, and PF [32] can be used for the object 2D-3D pose estimation system, and they can yield accurate results in 3D pose recognition from sequences of monocular images of the dynamics target.

6. Conclusions

The paper proposes a closed-form solution for estimating the accuracy of the circular feature's pose. The solution of circular feature's pose parameters is obtained using all detected 2D imaging points corresponding to the circular

feature in 3D. The algorithm presented the idea that the relation from the 2D imaging point to circular feature's pose is represented by an implicit function, and then the implicit function combined with the extraction error of 2D imaging point computes the accuracy of the pose variables, which is the covariance matrix of circular feature's pose. The main contributions of the present work can be summarized as follows: the algorithm defines a representation method for 5D circular pose parameters, and then the representation combines with the projection model of the 3D object circular feature to yield 2D virtual ellipse contour. An implicit function from 2D imaging point to the 5D circular feature's pose variable is defined using the points-to-contour distance, and then Gauss–Newton theorem is employed to compute the Jacobi matrix of the function with respect to such points; after that the Jacobi matrix is combined with 2D imaging point coordinates to estimate the accuracy of circular feature's pose variables. The covariance matrix of circular feature's 5D pose variables is studied, and the metric is introduced to guide the 2D imaging point selection. Integrating 2D imaging point selection into the pose estimation system leads to accuracy improvements of location parameter.

Appendix

A. The 2D Ellipse Equation Parameter Model with Pose Parameter

Let $(X, Y, Z)^T$ and $(l, m, n)^T$ denote the coordinate center and surface normal vector of the circular feature in 3D with radius R. Projection of the circular feature in 3D into the image plane to yield the 2-D ellipse curve $\Phi(\xi, (x, y))$ is represented in the following equation:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0\left(B^2 - 4AC = 1\right), \tag{A.1}$$

where we shall represent the parameters $[A, B, C, D, E, F]^T$ by the location $G = (X, Y, Z)^T$ and the 3D orientation vector $(l, m, n)^T$.

To find the 2-D ellipse curve, we will first form a cone having the projection center as vertex and which joins the vertex to every point on the circular whose center position is $G = (X, Y, Z)^T$ and surface normal vector is $(l, m, n)^T$ and intersect the cone with the image plane $Z = f$.

In order to find the equation of the cone S_{cone}, we need to construct the equation of the base circular G and the line that joins the vertex to the point on the circular G. The equation of the base circular G is obtained by intersecting the sphere S_G whose center is $G = (X, Y, Z)^T$ and radius is R with the plane π whose surface normal vector is $(l, m, n)^T$ as follows:

$$\begin{cases} (X_1 - X)^2 + (Y_1 - Y)^2 + (Z_1 - Z)^2 = R^2, \\ l(X_1 - X) + m(Y_1 - Y) + n(Z_1 - Z) = 0, \end{cases} \tag{A.2}$$

where the point $(X_1, Y_1, Z_1) \in S_G$.

For $\forall P(X_2, Y_2, Z_2) \in S_{\text{cone}}$, the equation of line that joins the vertex to the point on the circular G and the point on the cone is represented as follows:

$$\frac{X_1 - X_2}{X_2 - 0} = \frac{Y_1 - Y_2}{Y_2 - 0} = \frac{Z_1 - Z_2}{Z_2 - 0} = t, \tag{A.3}$$

where we can obtain the coordinate of the points on circular G and it can be denoted as follows:

$$(X_G, Y_G, Z_G) = (tX_2 + X_2, tY_2 + Y_2, tZ_2 + Z_2). \tag{A.4}$$

Especially, $(X_G, Y_G, Z_G) \in \pi$ and $(X_G, Y_G, Z_G) \in S_G$, and by solving two simultaneous equations given by equations (A.1) and (A.3), the equation of cone S_{cone} can be written as follows:

$$\begin{aligned} &(mY + nZ)^2 X_2^2 - 2(mY + nZ)X_2 X(mY_2 + nZ_2) + X^2(mY_2 + nZ_2)^2 + (lX + nZ)^2 \\ &- 2(lX + nZ)Y_2 Y(lX_2 + nZ_2) + (lX_2 + nZ_2)^2 Y^2 + (lX + mY)^2 Z_2^2 - 2(lX + mY) \\ &Z_2 Z(lX_2 + mY_2) + (lX_2 + mY_2)^2 Z^2 = R^2\left(l^2 X_2^2 + m^2 Y_2^2 + n^2 Z_2^2 + 2lmX_2Y_2 + 2\ln X_2Z_2 + 2mnY_2Z_2\right). \end{aligned} \tag{A.5}$$

By replacing Z_2 with f, the parameter model of 2D ellipse curve equation with pose parameter ξ is expressed as follows:

$$\begin{cases} A = (mY + nZ)^2 + l^2Y^2 + l^2Z^2 - R^2l^2; \\ B = -2mX(mY + nZ) - 2lY(lX + nZ) + 2lmZ^2 - 2lmR^2; \\ C = m^2X^2 + (lX + nZ)^2 + m^2Z^2 - R^2m^2; \\ D = -2Xnf(mY + nZ) + 2nlfY^2 - 2flZ(lX + mY) - 2\ \ln\ fR^2; \\ E = 2mnfX^2 - 2nfY(lX + nZ) - 2mfZ(lX + mY) - 2mnfR^2; \\ F = X^2n^2f^2 + Y^2n^2f^2 + (lX + mY)^2f^2 - R^2n^2f^2. \end{cases} \tag{A.6}$$

where f is focal length of the camera. By replacing $(l, m, n)^T$ with $(\sin\beta\cos\alpha, \sin\beta\sin\alpha, \cos\beta)^T$, equation (A.5) can be written as

$$\begin{cases} A = ((\sin\beta\sin\alpha)Y + \cos\beta Z)^2 + (\sin\beta\cos\alpha)^2\left(Y^2 + Z^2 - R^2\right); \\ B = -2(\sin\beta)^2XY - 2(\sin\beta\sin\alpha)\cos\beta XZ \\ -2(\sin\beta\cos\alpha)\cos\beta YZ + 2((\sin\beta\cos\alpha)\cdot(\sin\beta\sin\alpha))\left(Z^2 - R^2\right); \\ C = (\sin\beta\sin\alpha)^2\left(X^2 + Z^2 - R^2\right)((\sin\beta\cos\alpha)X + \cos\beta Z)^2; \\ D = -2X\cos\beta f((\sin\beta\sin\alpha)Y + \cos\beta Z) + 2\cos\beta(\sin\beta\cos\alpha)fY^2 \\ -2f(\sin\beta\cos\alpha)\left(Z((\sin\beta\cos\alpha)X + (\sin\beta\sin\alpha)Y) - \cos\beta R^2\right); \\ E = 2(\sin\beta\sin\alpha)\cos\beta R^2 - 2\cos\beta fY((\sin\beta\cos\alpha)X + \cos\beta Z) \\ -2(\sin\beta\sin\alpha)f\left(Z((\sin\beta\cos\alpha)X + (\sin\beta\sin\alpha)Y) - \cos\beta R^2\right); \\ F = X^2(\cos\beta)^2 + Y^2(\cos\beta)^2 + ((\sin\beta\cos\alpha)X + (\sin\beta\sin\alpha)Y)^2f^2 - R^2(\cos\beta)^2f^2. \end{cases} \tag{A.7}$$

B. Objective Function for 2D-3D Object Pose Estimation Using Circular Feature

The ultimate goal of object 2D-3D pose estimation is to identify 3D pose of an object of interest from a 2D image or image sequence.

Given the observed image $\mathbf{z}$ (possibly consisting of several image sequences), we shall give the objective function using the maximum posterior probability estimation (MAP) for 2D-3D object pose estimation. By maximizing the conditional probability distribution of the pose ξ, the objective function for 2D-3D object pose estimation is defined as follows:

$$\arg\max_{\boldsymbol{\xi}}\left\{p\left(\frac{\boldsymbol{\xi}}{\mathbf{z}}\right)\right\}, \tag{B.1}$$

where $p(\xi/\mathbf{z})$ can be expanded as the following equation:

$$p\left(\frac{\boldsymbol{\xi}}{\mathbf{z}}\right) = \frac{1}{p(\mathbf{z})}p\left(\frac{\mathbf{z}}{\boldsymbol{\xi}}\right)\cdot p(\boldsymbol{\xi}), \tag{B.2}$$

where $p(\mathbf{z}/\xi)$ is the likelihood of the arrived observation $\mathbf{z}$ and $p(\xi)$ represents the prior information of pose ξ of the circular feature in 3D.

UKF, CKF, and PF can be employed in 2D-3D object pose estimation problems, where the overall objective is to estimate pose of a moving object from a collection of samples arriving sequentially. The UKF can be seen as population-based Monte Carlo algorithms, in which the distribution of the pose space is approximated using a minimal set of carefully chosen sample points, called sigma points. Each sigma point is composed by a single pose and has an associated weighting coefficient. Let χ^{k-1} be a set of $2L + 1$ sigma points in time step $k - 1$ (where L is the dimension of the state space), containing information about the pose ξ_j^{k-1} and their associated weights ω_j^{ξ} as follows:

$$\chi^{(k-1)} = \left\{\left(\boldsymbol{\xi}_j^{\xi}, \omega_j^{\xi}\right), \quad j = 1, 2, \ldots, 2L + 1\right\}. \tag{B.3}$$

For the 2D-3D object pose estimation using circular feature, the sigma points $\xi_j^{(k-1)}$ are given by the location coordinates '$X_j^{(k-1)}, Y_j^{(k-1)}, Z_j^{(k-1)}$', orientation angles '$\alpha_j^{(k-1)}, \beta_j^{(k-1)}$', and their velocity vector as follows:

$$\xi_j^{k-1} = \left[X_i^{k-1}, \dot{X}_j^{k-1}, Y_j^{k-1}, \dot{Y}_j^{k-1}, Z_j^{k-1}, \dot{Z}_j^{k-1}, \alpha_j^{k-1}, \dot{\alpha}_j^{k-1}, \beta_j^{k-1}, \dot{\beta}_j^{k-1}\right]. \tag{B.4}$$

Note that a sigma point described in equation (4) represents a single pose state in the time step $k - 1$.

To estimate the pose of the object with these sigma points, each sigma point is propagated through the nonlinear process model. The transformed points are used to compute the mean and covariance of the prediction value $\hat{\xi}_j^k$. We propagate then the sigma points through the nonlinear observation model equation (2) as follows:

$$\bar{\mathbf{z}}_j^k = \hat{\Phi}\left(\hat{\boldsymbol{\xi}}_j^k\right), \tag{B.5}$$

with the resulted transformed observations, their mean $\mu_{\mathbf{z}}^k$, covariance $\mathbf{P}_{\mathbf{z}}^k$, and the cross covariance $\mathbf{P}_{\xi\mathbf{z}}^k$ are computed, respectively. We combine the information obtained in the prediction step with the new observation measured $\mathbf{z}^k$, and the gain $\mathbf{K}^k$ is given, and then the posterior mean and covariance are updated.

Algorithm steps are as follows: Algorithm 2

Data Availability

The data used to support the findings of this study are available in the following website: https://github.com/licui2006060222/Database.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All the authors have participated in writing the manuscript. All authors read and approved the manuscript.

Acknowledgments

The authors would like to thank the First-Class Disciplines Foundation of Ningxia (contract no. NXYLXK2017B09) and the Major Project of North Minzu University (contract no. ZDZX201801) for supporting this work. This work was supported in part by the National Natural Science Foundation of China (no. 11961001), by the Natural Science Foundation of Ningxia (no. 2018AAC03126), and by the Open Project of the Key Laboratory of Intelligent Information and Big Data Processing of Ningxia Province (no. 2019KLBD004).

References

[1] B. Huang, Y. Sun, and Q. Zeng, "General fusion frame of circles and points in vision pose estimation," *Optik—International Journal for Light and Electron Optics*, vol. 154, pp. 47–57, 2018.

[2] J. Lee, R. Sandhu, and A. Tannenbaum, "Particle filters and occlusion handling for rigid 2D-3D pose tracking," *Computer Vision and Image Understanding*, vol. 117, no. 8, pp. 922–933, 2013.

[3] C. Meng and H. Sun, "Monocular pose measurement method based on circle and line features," in *Proceedings of the 2nd International Conference on Electronics, Network and Computer Engineering (ICENCE 2016)*, Yinchuan, China, August 2016.

[4] B. Huang, Y. Sun, Y. Zhu, Z. Xiong, and J. Liu, "Vision pose estimation from planar dual circles in a single image," *Optik*, vol. 127, no. 10, pp. 4275–4280, 2016.

[5] C. Wang, D. Chen, M. Li, and J. Gong, "Direct solution for pose estimation of single circle with detected centre," *Electronics Letters*, vol. 52, no. 21, pp. 1751–1753, 2016.

[6] S. Ghosh, R. Ray, S. R. K. Vadali, S. N. Shome, and S. Nandy, "Reliable pose estimation of underwater dock using single camera: a scene invariant approach," *Machine Vision and Applications*, vol. 27, no. 2, pp. 221–236, 2016.

[7] X. K. Miao and F. Zhu, "Monocular vision pose measurement based on docking ring component," *Acta Optica Sinica*, vol. 33, no. 4, pp. 1–9, 2013.

[8] Y. M. Hao, F. Zhu, and J. Ou, "Error analysis of P3P pose estimation," *Computer Engineering and Applications*, vol. 44, pp. 239–242, 2008.

[9] C Li, D. Chen, J. Gong, and Y. Wu, "A high-accuracy algorithm combing shape prior with algebraic detected elliptical arc distance for circle pose measurement," in *Proceedings of the 2019 IEEE International Conference on Unmanned Systems (ICUS)*, pp. 554–557, Beijing, China, October 2019.

[10] J. Huo, Z. Guiyang, J. Cui, and M. Yang, "A objective function with measuring error uncertainty weighted for pose estimation in stereo vision," *Optics and Precision Engineering*, vol. 26, no. 4, pp. 834–842, 2018.

[11] R. Hao, Z. Wang, and Y. Li, "Error analysis method for monocular vision pose measurement," *Journal of Applied Optics*, vol. 40, no. 1, pp. 79–85, 2019.

[12] J. Huo, J. Cui, and N. Yang, "Error analysis of rigid body posture measurement system based on circular feature points," in *Proceedings of the Seventh International Conference on Machine Vision (ICMV 2014)*, Milan, Italy, February 2015.

[13] J.-S. Cui, J. Huo, M. Yang, and Y.-K. Wang, "Research on the rigid body posture measurement using monocular vision by coplanar feature points," *Optik*, vol. 126, no. 24, pp. 5423–5429, 2015.

[14] R. Balasubramanian, S. Das, S. Udayabaskaran, and K. Swaminathan, "Quantization error in stereo imaging systems," *International Journal of Computer Mathematics*, vol. 79, pp. 671–691, 2002.

[15] Z. Sun, Z. Xu, Y. Wang, and C. Zhou, "Control and compensation of perspective projection error analysis in machine vision measurement," *Computer Engineering and Applications*, vol. 54, pp. 266–270, 2018.

[16] F. Zhu, H. Yingming, J. Zhou, and D. Tan, "The analysis of collaborative object attitude's effect on visual pose's accuracy," *Chinese Journal of Scientific Instrument*, vol. 28, pp. 130–134, 2007.

[17] L.-J. Zhao, E.-H. Liu, W.-M Zhang, and R.-.J Zhao, "Analysis of position estimation precision by cooperative target with three feature points," *Optics and Precision Engineering*, vol. 54, pp. 266–270, 2018.

[18] C. C. Yang, M. M. Marefat, and F. W. Ciarallo, "Error analysis and planning accuracy for dimensional measurement in active vision inspection," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 3, p. 476, 1998.

[19] B. Kamgar-Parsi, Kamgar-Parsi, and Behzad, "Evaluation of quantization error in computer vision," *Physics-Based Vision: Principles and Practice: Radiometry*, vol. 1, p. 293, 1993.

[20] J. Zhou, F. Zhu, Y. M. Hao, and L. He, "Effect of camera parameters error on the precision of visual pose estimation," in *Proceedings of the ACCV 07 Workshop on Multi-Dimensional and Multi-View Image Processing*, pp. 183–187, Tokyo, Japan, November 2007.

[21] S. Min, H. Rixin, and W. Daojun, "Precision analysis to 3D reconstruction from image sequences," in *Proceedings of the the 5th ISPRSWorkshop on DMGISs*, Xinjiang, China, August 2007.

[22] B. Guendouz, C. Eswaran, and S. V. Muniandy, "Error propagation and accurate calibration for camera model," in *Proceedings of the 2006 IEEE International Conference on*

Engineering of Intelligent Systems, pp. 1–5, Islamabad, Pakistan, April 2006.

[23] A. Censi, "An accurate closed-form estimate of ICP's covariance," in *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, pp. 3167–3172, Rome, Italy, April 2007.

[24] X. Li, W. Guo, M. Li, C. Chen, and L. Sun, "A closed-form solution for estimating the accuracy of depth camera's relative pose," *Robot*, vol. 36, pp. 194–202, 2014.

[25] Y. C. Shiu and S. Ahmad, "3D location of circular and spherical features by monocular model-based vision," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 576–581, Cambridge, MA, USA, November 1989.

[26] C. Li, D. Chen, J. Gong, and Y. Wu, "Elliptic shape prior for object 2D-3D pose estimation using circular feature," *EURASIP Journal on Advances in Signal Processing*, vol. 34, 2020.

[27] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Proceedings of the Dundee Conference on Numerical Analysis*, pp. 105–116, Dundee, UK, June 1977.

[28] Y. Zhao and P. A. Vela, "Good feature selection for least squares pose optimization in VO/VSLAM," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1183–1189, Madrid, Spain, October 2018.

[29] B. Mirzasoleiman, A. Badanidiyuru, A. Karbasi, J. Vondrak, and A. Krause, "Lazier than lazy greedy," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1–7, Austin, TX, USA, January 2015.

[30] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: leveraging submodularity," in *Proceedings of the 49th IEEE Conference on Decision and Control (CDC)*, pp. 2572–2577, Atlanta, GA, USA, December 2010.

[31] L. Carlone and S. Karaman, "Attention and anticipation in fast visualinertial navigation," in *Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3886–3893, Singapore, June 2017.

[32] K. Picos, V. H. Diaz-Ramirez, A. Sanz Montemayor, J. Jose Pantrigo, and V. Kober, "Three-dimensional pose tracking by image correlation and particle filtering," *Optical Engineering*, vol. 57, no. 7, 2018.

[33] G. Hu, B. Gao, Y. Zhong, and C. Gu, "Unscented kalman filter with process noise covariance estimation for vehicular INS/GPS integration system," *Information Fusion*, vol. 64, pp. 194–204, 2020.

[34] G. Hu, W. Wang, Y. Zhong, B. Gao, and C. Gu, "A new direct filtering approach to INS/GNSS integration," *Aerospace Science and Technology*, vol. 77, pp. 755–764, 2018.

[35] B. Gao, G. Hu, Y. Zhong, and X. Zhu, "Cubature rule-based distributed optimal fusion with identification and prediction of kinematic model error for integrated UAV navigation," *Aerospace Science and Technology*, vol. 109, Article ID 106447, 2020.