

RESEARCH

Open Access



# Elliptic shape prior for object 2D-3D pose estimation using circular feature

Cui Li<sup>1</sup>, Derong Chen<sup>1</sup>, Jiulu Gong<sup>1\*</sup>  and Yangyu Wu<sup>1,2</sup>

\*Correspondence:

lujiugong@bit.edu.cn

<sup>1</sup>The National Laboratory for  
Mechatronic and Control, Beijing  
Institute of Technology, Beijing,  
100081 China  
Full list of author information is  
available at the end of the article

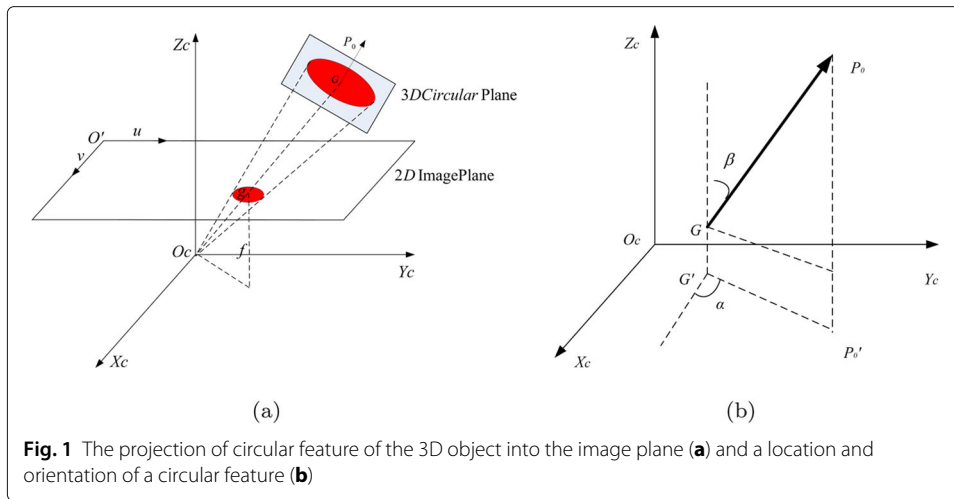
## Abstract

Many objects in real world have circular feature. It is a difficult task to obtain the 2D-3D pose estimation using circular feature in challenging scenarios. This paper proposes a method to incorporate elliptic shape prior for object pose estimation using a level set method. The relationship between the projection of the circular feature of a 3D object and the signed distance function corresponding to it is analyzed to yield a 2D elliptic shape prior. The method employs the combination of the grayscale histogram, the intensities of edge, and the smoothness distribution as main image feature descriptors that define the image statistical measure model. The elliptic shape prior combined with the image statistical measure energy model drives the elliptic shape contour to the projection of the circular feature of the 3D object with the current pose into the image plane. These works effectively reduce the impacts of the challenging scenarios on the pose estimate results. In addition, the method utilizes particle filters that take into account the motion dynamics of the object among scene frames, and this work provides the robust method for object 2D-3D pose estimation using circular feature in a challenging environment. Various numerical experiments are illustrated to show the performance and advantages of the proposed method.

**Keywords:** Pose estimation, Circular feature, Elliptic shape prior, Image statistical property

## 1 Introduction

Pose estimation is an essential step in many machine vision and photogrammetric applications; 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 Fig. 1b) [3–5]. These methods successfully applied to pose estimation of underwater dock [6]. However, since these methods rely on local features, the resulting solutions may yield unsatisfactory results in a challenging environment, such as the higher noise condition, complex background, and partial occlusions. To overcome this, the paper proposes an algorithm to combine the elliptic shape prior with the image data for object 2D-3D pose estimation using circular feature. However, before doing so, let us revisit several contributions related to the proposed method.



In order to perform accurate 3D pose estimation in a challenging environment, two approaches have been developed for the design of object 2D-3D pose estimation algorithms. A common approach was implemented for 3D pose estimation with shape prior information of a 3D model of the object and image statistics. It did not involve edge detection or image contour extraction. The shape of objects constrain the contour evolution to adopt familiar shapes to make up for poor segmentation and pose estimation results obtained in the presence of noise, clutter, or occlusion or when the statistics of the object and background are difficult to distinguish [2, 7, 8]. In [2, 7] and [8], projection of the 3D surface of an object to yield a 2D shape prior helps in a top-down manner to improve the extraction of the contour. The 3D pose is evolved to maximize the image statistical measure of discrepancy between its interior and exterior regions.

Another approach based on template matching filters has been proposed to solve 3D pose of an object: by generating a set of synthetic images of 3D model of the object as reference templates, a high matching score when the input and reference images are very similar. Given a known 3D model of target, this approach estimates its locations and orientation parameters by maximizing frequency response between the input and the current reference images [9, 10]. The input image is globally processed instead of processing only local feature, and it yields high accuracy of 3D pose estimation in comparison with the existing approaches based on segmentation in a challenging environment.

Although these approaches perform exceptionally well for many cases, the drawback of this algorithm is the input images are processed independently. So, they do not exploit the underlying dynamics inherent in a pose estimation task and they cannot handle erratic movements. In [2] and [11], not only to utilize both framework above but also to overcome their disadvantages, they extend them by incorporating a particle filter to exploit the underlying dynamics of system; this improvement provides the robust method for object 3D pose estimation in the presence of additive noise, complex background, and occlusion. Both methods rely on the 3D model of the object to obtain prior information and construct a 6D pose parameter model to estimate the object pose. For object 2D-3D pose estimation using circular feature, it is a difficult task to have representations for elliptic shape prior with good and fast numerical solutions using 5D pose parameters.

In this work, we propose an algorithm for object 2D-3D pose estimation using circular feature by exploiting elliptic shape constraint and image statistical measure. Given a 3D object with circular feature, the proposed algorithm estimates the pose of a moving object using a bank of elliptic shape templates, which is dynamically adapted by particle filters. By knowing that a spatial circular can appear in any pose configuration in the scene, the pose estimation problem can be stated as a search problem, in which the goal is to find the pose parameters of object by processing captured images of a scene. By generating a set of virtual elliptic shape contours as reference templates of the spatial circular feature of the 3D object, a comparison with the current view of the spatial circular in the scene can be performed. For this, we use particle filters as a reliable and adaptive search strategy in order to drive the elliptic shape contour to the projection of the circular feature of a 3D object with the current pose into the image. The main contributions of the present work can be summarized as follows:

- 1) The paper proposes a method to incorporate elliptic shape prior for object pose estimation using the level set method.
- 2) The relationship between the projection of the circular feature of the 3D object and the signed distance function corresponding to it is analyzed to yield a 2D elliptic shape prior.
- 3) The combination of the grayscale histogram, the intensities of edge, and the smoothness distribution as main image feature descriptors defines the image statistical measure model. The elliptic shape prior combined with the image statistical measure energy model drives the elliptic shape contour to the projection of the circular feature of the 3D object with the current pose into the image plane.
- 4) The proposed algorithm yields high accuracy of object 2D-3D pose estimation using circular feature by processing a sequence of 2D monocular images degraded with additive noise, complex background, and partially occlusion.

The paper is organized as follows. In Section 2, we define the representation method of the 5D circle pose parameters and give the objective function using the maximum posterior probability estimation (MAP) for 2D-3D object pose estimation using circular feature. In addition, we briefly explain an overview of the fundamental concepts used in the proposed method, particle filters presented in [11]. Section 3 explains the proposed algorithm for 2D-3D object pose estimation using circular feature. Specifically, we discuss the representation method of elliptic shape prior and the image statistical measure energy model. Section 4 presents experimental results obtained with the proposed algorithm when processing synthetic image sequences, which are discussed and compared with those obtained by existing object 2D-3D pose estimation using circular feature. The conclusions of the present work are summarized in Section 5.

## 2 Preliminaries

### 2.1 The position and orientation of a circular feature in 3D

As shown in Fig. 1a, 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 the direction it is pointed. The image frame  $u - v$  is a 2D frame with the  $u$  and  $v$  axes parallel to the  $Y_C$  and  $X_C$  of the camera frame, respectively. The projection of the circular feature with a radius of  $R$  into the image plane is represented as elliptic  $g$ .

As shown in Fig. 1b, the position and orientation of a circular feature in 3D is 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 [12]. Examples are shown in Fig. 1b. The direction angle  $\alpha$  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 counterclockwise rotation of  $\overrightarrow{GP_0}$ . The direction angle  $\beta$  indicates the angle between  $\overrightarrow{GP_0}$  and  $O_c Z_c$  axis. In addition,  $\|\overrightarrow{GP_0}\| = 1$  and  $\overrightarrow{GP_0} = (\sin\beta\cos\alpha, \sin\beta\sin\alpha, \cos\beta)^T$ . 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  $G = (X, Y, Z)^T$ .

## 2.2 Objective function for 2D-3D object pose estimation using circular feature

In this section, we shall give the objective function using the maximum posterior probability estimation (MAP) for 2D-3D object pose estimation using circular feature.

Given the observed image  $z$  (possibly consisting of several image sequences), by maximizing the conditional probability distribution of the pose  $\xi$ , the objective function for 2D-3D object pose estimation using circular feature is defined as follows:

$$\arg \max_{\xi} p(\xi|z) \quad (1)$$

where  $p(\xi|z)$  can be expanded as Eq. 2 [13]:

$$p(\xi|z) = \frac{1}{p(z)} p(z|\xi) \cdot p(\xi) \quad (2)$$

where  $p(z|\xi)$  is the likelihood of the arrived observation  $z$  and  $p(\xi)$  represents the prior information of the spatial circle pose  $\xi$  of the circular feature in 3D pose.

Particle filtering is widely employed in object pose estimation problems, where the overall objective is to estimate pose of a moving object from a collection of samples arriving sequentially. Particle filters can be seen as population-based Monte Carlo algorithms, in which the distribution of the pose space is approximated by random measures, called particles [11]. Each particle is composed of a single pose and has an associated weighting coefficient. Let  $Q^k$  be a set of particles in time step  $k$ , containing information about the pose  $\xi_i^k$  and with an associated weight  $J_i^k$  as follows:

$$Q^k = \{q_i^k, i = 1, \dots, N\} \quad (3)$$

where  $q_i^k = \{(\xi_i^k, J_i^k)\}$  represents the current particles. For the 2D-3D object pose estimation using circular feature, the pose  $\xi_i^k$  is given by the location coordinates  $X_i^k, Y_i^k, Z_i^k$  and orientation angles  $\alpha_i^k, \beta_i^k$  as follows:

$$\xi_i^k = [X_i^k, Y_i^k, Z_i^k, \alpha_i^k, \beta_i^k]^T \quad (4)$$

Note that an object's pose described in Eq. 4 represents a single pose in the time step  $t$ . Furthermore, the weighting coefficients associated to the object's pose is determined by a fitness function:

$$L(\xi_i^k) = J_i^k \quad (5)$$

To estimate the pose of the object with particle filters, the particles  $\{q_i^k\}$  are selected according to their weights [11], determined by the fitness function of Eq. 5, in which the

particles with more probability to be selected are those that contain higher weight values. Then, the diffusion process propagates the selected particles, which favors the diversity by randomly diffusing a selected particle. The prediction process can estimate the particle behavior, given by the observation of the motion dynamics of the object. In the paper, maximum likelihood is utilized as an evaluation method to determine the weight of each particle. The likelihood function can be obtained by using a different image statistical measure.

### 3 Method description

In this section, we present the method for pose estimation of object with circular feature using a level set function under the elliptic shape prior. First, we shall discuss the representation method of elliptic shape prior by a level set function. Later, we shall incorporate such representation with the image statistical measure energy model.

#### 3.1 Projection of the 3D object circular feature to yield a 2D elliptic shape prior

To interact with the ellipse contour in the image, the circular feature in 3D has to be projected to the image plane and then the result yields a 2D virtual elliptic shape prior. Moreover, the projected shape  $\Phi(\xi)$  is assumed to be represented by the signed Euclidean distance function, i.e.,  $\Phi(\xi, (x, y))$  yields the Euclidean distance of point  $(x, y)$  to the silhouette of the projected circular feature.

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 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 2D ellipse curve  $\Phi(\xi, (x, y))$  that equation is: [12]

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (B^2 - 4AC = 1) \quad (6)$$

where the parameters  $[A, B, C, D, E, F]^T$  are 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 them from the reference [12]. The derivation of the parameter is given in the Appendix. Moreover, these parameters yield the geometric parameters of the 2D ellipse curve  $\Phi(\xi, (x, y))$ , which can be denoted as:

$$\begin{cases} x_0 = \frac{2CD - BE}{B^2 - 4AC} \\ y_0 = \frac{2AE - BD}{B^2 - 4AC} \\ a = \sqrt{\frac{2(x_0^2 + Bx_0y_0 + Cy_0^2 - f)}{A + D - \sqrt{(A - C)^2 + D^2}}} \\ b = \sqrt{\frac{2(x_0^2 + Bx_0y_0 + Cy_0^2 - f)}{A + D + \sqrt{(A - C)^2 + D^2}}} \\ \theta = \frac{1}{2} \arctan\left(\frac{B}{A - C}\right) \end{cases} \quad (7)$$

where  $a$  and  $b$  are the semi-major axis and the semi-minor axis of the ellipse  $\Phi(\xi, (x, y))$ , respectively;  $\theta$  is the angle between major axis and horizontal direction; and  $(x_0, y_0)$  is the ellipse center. In Eq. 6, all 2D points  $(x, y)$  correspond to the sample pixel point set  $g_s$  which is collected via the equation  $\begin{cases} x' = a\cos\lambda\cos\theta - b\sin\theta\sin\lambda + x_0 \\ y' = a\cos\lambda\sin\theta + b\cos\theta\sin\lambda + y_0 \end{cases}, \lambda \in [0, 2\pi)$ . This set of points divides the domain of an image into the regions  $\text{inside}(g_s)$  and  $\text{outside}(g_s)$ , respectively. By applying the signed distance, we obtain elliptic shape representation by a

level set function, which is defined as follows:

$$\Phi(\xi, (x, y)) = \begin{cases} -\text{dist}((x, y), g_s), & (x, y) \in \text{inside}(g_s) \\ 0, & (x, y) \in g_s \\ \text{dist}((x, y), g_s), & (x, y) \in \text{outside}(g_s) \end{cases} \quad (8)$$

$$\text{where } \text{dist}((x, y), g_s) = \min_{(x_s, y_s) \in g_s} \| (x - x_s, y - y_s)^T \|_2.$$

### 3.2 Image statistical measure energy model with elliptic shape constraint

In this section, we describe the image statistical measure energy model for 2D-3D object pose estimation using circular feature. A variety of image information, such as intensity, edge, or texture, can be used to define an image statistical measure energy functional. Here, we employ the combination of the gray value, edge, and smoothness information as main image feature that drives the elliptic shape to the desired boundary. The image statistical measure model is defined as follows:

$$L(\xi) = e^{-(\lambda_1 L_1(\Phi(\xi), P_1) + \lambda_2 L_2(\Phi(\xi), P_2) + \lambda_3 L_3(\Phi(\xi), P_3))} \quad (9)$$

where the Bhattacharyya coefficient  $L_1$  is a divergence-type measure for the grayscale histogram distribution  $P_1$ . The lower the Bhattacharyya coefficient between the interior and exterior of the elliptic silhouette, the lower the similarity between them. The log-likelihood coefficient  $L_2$  is a divergence-type measure for the smoothness distribution  $P_2$ . The lower the log-likelihood coefficient between the interior and exterior of the elliptic silhouette, the lower the similarity between them. Term  $L_3$  is related to the energy along the length of the elliptic silhouette and the energy of the area inside of it. These three energy terms be defined so that the overall energy is minimized at the desired elliptic silhouette. That  $\lambda_i > 0$  is a constant that determined the weight of  $L_i(\Phi(\xi), P_i)$ . We acquire information about the grayscale histogram, smoothness distribution, and the intensities of edge in the following sections.

#### 3.2.1 Grayscale histogram

The region where the projection of the spatial circular feature of 3D object with the current poses into the image plane can be represented as an ellipse. Here, once the region is fixed, we select the grayscale histogram to model the interior and exterior of the ellipse contour because of its merits, such as robustness to highly noisy conditions and partial occlusion, and low computation cost [14, 15].

Assume an image  $I: \tilde{m} \times \tilde{n}$ ,  $\{\tilde{g}_{uv}\}_{u \in [1, \tilde{m}], v \in [1, \tilde{n}]}$  represents the grayscale value of each pixel, and the grayscale histogram is determined by dividing the grayscale values into different intervals. Each interval is indicated by  $\{t_k\}_{k \in [1, \tilde{l}]}$ , whereas  $\Delta \tilde{g}$  indicates the distance between the intervals, which is calculated using Eq. 10:

$$t_k = \left\lceil \frac{\tilde{g}_{uv}}{\Delta \tilde{g}} \right\rceil \quad (10)$$

A grayscale histogram counts the probability of each grayscale level code  $t_k$  occurring in the region.

Equation 11 represents the probability of the  $\tilde{k}$ th interval of the grayscale histogram occurring inside or outside the contour, respectively:

$$\begin{aligned} p_{\text{in}}^{\tilde{k}} &= \frac{\sum_{(u,v) \in \Omega_{\text{in}}} \delta(\tilde{g}(u,v) - \tilde{g}(t_{\tilde{k}}))}{n_{\text{in}}} \\ p_{\text{out}}^{\tilde{k}} &= \frac{\sum_{(u,v) \in \Omega_{\text{out}}} \delta(\tilde{g}(u,v) - \tilde{g}(t_{\tilde{k}}))}{n_{\text{out}}} \end{aligned} \quad (11)$$

where  $\delta(\tilde{g}(u,v) - \tilde{g}(t_{\tilde{k}}))$  determines the interval attribute of the grayscale value of pixel  $(u,v)$ ,  $\delta(x) = \begin{cases} 1, x=0 \\ 0, \text{other} \end{cases}$  is the Dirac function, and  $n_{\text{in}}, n_{\text{out}}$  represents the number of pixels in the regions inside and outside the contour  $\Omega_{\text{in}}, \Omega_{\text{out}}$ . According to Eq. 11, identifying the region attributes of pixel locations is essential to establish  $p_{\text{in}}^{\tilde{k}}, p_{\text{out}}^{\tilde{k}}$ . This paper implements the identification using Heaviside conversion, as expressed in Eq. 12:

$$H(\Phi(u,v)) = \begin{cases} 1, \Phi(u,v) > 0 \\ \frac{1}{2}, \Phi(u,v) = 0 \\ 0, \Phi(u,v) < 0 \end{cases} \quad (12)$$

The term  $H(\Phi(u,v))$  can normalize an arbitrary input value. According to Eqs. 8 and 12,  $1 - H(\Phi(u,v))$  and  $H(\Phi(u,v))$  can effectively identify the region attribute of pixel point  $(u,v)$ . Equation 11 can then be updated as:

$$\begin{aligned} p_{\text{in}}^{\tilde{k}} &= \frac{\sum_{(u,v) \in \Omega} (1 - H(\Phi(u,v))) \delta(\tilde{g}(u,v) - \tilde{g}(t_{\tilde{k}}))}{\sum_{(u,v) \in \Omega} (1 - H(\Phi(u,v)))} \\ p_{\text{out}}^{\tilde{k}} &= \frac{\sum_{(u,v) \in \Omega} H(\Phi(u,v)) \delta(\tilde{g}(u,v) - \tilde{g}(t_{\tilde{k}}))}{\sum_{(u,v) \in \Omega} H(\Phi(u,v))} \end{aligned} \quad (13)$$

where  $\Omega$  represents the image domain and  $\sum_{(u,v) \in \Omega} (1 - H(\Phi(u,v)))$  and  $\sum_{(u,v) \in \Omega} H(\Phi(u,v))$  represent the number of pixels inside and outside the contour, respectively. Finally, each interval probability is connected to build the region feature descriptor. The statistical properties of the grayscale values in the inner and outer regions of the contour in the interval  $t_{\tilde{k}}$  are expressed in Eq. 14

$$\begin{cases} p_{\text{in}} = \{p_{\text{in}}^{\tilde{k}}\}_{\tilde{k}=1, \dots, \tilde{I}}, \sum_{\tilde{k}=1}^{\tilde{I}} p_{\text{in}}^{\tilde{k}} = 1 \\ p_{\text{out}} = \{p_{\text{out}}^{\tilde{k}}\}_{\tilde{k}=1, \dots, \tilde{I}}, \sum_{\tilde{k}=1}^{\tilde{I}} p_{\text{out}}^{\tilde{k}} = 1 \end{cases} \quad (14)$$

After the grayscale histogram distributions are established for each region, there are many kinds of criteria that can be used to compare the similarity of these distributions. We adopt the Bhattacharyya similarity to measure the image statistical discrepancy between the interior and exterior regions of the elliptic shape prior contour. We define the similarity distance measure as follows:

$$L_1(\Phi(\xi), P_1) = \sum_{\tilde{k}=1}^{\tilde{I}} \sqrt{p_{\text{in}}^{\tilde{k}}(\Phi) \times p_{\text{out}}^{\tilde{k}}(\Phi)} \quad (15)$$



where  $p_{in}^{t_k}, p_{out}^{t_k} \in P_1$ ; the lower the Bhattacharyya coefficient  $L_1 \rightarrow 0$  can push the elliptic shape  $\Phi(\xi)$  toward the projection of the circular feature of 3D object with the current pose into the image plane.

### 3.2.2 Smoothness distribution

When the object and background differ much from each other on smoothness features, we select the smoothness distribution to model the interior and exterior regions of the ellipse contour. We define the smoothness distribution of the projection of the spatial circular feature of 3D object with the current pose into the image plane and foreground smoothness distribution as follows:

Assume an image  $I, \Omega \in R^2$  is the domain of an image  $I$ , suppose the values  $\{grad_{uv}\}_{u \in [1, \tilde{m}], v \in [1, \tilde{n}]}$  represent the gradient value of each pixel, obey the Gaussian distribution  $G(\mu, \Sigma)$ , and denote the probability density function by

$$p(grad(u, v), \mu, \Sigma) = \frac{1}{A} e^{-\frac{\|grad(u, v) - \mu\|_{\Sigma^{-1}}^2}{2}} \quad (16)$$

where  $A = \sqrt{2\pi \cdot \det(\Sigma)}$ . In this work, the probabilities of point belonging to the exterior and interior regions are

$$p_{out}(grad(u, v), \mu_{out}, \Sigma_{out}) = \frac{1}{A} e^{-\frac{\|grad(u, v) - \mu_{out}\|_{\Sigma_{out}^{-1}}^2}{2}} \quad (17)$$

and  $p_{in} = 1 - p_{out}$ , respectively. We have

$$p(\Phi) = \prod_{(u, v) \in I} [p_{out}(u, v)]^{H(\Phi)} [p_{in}(u, v)]^{(1-H(\Phi))} \quad (18)$$

Discarding the constant term, we can get the log-likelihood functional [16]:

$$\begin{aligned} L_2(\Phi) = & - \int_{\Omega} [\ln(1 - p_{out}(u, v))] H(\Phi(u, v)) d\Omega \\ & + \int_{\Omega} [\ln p_{out}(u, v)] H(\Phi(u, v)) d\Omega \end{aligned} \quad (19)$$

The nonnegative weighted parameters are used as the region force term of hood functional

$$\begin{aligned} L_2(\Phi) = & \int_{\Omega} [-\lambda_{in} \ln(1 - p_{out}(u, v))] H(\Phi(u, v)) d\Omega \\ & + \int_{\Omega} [\lambda_{out} \ln p_{out}(u, v)] H(\Phi(u, v)) d\Omega \end{aligned} \quad (20)$$

where  $\lambda_{in}$  and  $\lambda_{out} > 0$  are balance parameters.

Note that in Eq. 20, the lower *log* - likelihood value induces the elliptic shape  $\Phi(u, v)$  to approximate the projection of the circular feature of 3D object with the current pose into the image plane.

### 3.2.3 The image gradient measurement of shape priors

Here, we employ edge information that drives the elliptic shape  $\Phi(u, v)$  to the projection of the spatial circular feature of 3D object with the current pose into the image plane. We use the following edge indicator to acquire information about the intensities of edges:

$$\tilde{f} = \frac{1}{1 + |\nabla G_{\sigma}(u, v) * I(u, v)|^{\tilde{p}}} \quad (21)$$



where  $\tilde{f} \in [0, 1]$ ,  $\tilde{p} > 1$ ,  $G_\sigma$  is a Gaussian kernel with a standard deviation  $\sigma$ , and  $*$  denotes a convolution operation. Function usually takes smaller values at circle boundaries than at smooth regions. Based on  $\tilde{f}$ , we define the following basic energy functional for shape prior  $\Phi(u, v)$ :

$$L_3(\Phi) = Length(\Phi) + Area(\Phi) \quad (22)$$

The term *Length* is related to the energy along the length of the contour  $\Phi(u, v)$ , while the term *Area* is related to the energy of the area inside of  $\Phi(u, v)$ . These two energy terms can be defined so that the overall energy is minimized at the desired boundaries according to the edge indicator in Eq. 22:

$$Length(\Phi) = \int_{\Omega} g\delta(\Phi) |\nabla \Phi| d\Omega \quad (23)$$

and

$$Area(\Phi) = \int_{\Omega} gH(\Phi) d\Omega \quad (24)$$

Note that according to Eqs. 23 and 24, the minimization of these two energy terms depends heavily on the amount of edge information in the image. *Length*( $\Phi$ ) is then minimized when the elliptic shape  $\Phi(u, v)$  is located at the projection of the spatial circular feature of the 3D object with the current pose into the image plane.

$$\begin{aligned} L(\xi) = & \lambda_1 \sum_{k=1}^{\tilde{l}} \sqrt{p_{in}^{t_k}(\Phi) \times p_{out}^{t_k}(\Phi)} + \\ & \lambda_2 \int_{\Omega} [-\lambda_{in} \ln(1 - p_{out}(u, v)) + \lambda_{out} \ln p_{out}(u, v)] H(\Phi) d\Omega \\ & + \lambda_3 \left( \int_{\Omega} g\delta(\Phi) |\nabla \Phi| d\Omega + \int_{\Omega} gH(\Phi) d\Omega \right) \end{aligned} \quad (25)$$

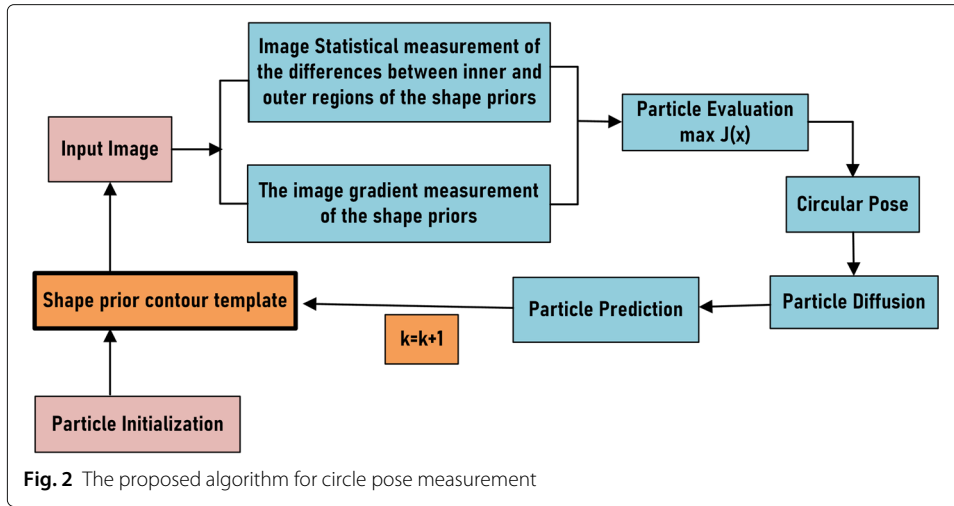
where  $\lambda_1, \lambda_2, \lambda_3 \geq 0$ ,  $\lambda_{in}, \lambda_{out} > 0$  are the constant user-specified parameters, which may vary for different images. As for the choice of  $\lambda_1, \lambda_2, \lambda_3, \lambda_{in}, \lambda_{out}$ , we select these parameters deliberately to get a desired result. Specially, the parameters  $\lambda_{in}, \lambda_{out}$  are not only used as the region force term of the image statistical measure model, they unify the order of the magnitude of each energy term.

### 3.3 Flowchart of our method

Given the observed image sequence  $z_{1:k} = \{z_1, z_2, \dots, z_k\}$  from time 1 to time  $k$ , the prior information of the spatial circle pose  $\xi_k$  at time  $k$  is provided by the inter-frame motion information. This paper establishes an objective function  $\arg \max_{\xi_k} \{p(\xi_k | z_{1:k})\}$  for the spatial circle pose measurement based on the video sequence. The particle-filtering method is adopted in the algorithm design, and the resulting algorithm flow is shown in Fig. 2.

## 4 Experiment results and analysis

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




---

**Algorithm 1** proposed method for circle pose measurement
 

---

**Step 1—initialize the particle set:**

$q_i^0 \in N(\hat{\xi}_0, v_k v_k^T), i = 0, \dots, N-1, \hat{\xi}_0$  indicates the initial pose measurement value and  $v_k$  indicates system noise.

**Step 2—particle evaluation:****Particle weight calculation:**

$$J_i^k = e^{-(\lambda_1 L_1(\Phi(\xi), P_1) + \lambda_2 L_2(\Phi(\xi), P_2) + \lambda_3 L_3(\Phi(\xi), P_3))}, \hat{J}^k = \max\{J_i^{k+1}, i = 1, \dots, N\};$$

**Best particle set:**  $\{q_{i\_index}^k\}$  comprise half of the particles with larger weights, the best particle  $\hat{q}^k = (\hat{\xi}^k, \hat{J}^k)$ .

**Step 3—refinement process:**

$Q_{nhb} = \delta(\hat{q}^k, \varepsilon), \hat{J}_{nhb}^k = \max\{J_{nhb_i}^k, i = 1, \dots, N_{nhb}\};$  if  $\hat{J}_{nhb}^k < \hat{J}^k$ , then  $\hat{\xi}^k = \hat{q}^k$ , otherwise  $\hat{J}^k = \hat{J}_{nhb}^k$  and  $\hat{\xi}^k = \hat{\xi}_{nhb}^k$ . Repeat the refinement process until the maximum number of iterations is reached. Observation noise is denoted as  $\varepsilon_k$ .

**Step 4—particle prediction:**

**Particle diffusion:**  $\{q_{index}^k\}$  and  $q_{index}^k \sim N(\hat{\xi}^k, ((1 - \hat{J}^k) \cdot v)^2)$  provide a particle sample set  $\{q_i^k\}, i = 0, \dots, N-1$  for the subsequent frames;

**Particle prediction:** the particle sample set of the subsequent frames  $\{q_i^{k+1}\}$  can be obtained by predicting  $\{q_i^k\}$ .

**Step 5—output:**

status estimation value  $\hat{\xi}^k$  can be used to determine whether it is over; if so, then exit the algorithm, otherwise,  $k = k + 1$ ; return to **Step 2**.

---

processing synthetic image sequences. All these synthetic image sequences are rendered using 3ds Max software; each frame of an input sequence is composed of a 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 100 scene frames consisting of monochrome image with  $256 \times 256$  pixels, and the effective focal length was  $f = 20$  mm, each circular feature with the radius  $R = 100$  mm, the appearance of the object during scene frames is dynamically modified by changing their orientation angles and location coordinates. A performance comparison

of the proposed algorithm with respect to existing algorithms is presented and discussed. We compared the following methods:

- (1) The algorithm  $(\cdot)^{[i]}$  in [12]: the ellipse detection [17] 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 proposed algorithm  $(\cdot)^{[ii]}$  in [5]: similar to the algorithm  $(\cdot)^{[i]}$ , the 2D ellipse parameters were given by the ellipse detection. The center of the circular feature in 3D is marked, and a virtual diameter parallel to the image plane that combines with the re-projection of the center is used to estimate the location. And then, a virtual chord parallel to the diameter is used to estimate the normal vector. The major contribution is the method formulates the problem with solving equations instead of matrix transformations.
- (3) The proposed algorithm  $(\cdot)^{[iii]}$  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 re-projection error for circles and points is defined to determine the optimal pose solution.
- (4) The proposed algorithm  $(\cdot)^{[iv]}$ .

The location error (LE) is given by [11]

$$LE = \|\xi_L - \hat{\xi}_L\| \quad (26)$$

where  $\xi_L$  and  $\hat{\xi}_L$  are the true and estimated coordinates of the object in the scene, respectively, given in millimeters. Moreover, the orientation error (OE) is given by

$$OE = \|\xi_0 - \hat{\xi}_0\| \quad (27)$$

where  $\xi_0$  and  $\hat{\xi}_0$  are the true and estimated rotation angles of the object with respect of the observer, respectively, given in degrees.

The performance of the tested algorithms is quantified in terms of percentages of normalized absolute errors (NAE), between the real  $\xi_{\text{real}}$  and estimated  $\xi_{\text{est}}$  pose parameters as follows: [11]

$$NAE = \frac{\|\xi_{\text{est}} - \xi_{\text{real}}\|}{\xi_{\text{real}}} \times 100 \quad (28)$$

The accuracy of location estimation of object is denoted by  $NAE_L$  and the accuracy of orientation estimation of the object is denoted by  $NAE_O$ ; both computed with Eq. 28.

Figure 6a presents the results of location estimation of the object obtained with the tested algorithms in processing sequences of synthetic images and while varying the variance of the additive noise. The means and standard deviations of the  $NAE_L$  and  $NAE_O$  of the four algorithms are shown in Tables 1 and 2, which are marked as  $T_{avg}^{(\cdot)}$ ,  $T_{std}^{(\cdot)}$ ,  $R_{avg}^{(\cdot)}$  and  $R_{std}^{(\cdot)}$ , respectively. The algorithm  $(\cdot)^{[ii]}$  yields better performance in terms of location estimation of the object than the algorithms  $(\cdot)^{[i]}$  and  $(\cdot)^{[iii]}$  in lower noisy conditions, such as  $\sigma^2 \leq 2\%$ . Because the algorithm  $(\cdot)^{[ii]}$  obtains the location estimation from the detected center in 2D image plane, the detected circular center is approaching the projection of the marked circular center with the true pose into the image plane in lower noisy conditions.

**Table 1** The means and standard deviations of  $NAE_L$  of the four algorithms

Noise level (%)	Normalized absolute error (%) <sup>*</sup>							
	$T_{avg}^{[i]}$	$T_{avg}^{[ii]}$	$T_{avg}^{[iii]}$	$T_{avg}^{[iv]}$	$T_{std}^{[i]}$	$T_{std}^{[ii]}$	$T_{std}^{[iii]}$	$T_{std}^{[iv]}$
0	0.50	0.33	0.50	<b>0.19</b>	0.01	0.03	0.01	<b>0.01</b>
2	0.46	0.39	0.46	<b>0.22</b>	0.05	0.29	0.05	<b>0.01</b>
4	3.34	5.90	3.37	<b>0.33</b>	311.95	733.08	322	<b>0.03</b>
6	12.40	8.42	13.32	<b>0.36</b>	2870	618	3072	<b>0.05</b>
8	13.58	14.51	14.02	<b>0.37</b>	1302	942	1508	<b>0.09</b>
10	25.36	28.39	25.76	<b>0.44</b>	7982	13733	8155	<b>0.22</b>

But among these algorithms' yield, there were high error levels in terms of location estimation in highly noisy conditions, and standard deviations of the  $NAE_L$  and  $NAE_O$  are very high. It means that the location estimation is incorrect in some scenarios. It can be seen that the proposed algorithm  $(\cdot)^{[iv]}$  yields the best performance in location estimation of the target among all considered. Also, in this comparison, the algorithm  $(\cdot)^{[iv]}$  presents robustness under different noise variance.

Figure 6b presents the orientation performance obtained with different values of additive noise SNR. Note that the algorithms  $(\cdot)^{[i]}$  and  $(\cdot)^{[iii]}$  yield good results in lower noisy conditions. Also, the algorithm  $(\cdot)^{[ii]}$  produces good results in terms of location but with high error levels in terms of orientation estimation. Because the algorithm  $(\cdot)^{[ii]}$  obtains the orientation estimation from the cross product of two special vectors, in which one vector consists of circular center and the projection of the center of the ellipse in 3D circular feature, and the another is obtained from the virtual chord of 3D circular feature, when the ellipse is approaching the circular, the high error level of the orientation estimation will be obtained. It can be seen that the proposed algorithm yields the lowest  $NAE_O$ , because the method  $(\cdot)^{[iv]}$  is the region-based algorithm to obtain the 3D pose parameters, in which the proposed algorithm does not require the ellipse detection. Furthermore, the proposed algorithm takes into account the motion dynamics of the object among scene frames, and better performance in pose estimation is obtained in comparison with the other tested algorithms.

Note that the accuracy of the previous algorithm  $(\cdot)^{[i]}$  has been proved in Fig. 7, and it produces good result in terms of location and orientation estimation among all previous algorithms in lower noisy conditions. According to our tests, the proposed algorithm yields excellent results in pose estimation of object from monocular images. The obtained results show that the proposed algorithm is highly accurate in estimation of 3D pose of the object. Also, the proposed algorithm yields robustness to the presence of additive noise.

**Table 2** The means and standard deviations of  $NAE_O$  of the four algorithms

Noise level (%)	Normalized absolute error (%) <sup>*</sup>							
	$R_{avg}^{[i]}$	$R_{avg}^{[ii]}$	$R_{avg}^{[iii]}$	$R_{avg}^{[iv]}$	$R_{std}^{[i]}$	$R_{std}^{[ii]}$	$R_{std}^{[iii]}$	$R_{std}^{[iv]}$
0	0.47	11.01	0.47	<b>0.33</b>	0.03	9.51	0.03	<b>0.03</b>
2	0.56	10.81	0.57	<b>0.37</b>	0.11	12.51	0.12	<b>0.11</b>
4	1.79	16.47	1.95	<b>0.47</b>	60.82	95.46	91.47	<b>0.09</b>
6	3.85	24.35	7.12	<b>0.68</b>	120	944	1520	<b>0.12</b>
8	5.22	29.78	21.02	<b>0.71</b>	179	1199	4629	<b>0.23</b>
10	7.63	27.26	15.88	<b>0.90</b>	268	1002	2544	<b>0.63</b>

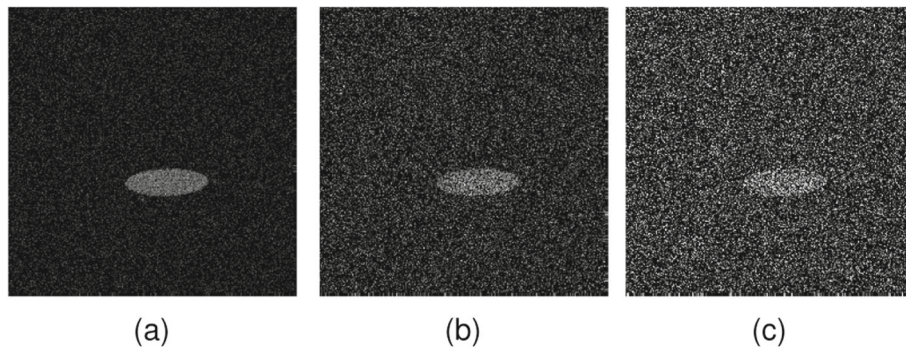
In the experiment depicted in Fig. 7, we evaluate the performance of the proposed algorithm in terms of LE and OE measures by processing sequences of synthetic images corrupted with zero-mean additive Gaussian noise with the variance  $\sigma_n^2 = 2\%, 6\%, 10\%$ . Figure 3 illustrates examples of noisy scene frames for different values of  $\sigma_n^2$ .

In Fig. 7, the estimated object's pose obtained with the algorithm  $(\cdot)^{[iv]}$  is indicated with dot lines; the estimated object's pose obtained with the algorithm  $(\cdot)^{[i]}$  is indicated with solid lines. We can see that after processing several frames of input sequence with highly noisy conditions, the algorithm  $(\cdot)^{[i]}$  fails; high LE and OE values are obtained and LE and OE lines are missing in Fig. 7. However, we can observe that the proposed algorithm is able to estimate the pose of object with good accuracy even in highly noisy conditions. The means of normalized absolute error (NAE) of LE and OE are no more than 0.5% and 1%, respectively. This is because the elliptic shape prior and the image feature drive the virtual elliptic shape contour to approximate the projection of circular feature in 3D into the image plane to measure the circular pose, in which the proposed algorithm does not require to the edge detection the co-elliptic arc matching and ellipse fitting.

Note that in Fig. 7, low LE and OE values are obtained when the additive noise variance is  $\sigma^2 \leq 2\%$ . This is because the global information of the image rather than the local features is considered in the proposed algorithm. The stability of the image statistical measurement (grayscale histogram, smoothness distribution) is utilized to overcome the impacts of noise on the measurement results. In addition, the temporal information among the frames is considered in the proposed algorithm. It can be shown that the proposed algorithm is very robust to the presence of additive noise in the scene.

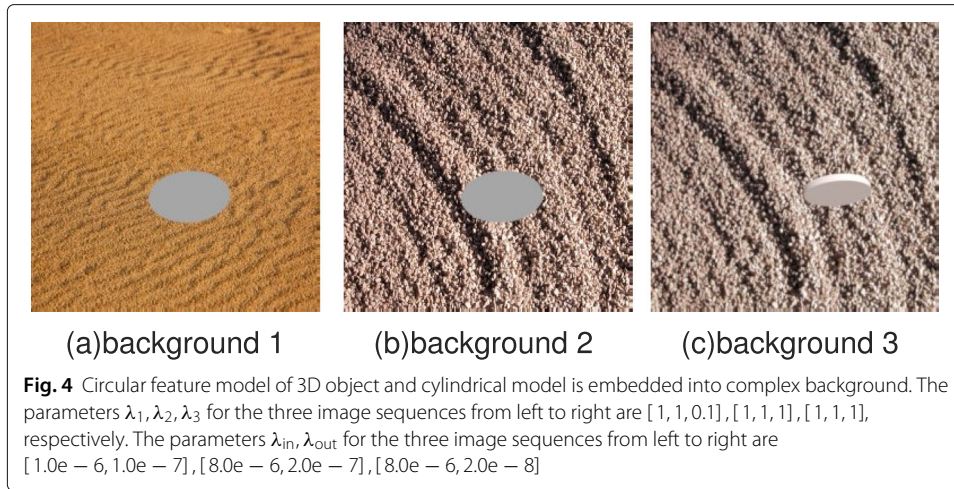
In the experiment depicted in Fig. 8, the performance of the proposed algorithm for pose estimation is evaluated and discussed by processing sequences of synthetic images with a challenging background. Each test sequence contains the objects with an unknown pose trajectory and embedded into a disjoint background depicted in Fig. 4.

Figure 8 shows the obtained results with the two algorithms when processing 100 scene frames while varying background. The estimated object's pose obtained for varying background is indicated with red lines, green lines, and blue lines, respectively. We can see that after processing several frames of input sequence with background, the algorithm  $(\cdot)^{[i]}$  fails; LE and OE lines are missing in Fig. 8. Especially, the algorithm  $(\cdot)^{[i]}$  fails when processing the first frame of input sequence that cylindrical model is embedded into



**Fig. 3** Synthetic image sequences degraded with additive noise with the variance (a)  $\sigma_n^2 = 2\%$ , (b)  $\sigma_n^2 = 6\%$ , and (c)  $\sigma_n^2 = 10\%$ . The parameters  $\lambda_1 = \lambda_3 = 1, \lambda_2 = 0$  for all three image sequences

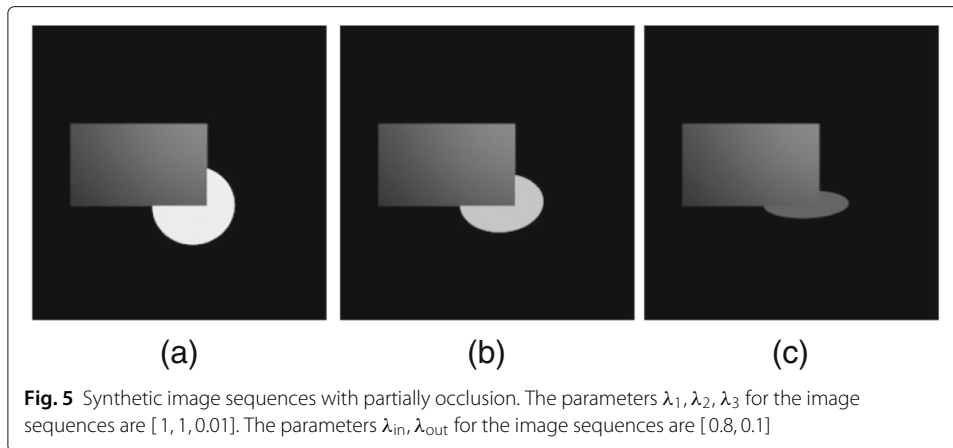




complex background 2; LE and OE lines are missing in Fig. 8. Whereas the algorithm  $(\cdot)^{[iv]}$  maintained reliable measurements throughout all the sequences, we can observe that the proposed algorithm is able to estimate the pose of object with good accuracy even in several challenging backgrounds. This is because the proposed algorithm takes into account the image smoothness of discrepancy between the object region and natural background. The key image feature descriptors, such as smoothness, gray histogram, and intensities of edge, are considered in the image statistical measurement, and the proposed algorithm set proper weighting value for each energy term. Figure 4 illustrates the weighting parameter of each energy term for varying scene frames. The parameters  $\lambda_1, \lambda_2, \lambda_3$  for the three image sequences with varying background from left to right are  $[1, 1, 0.1], [1, 1, 1], [1, 1, 1]$ , respectively. The parameters  $\lambda_{in}, \lambda_{out}$  for the three image sequences from left to right are  $[1.0e-6, 1.0e-7], [8.0e-6, 2.0e-7], [8.0e-6, 2.0e-8]$ .

Such experiment shows that the proposed algorithm that incorporates elliptic shape prior representation with the image statistical measure energy model is more stable to complex background pollution. These works induce the virtual elliptic shape contour to approximate the projection of circular feature in 3D into the image plan by maximizing the image statistical measurement of discrepancy between its interior and exterior regions and minimizing the intensities of edge at the desired of boundaries. One can obtain a good accuracy in the circular pose results.

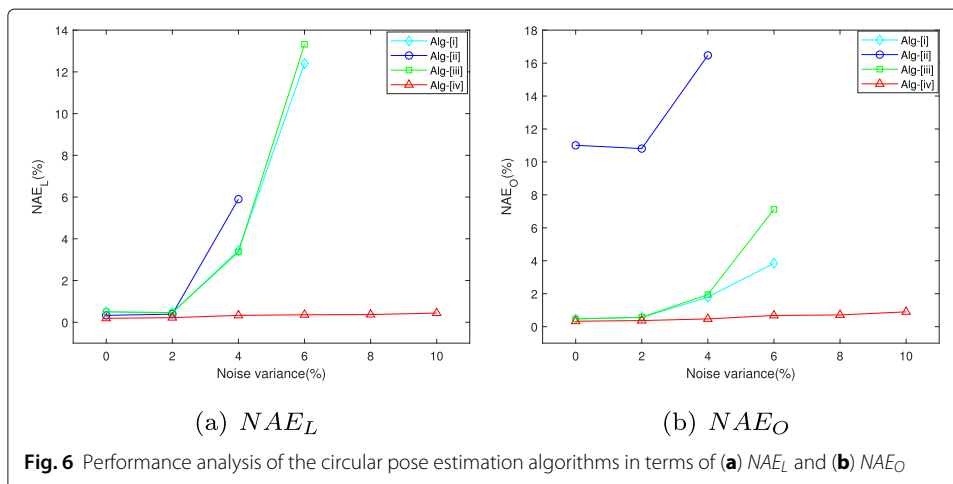
In the experiment depicted in Fig. 9, we tested the influence of partial occlusion. The motion of the object causes partially severe occlusion as depicted in Fig. 5. In this case, another significant advantage of using the elliptic shape prior becomes apparent. The estimated object's pose obtained with the proposed algorithm  $(\cdot)^{[iv]}$  is indicated with red line as depicted in Fig. 9; the estimated object's pose obtained with the proposed algorithm  $(\cdot)^{[iv]}$  is indicated with blue line as depicted in Fig. 9. We can see that after processing several frames of input sequence, the pose given by the algorithm  $(\cdot)^{[i]}$  is incorrect, the high LE and OE values are obtained, and red lines are missing in Fig. 9. This is because a local minimal value can appear in using the least-squares method, resulting in incorrect ellipse fitting and a bad spatial circle pose estimation. Despite the change of the partial occlusions, the proposed algorithm is able to estimate the pose of target with a good accuracy even in severe occlusion. This is because the information from the combined information



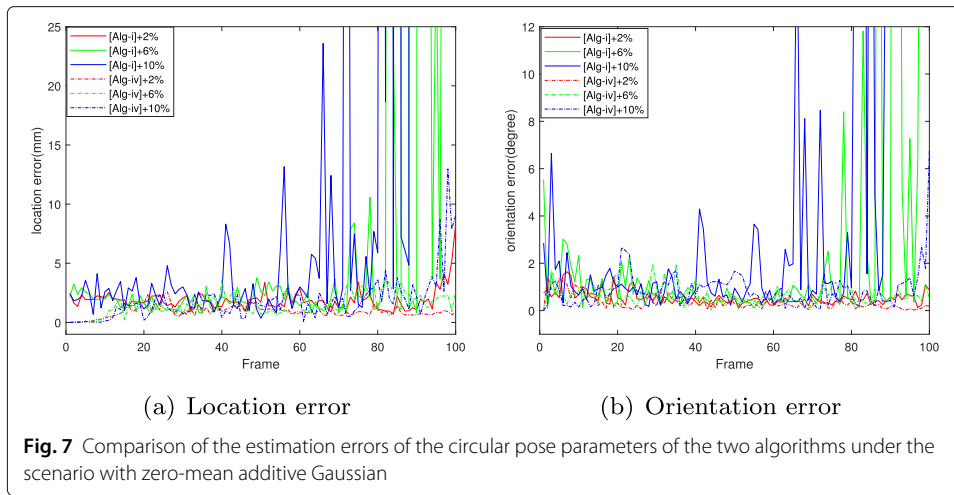
from both the elliptic shape prior and image data can be still sufficient for a reliable pose estimation. The elliptic shape prior can constrain the projection contour of the spatial circular to the vicinity of the edge of the spatial circle image. Note that in Fig. 9, low OE values are obtained when processing few more frames. This is because the slight occlusion does not harm the pose estimation. The nearly constant values of the LE indicate a stable result. The values of the partially severe occluded sequence have a higher deviation (up to 4 cm), but it is still possible to reliably estimate the object.

The overall computation time depends on the number of particle for the method to converge. For each sequence that includes the disturbances by noise, background, and occlusion, the computation time per stereo pair was approximately 2 min (1 min and 50 s to 2 min and 2 s) on a 1.8-GHz Inter(R)Core(TM)i7-8565U window10 machine. The computation time is significantly larger than that with other pose estimate models that often achieve real-time performance. However, in contrast to these approaches, our model includes a sophisticated interlocking of image statistical measure-based virtual elliptic contour matching and pose estimation that allow for good results in situations where current real-time.

As can be seen from Figs. 6, 7, 8, and 9, the proposed algorithm has excellent pose estimation performance under various application scenarios. At the same time, the proposed



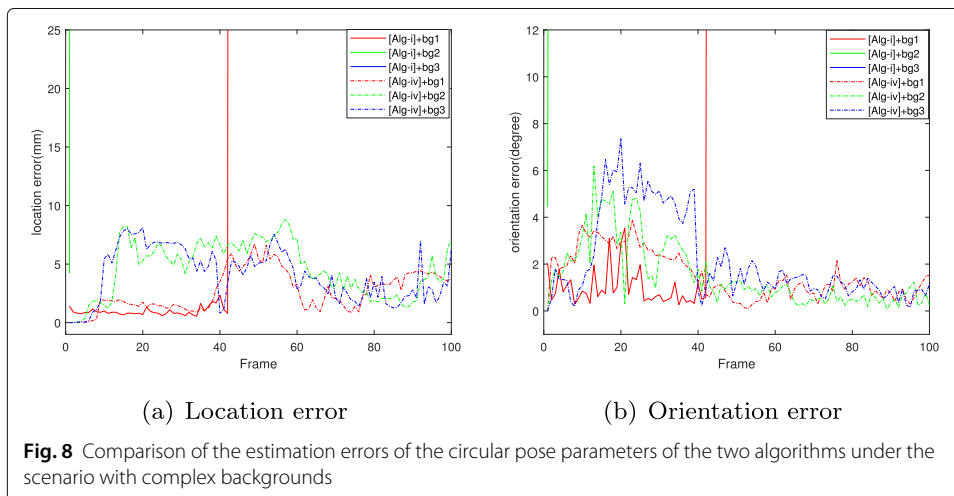


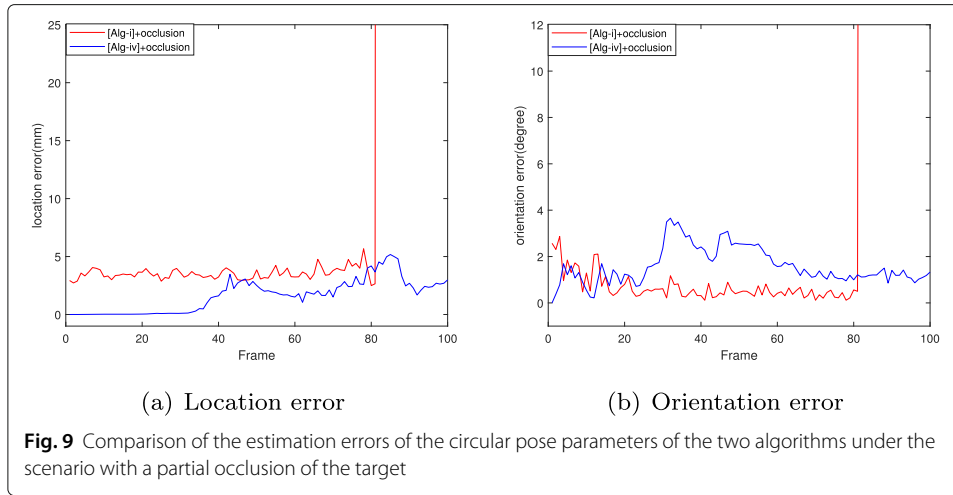


algorithm demonstrates robustness and provides accurate and reliable pose estimation results under challenging scenarios. The reason for this is that the proposed algorithm employs the elliptic shape prior information of the circular feature in 3D and the combination of image feature statistical measures, such as the grayscale histogram, the intensities of edge, and smoothness distribution to drive the virtual elliptic shape contour to the projection of the circular feature of 3D object with the current pose into the image plane. These works utilizes the elliptic shape constraint and the stability of the image statistical measurement, effectively reducing the impacts of highly noisy conditions, complex background, and partial severe occlusions on object pose estimation results. The proposed algorithm is able to estimate the pose of object with a good accuracy.

## 5 Conclusions

In order to resolve low accuracy in 3D object pose estimation or estimation failure under the scenarios with highly noisy conditions, complex backgrounds, and partial severe occlusions, this paper proposes an algorithm combing elliptic shape priors with the image statistical measure for 2D-3D object pose estimation using circular feature. The algorithm





defines a representation method for 5D circular pose parameters, constructs the elliptic shape prior model for the circular feature in 3D, and selects the grayscale histogram, smoothness distribution, and the intensities of edge as main image feature descriptors that define the image statistical measure model. Then, the algorithm incorporates the elliptic shape priors representation with the image statistical measure energy model to construct the likelihood function. The image feature statistical measure drives the virtual elliptic shape prior contour to approximate the projection of the circular feature of 3D object with the current pose into the image plane. A good accuracy pose estimation can be obtained. The algorithm is based on the global information of the image rather than on the local features. It utilizes the elliptic shape priors and the image statistical measurement to effectively reduce the impacts of noise, complex backgrounds, and partial severe occlusions on the pose estimation result. The simulation experiment demonstrates that the proposed algorithm provides reliable and accurate pose estimation results for 2D-3D object pose estimation using circular feature under challenge scenarios.

## Appendix

### 5.1 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 2D ellipse curve  $\Phi(\xi, (x, y))$  that equation is represented in Eq. 28

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (B^2 - 4AC = 1) \quad (29)$$

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

To find the 2D 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  $S_{\text{cone}}$ , 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 is obtained by intersecting the sphere  $S_G$  whose center is

$(X, Y, Z)^T$  and radius is  $R$  with the plane  $\pi$  whose surface normal vector is  $(l, m, n)^T$  as follows:

$$\begin{cases} (X_1 - X)^2 + (Y_1 - Y)^2 + (Z_1 - Z)^2 = R^2 \\ l(X_1 - X) + m(Y_1 - Y) + n(Z_1 - Z) = 0 \end{cases} \quad (30)$$

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 the line that joins the vertex to the point on the circular  $G$  and the point on the cone is represented as:

$$\frac{X_1 - X_2}{X_2 - 0} = \frac{Y_1 - Y_2}{Y_2 - 0} = \frac{Z_1 - Z_2}{Z_2 - 0} = t \quad (31)$$

where we can obtain the coordinate of the points on circular  $G$  which can be denoted as:

$$(X, Y, Z) = (tX_2 + X_2, tY_2 + Y_2, tZ_2 + Z_2) \quad (32)$$

Especially,  $(X, Y, Z) \in \pi$  and  $(X, Y, Z) \in S_G$ , by solving two simultaneous equations given by Eqs. 28 and 30, the equation of cone  $S_{\text{cone}}$  can be written as:

$$\begin{aligned} & (mY + nZ)^2 X_2^2 - 2(mY + nZ)X_2X(mY_2 + nZ_2) + \\ & X^2(mY_2 + nZ_2)^2 + (lX + nZ)^2 - 2(lX + nZ)Y_2Y(lX_2 + nZ_2) + \\ & (lX_2 + nZ_2)^2 Y^2 + (lX + mY)^2 Z_2^2 - \\ & 2(lX + mY)Z_2Z(lX_2 + mY_2) + (lX_2 + mY_2)^2 Z^2 \\ & = R^2 (l^2 X_2^2 + m^2 Y_2^2 + n^2 Z_2^2 + 2lmX_2Y_2 + 2lnX_2Z_2 + 2mnY_2Z_2) \end{aligned} \quad (33)$$

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

$$\begin{cases} A = (mY + nZ)^2 + l^2 Y^2 + l^2 Z^2 - R^2 l^2 \\ B = -2mX(mY + nZ) - 2lY(lX + nZ) + 2lmZ^2 - 2lmR^2 \\ C = m^2 X^2 + (lX + nZ)^2 + m^2 Z^2 - m^2 R^2 \\ D = -2nfX(mY + nZ) + 2nlfY^2 - 2fZ(lX + mY) - 2lnfR^2 \\ E = 2mnfX^2 - 2nfY(lX + nZ) - 2mfZ(lX + mY) - 2mnfR^2 \\ F = X^2 n^2 f^2 + Y^2 n^2 f^2 + (lX + mY)^2 f^2 - R^2 n^2 f^2 \end{cases} \quad (34)$$

where  $f$  is the focal length of the camera. By replacing  $(l, m, n)^T$  with  $(\sin\beta\cos\alpha, \sin\beta\sin\alpha, \cos\beta)^T$ , Eq. 33 can be written as

$$\begin{cases} A = ((\sin\beta\sin\alpha)Y + \cos\beta Z)^2 + (\sin\beta\cos\alpha)^2 Y^2 + (\sin\beta\cos\alpha)^2 Z^2 - R^2 (\sin\beta\cos\alpha)^2 \\ B = -2(\sin\beta\sin\alpha)X(\sin\beta\sin\alpha Y + \cos\beta Z) - 2(\sin\beta\cos\alpha)Y((\sin\beta\cos\alpha)X + \cos\beta Z) \\ \quad + 2(\sin\beta\cos\alpha)(\sin\beta\sin\alpha)Z^2 - 2(\sin\beta\cos\alpha)(\sin\beta\sin\alpha)R^2 \\ C = (\sin\beta\sin\alpha)^2 X^2 + ((\sin\beta\cos\alpha)X + \cos\beta Z)^2 + (\sin\beta\sin\alpha)^2 Z^2 - (\sin\beta\sin\alpha)^2 R^2 \\ D = -2\cos\beta fX((\sin\beta\sin\alpha)Y + \cos\beta Z) + 2\cos\beta(\sin\beta\cos\alpha)fY^2 \\ \quad - 2f(\sin\beta\cos\alpha)Z((\sin\beta\cos\alpha)X + \sin\beta\sin\alpha Y) - 2lnfR^2 \\ E = 2(\sin\beta\sin\alpha)\cos\beta fX^2 - 2\cos\beta fY((\sin\beta\cos\alpha)X + \cos\beta Z) \\ \quad - 2\sin\beta\sin\alpha fZ((\sin\beta\cos\alpha)X + (\sin\beta\sin\alpha)Y) - 2(\sin\beta\sin\alpha)\cos\beta fR^2 \\ F = X^2 (\cos\beta)^2 f^2 + Y^2 (\cos\beta)^2 f^2 + ((\sin\beta\cos\alpha)X + (\sin\beta\sin\alpha)Y)^2 f^2 - R^2 (\cos\beta)^2 f^2 \end{cases} \quad (35)$$

#### Acknowledgements

The authors thank for the valuable and constructive comments from the editor and reviewers. 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.

**Authors' contributions**

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

**Funding**

This work was supported in part by the Natural Science Foundation of Ningxia (No.2018AAC03126), the First-Class Disciplines Foundation of Ningxia (No.NXYLXK2017B09), the Major Project of North Minzu University (No.ZDZX201801).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>The National Laboratory for Mechatronic and Control, Beijing Institute of Technology, Beijing, 100081 China. <sup>2</sup>The Key Laboratory of Intelligent Information and Big Data Processing of Ningxia Province, North Minzu University, Yinchuan, 750021 China.

Received: 25 October 2019 Accepted: 23 June 2020

Published online: 17 July 2020

**References**

1. B. Huang, Y. Sun, Q. Zeng, General fusion frame of circles and points in vision pose estimation. *Optik*. **154**, 47–57 (2018)
2. J. Lee, R. Sandhu, A. Tannenbaum, Particle filters and occlusion handling for rigid 2d–3d pose tracking. *Computer Vision and Image Understanding*. **117**(8), 922–933 (2013)
3. C. Meng, H. Sun, in *Proceedings of the 2nd International Conference on Electronics, Network and Computer Engineering (ICENCE 2016)*, Monocular pose measurement method based on circle and line features (Atlantis Press, 2016). <https://doi.org/10.2991/icence-16.2016.149>
4. B. Huang, Y. Sun, Y. Zhu, Z. Xiong, J. Liu, Vision pose estimation from planar dual circles in a single image. *Optik*. **127**(10), 4275–4280 (2016)
5. C. Wang, D. Chen, M. Li, J. Gong, Direct solution for pose estimation of single circle with detected centre. *Electronics Letters*. **52**(21), 1751–1753 (2016)
6. S. Ghosh, R. Ray, S. R. K. Vadali, S. N. Shome, S. Nandy, Reliable pose estimation of underwater dock using single camera: a scene invariant approach. *Machine Vision and Applications*. **27**(2), 221–236 (2016)
7. T. Brox, B. Rosenhahn, J. Weickert, in *Joint Pattern Recognition Symposium*, Three-dimensional shape knowledge for joint image segmentation and pose estimation (Springer, 2005), pp. 109–116. [https://doi.org/10.1007/11550518\\_14](https://doi.org/10.1007/11550518_14)
8. S. Dambreville, R. Sandhu, A. Yezzi, A. Tannenbaum, A geometric approach to joint 2D region-based segmentation and 3D pose estimation using a 3D shape prior. *SIAM journal on imaging sciences*. **3**(1), 110–132 (2010)
9. V. H. Diaz-Ramirez, K. Picos, V. Kober, Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*. **323**, 32–43 (2014)
10. K. Picos, V. H. Diaz-Ramirez, V. Kober, A. S. Montemayor, J. J. Pantrigo, Accurate three-dimensional pose recognition from monocular images using template matched filtering. *Optical Engineering*. **55**(6), 063102 (2016)
11. K. Picos, V. H. Diaz-Ramirez, A. S. Montemayor, J. J. Pantrigo, V. Kober, Three-dimensional pose tracking by image correlation and particle filtering. *Optical Engineering*. **57**(7), 073108 (2018)
12. Y. C. Shiu, S. Ahmad, in *Conference Proceedings, IEEE International Conference on Systems, Man and Cybernetics*, 3D location of circular and spherical features by monocular model-based vision (IEEE, 1989), pp. 576–581. <https://doi.org/10.1109/icsmc.1989.71362>
13. S. Challa, M. R. Morelande, D. Mušicki, R. J. Evans, *Fundamentals of Object Tracking*. (Cambridge University Press, Cambridge, 2011)
14. J. Ning, L. Zhang, D. Zhang, W. Yu, Joint registration and active contour segmentation for object tracking. *IEEE transactions on circuits and systems for video technology*. **23**(9), 1589–1597 (2013)
15. Y. Hang, C. Derong, G. Jiulu, Object tracking using both a kernel and a non-parametric active contour model. *Neurocomputing*. **295**, 108–117 (2018)
16. S. Luo, X.-C. Tai, L. Huo, Y. Wang, R. Glowinski, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Convex shape prior for multi-object segmentation using a single level set function (IEEE, 2019), pp. 613–621. <https://doi.org/10.1109/iccv.2019.00070>
17. C. Lu, S. Xia, M. Shao, Y. Fu, Arc-support line segments revisited: An efficient high-quality ellipse detection. *IEEE Transactions on Image Processing*. **29**, 768–781 (2019)

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.