SCIENTIFIC REPORTS

OPEN

A 6-DOF Navigation Method based on Iterative Closest Imaging Point Algorithm

Shuai Shi [1,2], Zheng You[1,2], Kaichun Zhao[1,2], Zhaoyao Wang[1,3], Chenguang Ouyang[1] & Yongkui Cao[1]

To achieve six degree-of-freedom autonomous navigation of an inboard spacecraft, a novel algorithm called iterative closest imaging point (ICIP) is proposed, which deals with the pose estimation problem of a vision navigation system (VNS). This paper introduces the basics of the ICIP algorithm, including mathematical model, algorithm architecture, and convergence theory. On this basis, a navigation method is proposed. This method realizes its initialization using a Gaussian mixture model-based Kalman filter, which simultaneously solves the 3D-to-2D point correspondences and the camera pose. The initial value sensitivity, computational efficiency, robustness, and accuracy of the proposed navigation method are discussed based on simulation results. A navigation experiment verifies that the proposed method works effectively. The three-axis Euler angle accuracy is within 0.19° (1σ), and the three-axis position accuracy is within 1.87 mm (1σ). The ICIP algorithm estimates the full-state pose by merely finding the closest point couples respectively form the images obtained by the VNS and predicted at an initial value. Then the optimized solution of the imaging model is iteratively calculated and the full-state pose is obtained. Benefiting from the absence of a requirement for feature matching, the proposed navigation method offers advantages of low computational complexity, favorable stability, and applicability in an extremely simple environment in comparison with conventional methods.

The inboard spacecraft (IS) of Chinese Tiangong space station (Fig. 1) was proposed by Tsinghua University and China Manned Space Agency in 2015 to conduct algorithm evaluation and on-orbit service demonstration[1]. As the requirements of high-precision navigation and detection capability are considered, computer vision is an excellent choice for IS compared with other available navigation methods, such as infrared[2], ultrasound[3], UWB[4], and Wi-Fi[5]. Navigation is one of the most important applications of computer vision and has been the focus of research since its first proposal. The existing vision navigation methods can be simply classified into two categories according to the known/unknown environment where the user is situated. Research on navigating in known environments is discussed by Rudol P[6], tawdross P[7] and Serrão M[8], whereby the features of the known environment are previously obtained and used as navigation references. Other researchers have attempted to solve navigation problems in an unknown environment. Methods based on optic flow[9,10] and SLAM[11,12] are explored in depth.

On the basis of the IS mission plan, navigation inside the space station aims to achieve the following:

1) use of the cabin frame, which is attached to the space station, as the reference frame;
2) 6-degree of freedom (DOF) full-state parameter estimation throughout the flight of the IS; and
3) low computational complexity.

The IS presents 6-DOF free flight inside the space station, where the surrounding area is known and stable (Fig. 1(b)). Thus, the features attached inside the cabin can be used as navigation references. The coordinates of the features in the cabin frame are essential to achieve the aforementioned requirement 1). Point features are good choices for navigation references because they are easy to deploy inside the cabin and their 3D coordinates are

[1]Department of Precision Instrument, Tsinghua University, Beijing, China. [2]State Key Laboratory of Precision Measurement Technology and Instruments, Tsinghua University, Beijing, China. [3]China Manned Space Agency, Beijing, China. Correspondence and requests for materials should be addressed to Z.Y. (email: yz-dpi@mail.tsinghua.edu.cn) or K.Z. (email: kaichunz@mail.tsinghua.edu.cn)
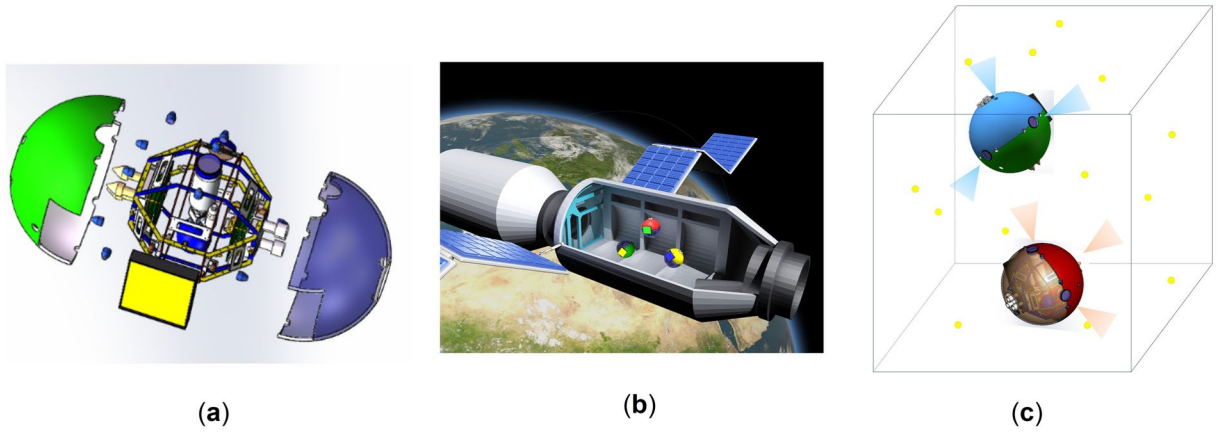
Figure 1. (**a**) 3D model, (**b**) flying sketch and (**c**) navigation sketch of the IS.

convenient to measure (Fig. 1(c)). These points, which are vision navigation beacons, can be artificially designed to become easily distinguishable from the environment through features such as shape, gray scale and polarization degree. Therefore, their images can be efficiently extracted from the pictures. The database of the beacon coordinates and features can also be uploaded to the IS before its flight and learning the cabin environment is not necessary, thereby meeting requirement 3). To address requirement 2), a sufficient number of beacons should be deployed reasonably.

Unfortunately, although the cooperative beacon points meet the basic requirements, challenges to realize IS navigation remain with existing methods. The number of beacons in sight during the IS flight are stochastic because many of the same beacons are deployed inside the space station cabin for good coverage. This problem seems to prevent the conventional perspective-N-point method[13,14] from directly solving the pose estimation problem of IS. Pattern recognition methods based on features of the images, such as shape context[15,16] and spectral graph theory[17,18], are reasonable ways to determine the corresponding relationship between two point sets of different images in most cases. However, these methods seem incapable of solving the problem in the situation of IS navigation. The number of beacons in sight during the IS flight is much smaller than the feature point amount in an image commonly studied; thus, most methods are not robust enough.

Iterative closest point (ICP), which was first introduced in 1992[19], is an algorithm employed to minimize the difference between two clouds of points. This algorithm is often conducted for point set registration or transformation estimation[20,21]. Although ICP does not directly solve the navigation problem of IS, the core idea of ICP provides an important reference. In the present paper, the imaging model for a multiple field-of-view vision system is established. Based on the imaging model, a novel algorithm called iterative closest imaging point (ICIP), which estimates the attitude and position of the vision system, is proposed. Furthermore, considering the sensitivity of ICP algorithms, a Gaussian mixture model (GMM)-based initialization[22–24] is utilized. On this basis, a continuous autonomous navigation method using the ICIP algorithm is designed and analyzed by numerical simulation. Experimental results show that ICIP-based navigation operates effectively. This method suits not only the IS navigation inside the space station but also similar application situations, such as indoor, inside tunnels, and underwater.

ICIP-based Autonomous Navigation Method

Imaging model. The ideal pinhole model of mapping of a 3D point P (Xw, Yw, Zw) to a 2D image p (u, v) is given by

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & s & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & \gamma & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix},$$

(1)

where Z_c is the optic axis coordinate of point P. dx and dy are the respective ratio coefficients in the x and y directions, respectively. s and γ are the non-orthogonal factors of axes of the image coordinate. (u_0, v_0), f, $\boldsymbol{R}$, and $\boldsymbol{T}$ are the pixel coordinate of the camera principal point, principal distance of the camera, 3×3 rotation matrix, and 3D translation vector, respectively. $\alpha_x = f/dx$ and $\alpha_y = f/dy$ are the scale factors of the u-axis and v-axis of the image coordinate, respectively.

On the basis of an ideal pinhole imaging model, the displacement of the principal point, focal length deviation, distortion, and other error factors must be considered in the actual imaging system. In practice, several cameras are attached to form a vision navigation system (VNS) in many cases so that the field of view is extended and the measurement model is optimized. The imaging model is extended based on equation (1) and can be written as

$$Z_{ci}\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_{xk} & \gamma_k & u_{0k} & 0 \\ 0 & \alpha_{yk} & v_{0k} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} C_b^k & \boldsymbol{0} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \begin{bmatrix} C_w^b & r_{bw}^b - r_{bk}^b \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_{Wi} \\ Y_{Wi} \\ Z_{Wi} \\ 1 \end{bmatrix}$$

(2)

where $k = 1, 2, 3, \ldots$ represents the number of cameras; $i = 1, 2, 3, \ldots$ represents the number of beacons observed, C_b^k is the rotation matrix from the VNS body coordinate to the camera k coordinate, C_w^b is the rotation matrix from the world coordinate to the VNS body coordinate, and $r_{bw}^b = (t_x, t_y, t_z)^T$ and $r_{bk}^b = (r_{xk}, r_{yk}, r_{zk})^T$ are the translation vectors from the origin of the VNS body coordinate to the origins of the world coordinate and camera k coordinates, respectively, which are expressed in the VNS body coordinate.

The 6-DOF navigation parameters are the three Euler angles related to C_w^b and the three components of r_{bw}^b. Let $A^k = \begin{bmatrix} \alpha_{xk} & \gamma_k & u_{0k} & 0 \\ 0 & \alpha_{yk} & v_{0k} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} C_b^k & \boldsymbol{0} \\ \boldsymbol{0}^T & 1 \end{bmatrix} = (a_{ij}^k)_{3\times4}$ and $C_b^w = (r_{ij}^k)_{3\times3}$, and the following is obtained:

$$\begin{cases} (a_{11}^k - a_{31}^k u_k)(r_{11}X_{Wi} + r_{12}Y_{Wi} + r_{13}Z_{Wi} + t_x - r_{xk}) \\ + (a_{12}^k - a_{32}^k u_k)(r_{21}X_{Wi} + r_{22}Y_{Wi} + r_{23}Z_{Wi} + t_y - r_{yk}) \\ + (a_{13}^k - a_{33}^k u_k)(r_{31}X_{Wi} + r_{32}Y_{Wi} + r_{33}Z_{Wi} + t_z - r_{zk}) = 0 \\ (a_{21}^k - a_{31}^k v_k)(r_{11}X_{Wi} + r_{12}Y_{Wi} + r_{13}Z_{Wi} + t_x - r_{xk}) \\ + (a_{22}^k - a_{32}^k v_k)(r_{21}X_{Wi} + r_{22}Y_{Wi} + r_{23}Z_{Wi} + t_y - r_{yk}) \\ + (a_{23}^k - a_{33}^k v_k)(r_{31}X_{Wi} + r_{32}Y_{Wi} + r_{33}Z_{Wi} + t_z - r_{zk}) = 0 \end{cases}$$

(3)

Equation (3) is the practical imaging model of the VNS. The pose of the VNS is possibly obtained when many beacons are observed.

ICIP algorithm statement. Two possible ways exist for a VNS to realize the navigation based on reference beacons. First, when the VNS takes pictures of the environment, it searches for a pose where similar pictures will be observed. This pose is the solution of the navigation parameters. Second, the corresponding relationship between the beacons and the observed images are directly determined through feature matching, and then the pose of the vision system based on the imaging model is calculated. For beacon-based navigation, only several feature points are in the image after feature extraction, causing extreme difficulty in completing pattern matching that utilizes existing methods. Thus, the first method using pose searching is implementable, and the problem becomes proposing a searching strategy.

One of the most widely used point set registration algorithms, ICP has advantages of simple architecture and low computational complexity. The basic principle of ICP is finding the nearest point couples between two point sets, and then minimizing the distance between the point couples by estimating the transformation between them. This process is operated iteratively until locally optimal solutions are found[19]. Later research focuses on the improvement of the origin method through various ways, such as optimizing the point set matching process[25,26] and improving the optimal estimation algorithm[27].

Inspired by the core idea of ICP, a novel algorithm called ICIP is proposed. Once the beacon coordinates are known and the vision system parameter is calibrated, the images observed at any pose can be predicted using the imaging model (equation (3)). These predicted images are similar to the obtained real images if the pose is close to the true value. Thus, a roughly corresponding relationship between the beacons and the images is obtained by matching the closest point couples between a predicted image and a real image. In other words, each point couple is assumed as the same image of the beacon. Therefore, the 3D coordinates of beacons and their images are obtained so that an observable measurement equation can be built based on equation (3) once ≥ 3 beacons are in sight. The optimal solution of the pose can be found using the Levenberg–Marquardt (L-M) method, and then new predicted images are generated at that pose. The process composed of predicting images, matching closest couples, and L-M iteration is repeatedly operated until a locally optimal solution of the pose is found. The distance between two image points $p_1(u_1, v_1)$ and $p_2(u_2, v_2)$ is defined as $d(p_1, p_2) = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}$.

The ICIP algorithm is stated as follows:

- The pictures of the current surroundings are taken by the VNS, and the pixel coordinates of the beacon images, namely, data point set q, are recognized through simple feature extraction. The database of the beacons P, namely, the coordinate of each beacon, is given.
- An initial value of the VNS pose is estimated (see Initialization section), and the iteration is initialized by setting $k = 0$.

a) *Generating the model point set*: the visible beacon set $P_0^k \subseteq P$ and the corresponding image point set p_0^k at the initial pose (R_0^k, t_0^k) are predicted based on the beacon position database P and the imaging model $p_0^k = g(R_0^k, t_0^k, P_0^k)$, where $g()$ stands for equation (3).

b) *Matching the nearest couples*: for each data point $q_i(u', v') \in q$ where $i = 1, 2, \ldots$, and n stands for the

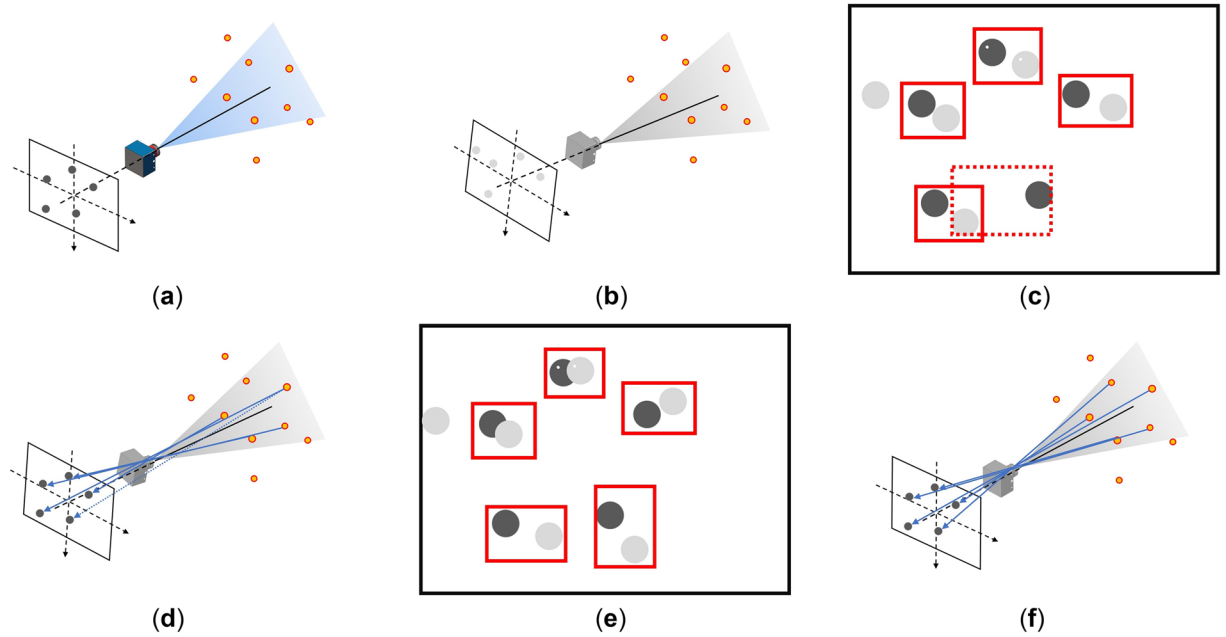**Figure 2.** ICIP algorithm process: (**a**)–(**f**) are processed circularly until convergent. (**a**) VNS captures a picture of current surroundings and beacon images are extracted. (**b**) Observed beacon images are predicted based on imaging model at current estimated initial pose. (**c**) Nearest point couples between real beacon images and predicted ones are matched, and a mismatch may occur during this step, which is marked by a dashed frame. (**d**) Measurement equation is formed based on imaging model by assuming that each point couple is the same image of beacon. Optimized solution of the pose is then calculated, and mismatched image is shown by a dashed-line arrow. (**e**) A newly observed beacon image is predicted at the previous optimized pose, and new nearest point couples are matched. As the optimized pose is closer to the true pose, a mismatch may be corrected, and (**f**) a new, more accurate optimized pose is calculated based on previous new point couples.

number of the visible beacon, as well as its image in the picture, find a $p_i^k(u, v) \in \boldsymbol{p}_0^k$ such that (s.t.) $d(p_i^k, q_i) = \min$.

c) *Estimating the transformation*: the rotation matrix $\boldsymbol{R}^k$ and the translation vector $\boldsymbol{t}^k$, s.t. $\sum_{i=1}^{n} d(\boldsymbol{g}(\boldsymbol{R}^k, \boldsymbol{t}^k, P_i^k), q_i^k) = \min$ are calculated, where $P_i^k(X_w, Y_w, Z_w)$ is the 3D coordinate of the corresponding beacon of $p_i^k(u, v)$.

d) *Applying the transformation*: $\boldsymbol{p}^k = \boldsymbol{g}(\boldsymbol{R}^k, \boldsymbol{t}^k, \boldsymbol{P}^k)$, where $\boldsymbol{P}^k = \{P_i^k, i = 1, 2, ..., n\}$.

e) *Examining convergence*: the iteration if $d^k - d^{k+1} < \tau$ is terminated, where $\tau > 0$ is a present threshold, and

$$d^k = \begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} d(p_i^0, q_i), \ k = 0 \\ \dfrac{1}{n}\sum_{i=1}^{n} d(\boldsymbol{g}(\boldsymbol{R}^k, \boldsymbol{t}^k, P_i^k), q_i), \ k = 1, 2, ..., n \end{cases}. \tag{4}$$

Otherwise, the transformation in step c is taken as a new initial value: $(\boldsymbol{R}^k, \boldsymbol{t}^k) = (\boldsymbol{R}_0^{k+1}, \boldsymbol{t}_0^{k+1})$, $k = k + 1$ and the process returns to step a. The process of the algorithm is shown in Fig. 2.

Convergence theory. *Theorem*: The iterative closest imaging point algorithm always converges monotonically to a local minimum estimated pose.

Proof: For the kth iteration, after ICIP step b, the mean pixel distance error of the closest image point couples e^k is represented in the following:

$$e^k = \frac{1}{n}\sum_{i=1}^{n} d(p_i^k, q_i). \tag{5}$$

The optimized mean distance between point couples d^k after step c is shown in equation (4). The case $d^k \leq e^k$ always occurs because the transformation is estimated to minimize the distance between the closest image point couples.
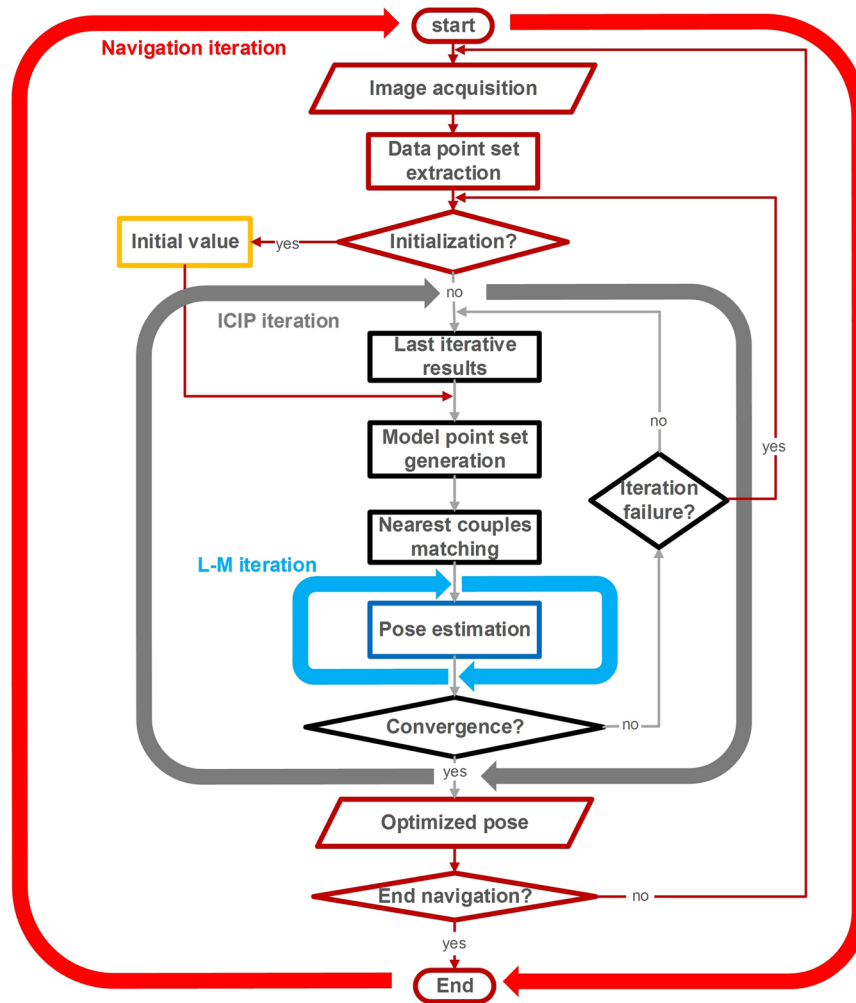
Figure 3. Navigation process based on ICIP algorithm.

During the subsequent iteration, a new point set $p_i^{k+1}(u, v)$ is obtained after step b, forming a new closest corresponding point set of q. Thus, the following is clear: $d(p_i^{k+1}, q_i) \leq d(g(R^k, t^k, P_i^k), q_i)$, $i = 1, 2, ..., n$. According to the definition, the image point distance are positive numbers. Therefore, $e^{k+1} \leq d^k$, and the following inequality is obtained:

$$0 \leq d^{k+1} \leq e^{k+1} \leq d^k \leq e^k \text{ for all } k. \tag{6}$$

The iteration will converge to the true value as long as the initial estimation is within tolerance.

ICIP-based navigation method. A continuous autonomous navigation method utilizing the ICIP algorithm is designed as shown in Fig. 3.

The entire navigation process consists of three levels of iterations. The third level is *L-M iteration* that conducts the pose estimation (marked in blue in Fig. 3). As the estimated pose is not accurate enough, it is taken as the initial value, and another loop containing model point set generation, nearest point couple matching, and transformation estimation is conducted until convergence, which is the second-level iteration: *ICIP iteration* (marked in gray in Fig. 3). The convergent pose estimation is considered the initial value for the next moment, and a new process of image acquisition, data point set extraction, and ICIP iteration is performed. This process is the first-level iteration, namely, *navigation iteration* (marked in red in Fig. 3).

For the first step of the navigation iteration, the initial value is required to estimate an approximate pose. Once the ICIP iteration fails due to accidental factors, the loop breaks and a new one starts from the initialization step. Note that the failure can be determined by the distance threshold and the iteration times. The main influencing factor for convergence of the navigation method is whether the initial value is within tolerance. For the ICIP iteration, if the initial value is not reasonable, falling into a local optimal solution is easy. For the navigation iteration, if the computation speed is slower than the moving speed of the system, the navigation may fail to continue. These problems are discussed in detail in the two following sections.

Initialization for ICIP.   An extended Kalman filter (EKF) framework proposed by Moreno-Noguer *et al*.[22] is employed to initialize the ICIP. The VNS pose with respect to the beacons are represented as a GMM. Constraining the initial pose of the IS within a certain range is reasonable, and the pose samples can be generated to train the GMM offline. Thereafter, 6-vector Gaussian components $\{\mathbf{x}_1, ..., \mathbf{x}_G\}$, which represents the pose and corresponding covariance matrix $\{\Sigma_1^x, ..., \Sigma_G^x\}$ are obtained. The initial pose $\boldsymbol{x}$ can be obtained by minimizing the following objective function:

$$Error(\boldsymbol{x}) = \sum_{(\boldsymbol{x}, \boldsymbol{P}) \in Matches} \|\boldsymbol{q} - \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{P})\| + \tau N_{nd}, \tag{7}$$

where *Matches* is a set of $(\boldsymbol{P}, \boldsymbol{q})$ pairs that represents the correspondence between the 3D beacons $\boldsymbol{P}$ and their image $\boldsymbol{q}$. $N_{nd} = |NotDetected|$ denotes beacons for which no match can be established. τ is a penalty term.

The pose search strategy is as follows: for each Gaussian component, the 2D projections $\boldsymbol{p}$ of the 3D beacons $\boldsymbol{P}$ are calculated. The potential match is considered in turn such that

$$(\boldsymbol{p}_i - \boldsymbol{q}_i)^T (\Sigma_i^p)^{-1}(\boldsymbol{p}_i - \boldsymbol{q}_i) \leq M^2 \tag{8}$$

where $\Sigma_i^p = \boldsymbol{J}(\boldsymbol{P}_i)\Sigma_g^x(\boldsymbol{J}(\boldsymbol{P}_i))^T$ indicates the error propagation, $\boldsymbol{J}(\boldsymbol{P}_i)$ is the Jacobian matrix of the projection $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{P}_i)$, and M is a confidence threshold. An EKF framework is then utilized to update the pose estimate and its covariance:

$$\boldsymbol{x}_g^+ = \boldsymbol{x}_g + \boldsymbol{K}(\boldsymbol{q}_j - \boldsymbol{g}(\boldsymbol{x}_g, \boldsymbol{P}_i))$$
$$\Sigma_g^{x+} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{J}(\boldsymbol{P}_i))\Sigma_g^x \tag{9}$$

Only three 3D-to-2D matches are needed to search the pose based on the imaging model. As the three orthogonal cameras form an optimal model, one match from each camera is selected to participate in the update in equation (9). After three match hypotheses, all the 3D beacons in sight are projected onto the image. The *Error* in equation (7) is computed by setting a distance on the image τ as the threshold for undetected beacons. The 3D beacons whose projection is farther apart from any 2D features than τ are considered as not detected.

This preceding process is repeated over the Gaussian components until a sufficiently small *Error* is obtained. In addition, as shown in Fig. 3, initialization is conducted again while ICIP fails to converge. This process is performed by training a new GMM online based on the recorded previous IS motion parameters. For instance, under non-highly dynamic conditions, the previous pose can be used to estimate a small region where the IS is most likely currently located. This region is regarded as a new, constrained initialization space for the IS. Pose sample within this space is randomly generated to train the GMM and the current pose initial can be found to continue the ICIP iteration.

Simulation Analysis.   To analyze the performance of the proposed navigation method, we conducted numerical simulation using MATLAB. Note that the performance varies under different beacon deployments and movement space conditions. Without loss of generality, the following discussion is based on the same beacon deployment as the IS ground testing system. The VNS is also limited to move inside this system, which is a simulated environment of the space station cabin of $1 \times 1 \times 1.5$ m³. The VNS utilizes a configuration of three orthogonal cameras. The system moves inside the given space, takes pictures of the environment, and calculates its own pose utilizing the proposed method, as shown in Fig. 1(c). The simulation conditions are provided in Supplementary Table S1.

Initial value sensitivity.   The initial value sensitivity of the ICIP algorithm presents two practical significances as previously mentioned. First, at the beginning of the navigation iteration, a reasonable initial estimated pose is required. Thus, the initial value sensitivity submits a request to the accuracy of that estimation. Second, as the optimized solution of the previous navigation iteration is the initial value of the current navigation iteration, the moving and rotational speed of the VNS are limited by the initial value sensitivity. In other words, the displacement of the VNS during the computing time must be within the initial value of allowable range.

Monte Carlo (M-C) simulation is conducted to obtain statistical results under different initial value error conditions. The "initial value error" refers to the difference in both attitude angle and position between the initial value and the true value, that is, $\Delta\Psi$ and Δr. As shown in the simulation result in Fig. 4, the success rate η decreases and the number of ICIP iterations, N_{ICIP}, increases as $\Delta\Psi$ and Δr increase. According to Fig. 4(a), a 20-component GMM is trained. The initialization results are shown in Fig. 4(c), revealing that the success rate of the initialization reaches 99.6% in a 10,000-trail M-C simulation. The same conclusion[28] that the initialization method is robust can be presented based on the simulation. After proper initialization, the pose is refined by ICIP, and the optimized solution is taken as the initial value of the next navigation iteration. Whether the navigation can be continuous mainly depends on the onboard operating speed of the algorithm, which is discussed in the following subsection.

Operating speed.   A typical process of the navigation iteration after image acquisition is composed of image transmission, image processing and feature extraction, model point set generation, closest point couple matching, and L-M iteration. The operating speed is evaluated by MATLAB R2014a running on a PC (Intel (R) Core (TM) i7-4900 CPU at 3.60 GHz, 8.00 GB RAM).

As shown in Fig. 5(d), the main time cost lies in the process of image processing and the L-M iteration. For each loop of the navigation iteration, the first two steps in Fig. 5 are performed once, and the last three steps that form the
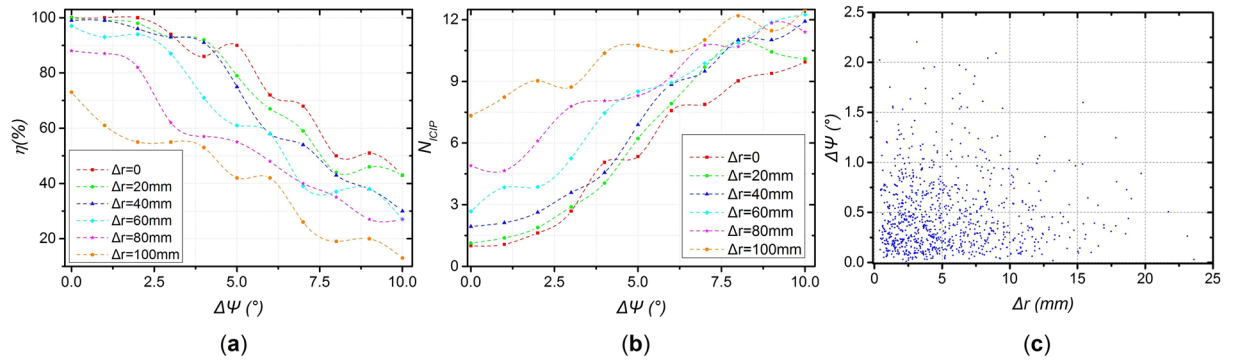
Figure 4. (**a**) Success rate η and (**b**) number of ICIP iterations, N_{ICIP}, vs. displacement of VNS. (**c**) Initialization M-C simulation results.
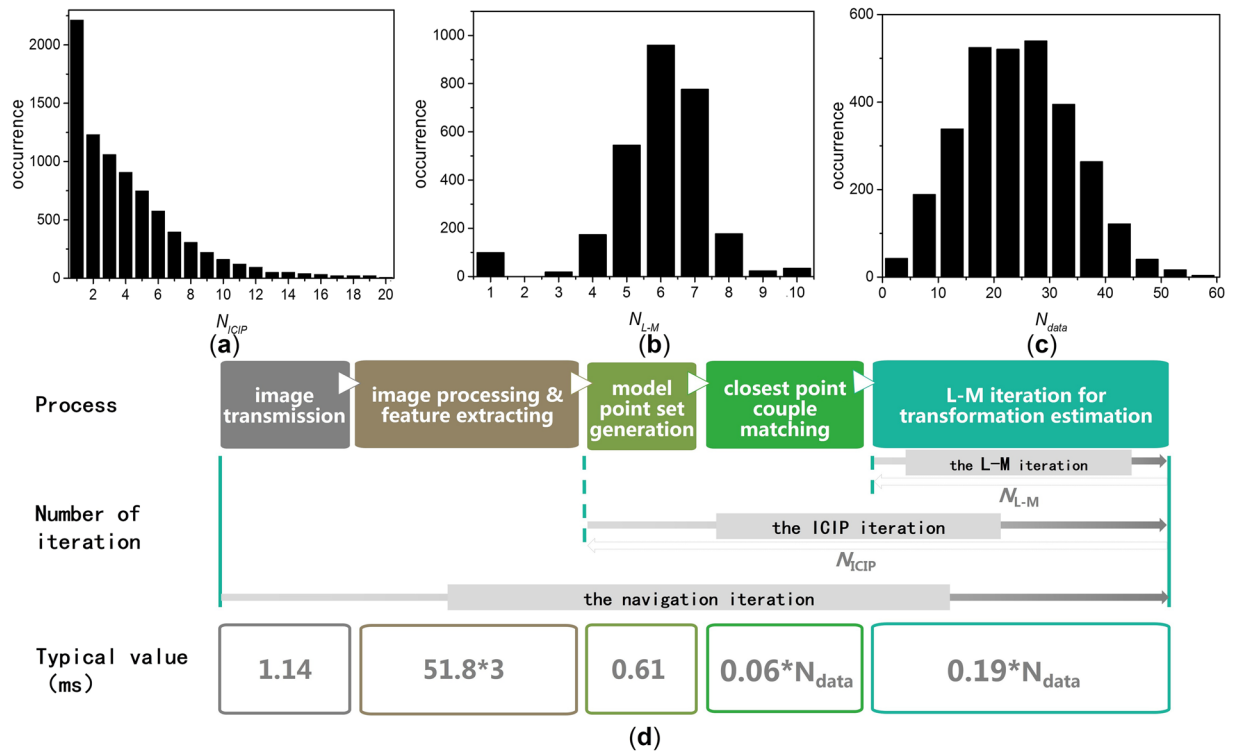

Figure 5. Statistics of (**a**) number of ICIP iteration time N_{ICIP}, (**b**) number of L-M iteration time N_{L-M}, and (**c**) number of in-sight beacons N_{data}. (**d**) is the time taken by each process of ICIP algorithm-based navigation.

ICIP iteration are performed N_{ICIP} times depending on the initial value accuracy as shown in Fig. 4(b). The time cost of the closest point couple matching step is proportional to the number of in-sight beacons N_{data}. The duration time of the L-M iteration is proportional to N_{data} and the number of L-M iteration time N_{L-M}. Under the given beacon deployment, the mean values of N_{data}, N_{ICIP}, and N_{L-M} are 24.64, 4.06, and 6.01, respectively. The statistical results of these numbers are presented in Fig. 5(a–c). Thus, the time of one navigation iteration loop is approximately 279.25 ms. Under this condition, the navigation system can provide continuous service when the movement is within 7 deg/s and 70 mm/s. The L-M and ICIP iterations consume 40.9% and 43.9% of the total time, respectively.

Accuracy. Accuracy is one of the most important performance indices of a navigation system. In this study, the navigation accuracy is mainly determined by the camera configuration and beacon deployment. Autonomous navigation of IS is simulated while the $\Delta\Psi$ and Δr are within the allowable range. In the simulation results shown in Fig. 6, the accuracy of the three-axis Euler angles is within 0.06° (1σ) and the three-axis position accuracy is within 1.8 mm (1σ). This result indicates that the proposed method fits the requirement of the IS navigation.
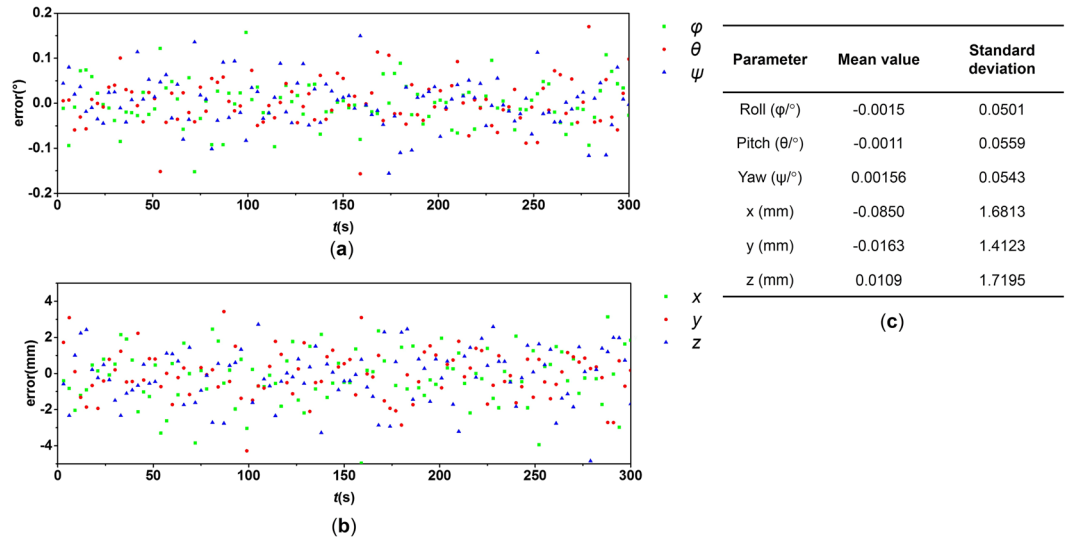
Figure 6. Simulation results of ICIP method: (**a**) attitude error, (**b**) position error, and (**c**) result datasheet.
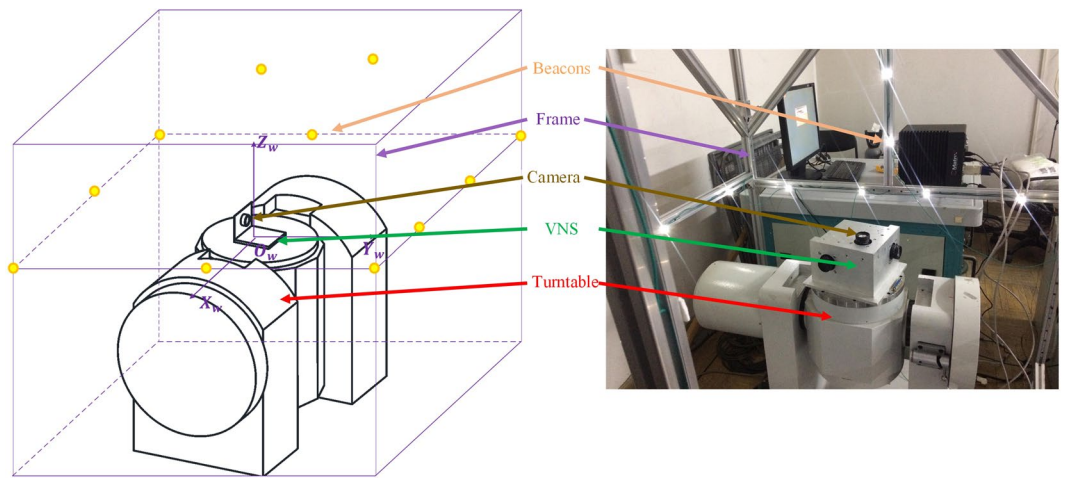

Figure 7. Architecture of navigation experiment.

Robustness. On the basis of the imaging model equation (2), the 6-DOF pose can be calculated as long as ≥ 3 beacons are observed. Simulation results show that the navigation method works properly when manually blocking some of the beacons. Thus, the proposed ICIP is robust to occlusion. This property can be considered as an advantage to reduce the computation load by actively choosing only some of the observed beacons to form the imaging model. As shown in Fig. 5(d), the time cost can be reduced when N_{data} is reduced. However, simulation results also indicate that this improvement in efficiency is at the cost of accuracy because lesser beacons in Eq. 2 lead to larger error propagation. The outliers may cause a convergence problem for the ICP-based algorithm. Therefore, proper design and extraction of beacons is necessary, as previously discussed. Finally, continuous navigation is achieved by obtaining the last estimated pose as the initial value of the current iteration. Thus, once the pose estimation fails, the navigation has to break off. To solve this problem, we design the proposed method to immediately start a new cycle by estimating a new initial value for the current pose and run into the iteration loop (see Fig. 3). As a high initialization success rate can be achieved using the GMM, the navigation can continue in a stable manner. In summary, ICIP algorithm-based navigation is robust to beacon blocking, feature extraction error, and cyclic failure.

Navigation Experiment. A demonstration experiment is conducted for further verification. The architecture is shown in Fig. 7.

All coordinates involved in the navigation experiment are right-handed Cartesian coordinates described as follows:

1. The VNS body coordinate $O_b X_b Y_b Z_b$ is fixed on the system frame and defined for easy use.
2. In the turntable coordinate $O_t X_t Y_t Z_t$, O_t is the center of the turntable and Z_t and X_t point to the forward of the turntable main and auxiliary axes, respectively.
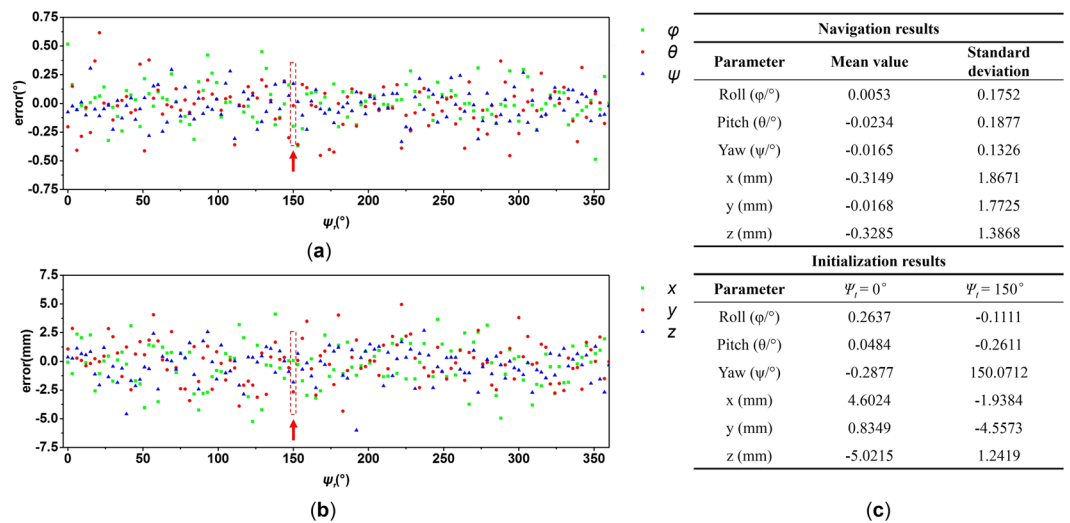
Figure 8. Results of navigation experiment: (**a**) attitude error, (**b**) position error, and (**c**) result datasheet. The red arrow in (**a**) and (**b**) denotes the moment when pose estimation failure is manually added to start and test a new initialization. The corresponding results can be seen in (**c**).

3. The world coordinate $O_wX_wY_wZ_w$ is defined by the beacons and represents the cabin coordinate where the IS operates.
4. The camera coordinates X, Y, and Z are defined based on the imaging model in the *imaging model* section and recorded as $O_xX_xY_xZ_x$, $O_yX_yY_yZ_y$, and $O_zX_zY_zZ_z$, respectively.

LED beacons are fixed on the steel frame surrounding the turntable. The experiment is designed as follows. First, the origin of the world coordinate is defined as the center of the turntable platform, which is the origin of the VNS body coordinate in the original pose. The three axes of the world coordinates X_w, Y_w, and Z_w are defined parallel to the axes of the turntable coordinates X_t, Y_t, and Z_t, respectively. Subsequently, the position of the beacons is measured using a theodolite. Second, the turntable is controlled to rotate around the Z_t axis, whereas the VNS moves with the turntable, takes pictures of the surrounding environment, and calculates the pose based on the beacons observed. Finally, on the basis of the definition of the coordinates, the position of the turntable platform center does not move during the rotation. The true value of the VNS position is (0, 0, 0) and the true attitude is (0, 0, ψ_t), where ψ_t is the angle given by the turntable around its Z_t axis. Therefore, the pose error is obtained as the difference between the calculated pose and the true value.

The VNS parameter calibration is conducted using a checkerboard-fixed post-processing calibration, details of which are presented in our previous study[1]. The calibrated parameters are shown in the Supplementary Tables S2 and S3. We use the 902E-1 two-axis testing turntable produced by Beijing Precision Engineering Institute for Aircraft Industry (Beijing, China). The testing turntable is aligned before the experiment, and the accuracy of its angular position is 8″ in the direction of both axes. The cameras integrated in the VNS are Daheng Image DH-HV1310FM (with resolution of 1280 × 1024, and 18.6 frames under the highest resolution). The theodolite used is Leica TM6100A with an accuracy of 5″ for both horizontal and vertical angles. Navigation results are shown in Fig. 8.

Discussion
Experiment results, ways of improvement, and future work are discussed as follows:

1) The experiment results show that the navigation method works properly during the validation. The stand deviations of the three-axis position error are 1.87, 1.77, and 1.39 mm, and the stand deviations of the three-axis attitude error are 0.18, 0.19, and 0.13°, respectively. In this study, only the connective region size and roundness are utilized to justify the reality of the beacon image. Compared with conventional vision navigation based on feature matching, the feature extraction and match process is much simpler, and the required number of beacon images is much smaller than that of the feature points. In particular, pose estimation failure is manually added when the turntable rotates to $\psi_t = 150°$. Thereafter, another initialization is conducted and the system continues to navigate properly. Therefore, the advantages of low computational complexity, favorable stability, and applicability in an extremely simple environment are verified.
2) The deployment of the navigation beacons significantly influences the accuracy of the system. Thus, future work aims to focus on the beacon deployment strategy and optimization methods. Although theodolite is used because of its high precision, the coordinate error of the beacon cannot be ignored and leads to a measurement error. The concrete operation of deployment and measurement of the beacons will also be the focus. The 3D coordinates of the beacons are essential in this case because the pose with

respect to the given frame, that is, the space station cabin frame, is required. As the beacons are measured in the given frame, they establish contacts with the VNS through the imaging process. However, in many applications, only relative positioning is required. The features of the environment need not be measured beforehand but recognized and recorded through machine learning by the VNS. The proposed ICIP algorithm also offers potential application in these cases by using the environment features as the navigation beacons.

3) A promising way to improve the system performance is to apply integrated navigation technology, which we will study in future work. Through a combination of a micro inertial measurement unit (MIMU), the dynamic performance and refresh rate can be improved. Therefore, the movement permissible range of the VNS will be widely extended. The vision navigation results can help control the drift error of the MIMU. Furthermore, the inertial parameters can also help predict the position of the beacon images, thereby hastening the search process and reducing the image process complexity.

Conclusion

Inspired by the requirement of IS navigation, the ICIP algorithm is proposed in this paper. Based on the imaging model and beacon database, the optimized pose of the VNS can be obtained by the algorithm. The process is similar to adjusting the VNS pose so that it observes images that are most analogous to the real scene. A navigation method is introduced, which utilizes the ICIP algorithm to calculate the current pose and takes it as the initial value of the subsequent iteration. A GMM-based EKF framework is employed to initialize the navigation at the beginning and the failure occurrence moment. Thus, a continuous autonomous navigation is realized. A navigation experiment conducted in the IS ground testing system validates the method. The proposed work fits the navigation inside a cooperative space and has potential for application in unknown environments.

Data availability statement. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Shi, S. *et al*. Error analysis and calibration method of a multiple field-of-view navigation system. *Sensors*. **17**(3), 655 (2017).
2. Want, R., Hopper, A., Falcao, V. & Gibbons, J. The active badge location system. *ACM T. Inform. Syst.* **10**(1), 91–102 (1992).
3. Nolet S. The SPHERES navigation system: from early development to on-orbit testing. AIAA Guidance, Navigation and Control Conference and Exhibit. 20–23 (2007).
4. Curran, K. *et al*. An evaluation of indoor location determination technologies. *J. Location Based Serv.* **5**(2), 61–78 (2011).
5. Castro P., Chiu P., Kremenek T. & Muntz R. A probabilistic room location service for wireless networked environments. Ubicomp 2001: Ubiquitous Computing. Springer Berlin/Heidelberg (2001).
6. Rudol P., Wzorek M. & Doherty P. Vision-based pose estimation for autonomous indoor navigation of micro-scale unmanned aircraft systems. Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE: 1913–1920 (2010).
7. Tawdross P. & König A. Feasibility study of a novel bio-inspired location sensor concept for indoor location based services in ambient intelligence applications. Intelligent Solutions in Embedded Systems, 2005. Third International Workshop on. IEEE: 62–69 (2005).
8. Serrão, M., Rodrigues, J. M. F. & Buf, J. M. H. D. Navigation framework using visual landmarks and a GIS. *Procedia Comput. Sci.* **27**, 28–37 (2014).
9. Horn B. *Robot Vision*. (MIT press, 1986).
10. Mccarthy C. & Bames N. Performance of optical flow techniques for indoor navigation with a mobile robot. Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on. IEEE. 5: 5093–5098 (2004).
11. Paz, L. M., Piniés, P., Tardós, J. D. & Neira, J. Large-scale 6-DOF SLAM with stereo-in-hand. *IEEE T. Robot.* **24**(5), 946–957 (2008).
12. Fei, W., Jin-Qiang, C., Ben-Mei, C. & Tong, H. L. A comprehensive UAV indoor navigation system based on vision optical flow and laser FastSLAM. *Acta Automatica Sin.* **39**(11), 1889–1899 (2013).
13. Lepetit, V., Moreno-Noguer, F. & Fua, P. Epnp: An accurate O(n) solution to the pnp problem. *Int. J. Comput. Vis.* **81**, 155 (2009).
14. Gao, X. S., Hou, X. R., Tang, J. & Cheng, H. F. Complete solution classification for the perspective-three-point problem. *IEEE trans. Pattern Anal. Mach. Intell.* **25**, 930–943 (2003).
15. Belongie, S., Malik, J. & Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE trans. Pattern Anal. Mach. Intell.* **24**(4), 509–522 (2002).
16. Xiao, R. *et al*. Shape context and projection geometry constrained vasculature matching for 3D reconstruction of coronary artery. *Neurocomputing*. **195**, 65–73 (2016).
17. Scott, G. L. & Longuet-higgins, H. An algorithm for associating the features of two patterns. *Proceedings: Biological Sciences. The Royal Society, London.* **244**, 21–26 (1991).
18. Leordeanu, M. & Hebert, M. A spectral technique for correspondence problems using pairwise constraints. *The Tenth IEEE International Conference on Computer Vision.* **2**, 1482–1489 (2005).
19. Besl, P. J. & McKay, N. D. A method for registration of 3-D shapes. *IEEE trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992).
20. Pomerleau, F., Colas, F. & Siegwart, R. A review of point cloud registration algorithms for mobile robotics. *Found. Trends. Robot.* **4**(1), 1–104 (2015).
21. Low K. L. *Linear Least-Squares Optimization for Point-to-Plane ICP Surface Registration*. Chapel Hill, University of North Carolina 4 (2004).
22. Moreno-Noguer, F., Lepetit, V. & Fua, P. Pose priors for simultaneously solving alignment and correspondence. *Computer Vision–ECCV* **2008**, 405–418 (2008).
23. Serradell E., Özuysal M., Lepetit V., Fua P. & Moreno-Noguer F. Combining geometric and appearance priors for robust homography estimation. European Conference on Computer Vision. Springer, Berlin, Heidelberg: 58–72 (2010).
24. Sánchez-Riera J., Östlund J., Fua P. & Moreno-Noguer F. Simultaneous pose, correspondence and non-rigid shape. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE: 1189–1196 (2010).
25. Rusinkiewicz S. & Levoy M. Efficient variants of the ICP algorithm. International Conference on 3D Digital Imaging and Modeling (3DIM): 145~152 (2001).

26. Gold, S. *et al*. New algorithms for 2D and 3D point matching: Pose estimation and correspondence. *Pattern Rec.* **31**(8), 1019–1031 (1996).
27. Fitzgibbon, A. W. Robust registration of 2D and 3D point sets. *Image Vision. Comput.* **21**(13), 1145–1153 (2003).
28. Jayarathne U. L., McLeod A. J., Peters T. M., & Chen E. C. S. Robust intraoperative US probe tracking using a monocular endoscopic camera. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg: 363–370 (2013).

Acknowledgements

Author Contributions

S.S. proposed the idea of the ICIP algorithm, the navigation method and the simulation analysis; S.S., C.O. and Y.C. performed the experiments and analyzed the data; K.Z. and Z.Y. conceived and supervised the experiments; S.S. wrote the paper; Z.W. worked on the navigation initialization and the proofreading.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-17768-2.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.