An Automatic Tracking Framework for Augmented Reality Induced Maintenance Applications of Chinese Space Station

ZHAO Xin[1,3], *WANG Yue*[2*], *FU Hongyong*[1*]

1. Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, P. R. China;
2. School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China;
3. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, P. R. China

(Received 19 December 2023; revised 3 April 2024; accepted 20 April 2024)

Abstract: Tracking registration is a critical technology of augmented reality (AR). Marker-based and image-based registration methods have been widely used in augmented assembly systems. However, the space station encounters distinctive challenges due to weak product texture, symmetrical structure, limited data, and difficulty in attaching markers. This paper presents a novel 3D object tracking method specifically designed for induced maintenance applications in the Chinese space station, aiming to replace traditional paper manuals with the AR technology to provide astronauts with more intuitive operational guidance. We propose a marker-less approach for intelligent maintenance of the Chinese space station. A point pair feature method that combines curvature information for improving efficiency is employed to estimate the initial frame pose, obviating the need for manual adjustment to achieve the corresponding pose. This is crucial for astronauts, as precise movement in the microgravity environment is a significant challenge. Precise restrictions on the position of astronauts are eliminated, and the tracking robustness of space symmetric products is further enhanced by incorporating both texture and region information. The method utilizes both point cloud and 2D image. It leverages point cloud matching to estimate the initial pose of the first frame and recalculates the pose after loss of tracking. Once the pose of the previous frame is obtained, the tracking is calculated solely according to the region and texture information of the 2D image to obtain real-time tracking. The experimental results show that our method has the same pose trend as the marked-based method. And the error difference between the measurements based on electronic vernier calipers is in the millimeter level. The successful replacement of filters using specialized tools designed specifically for space products demonstrates the practicality and potential of implementing induced maintenance procedures on space station.

Key words: augmented reality (AR); pose tracking; CAD model; Chinese space station

CLC number: V267.9 **Document code**: A **Article ID**: 1005-1120(2024)03-0387-16

0 Introduction

The Chinese space station has established 8 strategic directions and 31 research topics for space science and applications, along with 13 scientific experiment cabinet and other tasks. There is a need for on-orbit assembly, on-orbit deep maintenance, and complex scientific experiment operations. The complexity, diversity, and rapid response of modern manned space equipment technology have made to the possibility that on-orbit maintenance and scientific experiment operations may need to be completed accurately and quickly in the absence of sufficient technical support conditions.

At the same time, the complexity and diversity of manned space equipment, extreme space environments, and long-term residency in orbit impose additional requirements for decision-making in mainte-

*Corresponding author, E-mail addresses: wang.yue@cqupt.edu.cn; fuhongyong@csu.ac.cn.

How to cite this article: ZHAO Xin, WANG Yue, FU Hongyong. An automatic tracking framework for augmented reality induced maintenance applications of chinese space station[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2024, 41(3): 387-402.

http://dx.doi.org/10.16356/j.1005-1120.2024.03.010

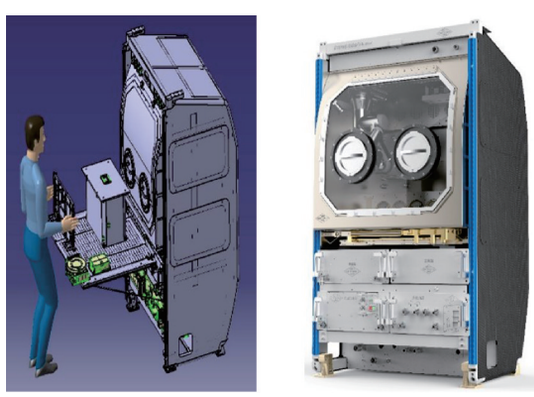
nance and scientific experimental operations. To effectively tackle these challenges, a combination of information augmentation technology and intelligent technology can be used to provide maintenance operator with direct, real-time synchronization, and precise operational guidance. Induced maintenance technology involves the use of augmented reality (AR) and information technology to efficiently guide maintenance.

The Chinese space station has been completed in 2022 and will be in orbit for a period of ten years. During this period, a large number of scientific experiments and space applications will be conducted. The prolonged operation missions of astronauts in orbit inevitably encounter equipment failures and engage in maintenance operations. In the realm of maintaining manned space equipment and conducting scientific experiments, astronauts have traditionally relied on offline guidance such as paper manuals, electronic manuals, and images for in-orbit training and operational task completion. However, this approach has resulted in low operational efficiency, high error rates, and significant psychological pressure on astronauts. Consequently, it seriously impacts the accuracy, timeliness, and reliability of astronaut operations.

Induced maintenance refers to the process of perceiving the changes in the surrounding scene and maintenance object state through natural interactive means, and real-time overlaying digital maintenance instructions onto physical objects and processes. This enables personnel to receive operational guidance and warnings through natural perception.

AR technology has emerged since the 1960s, and with the rapid development of computer processing and imaging capabilities, wearable augmented reality devices have become feasible. They have been widely applied in maintenance, assembly, and training fields, as well as making significant strides in aerospace domain. For instance, the European Space Agency (ESA) proposed the wearable augmented reality (WEAR) [1] system. The German Aerospace Center has developed the mobile AR for space operation procedure (MARSOP) system and Microsoft collaborated with NASA on the Sidekick project, all of which have been or are about to be tested and validated on the International Space Station.

To facilitate the operation of scientific experiments and perform maintenance and other tasks in orbit, our department has specially deployed an on-orbit maintenance and manipulation workbench (MMW)[2], as shown in Fig.1. At present, the maintenance of Chinese space stations still relies on traditional methods. Our team has also proposed the use of AR for induced maintenance on space stations and verified the superiority[3]. The tracking registration technology of 3D targets is one of the key technologies. However, several challenges persist in its application within the space station.



(a) The usage status of MMW (b) The folding status of MMW

Fig.1 Usage status of MMW and folding status of MMW

Firstly, the marker-based approaches necessitate artificially designed identifiers which are challenging to implement and environmentally damaging in practical applications. Moreover, the spatial devices exhibit weak texture surfaces, which result in suboptimal tracking outcomes when relying on feature-point based tracking methods. Additionally, edge-based approaches tend to fail when the background is cluttered. Subsequently, deep learning-based approaches necessitate a large amount of training data, whereas our space station dataset is limited in size and is not readily available. Finally, the utilization of foreign development kits has resulted in the disclosure of product information.

To address the current challenges, we propose an automated tracking framework for space station

maintenance. The only requirement is the CAD model of the object, without a large amount of training data. Furthermore, there is no necessity for any preprocessing of the scene. The main contributions of the proposed method are:

(1) A new point cloud matching method is employed to estimate the initial frame pose, obviating the need for manual adjustment to achieve the corresponding pose. Precise restrictions on the position of astronauts are eliminated, which is crucial for astronauts, as precise movement in the microgravity environment is significant challenge. And the tracking robustness of space symmetric products is further enhanced by incorporating both texture and region information.

(2) We propose a marker-less approach for intelligent maintenance of the Chinese space station, marking the pioneering utilization of such an approach in Chinese space stations. Also, we demonstrate the practicality of implementing induced maintenance procedures on the space station by replacing the filter.

1 Related Work

1.1 Induced maintenance application

Aviation equipment is a technology intensive product with a sophisticated internal structure. Traditionally, relying on paper or electronic manuals for training, maintenance, assembly, disassembly, and other tasks has been time-consuming, error prone, and costly. However, the emergence and development of augmented reality technology have made it possible to integrate AR characteristics into induced maintenance.

Induced maintenance utilizes AR technology to superimpose induced information onto real scenarios, thereby providing enhanced guidance and support to maintenance operators, whether they are professionals or untrained personnel. This eliminates the need for operators to consult paper or electronic manuals by offering intuitive 2D/3D information, resulting in an improved operating experience for maintenance personnel. The benefits encompass reducing content consultation time, minimizing manual and equipment switching, heightening focus on maintenance tasks, diminishing errors resulting from human factors such as missed operations, enhancing efficiency and operational quality.

In 1992, Boeing developed an augmented reality application for manual manufacturing processes[4]. In 1993, Feiner et al.[5] introduced the KARMA system, an augmented reality solution that utilizes virtual information to facilitate printer maintenance and the installation and operation of automobile door lock. The maintenance mission of the Hubble telescope was successfully executed in the same year. Subsequently, analysis of the survey data collected from the trainees demonstrated that training using virtual reality simulators significantly enhanced their operational efficiency during real flight missions. In 2003, the successful development of the German Starmate[6] system Arvika[7] heralded the immense potential of AR technology in maintaining and assembling complex electromechanical systems. In 2009, WEAR was developed by the ESA, facilitating astronauts in conducting air filter inspections through augmented reality assistance. In 2012, the German Aerospace Center developed a system called MARSOP for payload activities within the space station module. The maintenance and disassembly experiment conducted at the European Astronaut Center in Cologne, Germany, using the MARSOP system bioscience experimental cabinet demonstrated that employing an augmented reality maintenance induction system can significantly enhance astronauts' maintenance efficiency. Moreover, astronauts experience reduced physiological burden themselves, while complex maintenance tasks exhibit substantial reduction in completion time. In 2007, a collaborative research was taken by six organizations, including the United States Air Force Joint Air Force Research Laboratory and Columbia University, to investigate augmented reality for maintenance and repair (ARMAR) projects. The prototype system completed in 2011 can be utilized for auxiliary maintenance of wheeled armored vehicles[8]. In 2013, Benbelkacem et al.[9] from the Algerian Advanced Technology Development Center developed a photovoltaic water pump mainte-

nance system based on augmented reality technology to address the challenges of water pump maintenance in photovoltaic solar systems in re-mote areas. In 2015, Boeing conducted a controlled experiment to analyze the impact of AR technology on mitigating assembly errors. Through this empirical study, Boeing firmly posited that AR technology holds potential in reducing training time, expediting manufacturing speed, and alleviating task-switching difficulty for workers. NASA initiated the "Project Sidekick" project in 2016 with the objective of providing support to maintenance personnel for executing associated tasks aboard the International Space Station. In 2021, NASA utilized the Microsoft Hololens in conjunction with Sidekick to provide astronauts with augmented reality assistance for space equipment maintenance and repair tasks, enabling them to independently maintain the T2 treadmill without ground support.

Although domestic research on related content started relatively late, it has currently flourished. In 2008, Nanjing University of Aeronautics and Astronautics[10] conducted a series of comprehensive studies on maintenance guidance systems based on augmented reality. They successfully developed a prototype system for augmented reality maintenance guidance using the WDZ-1 turbine generator. In 2016, Cui et al.[11] employed a modular design approach to develop an induced maintenance prototype system, which was utilized for inducing two maintenance tasks involving engine piston ring and clutch plate replacement, thereby substantiating the system's functionality. In 2017, Beijing Jiaotong University implemented augmented reality technology in the auxiliary maintenance of high-speed trains and developed the CRH380BL high-speed train augmented reality auxiliary maintenance system to facilitate maintenance personnel in accomplishing their tasks and enhance daily work efficiency. In 2019, Lenovo implemented AR-assisted assembly solutions for large aircraft manufacturing of Commercial Aircraft Corporation of China Ltd (COMAC). This solution is currently operational on the production line of the fourth C919. In 2021, the "5G+AR Remote Technical Support Platform", jointly developed by China Southern Airlines and China Telecom was officially launched in Chengdu and Kunming. This intelligent platform facilitates expert consultations through both individual connections and multi-person conversations. During the troubleshooting guidance process, on-site maintenance personnel exclusively utilize AR glasses, ensuring seamless real-time communication without the need for phone distractions. Consequently, hands are freed up while focus and troubleshooting efficiency are significantly enhanced.

1.2 Augmented reality tracking technologies

In 2005, Regenbrecht et al.[12] used a marker-based approach to track the user's position and orientation for engine maintenance. They attached markers to the location near the engine. The use of multiple markers at well-defined positions will provide precise tracking and the mobile AR system at Embry-Riddle Aeronautical University (ERAU) uses markers that are placed on the aircraft and on aircraft components[13]. In 2011, Crescenzio et al.[14] developed an AR interface that can obtain reference images of objects and extract local invariant features. Then, they store the extracted features in the system for future use. During the online process, the system continuously processes the video stream from the camera and processes each incoming frame. If it extracts local invariant features, such as scale-invariant feature transform (SIFT) and speeded-up robust features (SURF), from frames and matches them with the features of the reference image and the system finds a sufficient number of matches, it concludes that the camera can see the object[14]. In 2013, Radkowski et al.[15] associated the visual features with the constructed scene template library, then the pose calculation of the mobile AR system was achieved. In 2014, Jo et al.[16] proposed a framework for augmented reality and knowledge-based systems in maintaining aircraft. They adopted three algorithms, the augmented feature (AF), the self-similarity image matching, the geometry-pixels method to provide real-time and robust performance. In 2021, Zhang et al.[17] used image edge inclination to weight the pixel values of the

distance image, improved the accuracy of image matching calculation, and used taboo search algorithm to continuously track camera pose, ensuring the real-time performance of enhanced assembly tracking and registration. In 2024, Nazri et al.[18] used the Vuforia SDK for repairing Boeing 737.

The task of feature point extraction and description is to identify pixels that exhibit significant dissimilarities in appearance compared to other pixels. Feature point extraction encompasses techniques such as the Harris[19] corner detector and FAST[20] corner detector, while feature point descriptions involve methods like SIFT[21], SURF[22], ORB[23], BRISK[24], and KAZE[25].

The marker-based method is easy, but not applicable. Space devices exhibit weak textured surfaces, which leads to suboptimal tracking results using feature-based methods. In addition, edge-based methods are efficient, but they often fail when the background is cluttered. Finally, there are device information security issues when using foreign development kits and it does not have initial pose.

Currently, region-based approaches are gaining popularity due to their ability to achieve robust tracking in complex backgrounds using only monocular images. However, the performance of these algorithms is adversely affected when dealing with symmetrical objects due to their reliance on model contours. To address this limitation, we propose a novel approach that combines texture and region features to enhance tracking performance. In this paper, we present an extensive overview of commonly employed features and provide a comprehensive introduction to state-of-the-art region-based methods.

The first real-time region-based 3D object tracking method, PWP3D[26], was implemented by Prisacarius and Reid in 2012. In 2016, Tjaden et al.[27] employed a Gaussian Newton-like pose optimization strategy to effectively address rapid rotation and scaling changes, as opposed to utilizing a step-wise optimization approach. This novel strategy resulted in significant enhancements in terms of tracking robustness. Hexner et al.[28] proposed employing multiple local background color statistical models instead of the global background color statistical model utilized in PWP3D, aiming to enhance the accuracy of foreground and background discrimination and improve the tracking stability of PWP3D in complex backgrounds. In 2017, Tjaden et al.[29] proposed a local color histogram based on temporal consistency to enhance the accuracy of the foreground probability map. Building upon this, they developed an error recovery method utilizing fast template matching. This approach enables rapid repositioning and restoration of the system to achieve precise pose tracking when an object reappears in the camera's field of view due to either leaving or being occluded. This further enhances the robustness of 3D tracking. In 2018, Tjaden et al. proposed RBOT[30] and systematically derived the Gaussian-Newton optimization scheme. This approach exhibits significant potential for accelerating convergence speed and represents the first solution to address simultaneous camera and object motion in clutter scenes across common and critical scenarios. In 2020, Stoiber et al.[31] proposed a novel and highly efficient sparse method RBGT, which exhibits an impressive one order of magnitude improvement in region-based 6DoF target tracking speed. In 2022, Stoiber et al. developed SRT3D[32], which introduced smooth step functions incorporating global and local uncertainties. Additionally, a pre-rendered sparse viewpoint model was utilized to establish a joint posterior probability of object pose. The maximization of the function was achieved through second-order Newton optimization and Tikhonov regularization techniques. Through multiple experiments, this algorithm has demonstrated significant advancements in terms of both runtime efficiency and quality, particularly excelling in handling noisy and clutter images encountered in real-world scenarios. They also developed iterative corresponding geometry (ICG) by incorporating depth information[33], although the inclusion of additional depth information posed limitations on its applicability across diverse scenarios. Tian et al.[34] proposed an adaptive local optimization method and a non-local optimization method for enhancing the estimation of object pose in scenarios involving large displacements. The ICG algorithm was further extended by Stoiber et al. in 2023, incorporat-

ing model texture information and introducing ICG+[35], which resulted in significant performance enhancements for the algorithm. And Wang et al.[36] proposed a learning-based active contour model to achieve the state of art tracking.

2 Proposed Approach

The tracking method for AR-induced maintenance is described in this section.

2.1 Overview

The proposed algorithm's overall workflow is illustrated in Fig.2, which comprises three steps: Input, pose estimation, and pose tracking. The input includes image frames captured by the depth camera and the CAD model in need of repairing. In the pose estimation phase, color images and depth images obtained from the depth camera are utilized to generate the point cloud of the real scene. Subsequently, point pair features are employed to compute the transformation between the environment point cloud and CAD model converted point cloud. During pose tracking phase, only color images from the depth camera and previously estimated pose that comes from the initialization phase are used, and region and texture information is leveraged for calculating poses. Initialization is applied when either it is first frame or tracking has been lost, otherwise, tracking method is implemented.

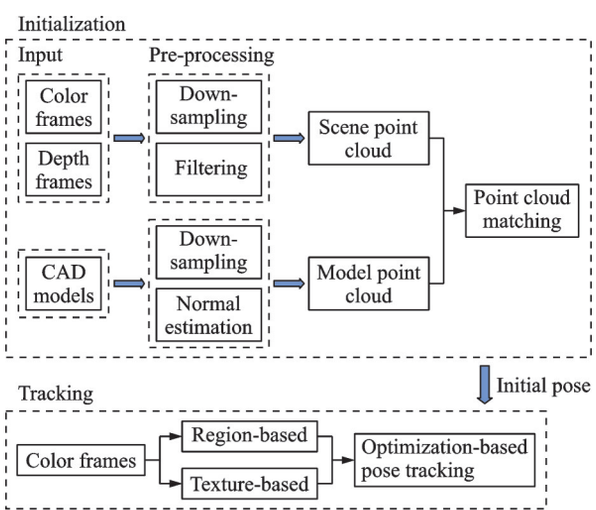
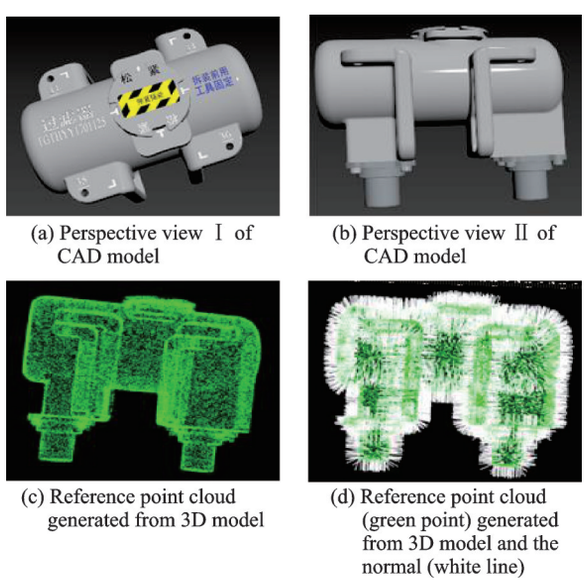


Fig.2 Overall workflow of the proposed algorithm

2.2 Initialization

2.2.1 Maintenance object model point cloud generation

A CAD model is a geometry composed of points and their triangular surfaces. The maintenance object model point cloud $P=\{P_0, P_1, P_2, \cdots, P_m, \boldsymbol{n}_0, \boldsymbol{n}_1, \boldsymbol{n}_2, \cdots, \boldsymbol{n}_n\}$ is generated from its 3D model in a computer-aided design system, where $\{P_0, P_1, P_2, \cdots, P_m\}$ and $\{\boldsymbol{n}_0, \boldsymbol{n}_1, \boldsymbol{n}_2, \cdots, \boldsymbol{n}_n\}$ represent the 3D sampling points and their corresponding normal vectors, respectively. We uniformly sample the model at a fixed distance step using the PCL Library's pcl_mesh_sampling function. In this way, we can obtain a point cloud with the geometric information of the model. The normal vector at each sampled point is determined based on the fitting plane formed by the point and its neighboring points. At the same time, by adjusting the distance step, the number of model point clouds can be reduced to facilitate subsequent calculations. Taking filter as examples, sample evenly on the surface of the model to obtain the sampling points and corresponding normal vectors are shown in Fig.3.



(a) Perspective view Ⅰ of CAD model

(b) Perspective view Ⅱ of CAD model

(c) Reference point cloud generated from 3D model

(d) Reference point cloud (green point) generated from 3D model and the normal (white line)

Fig.3 Point cloud of filter generation

2.2.2 Maintenance scene point cloud generation

The advantage of the depth camera is that it can obtain the depth image corresponding to the color image. The value of each pixel in the depth image is the actual distance from the camera to the object.

At any time k, we obtain a depth image $D_k(u)$, and the value of each pixel $u_D=(x_d, y_d)$ in the depth image $D_k(u)$ is the actual distance from the camera to the object. To obtain the environment scene point cloud $V_k(u)$, we back-project the depth values into the sensor's image frame according to

$$V_k(u) = D_k(u)\boldsymbol{K}^{-1}[u_D, 1] \tag{1}$$

where $\boldsymbol{K}$ is the camera's intrinsic matrix, and $\boldsymbol{K}^{-1}$ the inverse of $\boldsymbol{K}$. The method for calculating the normal vector of environment scene point cloud is identical to that used for model point cloud.

2.2.3 Point cloud feature matching

As the point pair feature algorithm is highly efficient and outperforms other methods in occluded and cluttered environments[37], our approach is built upon the original point pair feature (PPF) method[38], which we refer to as Drost-PPF in this paper. In this section, to enhance comprehension of this article, we begin by providing a comprehensive overview of the Drost-PPF approach and subsequently introducing our major ideas.

2.2.3.1 *Drost-PPF method*

(1) Point pair feature

The Drost-PPF utilizes four parameters to describe the relative position and direction of two orientation points, as illustrated in Fig.4. The points are represented by m_1 and m_2, while $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ represent their respective normal vectors. F_1 denotes the distance between the two points, whereas F_2 and F_3 correspond to the angles formed between the normal vector of each point and the distance vector defined by the two points. F_4 represents the angle between the two normal vectors. The equation $d = m_2 - m_1$ signifies the difference between m_2 and m_1. The feature $\boldsymbol{F}$ is defined as follows

$$\boldsymbol{F}(m_1, m_2) = (F_1, F_2, F_3, F_4) = \left(\|d\|_2, \angle(\boldsymbol{n}_1, d), \angle(\boldsymbol{n}_2, d), \angle(\boldsymbol{n}_1, \boldsymbol{n}_2)\right) \tag{2}$$

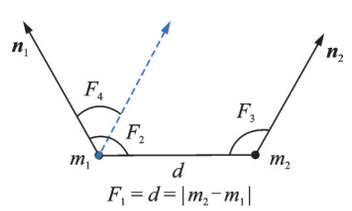


Fig.4 Point pair feature F of two oriented points

where $\angle$ represents the angle between vectors and d the diameter of the CAD model.

(2) Global feature description

Based on the Drost-PPF, the representation of the model consists of a collection of point pair features that are grouped based on their similar feature vectors. To accomplish this, it is necessary to calculate the feature vector $\boldsymbol{F}$ of Eq.(2) for every point pair (m_i, m_j) on the model surface. In practice, this model representation is stored in a hash table, where the features $\boldsymbol{F}$ are used as indices. By using F_s as a key to access the hash table, all model features $F_m(m_i, m_j)$ that are similar to a given scene feature $F_s(s_i, s_j)$ can be easily retrieved. The distances and angles are sampled at intervals of d_{dist} and d_{angle}, respectively. Feature vectors that have the same discrete versions are then grouped together. To illustrate this concept, Fig.5 provides an example of point pairs with similar features on a single object. These point pairs are collected within set A.

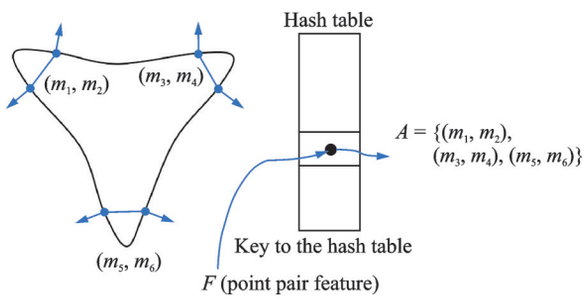


Fig.5 Model point pairs with similar features

(3) Feature matching

Feature matching is the process of identifying the most similar point pair features of model points and scene points in a hash table. This is done by calculating the point pair feature for a given scene point pair (s_r, s_j), and using this feature as a key value to find the corresponding model point pair (m_r, m_j) in the hash table. Once these pairs are identified, the reference points s_r and m_r are moved to the origin of the local coordinate system, and their normal vectors are aligned with the x-axis. This allows for rotation around the normal vector, aligning the model with the scene, as depicted in Fig.6. The transformation from the model to the scene can be represented by a point and a rotation angle α, denoted as

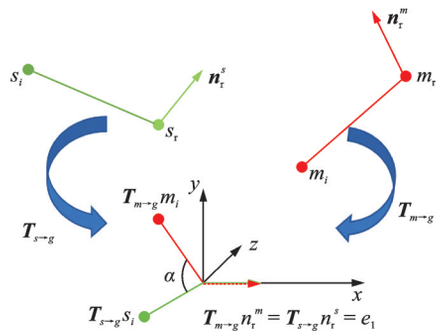


Fig.6 Transformation between model point pairs and scene point pairs

(m_r, α). If the model point pair (m_r, m_j) and the scene point pair (s_r, s_j) have similar point pair features, the conversion relationship between the two points can be calculated using Eq.(3). In Fig.6, $T_{m\to g}$ represents a transformation that translates the reference point m_r to the origin of the coordinate system and rotates the normal vector n_r^m to align with the x-axis. Similarly, $T_{s\to g}$ is a transformation that translates the reference point s_r to the origin of the coordinate system and rotates the normal vector n_r^s to align with the x-axis. $T_{s\to g}^{-1}$ denotes the inverse of $T_{s\to g}$, and $R_X(\alpha)$ represents a rotation around the x-axis with an angle α.

$$s_i = T_{s\to g}^{-1} R_X(\alpha) T_{m\to g} m_i \tag{3}$$

(4) Voting scheme

Once the scene point pair (s_r, s_j) in the scene and the model point pair (m_r, m_j) have similar features, the transformation between them can be represented by a point and a rotation angle α, which is denoted as (m_r, α). Once the optimal transformation is found, the global pose of the object can be recovered. To find the optimal transformation, a two-dimensional accumulator array is created for voting. The number of accumulator's rows, N_m, is equal to the number of model sample points $|M|$. The number of accumulator's columns N_{angle} corresponds to the number of sample steps n_{angle} of the rotation angle α. This accumulator array represents the discrete space of transformation for a fixed reference point. For the actual voting, the reference point s_r is paired with every other point from the scene, and the model surface is searched for point pairs (m_r, m_i) that have a similar distance and normal orientation to (s_r, s_i). This search answers the question of where on the model the pair of scene points (s_r, s_i) could be, and is performed using the pre-computed model description: Feature $F_s(s_r, s_i)$ is calculated and used as the key to the hash table of the global model description, which returns the set of similar features in the model. For each match (m_r, m_i), i.e., for every possible position of (s_r, s_i) on the model surface, the rotation angle α is calculated using Eq.(3) which corresponds to the transformation that maps (m_r, m_i) to (s_r, s_i), as shown in Fig.7. A vote is cast for the (m_r, α). After all points s_i are processed, the peak in the accumulator array corresponds to the global rigid movement can be calculated.

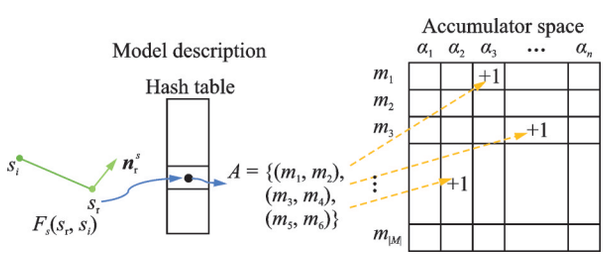


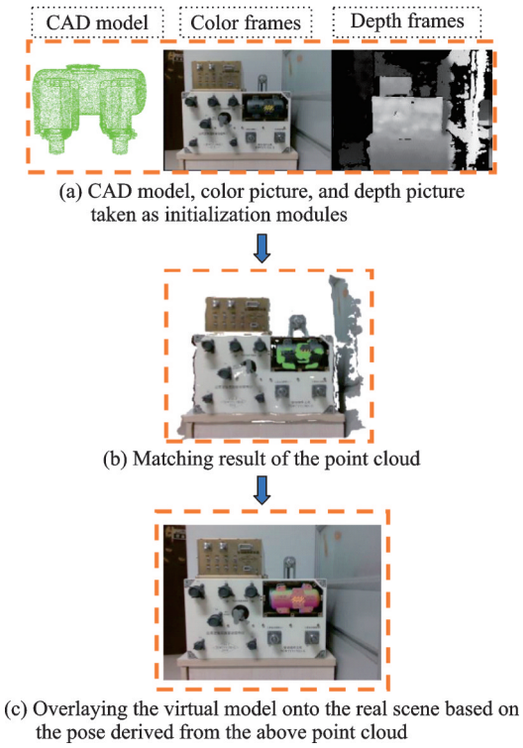
Fig.7 Visualization of the voting scheme

(5) Pose clustering

Each scene reference point can obtain a pose using the voting scheme mentioned earlier. The consistency of results across multiple reference points ensures the correct pose is achieved. However, there is a possibility of mismatching, resulting in an incorrect pose that deviates from the intended one. To overcome this issue, Drost-PPF incorporates a pose cluster step to filter out erroneous poses and improve the accuracy of the final result. In this step, Drost-PPF considers retrieved poses if their translations and rotations do not exceed a predetermined threshold. The pose of each cluster is then calculated by averaging its constituent poses. Finally, the optimal pose is determined based on the cluster pose with the highest number of votes from reference points.

2.2.3.2 *Our point cloud feature matching method*

The process of point cloud matching comprises the follow steps and the initialization result is shown in Fig.8.



(a) CAD model, color picture, and depth picture taken as initialization modules

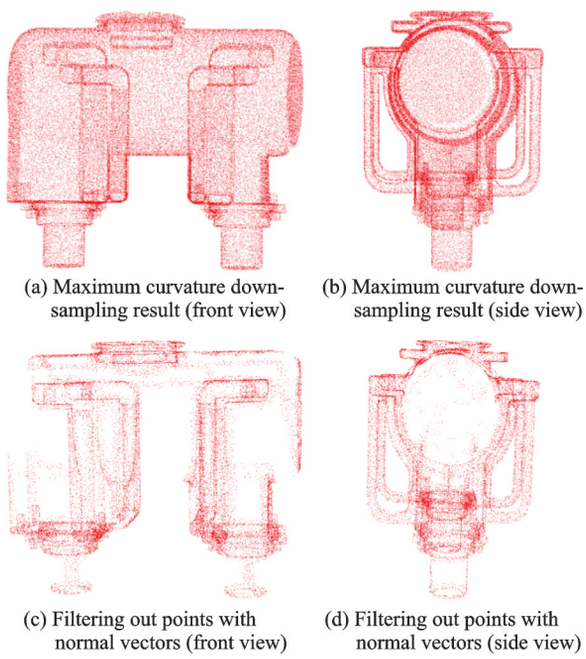
(b) Matching result of the point cloud

(c) Overlaying the virtual model onto the real scene based on the pose derived from the above point cloud

Fig.8 Initialization result

(1) Due to the maintenance operations need to be performed on the MMW within a range of approximately one meter. To achieve this, we use pass-through filtering based on the depth of the generated point cloud to eliminate background clutter.

(2) In order to enhance computational efficiency, down-sampling is applied to both the model and scene cloud. The sampling interval is determined based on the diameter of the model, ensuring that there is no need for readjusting sampling parameters across different models. Additionally, in the Drost-PPF a significant redundancy of features exists between each pair of points. To further improve computational efficiency and accuracy, we propose a novel down-sampling method. Step 1, this method first samples the grid based on maximum curvature, specifically retaining points with the greatest curvature in the grid. Step 2, points with less prominent geometry are filtered out based on the angle between their normal vectors. The point pair features are exclusively computed between the point cloud acquired in Step 2 and the point cloud obtained in Step 1. Given the predominantly planar structure of most industrial products, this method significantly reduces the number of feature point pairs, eliminates redundant features, and enhances computational efficiency. The sampling results of the model are shown in Fig.9.



(a) Maximum curvature down-sampling result (front view)

(b) Maximum curvature down-sampling result (side view)

(c) Filtering out points with normal vectors (front view)

(d) Filtering out points with normal vectors (side view)

Fig.9 Down sampling

(3) In the online stage, the scene reference point is selected based on a curvature value higher than the minimum curvature value of the model sampling point, replacing the previous random sampling approach.

(4) Feature matching. The feature matching is in accordance with the Drost-PPF algorithm.

The main three parameters are the sampling rate τ_d for down-sampling point clouds, the quantization step sizes Δ_{angle} and the radius r for computing the normal of fitting a plane for computing quantized PPFs. We select $\Delta_{\text{angle}}=30$, $\tau_d = 0.05d$, $r = 2.5\tau_d$. The selection of parameters comes from Ref.[37].

2.3 Pose tracking

We obtain the initial pose of the model from the pose estimation section, and subsequently start calculating the pose transformation between consecutive frames based on the acquired pose. Building upon the widely adopted region-based approach, our study incorporates texture feature of space station products to enhance the estimation of inter-frame transformation.

2.3.1 Sparse viewpoint model

We employ the identical sparse viewpoint model as utilized in SRT3D[32], where the virtual camera is positioned at the vertex of the geodesic to generate pose projections from 2 562 distinct perspectives. The geodesic vertices are illustrated in Fig.10, which is achieved by subdividing each triangle of the icosahedron four times. For each rendered image's model contour, 200 points are randomly sampled along its edge, with their normal vectors approximately perpendicular to the respective point's contour, some tracking results of the filter is shown in Fig.11, where different rows represent results under varying pose. In the first row, the model textures can be well detected and have a high number of region points, which contributes to the stability of the tracking. However, in the second and third row poses, there are fewer features, leading to a decrease in tracking stability.

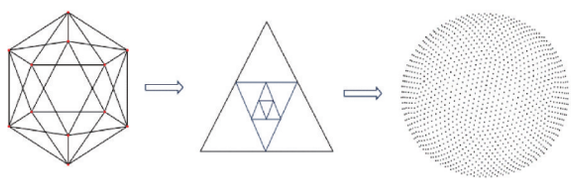


Fig.10 Sparse viewpoint division figure

(a) Pose tracking result (b) Region point (c) Texture point

Fig.11 Pose tracking results

2.3.2 Region-based and texture-based tracking

The original region-based method requires the calculation of the entire area, while the sparse region-based method only necessitates calculating elements on the corresponding line. Sparse region-based methods not only achieve higher accuracy but also significantly enhance computational efficiency. The sparse region-based approach aims to minimize the distance along the corresponding line to obtain pose[32]. The region modality calculation as in Ref.[32] was utilized, as indicated by Eqs.(4—6).

$$d_{ri}(\boldsymbol{\theta}) = \boldsymbol{n}_i^{\mathrm{T}}(\pi(cX_i(\boldsymbol{\theta})) - c_i) \tag{4}$$

$$l(r_i) = I(c_i + rn_i) \tag{5}$$

$$p(\boldsymbol{\theta}|D_r) \propto \prod_{i=1}^{n_c} p\left(d_{si}(\boldsymbol{\theta}) | \omega_{si}, l_{si}\right) \tag{6}$$

where $\boldsymbol{\theta}$ is the pose variation vector, c_i the ith sampling point on the projected contour of the model, n_i the normal of the point c_i, cX_i the 3D corresponding point of c_i in the model, $\pi(\cdot)$ the projection function of the pinhole camera that projects 3D spatial points onto a two-dimensional image, $l(r_i)$ the corresponding line about c_i, I the image, D_r the joint posterior probability of the region-based method, and n_c the number of the sampling points; d_{si}, ω_{si}, l_{si} are the discrete point values of the corresponding line.

Texture-based methods aim to minimize point-

to-point distances for obtaining pose[35]. The same texture modality calculation as in Ref.[35] was utilized, as indicated by Eqs.(7—9).

$$p(\boldsymbol{\theta}|D_{\mathrm{t}}) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma_{\mathrm{t}}^2}\rho_{\mathrm{tuk}}\left(\left\| x_i' - x_i(\boldsymbol{\theta})\right\|_2\right)\right) \tag{7}$$

$$x_i(\boldsymbol{\theta}) = \pi({}_C\boldsymbol{T}_{MM}\boldsymbol{T}(\boldsymbol{\theta})_M X_i) \tag{8}$$

$$\rho_{\mathrm{tuk}} = \begin{cases} \dfrac{c^2}{6}\left(1-\left(1-\left(\dfrac{x}{c}^2\right)\right)^3\right) & |x| \leqslant c \\ \dfrac{c^2}{6} & \text{Otherwise} \end{cases} \tag{9}$$

where D_{t} represents the joint posterior probability of the texture-based method, and x_i' the detected feature point in the current frame that corresponding to $({}_M X_i)$ in the key frame. The homogeneous transformation matrix $({}_C\boldsymbol{T}_M)$ represents the relative pose between the model and camera coordinate frames M and C. $({}_M\boldsymbol{T}(\boldsymbol{\theta}))$ represents using axis-angle representation to obtain the matrix representation in the model coordinate system. σ_{t} is user-defined to indicates the uncertainty. And the term ρ_{tuk} is the Tukey norm with the aim to minimize the effect of outliers. Also, c is a user-defined constant with the aim to maximize the residual error expected for inliers.

The points in each new frame within a nearby rectangular region of the previous pose estimation are detected. The descriptors from these points are subsequently matched to features from keyframes if the difference in orientation compared to existing keyframes exceeds a specific threshold. The incorporation of keyframes aids in reducing drift and enhancing the overall tracking robustness. Subsequently, if a frame qualifies as a keyframe, a depth rendering is produced, and 3D model points are reconstructed for each keypoint falling on the silhouette. Along with their descriptors, unobstructed 3D points are then stored for the keyframe.

2.3.3 Optimization

In this article, we employed the probability density function (PDF) as a representation of the likelihood for a specific pose given multimodal data, and subsequently estimated the posture by maximizing this function. The joint PDF is defined as follows

$$p(\boldsymbol{\theta}|D) = p(\boldsymbol{\theta}|D_{\mathrm{r}})\,p(\boldsymbol{\theta}|D_{\mathrm{t}}) \tag{10}$$

We find the optimal pose by estimating variables. We employ Gaussian-Newton optimization with the Tikhonov regularization to maximize the PDF for estimating the pose variable $\hat{\boldsymbol{\theta}}$, shown as

$$\hat{\boldsymbol{\theta}} = \left(-\boldsymbol{H} + \begin{bmatrix} \lambda_{\mathrm{r}} I_3 & 0 \\ 0 & \lambda_{\mathrm{t}} I_3 \end{bmatrix}\right)^{-1} \boldsymbol{g} \tag{11}$$

where λ_{r} and λ_{t} represent the regularization terms for rotational and translational components, respectively. The gradient $\boldsymbol{g}$ and Hessian $\boldsymbol{H}$ represent the first- and second-order partial derivatives of the natural logarithm of the joint PDF $p(\boldsymbol{\theta}|D)$.

$$\boldsymbol{g}_{\mathrm{r}}^{\mathrm{T}} = \frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\boldsymbol{\theta}|D_{\mathrm{r}})\bigg|_{\theta=0} \tag{12}$$

$$\boldsymbol{H}_{\mathrm{r}} = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln p(\boldsymbol{\theta}|D_{\mathrm{r}})\bigg|_{\theta=0} \tag{13}$$

$$\boldsymbol{g}_{\mathrm{t}}^{\mathrm{T}} = \frac{\partial}{\partial \boldsymbol{\theta}} \ln p(\boldsymbol{\theta}|D_{\mathrm{t}})\bigg|_{\theta=0} \tag{14}$$

$$\boldsymbol{H}_{\mathrm{t}} = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \ln p(\boldsymbol{\theta}|D_{\mathrm{t}})\bigg|_{\theta=0} \tag{15}$$

$$\boldsymbol{g} = \boldsymbol{g}_{\mathrm{r}} + \boldsymbol{g}_{\mathrm{t}} \tag{16}$$

$$\boldsymbol{H} = \boldsymbol{H}_{\mathrm{r}} + \boldsymbol{H}_{\mathrm{t}} \tag{17}$$

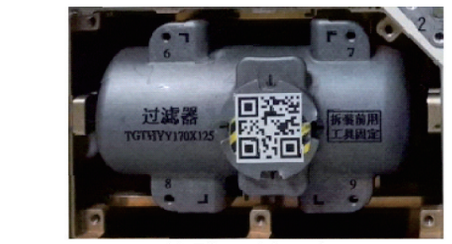
where $\boldsymbol{g}_{\mathrm{r}}$, $\boldsymbol{H}_{\mathrm{r}}$ represent the first- and second-order partial derivatives of the natural logarithm of the $p(\boldsymbol{\theta}|D_{\mathrm{r}})$, and $\boldsymbol{g}_{\mathrm{t}}$, $\boldsymbol{H}_{\mathrm{t}}$ the first- and second-order partial derivatives of the natural logarithm of the $p(\boldsymbol{\theta}|D_{\mathrm{t}})$.

By iteratively repeating this process, a pose can be obtained that maximizes a joint probability density function incorporating both texture and region information.

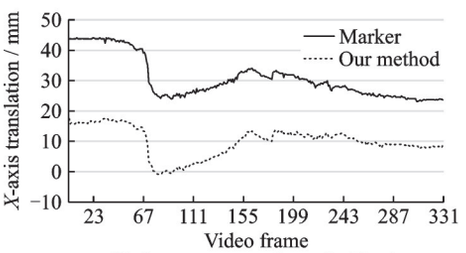
3 Experiment and Evaluation

3.1 Comparison with marker-based method

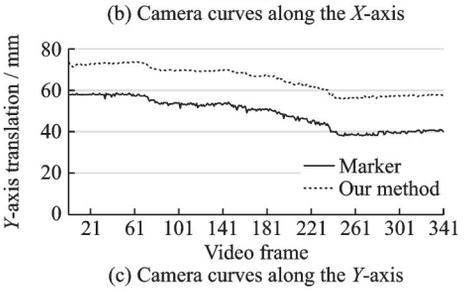
In this section, we initially verified the tracking performance of our algorithm by comparing it with the marker-based method, and the result is shown in Fig.12. Subsequently, we compare the measured values from the electronic caliper with the estimated values from our method. Finally, we present a pro-



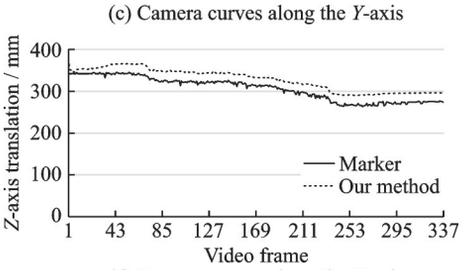
(a) Filter with marker



(b) Camera curves along the *X*-axis



(c) Camera curves along the *Y*-axis



(d) Camera curves along the *Z*-axis

Fig.12 Camera motion coordinate curves

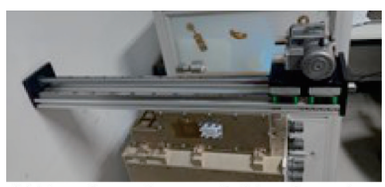
totype system developed for the specific maintenance task of the fluid circuit in space stations to demonstrate the practicality and versatility of our framework. All experiments used the same configuration, and we used the Intel® RealSense™ Depth Camera D435i and a laptop equipped with a 2.50 GHz Intel(R) Core (TM) i9-12900H CPU. The tracking algorithms are implemented using C++ language under the Visual Studio 2022 platform, while the augmented information is implemented utilizing OpenGL. The camera's resolution is set to 640 pixel×480 pixel.

3.2 Comparison with real value

We slid the model along the sliding rod, and the displacement of the movement is measured using an electronic caliper. This measurement is then compared with the actual value, and the result and the scence are shown in the Table 1 and Fig.13.

Table 1 Displacement comparison mm

Figure	Real	Estimate	Error
Fig.13(a)	136.91	138.293	1.383
Fig.13(b)	114.14	112.826	1.314
Fig.13(c)	128.45	128.402	0.048



(a) Experimental scene of object front view

(b) Experimental scene of object side view

(c) Experimental scene of object top view

Fig.13 Experimental scenario

3.3 AR assistance assembly case

3.3.1 AR assembly visualization

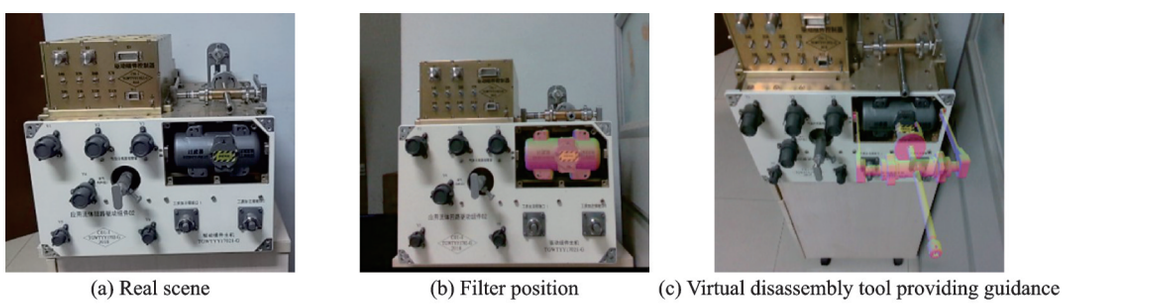
The AR auxiliary guidance is provided to facilitate the replacement of the com-ponent filter within the fluid circuit of space equipment. The fluid circuit, the filter and the specialized disassembly tool are illustrated in Fig.14. Fig.15 illustrates virtual guidance information and corresponding actual operations.

3.3.2 Time consumption experiment

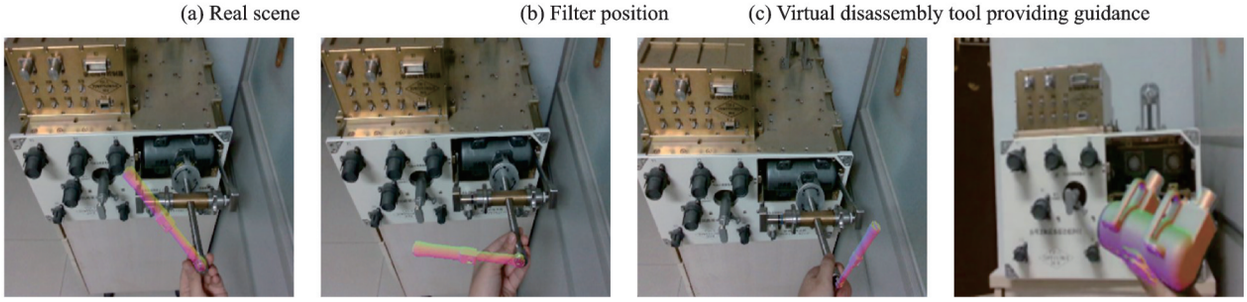
The time required for each stage in our automatic tracking framework is presented in Table 2. It is evident that, except for the time-consuming of pose estimation through point cloud matching, our method demonstrates real-time performance during tracking, achieving approximately 30 frame/s on the CPU. Despite the longer duration of pose estimation, we solely employ it after the initial frame and trace loss occurrences. Thus, we still consider our algorithm to be real-time.

(a) Fluid circuit (b) Filter (c) Specialized disassembly tool

Fig.14 Fluid circuit, filter and the specialized disassembly tool



(a) Real scene (b) Filter position (c) Virtual disassembly tool providing guidance



(d) Position of torque wrench before tightening the special tool (e) Position of torque wrench after tightening the special tool (f) Using the torque wrench to loosen the special tool (g) Removing the filter

Fig.15 Virtual guidance information corresponding to actual operations

Table 2 Time consumption of various system modules

Module	Time/s
Initialization	0.22
Tracking	0.032

4 Conclusions

We propose an automatic tracking framework for induced maintenance applications of Chinese Space Stations. Because it combines with the point cloud matching method, this system can automatically obtain the initial pose without requiring manual specification. This eliminates the need for astronauts to precisely align themselves when performing operational tasks in microgravity, making it valuable on the space stations. Meanwhile, considering the weak texture property of space products, the tracking algorithm's performance is further enhanced by combing texture and region methods, achieving real-time execution on a CPU. The accuracy of our method is demonstrated by the consistent pose transformation trend with marker-based algorithms, because the maker-based approach is very mature.

Our method does not rely on marker, making it more suitable for the operational of space stations in comparison to marker-based approaches. Additionally, it is better suited for the limited features of space products compared with feature-based methods. Furthermore, it enhances robustness by exclusively utilizing region-based methods. However, for some objects with strong textures, incorporating more texture features may degrade the poorer tracking performance.

In the future work, our research also has some aspects to be improved. (1) The introduction of a viewpoint restriction allows for the exclusion of point pair features that cannot coexist within the

same perspective, thereby enhancing the computational efficiency of point cloud algorithms. (2) The enhancing real-time performance and stability can be achieved by leveraging previously lost pose information for rapid recovery, instead of relying on point cloud recalculations when tracking is disrupted.

References

[1] SCHEID F, NITSCH A, KOENIG H, et al. European SDTO operation at Col-CC[C]//Proceedings of SpaceOps 2010 Conference Delivering on the Dream Hosted by NASA Marshall Space Flight Center and Organized by AIAA. [S.l.]: AAAI, 2010.

[2] WANG F, ZHANG L, XU Y, et al. Development of the on-orbit maintenance and manipulation workbench (MMW) for the Chinese space station[J]. Acta Astronautica, 2023, 214: 366-379.

[3] ZHANG Y, ZHANG L, FU H, et al. A system and method for intelligent induced maintenance of space application facilities[J]. Mobile Information Systems, 2022. DOI: 10.1155/2022/5331280.

[4] CAUDELL T P, MIZELL D W. Augmented reality: An application of heads-up display technology to manual manufacturing processes [C]//Proceedings of the 25th Hawaii International Conference on System Sciences. Piscataway, NJ, USA: IEEE, 1992: 659-669.

[5] FEINER S, MACINTYRE B, SELIGMANN D. Knowledge-based augmented reality[J]. Communications of the ACM, 1993, 36(7): 53-62.

[6] SCHWALD B, LAVAL B D, SA T O, et al. An augmented reality system for training and assistance to maintenance in the industrial context [J]. Journal of WSCG, 2003, 11(3): 1791904.

[7] STRIEKER D, WEIDENHAUSEN J. ARVIKA: Augmented-reality for development production, and service[J]. Computer Graphik Topics, 2003(6): 15.

[8] HENDERSON S J, FEINER S K. Augmented reality in the psychomotor phase of a procedural task[C]// Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR). [S.l.]: IEEE, 2011: 191-200.

[9] BENBELKACEM S, BELHOCINE M, BELLARBI A, et al. Augmented reality for photovoltaic pumping systems maintenance tasks[J]. Renewable Energy, 2013, 55: 428-437.

[10] ZHAO Xincan. Research on key technologies of augmented reality maintenance guidance system[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2008. (in Chinese)

[11] CUI Bo, WANG Wei, QU Yu. Design and implementation of equipment induced maintenance system based on augmented reality[J]. Fire Control & Command Control, 2016, 41(11): 176-181. (in Chinese)

[12] REGENBRECHT H, BARATOFF G, WILKE W. Augmented reality projects in the automotive and aerospace industries[J]. IEEE Computer Graphics and Applications, 2005, 25(6): 48-56.

[13] HARITOS T, MACCHIARELLA N D. A mobile application of augmented reality for aerospace maintenance training[C]//Proceedings of the 24th Digital Avionics Systems Conference. [S.l.]: IEEE, 2005.

[14] CRESCENZIO F, DE FANTINI M, PERSIANI F, et al. Augmented reality for aircraft maintenance training and operations support[J]. IEEE Computer Graphics and Applications, 2011, 31(1): 96-101.

[15] RADKOWSKI R, OLIVER J. Natural feature tracking augmented reality for on-site assembly assistance systems[C]//Proceedings of Conference on Virtual, Augmented and Mixed Reality Systems and Applications. Berlin, Heidelberg: Springer, 2013: 281-290.

[16] JO G S, OH K J, HA I, et al. A unified framework for augmented reality and knowledge-based systems in maintaining aircraft[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI Press, 2014, 28(2): 2990-2997.

[17] ZHANG H, GUO YU, TANG P, et al. Tracking and registration method based on image matching for augmented reality aided assembly system[J]. Computer Integrated Manufacturing Systems, 2021, 27(5): 1281-1291.

[18] NAZRI N A M, ABDULLAH M N, MUSTAPHA F. Implementation of Augmented Reality in Aircraft Inspection Process[C]//Proceedings of E3S Web of Conferences. [S.l.]: EDP Sciences, 2024, 477: 00067.

[19] HARRIS C G, STEPHENS M J. A combined corner and edge detector[C]//Proceedings of Alvey Vision Conference. Manchester, UK: [s.n.], 1988.

[20] VISWANATHAN D G. Features from accelerated segment test (fast)[C]//Proceedings of the 10th Workshop on Image Analysis for Multimedia Interactive Services. London, UK: [s.n.], 2009: 6-8.

[21] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60: 91-110.

[22] BAY H, ESS A, TUYTELAARS T. Speeded-up

robust features (SURF)[J]. Computer Vision and Image Understanding, 2008,110(3): 346-359.

[23] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011: 2564-2571.

[24] LEUTENEGGER S, CHLI M, SIEGWART R Y. BRISK: Binary robust invariant scalable keypoints[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011: 2548-2555.

[25] ALCANTARILLA F P, BARTOLI A, DAVISON A J. KAZE features[C]//Proceedings of European Conference on Computer Vision (ECCV). Florence, Italy: Springer, 2012: 214-227.

[26] PRISACARIU V A, REID I D. PWP3D: Real-time segmentation and tracking of 3D objects[J]. International Journal of Computer Vision, 2012, 98(3): 335-354.

[27] TJADEN H, SCHWANECKE U, SCHÖMER E. Real-time monocular segmentation and pose tracking of multiple objects[C]//Proceedings of European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: [s.n.], 2016: 423-438.

[28] HEXNER J, HAGEGE R R. 2D-3D pose estimation of heterogeneous objects using a region based approach[J]. International Journal of Computer Vision, 2016, 118: 95-112.

[29] TJADEN H, SCHWANECKE U, SCHOMER E. Real-time monocular pose estimation of 3D objects using temporally consistent local color histograms[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 124-132.

[30] TJADEN H, SCHWANECKE U, SCHOMER E, et al. A region-based Gauss-Newton approach to real-time monocular multiple object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(8): 1797-1812.

[31] STOIBER M, PFANNE M, STROBL K H, et al. A sparse gaussian approach to region-based 6DoF object tracking[C]//Proceedings of the Asian Conference on Computer Vision (ACCV). [S.l.]: Springer, 2020: 666-682.

[32] STOIBER M, PFANNE M, STROBL K H, et al. SRT3D: A sparse region-based 3D object tracking approach for the real world[J]. International Journal of Computer Vision, 2022, 130(4): 1008-1030.

[33] STOIBER M, SUNDERMEYER M, TRIEBEL R. Iterative corresponding geometry: Fusing region and depth for highly efficient 3D tracking of texture less objects[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, Louisiana, USA: IEEE, 2022: 6855-6865.

[34] TIAN X, LIN X, ZHONG F, et al. Large-displacement 3D object tracking with hybrid non-local optimization[C]//Proceedings of European Conference on Computer Vision (ECCV). [S.l.]: Springer, 2022: 627-643.

[35] STOIBER M, ELSAYED M, REICHERT A E. Fusing visual appearance and geometry for multi-modality 6DoF object tracking[EB/OL]. (2023-01-01). https://doi.org/10.48550/arXiv.2302.11458.

[36] WANG L, YAN S, ZHEN J, et al. Deep active contours for real-time 6-DoF object tracking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023: 14034-14044.

[37] ZHAO X, CUI X, FU H. 6DOF pose estimation of a 3D rigid object based on an adaptive model curvature point pair features[C]//Proceedings of 2024 10th International Conference on Control, Automation and Robotics (ICCAR). Angers France: IEEE, 2024.

[38] DROST B, ULRICH M, NAVAB N, et al. Model globally, match locally: Efficient and Robust 3D object recognition[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, California, USA: IEEE, 2010.

Acknowledgements This research was supported by the National Key Research and Development Program (No. 2022YFB3306100), the Aeronautical Science Fund of China (No.2019ZE105001), and the General Project of Chongqing Natural Science Foundation (No.cstc2019jcyj-msxmX0530).

Authors Ms. **ZHAO Xin** received the B.E. degree in mechanical engineering from Dalian Maritime University, Dalian, China, in 2019, where she is now pursuing the Ph.D. degree. Her current research interests include point cloud registration, tracking and disassembly sequence generation for augmented reality.

Dr. **WANG Yue** received the Ph.D. degree from Northwestern Polytechnical University, Xi'an, China, in 2018. He is

currently an associate professor with the Chongqing University of Posts and Telecommunications. His current research interests include defect detection and augmented reality.

Prof. **FU Hongyong** received the Ph.D. degree from University of the Chinese Academy of Sciences, China, in 2008. He is currently a professor with the Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing. His current research interests include mechanism reliability, testability, fault diagnosis, and virtual/ augmented reality maintenance.

Author contributions Ms. **ZHAO Xin** designed the study, completed the experiment and wrote the manuscript. Dr. **WANG Yue** contributed to the discussion and manuscript revision. Prof. **FU Hongyong** contributed to CAD model support and background of the study. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor: ZHANG Huangqun)

一种用于中国空间站增强现实维护应用的自动跟踪框架

赵 鑫[1,3], 王 月[2], 伏洪勇[1]

(1.中国科学院空间应用工程与技术中心,北京 100094,中国;
2.重庆邮电大学先进制造工程学院,重庆 400065,中国;
3.中国科学院大学计算机科学与技术学院,北京 100049,中国)

摘要:跟踪注册是增强现实(Augmented reality, AR)中一项至关重要的技术。基于人工标识和基于图像的注册方法已广泛应用于增强装配系统中。然而,由于空间产品的弱纹理、结构对称、数据有限以及难以张贴标识等特点,空间站的跟踪注册仍面临挑战。本文提出了一种专门为中国空间站诱导维修应用设计的三维物体跟踪方法,旨在用AR技术取代传统的纸质手册,为宇航员提供更直观的操作指导。本文提出了一种无标记的中国空间站智能维护方法。采用点对特征方法来估计初始帧姿态,不需要手动调整来获得相应的姿态,这对宇航员来说至关重要,因为在微重力环境中精确运动是一项重大挑战,而且结合曲率信息可以提高采用点对特征方法估计初始帧姿态的效率。通过结合纹理和区域信息,消除了对宇航员位置的精确限制,进一步增强了空间对称产品的跟踪鲁棒性。该方法利用点云匹配来估计第一帧的初始姿态,并在失去跟踪后重新计算姿态。一旦获得了前一帧的姿态,则仅根据二维图像的区域和纹理信息来计算跟踪,以获得实时跟踪。实验结果表明:本文方法与基于标识的方法具有相同的姿态趋势;基于电子游标卡尺测量之间的误差在毫米级;使用专门为航天产品设计的专用工具成功更换过滤器,证明了在空间站上实施诱导维修程序的实用性和潜力。

关键词:增强现实;位姿跟踪;CAD模型;中国空间站