Equipment Maintenance Assistant Technology Based on Augmented Reality

Fan LIU1,* and Peng-yuan LIU1

1Forth Department, Mechanical Engineering College, No. 97, He-ping West Road, Shi Jia-zhuang, 050003, He Bei Province, PR China

Keywords: Augmented reality, Equipment maintenance, Computer vision, Virtual reality fusion, Human - computer interaction

Abstract. With the development of augmented reality technology, especially the advent of HOLOLENS launched by Microsoft, the equipment maintenance assistant system based on the augmented reality (AR) technology is becoming more mature and excellence. The system takes advantage of the interactive characteristics of AR to make the maintenance process of equipment become more intuitive, which can improve the maintenance efficiency significantly. In this paper, the system is divided into five modules according to their functions, the development of related technologies in these five modules is introduced, and the trend of current research is concluded, provide the corresponding technical reference for the workers who are engaged in the research of augmented reality.

Introduction

With the rapid development of information technology, augmented reality (AR) technology has been widely used. AR is developed from the virtual reality (VR) technology. AR technology is a means of mutual integration of the real scene and virtual objects, where people can interact with objects in the virtual world in a natural way. AR system is characterized by the consistency of the real space and real-time interaction, including the registration of three-dimensional objects, tracking, positioning and display technology. In 1997, Professor Ronald Azuma proposed a widely accepted definition of AR, which can be described as: three-dimensional registration, virtual reality and real-time interaction [1].

The maintenance technology [2] based on AR integrates the real maintenance environment and virtual maintenance environment. It is not isolated from the surroundings, and it can maintain the user's sense to the real environment as well. Moreover, it allows users to interact with virtual objects through a specific device in the scene. The entire maintenance process is completed in the real environment, while the virtual environment is mainly responsible for maintenance guide, tips and other functions. Compared with other maintenance technology, the maintenance technology based on AR is more vivid, intuitive and timely, so it can be regarded as the highest stage of the maintenance support technology. This technique can be used to guide the equipment maintenance training, providing better conditions of the maintenance assistant for equipment maintenance personnel, overcoming the problems cause by the complexity of equipment and lack of understanding of equipment, improving the quality and effect of maintenance.

In recent years, the domestic research on AR is gradually increasing. However, limited by the development level of hardware, these achievements only remain in laboratory level. In January 2015, Microsoft launched the HOLOLENS glasses, which provides an advanced hardware platform for AR research. Coupled with the Kinect released a few years ago, a new upsurge of research of
AR is set off. Based on the HOLOLENS hardware platform, this paper introduces and analyzes the current situation and development trend of AR-based maintenance assistant technology.

**Framework of Maintenance Assistant Technology Based on AR**

The equipment maintenance assistant system based on AR can be divided into five modules according to their functions: namely, maintenance object visual processing, viewpoint pose tracking, virtual environment generation, virtual reality fusion and human-computer interaction, is shown in Figure 1.

![Figure 1. Framework of maintenance assistant technology based on AR.](image-url)

The functions of the maintenance object visual processing module mainly include the identification and tracking of the target, the 3D scene reconstruction; the main contents of the viewpoint pose tracking module are fusing the information obtained from viewpoint pose tracking and camera pose estimation; the main content of virtual environment generation module are using specific tools to create a virtual object or environment according to virtual object registration pose; the main contents of virtual reality fusion are solving the matching problems between the virtual
entity and the real environment, including the spatial matching and visual consistency processing, so as to combine the virtual objects and the real environment realistically; the main contents of human-computer interaction module are completing the maintenance of auxiliary functions by the voice, gesture and other instructions operation virtual information.

Research on Maintenance Assistant Technology Based on Augmented Reality

Maintenance Object Visual Processing Module

The main function of the maintenance object vision processing module is using the natural feature points obtained by the camera to identify and track maintenance targets and maintenance scene 3D reconstruction. The main content of this chapter include the joint detection based on RGB-D (RGB three channel color image and depth image), fast index matching method for feature points of maintenance objects, video target tracking of maintenance object, three-dimensional reconstruction of maintenance scene. They will be discussed in detail as follows.

Joint Detection Based on RGB-D

HOLOLENS integrates RGB and depth camera, able to carry out joint detection based on RGB-D. Compared with the traditional single RGB detection or grayscale detection, joint detection based on RGB-D considers depth information, which can effectively improve the speed, accuracy and robustness of the detection and recognition. RGB-D joint detection belongs to multi-modal data fusion methods. The fusion methods can be divided into feature fusion and decision fusion.

Feature layer fusion [3], also known as early fusion, is a simple and direct method of fusion. It combines the vectors extracted from different modes to form a more descriptive vector. A simple but effective application is to reduce the search space by employing the depth information as the prior information, to separate the background and the target according to the difference of depth, in order to improve the speed of the target detection. Feature fusion, served as another strategy, is using the depth information as the additional feature vector of target detection. The feature vector is fused with the feature vector extracted from the RGB image, and the target detection is determined by using the fusion vector.

Different from the feature level fusion, the fusion of decision layer is also called the late fusion. The classifier is designed according to the different characteristics of each model and then combined with the output of the classifier to make the final decision. The existing research work can extract Harris 3D corner, autocorrelation feature of time-space gradient from the RGB image and can extract eigen-joints, HON4D features from the depth data to make the decision layer fusion as well.

Fast index Matching Method for Feature Points of Maintenance Object

In order to realize the maintenance object recognition and tracking based on the natural features of the image, it needs to extract the features from the image and match the feature points in the 3D feature database, so as to obtain the corresponding 2D image points and 3D feature points. This requires fast extraction and description of the image features. Usual methods include fast corner detection, binary description, SURF, Freak and so on. However, the number of descriptors in the library is also growing with the increasing of the scene space, In order to maintain a high matching precision, the speed of feature matching based on the above indexing method is reducing in a geometrical progression.

The main solutions of current research include: improving binary descriptor index method, taking advantage of characteristics of one-to-many between the three-dimensional feature points and the...
descriptors in the library, establishing a supervised feature indexing method. It can also improve the matching speed by establishing an index for the features in the library using hierarchical clustering tree[4] or locality sensitive hashing method.

**Video target tracking of maintenance object**

The essence of video object tracking is to find the location of the interested objects in the continuous video frame. How to accurately identify the target and how to conduct a long time in real-time and stable tracking are the key problems in the study.

Some common target tracking algorithms [5] include the particle filter based target tracking method, mean-shift based target tracking method and the Calman filter based target tracking method etc. In practical application, these methods are used in combination in order to take advantage of different algorithms. For example, the combined of mean-shift and particle filtering methods can improve the tracking accuracy using mean-shift algorithm for each particle to polish up the performance. The combination of mean-shift and Calman filtering method can improve tracking performance by Calman filtering method to predict the position of the target in the next frame or by using the Calman filter to update the target model.

In order to enhance the adaptability to the environment, many scholars carry out the adaptive setting of parameters and the description of the target by integrating the idea of machine learning in the target tracking method. Thus, a target tracking method based on machine learning is proposed. Usually the support vector machine (SVM) and AdaBoost[6] are used for learning algorithm of target tracking. The drawback of SVM tracking is that it cannot handle the disappearance or reappearance of the target and the partial or total occlusion of the target. The real-time performance of AdaBoost is also need to be improved.

With the development of the deep learning theory, the video object tracking system based on depth learning has been applied to the research gradually. CNN (Neural network Convolutional) is the most typical application in target recognition and tracking.

Deep learning model has some properties such as large quantities of data and long training time, but the model successfully trained has superior performance in recognition. In order to strive for more dominant rights in the growing industry of deep learning, attract more data scientists and researchers to use, Facebook, Apple, Google, Microsoft and other technology giants have opened source their own deep learning system. For instance, on November 9, 2015 Google opened source the artificial intelligence system that is called TensorFlow. In January 2016 Microsoft opened source artificial intelligence toolkit CNTK (Network Toolkit Computational). They are the latest software platform used for deep learning currently, which has integrated many classic algorithms that can be used to solve the problem of image recognition and video tracking. Among them, since CNTK toolkit and HOLOLENS glasses are both launched by Microsoft, they are compatible with each other well. CNTK is the fastest running deep learning software currently, it can satisfy the real-time requirement of maintenance support system based on AR.

**3D Scene Reconstruction**

The traditional 3D reconstruction methods usually need binocular vision theory to calculate the depth of the scene. The application of Kinect equipment has greatly simplified the process of 3D reconstruction. Using depth sensor, we can simultaneously capture the image of a single angle of view and the depth data corresponding to the angle of view. Through the fusion of depth sensor data collected from multiple perspectives, the 3D reconstruction of the real scene can be achieved.
Viewpoint Pose Tracking Module

Camera Pose Tracking

Camera pose estimation based on natural feature points can be seen as a perspective n point projection problem [7], abbreviated as PnP (n Point Perspective). Assume the camera model is a perspective projection model and has been calibrated, the model is shown in Figure 2. \( O_c \) is the origin of the camera coordinate system, \( O_s \) is the origin of the image coordinate system, and \( O_w \) is the origin of the world coordinate system.

Take a picture that include n spatial points (such as \( P_1, P_2, P_3 \)) that coordinates have been known under the world coordinate system, and n spatial points corresponding to the coordinates under the image coordinate system (\( P'_1, P'_2, P'_3 \)) is also known, calculating the coordinates of n points under camera coordinate system or the position \( R \) and angle \( t \) that world coordinate system transform to the camera coordinate system \( R \) and \( t \).

Figure 2. perspective projection model of feature point.

At present, the main solving methods contain analytical solution algorithm, iterative algorithm and global optimization algorithm [8]. The analytical solution algorithm is used to obtain the analytical expression of the position and attitude of a camera by using the geometric relations between the 3D feature points and the 2D feature points. Now, EPnP algorithm proposed in the literature is often regarded as one of the simplest analytical algorithm.

Compared with the analytical solution algorithm, the advantage of the iterative solution algorithm is calculation process is simple, and stronger to resist image noise interference, feature point error matching.

POSIT (Pose from Orthography and Scaling with Iteration) algorithm is an iterative algorithm designed according to weak perspective projection model, and there is a special function in OpenCV to implement this algorithm. The global optimization algorithm is proposed to solve the shortcomings of the iterative algorithm which can not guarantee the global optimal solution.

Among the three methods of solving the current position and attitude, the iterative algorithm has better real-time performance, the ability to resist noise is strong, and has a high estimation accuracy, so it is more widely used in 3D registration technology. However, due to the complexity of the position and attitude tracking algorithm, the algorithm needs to be improved and adjusted in practical application. Literature [9] proposed a fast camera pose estimation algorithm that is composed by the iterative part and fine adjustment process.
**Viewpoint Pose Tracking**

The main function of the position and pose tracking module is to fuse camera pose information based on natural feature recognition algorithm and data that from a position sensor. Achieve real-time acquire pose information of maintenance personnel in real physical world. On the one hand, the information can be used as a reference for the calculation of the positions of virtual object, on the other hand, it is an important input information of human computer interaction. The position and attitude sensors (such as gyroscope) can record and track the movements of maintenance personnel head, which are mainly used to supplement the information of the position and orientation based on natural feature recognition.

**Virtual Environment Generation Module**

The module is used to generate the virtual maintenance environment. Virtual maintenance environment information can be divided into general information (such as the maintenance of text, voice prompts, video presentations, etc.) and the depth of integration of information (the application of maintenance tools and equipment, the phenomenon of maintenance process).

For the general integration of information, achieve it by multimedia technology. For the depth of fusion information, virtual simulation technology can be used and based on Unity3D development tools to achieve. The concrete content includes the establishment of three-dimensional model of maintenance tools and equipment, the definition of behavior patterns and output under the virtual maintenance environment.

**Virtual Reality Fusion Module**

Virtual reality fusion technology is used to realize the 3D registration of virtual scene and real environment. The core is the viewpoint (camera) calibration and 3D reconstruction technique. For the general fusion information, first the virtual reality fusion module determines the position of the maintenance object in the world coordinate system.

For the general integration of information, achieve management based on the HOLOLENS development language. First the virtual reality fusion module determines the position of the maintenance object in the world coordinate system. After obtaining the projection displayed on the semi permeable membrane of the HOLOLENS, determine the location that fusion type information on the semi permeable membrane according to this value.

For the depth information, it is needed to solve the space time consistency problem. First the consistency of depth fusion space require to solve the consistency of the position and orientation matching among the view (camera), the maintenance object in the real maintenance environment, and the maintenance tools and equipment in the virtual maintenance assistant information environment. It need to achieve the precise positioning of the maintenance object and viewpoint (camera). Then it is also need to determine the relation of spatial interference between virtual maintenance tools and real maintenance objects.

Time consistency requires to use the same simulation time between two environments to ensure that the time is strictly synchronous. Otherwise it will cause two environmental imaging time inconsistency problem, impact the sense of reality of the augmented reality system seriously. The key problem of time synchronization is that the refresh rate of virtual maintenance assistant information environment is not less than 25 frames per second. At present, the Unity3D virtual reality development platform provides a good support for this.
Human Computer Interaction Module

At present, there are two forms of human-computer interaction, one is through voice commands to control the virtual information, the other is through the human action to perform related operations on virtual Information. HoloLens and Kinect are equipped with the corresponding voice operating system, for developers can be used directly. Current there are two kinds of human motion recognition [10], gesture recognition and body motion recognition.

Gesture recognition control is simple and easy to operate, it can be divided into two types of dynamic gesture recognition and static gesture recognition. The main difference between them is that dynamic gesture recognition is needed to track the trajectory of the hand. Dynamic gesture recognition is widely used in the field of research, and is more practical. Particle filter algorithm and CamShift algorithm [11], are two ways that commonly used in hand gesture tracking algorithm.

Dynamic gesture recognition added a tracking process than the static gesture recognition, the process is shown in Figure 3:

![Gesture recognition process](image)

Another type of recognition is motion recognition based on body movements, such as lifting arms, lifting leg, nodding, squat, etc. The discrimination of these actions is mainly based on the identification and tracking of human joint points, Kinect and other somatosensory equipment itself has the function of automatically capturing the point of human joints. Some simple actions can be identified by tracking the combined position of these joints.

Development trend

From the performance requirements of the point of view, the development trend of equipment maintenance technology based on augmented reality is mainly to improve the real-time (maintenance object recognition, 3D reconstruction of the maintenance scene, real-time solution of view position and attitude), robustness (video target tracking of long time and stable), realistic (Precise positioning of virtual object, compact fusion of virtual and reality) performance of the system and more intelligent human-computer interaction.

From the technical level, there are mainly the following trends:

Multimodal and high dimensional feature fusion. The current research main obtained the lower dimension of the vector after the fusion of the RGB-D images. It is an effective way to improve the real-time RGB images and depth robust performance of the system by make full use of the characteristics of RGB-D images and strong and effective fusion. The current research trend is fusion the RGB feature and the depth feature in the original image layer and decision layer respectively. A deep learning model is constructed in the literature [12], it propose an improved multi mode sparse self coding algorithm (MMSAE) based on the SAE algorithm. This algorithm
can extracted RGB features and the deep features distinguishing based on the characteristics of each class of objects, effective integration of RGB information and depth information. At present, there is still not an absolute advantage of the feature description algorithm can make different modes under a unified framework, so it is a significant challenge to design an adaptive feature detection and description algorithm for different modal data.

Artificial intelligence system based on deep learning. With the continuous improvement of deep learning theory and algorithm, deep learning has been more and more applied in speech recognition, computer vision and human motion recognition. Google TensorFlow deep learning system, Microsoft CNTK artificial intelligence tool is currently the latest, computing the fastest software application platform. The software platform can be transplanted to the auxiliary maintenance system based on augmented reality, which can optimize the system performance effectively and improve the degree of intelligence. Specific features can be used for feature fusion of RGB-D, recognition tracking of maintenance object, recognition of speech and human.

Pose tracking based on RGB-D camera. As Kinect equipment can easily get depth image sequences, this gives a new idea to solve the position and pose tracking. Traditional camera pose solution just use the mapping information of three-dimensional feature points to the two-dimensional plane, loss of the important depth data. Kinect can be used to restore the deep information, by tracking the distance from multiple natural feature points to the camera lens, base on the geometric relation to track the position of the camera, the displacement vector t is obtained, which avoids the complex matrix calculation of the original 2D feature points. For rotational variables R, one is based on the known position of the camera, use the imaging range of the camera and the corresponding image position of characteristic point solving it in the polar coordinate system, two is to add hardware sensors records the change of camera rotation angle. Such as the gyroscope that added in Hololens Microsoft glasses.

Complex motion recognition. Current motion recognition methods mainly focus on the recognition of gestures and simple action, which affect the experience of human-computer interaction. The certain development trend of human-computer interaction is able to identify more complex human actions, and conduct a richer behavior control for the augmented reality system. In addition, in the action of multiple human objects recognition and human interaction recognition, at present, there is almost no relevant research, an important reason is that the current depth of the sensor does not have the ability to handle large range of data.

Summary
Nowadays AR technology is in a period of vigorous development, With the development of the corresponding hardware devices and software platforms, the equipment maintenance assistant technology based on augmented reality has been turned from the simulation experiment to the practical application gradually. The main trend of the research is to improve the real time, robustness, real sense and intelligence of the system. Can be expected in the near future, the technology will bring to subversive change for the traditional maintenance mode, and bring vivid image of the maintenance experience for the vast number of maintenance technicians.

Reference


