Research on Voice Interaction for Augmented Reality assisted Maintenance

Ranran Wang [1]*, Chuan Lyu[2], Shunfeng Mei[3], Chunhui Guo[4], Jie Geng[5]** and Yu Xiao[6]

[1]* *School of Reliability and Systems Engineering, Beihang University in Beijing of China+ wrrbuaa@163.com*
[2] *School of Reliability and Systems Engineering, Beihang University in Beijing of China + lc@buaa.edu.cn*
[3] *School of Reliability and Systems Engineering, Beihang University in Beijing of China + meisf125@163.com*
[4] *School of Reliability and Systems Engineering, Beihang University in Beijing of China + guochunhui@buaa.edu.cn*
[5]** *School of Reliability and Systems Engineering, Beihang University in Beijing of China + wrrbuaa@163.com*
[6] *School of Reliability and Systems Engineering, Beihang University in Beijing of China + holomoon@me.com*

Abstract. This paper aims at the process of augmented reality maintenance, and studies the voice interaction method to realize a more natural and efficient human-computer interaction between maintenance personnel and virtual objects. Study the application and functional requirements of voice interaction in augmented reality assisted maintenance by analyzing the current interaction status. Determine the information transmission between computer and maintenance personnel and establish a knowledge base for maintenance process. Thus, a suitable semantic-based model of voice interaction mechanism is designed.

In this paper, speech signal and text information are converted by speech recognition and synthesis technology. And the semantic understanding method based on keyword matching is used to analyze the order of maintenance personnel, so as to convey the feedback information to maintenance personnel or make the operating system perform corresponding functions. Finally complete the communication between the maintenance personnel and the virtual objects.

1. Introduction

Augmented reality is a new technology developed on the virtual reality. With the help of computer image and data interaction technology, virtual information can be mapped in real time to real scene, and human perception of real scene can be increased. At present, it is widely used in various fields.

For the assembly system with a large number of components and complex constraints, the primary maintenance personnel have limited efficiency and long training time. Especially in the complex maintenance environment, the mis-operation probability is higher.

But different from virtual maintenance, a typical augmented reality system is user-centric which display virtual information through a headband display or projector and so on. Therefore, human-computer interaction is one of the key technologies, and we seek more possibilities for the communication between maintenance personnel and virtual objects. The traditional human-computer interaction technology is completed by the mouse and the keyboard, and now people want some simple and natural ways of information exchange, such as posture, facial expression, speech in interpersonal communication.

Currently, interactive research in augmented reality focuses more on human gestures, such as the HoloLens produced by Microsoft. Because of the particularity of the maintenance environment, the posture of maintenance personnel is relatively fixed, and their hands cannot completely free. The semantic meaning of the gesture is irrelevant to the movement of the finger, so the user's operation is not natural and inconvenient for memory. The application of voice interaction can overcome these difficulties. With the development of speech recognition and synthesis technology, voice interaction has become an important interactive mode of virtual assembly technology in the field of industrial design and simulation. To use voice guidance can greatly save time between man and machine. The maintenance personnel get rid of the constraint of fixed posture through the voice interaction mechanism. The system can quickly and accurately identify the user's intent and pass it to the corresponding application.

The research of speech interaction involves the application of speech recognition and synthesis technology, the research of semantic understanding, the design of speech interaction mechanism and the construction of speech interaction system. The research process of this paper is shown in the following figure 1.

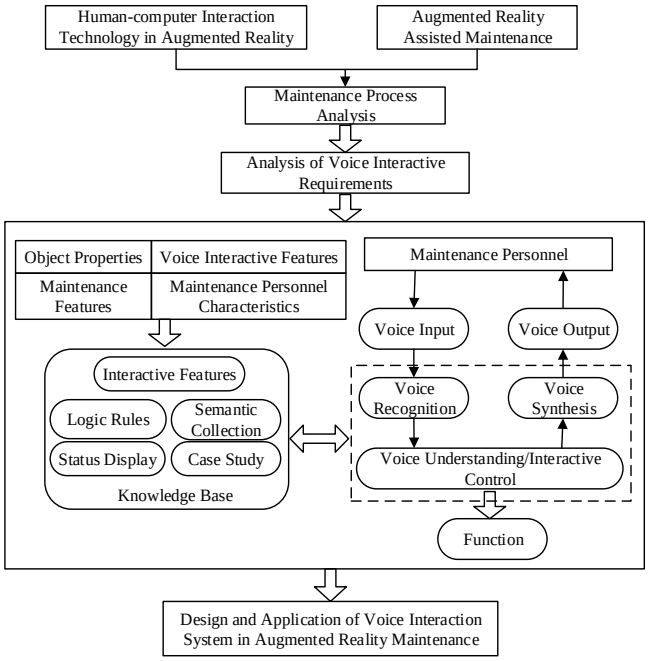

Figure 1. The research process of the thesis

2. Applicability of Speech Interaction and Functional Requirements Analysis in Augmented Reality Maintenance

2.1 The analysis of the interactive features in augmented reality assisted maintenance

The application of augmented reality technology in maintenance field can make up for the deficiency of virtual maintenance technology in the real maintenance scene. It provides more information to users in the complex equipment maintenance, and even manipulate the presentation intuitively. And it provides accurate and real-time help for professional users, and convenience to learning and training for beginner. It is of great significance to shorten the maintenance cycle of equipment and improve the efficiency of maintenance.

The difference between augmented reality maintenance and virtual maintenance is that the former users place in the real environment and have higher requirements of real-time capability. As a result, users have high demands for interactivity, and they hope a more natural way to interact with machines.

At present, interactive research in Augmented Reality focuses on gesture interaction. However, in the actual maintenance, especially in the case of battlefield repair, gestures have a long interaction time and the hands of the maintenance personnel are occupied which make the operation inconvenient. Through the above analysis, it is necessary to study the voice interaction function, and the voice interaction can make up for the shortcomings of the gesture interaction. After the voice interaction system has built, it can be combined with other ways to realize multi-channel interaction.

2.2Analysis of voice interactive requirements in augmented reality maintenance

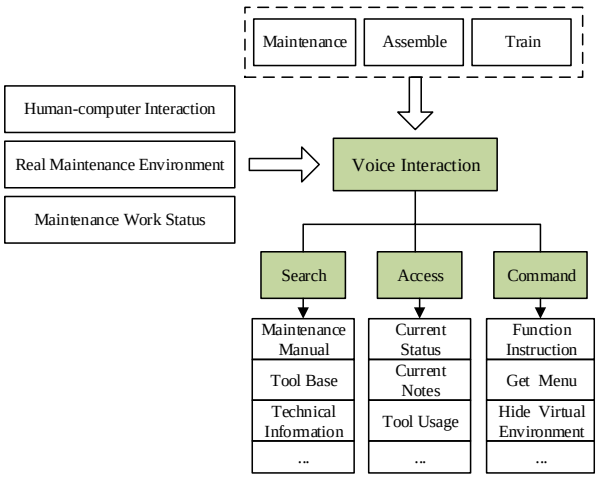

Figure 2. Analysis of voice interactive requirements for augmented reality maintenance

The influence factors in interactive process of augmented reality maintenance are analyzed comprehensively. It including the interactive mode, the real maintenance environment and the working state of the maintenance personnel. And then learn from the voice interaction requirements of virtual maintenance, assembly and personnel training to determine the interactive requirements in the augmented reality maintenance. The results are divided into three parts: information search, resource access and command interaction. As shown in the following figure 2.

- Users can search the maintenance information, such as the current maintenance status, attention, tool usage and other dynamic information. To realize the search function needs the auxiliary support of other technology modules in the augmented reality system, such as scene acquisition module and tracking registration technology. For example, if the maintenance personnel input " What is the next maintenance operation? " by voice, then the system can prompt it according to the current maintenance status.
- Resource access is a variety of technical data required for the maintenance process, including electronic maintenance manual, maintenance tool library and maintenance gesture library. For example, if the maintenance personnel input " Which kinds of maintenance tools will be used in the maintenance process for XX?" by voice, the system will provide the list of tools according to requirements to help the maintenance personnel prepare in advance.
- Command interaction includes function selection, instruction information confirmation and so on. In the operation process, the basic selection and determination function can be realized by voice interaction. For example, the system prompts "Please enter your selection.", then users can say the name or number of the object to be selected directly. Due to the different personal language habits or the diversity of the maintenance environment, it is necessary for the system to confirm the order to ensure correctness. Such as "whether to confirm the exit?".

3. research on voice interaction mechanism in augmented reality assisted maintenance

When analyzing the requirements of voice interaction functions in augmented reality maintenance, it is necessary to study the voice interaction mechanism, and determine the way of information transmission between computer and maintenance personnel. It is the basis for researching the voice interaction method and constructing interactive system. The conversion between command information and voice signal is through speech recognition and synthesis technology, and the communication between maintenance personnel and system is realized through semantic understanding.

Users can send the voice information to the voice recognition module by the hardware such as a microphone, and then extract the characteristic parameters of the identified text information to match the keyword with the knowledge base and the case base. Output the result with the highest matching degree to the dialogue management module, and confirm the completion of the appropriate control commands. The voice interaction mechanism for augmented reality maintenance is as shown in the figure 3.

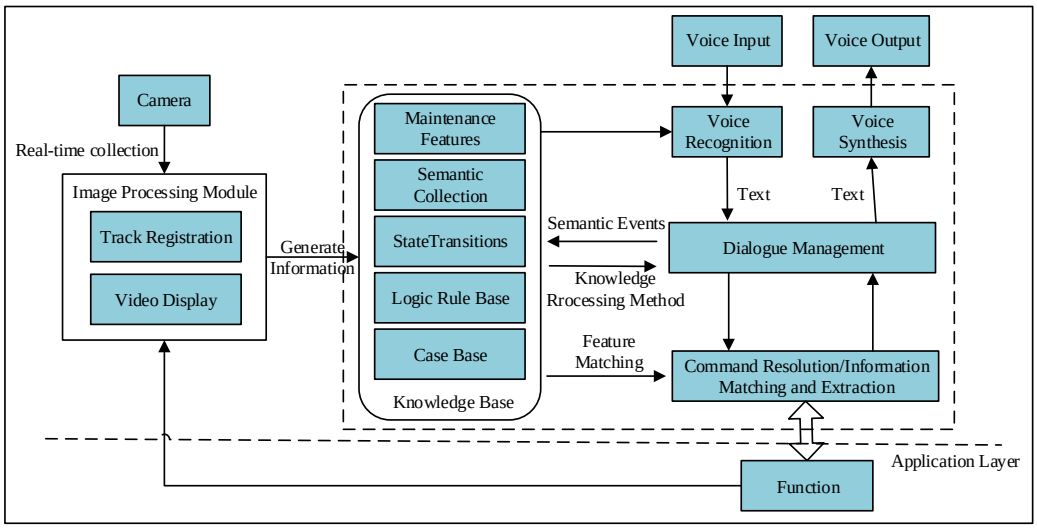

Figure 3. Analysis of voice interactive mechanism for augmented reality maintenance

It mainly includes speech recognition and synthesis module, semantic understanding module, dialogue management module and knowledge base. The knowledge base is the basis for understanding the user's intention, including maintenance features, semantic template library, logic rule base, status information and

so on. All the information needed in the maintenance process comes from the knowledge base. The case base which can record the information in each identification process and simplify the matching process to a certain extent. It can be updated through continuous learning.

Speech recognition and synthesis module relies on speech recognition technology and speech synthesis technology. It can convert the user's voice signal and text information. Through semantic understanding, it can parse the user's instructions and translate the identified text information into a language which can understand by machine. In the voice interaction system, it can convert information into commands understandable by the control module so that the function can be executed and output voice prompt; It can also analysis based on users requirements, and retrieve the information in the knowledge base. Convert it to a natural language text that the users can understand according to certain grammar file, and then converts the text into the voice signal to the user through the speech synthesis technology.

The dialog management module is the brain of the voice interaction system. It mainly includes two steps:

1) Remember the current state of the system runtime. Select the appropriate action after the conversation occurs, and switch to the next status according to the dynamic transition diagram.

2) Maintain and update the state of the dialogue. The dialogue state is the data representation that the machine can handle, including information that affects the decision making.

Maintenance status during the maintenance is relatively limited, it can directly define the state of the dialogue system based on Finite State Machine (FSM)

4. Speech recognition and semantic understanding

4.1 Semantic recognition and understanding

Speech recognition refers to the technology that the voice system can transform the natural speech signal into a text file in computer-readable or executable command in real time. Speech recognition module usually consists of three parts: speech feature extraction, pattern matching and reference pattern library. It needs to sense the scene actively.

During the maintenance process, maintenance personnel input the maintenance problems encountered by voice into the system. At present, the mature speech recognition technology can convert maintenance personnel commands into words. Using semantic comprehension method based on keyword matching analyze the maintenance personnel' intention, and match the extracted characteristic parameters with the maintenance semantic set. And output the result with the highest matching degree to the dialog management module, so as to realize operation or convey information to maintenance personnel.

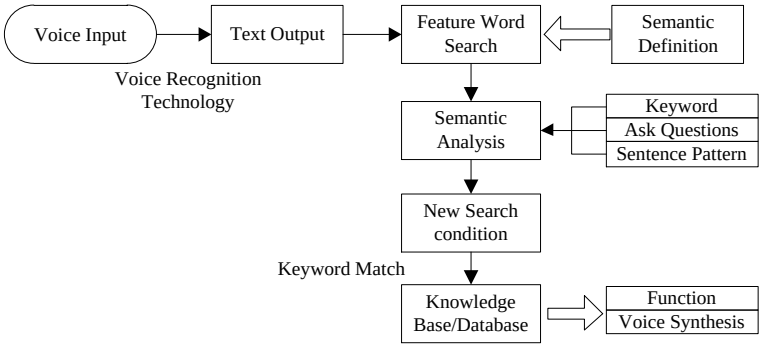

Figure 4. Speech recognition and understanding model

This process focuses on how to complete the semantic understanding of speech signals. First, the semantic definition of the user terms in the knowledge base is made. Due to the existence of a certain structure of Chinese information, we can obtain the characteristic words that can match the knowledge base by analyzing the natural language phrases of the user's expression search target. The characteristic words are extracted from the textual information so as to make a semantic analysis of the key words. The way of asking questions and the key points of the questions form a new retrieval condition , and use the keyword matching method to retrieve the required information in the knowledge base.

4.2 The establishment of the library

The construction of the knowledge base and case base in the semantic understanding model is a key issue. The third chapter introduces the contents of the knowledge base, and the specific analysis will be carried out as follows. They are the core components of the semantic understanding part, which support the user's intention recognition and implementation. The lack of comprehensive knowledge base construction directly leads to the speech interactive system cannot obtain the correct recognition result. Therefore, the knowledge base should be built on the basis of a comprehensive analysis of the requirements of language interaction for augmented reality maintenance. In addition, the structure of knowledge base and case base should be designed to be as complete and clear as possible for retrieval.

Knowledge base includes dictionary library, semantic template library, logical reasoning library and case base. The dictionary library is used to store the different expressions of the voice instruction vocabulary. The semantic template library is used to match the structure of the language when the user semantic parsing is performed, so as to obtain the resolution of the user's intention. Logical reasoning library is used to record the prerequisites and interaction events triggered by maintenance events to support the system to determine the application. The experience database is used to record the matching process and result of each interactive event processed by the system to improve the accuracy of system identification. The case base contains a record of each matching process. It learn each new case to complete the dynamic build by learning components.

The semantic collection provides support for speech recognition, and summarizes the required vocabulary. It completes the construction of the dictionary by integrating the vocabulary encoding, the meaning of the vocabulary, the description of the vocabulary and the description of the vocabulary text by creating a limited collection. Define the knowledge of maintenance feature words and functions used in the actual repair. The complete semantics set is a necessary condition for accurately identifying the content of the user's input voice instruction. On the basis of the complete analysis of the functional requirements of voice interaction, the user terms used in the maintenance process are analyzed and sorted out. Establish proper grammar files according to the features of the maintenance process function instruction and the communication characteristics between maintenance personnel and virtual environment in actual maintenance process. The user terms is integrated into the semantic collection through a certain grammatical format, forming a complete user instruction dictionary, as shown in figure 5.

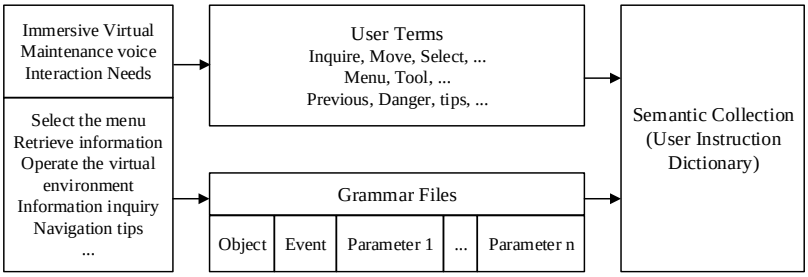

Figure 5. The process of establishing the library

5. Speech synthesis or system execution

This part contains natural language synthesis and speech signal synthesis. Natural language generation is a branch of artificial intelligence and computational linguistics, as opposed to natural language analysis. It divides into textual planning stage and surface form generation stage. Starting from the conceptual level of abstraction, the former is the information what should be told and be expressed. The latter produces text by selecting and executing certain semantics and grammar rules determine how the text is expressed. For the possible interaction of sentences in augmented reality system, corresponding sentence patterns are constructed which includes constants and variables. For example, "The tool used in this step is XX", " Next, please unscrew the screw " and so on. After the semantic understanding is completed and the information to be conveyed is confirmed, the text generator embeds the information as a string into the template instead of the variable to generate a text message that expresses reasonableness.

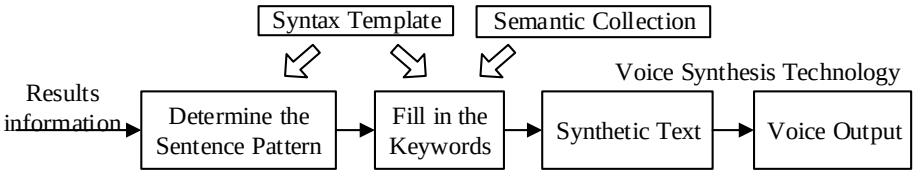

Figure 6. Speech synthesis model

The synthesized text is converted to standard natural spoken language in real time according to the voice signal processing rules. And system convey the voice signal to the maintenance personnel and accept the information feedback of maintenance personnel. Thus complete the interactive information process.

6.Conclusion

The application of voice interaction in augmented reality maintenance can enhance information exchange and give more operability for users, and improve the work efficiency of maintenance personnel. This paper mainly studied the speech interaction method in augmented reality maintenance, and the interaction mechanism is designed to lay the foundation for the realization of the voice interactive system. The main research significance is as follows: Promoting the applicability of augmented reality technology in the maintenance field; Provided a maintenance-oriented human-computer interaction mode; The role of voice interaction in virtual maintenance continues to be played in the field of augmented reality maintenance.

Semantic understanding is the core of interactive process. In this paper, the understanding method of keyword matching is proposed, but there is no specific research on it. This is also the direction of future research. After perfecting the research of above technology in the method, the system of voice interaction can be established. With the help of HoloLens and other hardware devices, it can be applied in the actual maintenance to test and improve.

References

[1]Guo, Z., Lv, C., Zhou, D., & Wang, Z. (2016). Human-computer interaction in immersive virtual maintenance. Intelligent Control and Automation (pp.2567-2573). IEEE.

[2]Ke, C., Kang, B., Chen, D., & Li, X. (2005). An augmented reality-based application for equipment maintenance. International Conference on Affective Computing and Intelligent Interaction (Vol.3784, pp.836-841). Springer-Verlag.

[3]Hamidia, M., Zenati, N., Belghit, H., Guetiteni, K., & Achour, N. (2016). Voice interaction using Gaussian Mixture Models for Augmented Reality applications. International Conference on Electrical Engineering. IEEE.

[4]Araújo, Tiago, et al. "Aspects of Voice Interaction on a Mobile Augmented Reality Application." (2016)..

[5]Xincan, Z., Zuo Hongfu|赵新灿, and 左洪福. "Research on Interaction Technique of Augmented Reality Maintenance Guiding System." China Mechanical Engineering 19.11(2008):1312-1316.

[6]Chen, Guiling, et al. "Virtual human control technology for immersed virtual maintenance system in DELMIA environment." International Conference on Audio Language and Image Processing IEEE, 2010:1352-1355.