Journal of information and communication convergence engineering
http://jicce.org

J. Inf. Commun. Converg. Eng. 21(3): 239-243, Sep. 2023 Regular paper

Design of an Automatic Summary System for Minutes Using Virtual Reality

Amsuk Oh*, ***Member*, *KIICE***

Department of Digital Contents, TongMyong University, Busan 48520, Korea

Abstract

Owing to the environment in which it has become difficult to face people, on-tact communication has been activated, and online video conferencing platforms have become indispensable collaboration tools. Although costs have dropped and productivity has improved, communication remains poor. Recently, various companies, including existing online videoconferencing companies, have attempted to solve communication problems by establishing a videoconferencing platform within the virtual reality (Virtual Reality) space. Although the VR videoconference platform has only improved upon the benefits of existing video conferences, the problem of manually summarizing minutes because there is no function to summarize minute documents still remains. Therefore, this study proposes a method for establishing a meeting minute summary system without applying cases to a VR videoconference platform. This study aims to solve the problem of communication difficulties by combining VR, a metaverse technology, with an existing online video conferencing platform.

Index Terms: Summary of Meeting Minutes, Video Conferencing Platform, Virtual Reality

I. INTRODUCTION

In November 2019, online videoconferencing platforms became an indispensable tool as non-face-to-face work became commonplace after the pandemic. It has many advantages over offline meetings, including not requiring a physical meeting space, receiving less external interference, and the ability to share various presentations and produce minutes [1]. However, unlike offline meetings, there were often problems that made communication difficult because it was difficult to explain abstract concepts or understand nonverbal expressions. Additionally, it was a problem that conference attendees' concentration was decreased because it was uncomfortable to continue to display their actual faces on the monitor screen [2].

Recently, global online videoconferencing service providers, such as ZOOM and Teams, have increased their efforts to incorporate metaverse technology to address the difficulty of interacting with existing online videoconferencing platforms [3-4]. The result is a virtual reality (VR)-based videoconferencing platform. VR is a core metaverse technology that simulates a realistic experience by implementing a virtual world. This technology is used in all areas of business-to-business (B2B) such as remote work, HR, corporate events, and collaboration with overseas partners or branches [5].

However, one drawback of existing online or VR videoconferencing platforms is that it is difficult to review the meeting contents. For example, when daily or business conversations that are not in the shared data come and go, they may be considered insignificant at the time of the conversation. Therefore, if you miss an important part of a conversation, you will have to re-listen to the recorded voice for reconfirmation. Therefore, summarization, which involves extracting and reconstructing keywords and sentences from

Received 21 June 2023, **Revised** 7 August 2023, **Accepted** 10 August 2023
***Corresponding Author** Amsuk Oh (E-mail: asoh@tu.ac.kr; Tel: +82-51-629-1211)
Department of Digital Contents, Tongmyong University, Busan, 48520 Korea

Open Access https://doi.org/10.56977/jicce.2023.21.3.239 print ISSN: 2234-8255 online ISSN: 2234-8883

sentences, can be a solution to the speech conversation replication problem [6].

Therefore, this study aimed to investigate how to incorporate a summary function into voice-recognition meeting minutes generated within a VR-based video conference platform. A meeting minute summary function was created using a natural language processing model and applied to a pre-produced VR video conferencing platform to ensure that meeting minute summarization was successfully performed within the system. Thus, the execution environment of the device was integrated into a single VR, which improved its ease of use.

II. SYSTEM MODEL AND METHODS

A. Summary of the VR Meeting Minutes

Recently, VR has drawn attention as a next-generation technology that can overcome the limitations of video conferencing while making telecommuting more permanent. VR is an artificially created virtual environment or technology that mimics reality. Users can directly control a typical online environment using a device called head mounted display (HMD) and a controller if it stops watching in a virtual environment. Thus, they interact with numerous elements implemented within VR [7]. Global companies such as Microsoft, Facebook, and Google are developing and releasing VR videoconferencing platforms that combine VR with videoconferencing [7].

B. Text Summary

A text summary is a method for converting a long document into a short sentence composed of key words. In summary, information loss should be minimized, and text summarization is generally divided into extractive and abstract summarization. Extractive summary is a method of extracting important key sentences or words from an original text and creating a summary. Because this method only summarizes sentences or words that are present in the original text, the model's language expression ability is limited, and the sentences in the output result may not be natural. Text rank is a representative machine learning algorithm for extractive summaries, of which the Naver News summary bot is an example. Abstract summary, which falls under the natural language generation (NLG) category of the natural language processing fields, is a method of interpreting the meaning of the original text to understand the context, generate new sentences, and summarize the original text [8].

C. Summary of the Meeting Minutes

A study embodying this concept has been presented [9]. In existing methodologies, probability graph models, the hidden Markov model (HMM), and the conditional random field (CRF), all of which have been quickly replaced by neural network structures, represent sentences. Until 2016, the proposed structures mainly included recurrent neural networks (RNNs), long short-term memory (LSTM), and gated recurrent units (GRUs). An RNN is a recursive structure in which the output enters an input, whereas a circular neural network is a structure in which the hidden state of the previous word is used to determine the hidden state of the word [10].

III. VR MEETING MINUTES SUMMARY SYSTEM DESIGN

A. Platform Design Considerations

The following issues must be considered before the proposed VR video conference platform and automatic summary system for meeting minutes are implemented:

First, for an effective interaction between the user's objects within the VR environment, unnecessary information should be removed and designed to have an esthetic and minimal design. Additionally, it supports avatars and must improve their usability by diversifying how the controller is operated.

Second, because of the nature of a VR platform with several small groups, it is necessary to build a server suitable for multiple communications to support smooth one-to-one (1:N), one-to-many (1:N), and many-to-many (M:N) communications between users.

Third, voice recognition is a technology that converts spoken words into text. It is commonly supported by online and VR video conferencing platforms. In this study, a voice recognition system that considers information exchange was designed to collect the minutes.

Fourth, the minutes must be automatically and accurately summarized. A minute summary is the task of generating accurate natural key sentences.

B. Configuration of the Entire System

The automatic summary system for VR meeting minutes proposed in this paper is demonstrated using four factors: usability, multicommunication, information exchange, and accuracy. Figure 1 shows the system configuration and its operation.

① The VR conference platform consists of a VR space

produced by VR devices and unity, a photon network, a voice recognition API, and a meeting minutes summary algorithm with a language summary model.

② When running the platform, it is linked to detect hardware behavior within the platform through the Oculus Integration SDK, and screens such as avatars, conference tools, and data are output within VR.

③ When all objects are loaded into the VR space, they can access the conference room with the support of a photon network.

④ The Photon server uses Photon Core (based on C++) to retrieve information about meeting attendees stored in the minutes management software.

⑤ When the server is connected, a meeting is held, and because it is difficult to use a keyboard in a VR environment, it communicates by adding a voice recognition function.

⑥ During the meeting, all voice-recognized text was stored in the minute management software connected to the Photon server using the minute summary algorithm.

⑦ Meeting presenters or participants will have free access to text documents containing summaries of meeting minutes stored in the meeting minute management software

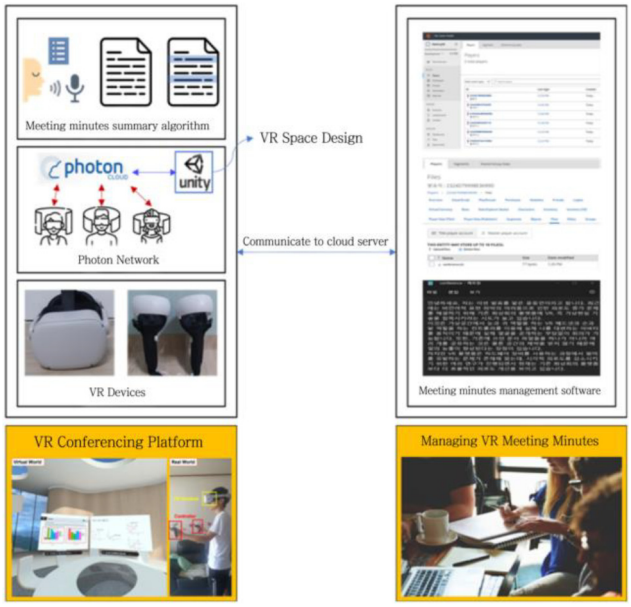

Fig. 1. Configuring the VR meeting minutes summary system

The VR meeting minute summary system presented in this paper operates as illustrated in Fig. 2. First, a VR HMD, which is a VR device, is installed and a VR conference platform is executed. The user's device is connected to the photon network, and the minute management software operates simultaneously. If one first accesses the videoconference server, a new conference room is automatically created; otherwise, one will access the previous conference room and retrieve the meeting records. When a meeting is conducted and terminated using a speech-recognition function known as speech-to-text (STT), the meeting minutes that convert voice into text are generated and stored in the meeting minute management software using the proposed meeting minute summary algorithm. The minutes saved can be retrieved at any time.

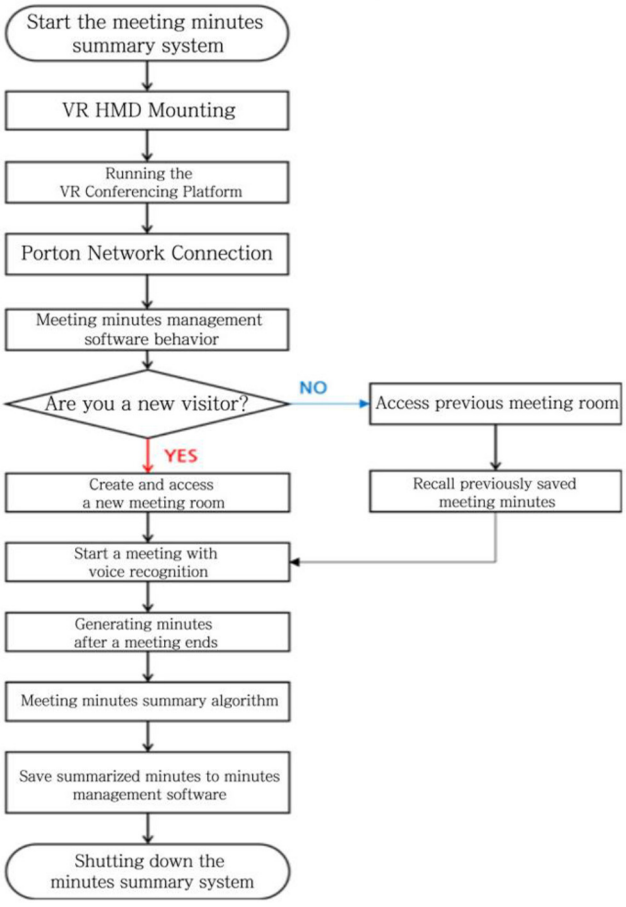

Fig. 2. Operation flowchart of the VR meeting summary system

C. VR Device

The entire system's VR device, which helps execute and manipulate content, consists of a VR headset that serves as a human eye in a virtual space and a touch controller that serves as a human hand and foot, as shown in Fig. 3. Although VR platforms can run within VR headsets rather than on PCs and mobile devices, repeatedly installing and uninstalling VR headsets is laborious. Therefore, system development and testing were performed in a PC environment using a game engine that supports a VR environment.

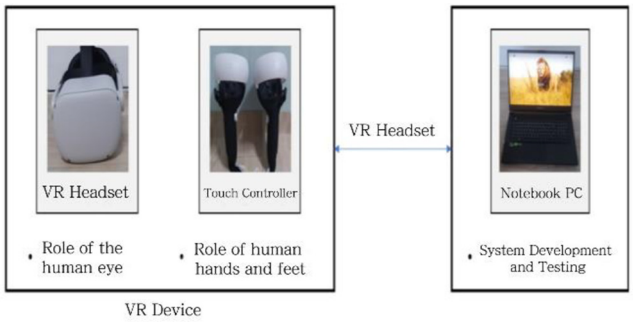

Fig. 3. Operation flowchart of the VR meeting summary system

D. Meeting Space

Figure 4 shows the structure of the VR meeting space in charge of the content, voice recognition function, and meeting minute-sharing function.

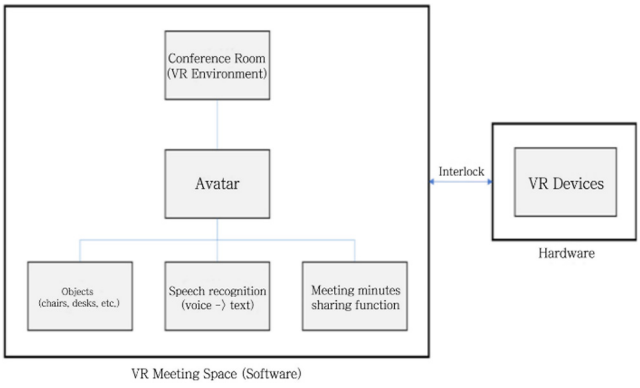

Fig. 4. Structure of the VR conference space

Table 1 describes each component. When one first accesses the conference room in the VR environment, the VR meeting space, avatar, and object are the first outputs. Following that, VR devices are connected to operate objects, such as chairs and desks, using avatars and voice recognition functions that convert voices into text, and minute sharing functions. Existing online conferencing platforms only use vision and hearing; however, VR provides a great sense of immersion because it also uses touch. Voice recognition is often used in VR environments where keyboard or mouse input is limited because the voice is translated into text.

Table 1. Components of the VR meeting space

Components	Feature Description
Conference Room (VR Environment)	Meeting space in a VR environment
Avatar	Virtual self working in VR
Objects (chairs, desks, etc.)	Aesthetic Elements of VR Conference Space
Voice recognition function (voice -> text)	Convert voice conversations to text
Meeting minutes sharing function	Sharing minutes

E. Meeting Minutes Summary Workflow

The meeting minute summary algorithm accurately summarizes voice conversations when they are converted into text using the voice recognition function. Figure because it is necessary to build a language model to have this summary function. Go through the same workflow as Fig. 5.

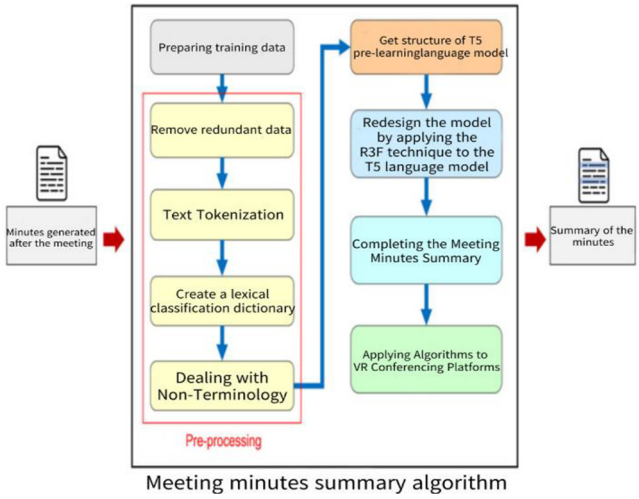

Fig. 5. Workflow of the minute summary algorithm

First, we prepared training data for model learning. Subsequently, we removed duplicate sentence data from the training data and tokenized the remaining sentences to improve the performance of the model. Here, tokenized data are solved by creating a vocabulary classification dictionary because even the same word has a different meaning. Subsequently, the meeting minute summary algorithm is preprocessed by performing a non-term processing order that is not of high importance in sentence analysis, such as investigation or exclamation. Additionally, only the structure of the T5 prelearning language model selected from related studies was introduced, and the model was redesigned by applying an R3F technique that reduces learning loss. The completed model was applied to a VR conference platform to automatically summarize voice-recognized text data in the VR conference space.

IV. CONCLUSION AND FUTURE RESEARCH WORKS

In this paper, a VR videoconferencing platform was proposed and implemented to solve the communication difficulties of existing online videoconferencing platforms. First, the avatar was designed using the Magica Voxel design tool to ensure that the actual face of the meeting participants remained visible on the screen. Several groups of small numbers of people created multiple conference rooms, enabling RPC communication using a Photon Server, which is advantageous for accessing them. The STT API provided by IBM was imported and used to implement voice recognition functions that converted the conference participants' voices into texts. Because of the implementation, when a voice was spoken within a video conference platform executed in an Oculus VR environment, the conversion of the voice into text was confirmed. Therefore, it is important to preprocess words according to experimental standards such as age, gender, and occupation so that abstract summaries can perform well for long sentences. In addition, the convenience of meeting participants is expected to increase with the minute summary system proposed in this study.

REFERENCES

[1] J. Edvardsson, H. W. Linderholm, B. Gunnarson, A. Hansson, T. T. Chen, and H. Gärtner, "To organize a conference under ever-changing conditions - Editorial to the special issue from the TRACE 2021 virtual meeting," *Dendrochronologia*, vol. 76, Dec. 2022. DOI: 10.1016/j.dendro.2022.126022.

[2] S. Bhargava, N. Negbenebor, R. Sadoughifar, S. Ahmad, and G. Kroumpouzos, "Virtual conferences and e-learning in dermatology during COVID-19 pandemic: Results of a web-based, global survey," *Clinics in Dermatology*, vol. 39, no. 3, pp. 461-466, May 2021. DOI: 10.1016/j.clindermatol.2021.06.002.

[3] A. Aghajanyan, A. Shrivastava, A. Gupta, and N. Goyal, "Better fine-tuning by reducing representational collapse," *arXiv preprint arXiv:2008.03156v1*, Aug. 2022. DOI: 10.48550/arXiv.2008.03156.

[4] J. Bradbury, S. Merity, C. Xiong, and R. Socher, "Quasi-recurrent neural networks," *arXiv preprint arXiv:1611.01576v2*, Nov. 2016. DOI: 10.48550/arXiv.1611.01576.

[5] H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and T. Zhao, "SMART: Robust and efficient fine-tuning for pre-trained natural language models through principled regularized optimization," in *The 58th annual meeting of the Association for Computational Linguistics (ACL 2020)*, *arXiv preprint arXiv:1911.03437v5*, pp. 2177-2190, Nov. 2019. DOI: 10.48550/arXiv.1911.03437.

[6] J. Devlin, M-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *arXiv preprint arXiv:1810.04805v2*, vol. 1, pp. 4171-4186, Oct. 2018. DOI: 10.48550/arXiv.1810.04805.

[7] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114v11*, Dec. 2013. DOI: 10.48550/arXiv.1312.6114.

[8] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding", *Neural Information Processing Systems 32 (NeurIPS 2019)*, *arXiv preprint arXiv:1906.08237v2*, pp. 57530-5763, Jul. 2019. DOI: 10.48550/arXiv.1906.08237.

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized bert pretraining approach," *arXiv preprint arXiv: 1907.11692*, Jul. 2019. DOI: 10.48550/arXiv.1907.11692.

[10] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *The Journal of Machine Learning Research*, *arXiv preprint arXiv:1910.10683v3*, vol. 21, no. 1, pp. 1-67, Jan. 2020. DOI: 10.48550/arXiv.1910.10683.

Amsuk Oh

Received his Ph.D. degree from the Department of Computer Engineering, University of Busan, KOREA, in 1997. He is currently a Professor at the Department of Digital Media Engineering, College of Technology, Tongmyong University, Busan, Korea. His research interests include databases, big data, the Internet of Things, healthcare systems, and medical information systems.