Development of Immersive Display System of Web Service in Living Space

Hiroshi SUGIMURA, Kazuki UTSUMI, Masayuki KANEKO, Kazuki ARIMA,
Masahiro SAKAMOTO, Yuki IMAIZUMI, Tomoki WATANABE and Masao ISSHIKI
Department of Home Electronics Development
Faculty of Creative Engineering
Kanagawa Institute of Technology
Shimo-ogino 1030, Atsugi, Kanagawa, JAPAN
Email: sugimura@he.kanagawa-it.ac.jp

***Abstract*—In this paper, we propose a system to superpose information of web services into living space. The system is controlled by natural actions which are gestures and voice in order to simple control apparatus. The system is assembled by three major modules which the sound recognition, the gesture recognition and a projector. We describe the entire process, the implementation and the evaluation of a prototype system, and confirmed an usefulness of proposal system by using a questionnaire.**

I. Introduction

Development of advanced display system is important to control smart house system or HEMS. Studies of intelligent television system collaborates network techniques to surf the Internet on television and the Skype conversation [1], [2]. A method which controls user's home by internet TV is proposed [3].

In order to simplify control apparatus, many studies propose methods using human natural actions. The studies of voice recognition control system already apply web search engines, control for smart phone, and so on [5], [6], [8]. A system to control a personal computer by using Kinect sensor is developed [4]. This paper develops a system that controls information of the Internet to harmonize both two techniques in living space. We thus propose a system that displays information of the Internet at will using gestures and voice in living space.

II. An outline of the system

Fig. 1 shows an overview of the system. The system has a camera and a microphone to obtain user's voice and gestures. A user is able to control the system without other input devices such as mouse, keyboard, and touch screen. A user instructs a function to the system. Information obtained from the function is projected onto a wall of room. A function corresponds to a web service and one or more terms, which are contained into a database. It seems ontological information. In this way, the system is able to operate by natural human speech. Displayed information is able to manipulate by user's gesture. The gesture is represented by a hand shape and a 3D coordinate value. The kind of hand shape is two types that are "grab" and "release". The "grab" represent as mouse button down and also "release" represent as mouse button up.

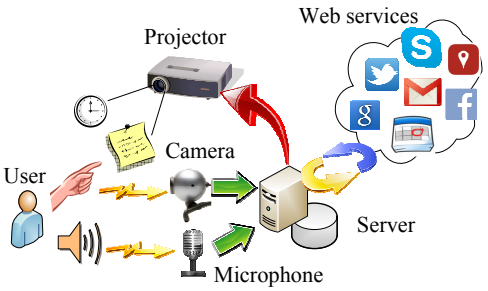

Fig. 1. An overview of the system

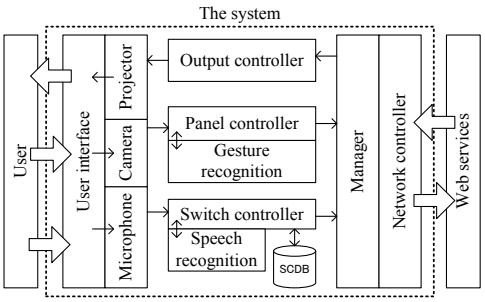

Fig. 2. An outline of the system

III. Implementation

Fig 2 illustrates an outline of the system. The part of surrounded with a broken line is the entire system. White arrows indicate the external interface. A user controls the system through user interface that contains a microphone, a camera, and a projector. Network controller manages user accounts for web services and obtains information from web services. Small arrows in a broken line indicate data flows between the modules. Data flows are shown by Fig. 3 and Fig. 4.

Fig. 3 illustrates a flow of the voice control. Voice data obtained by microphone is forwarded to switch controller module and transcribed to a text string by speech recognition module. A search is carried out into SCDB by using each term in the text string. When any term is fetched, network controller module obtains information from web service in correspondence with the term. As shown table 1, SCDB consists

978-1-4799-05145-1/14/$31.00 ©2014 IEEE

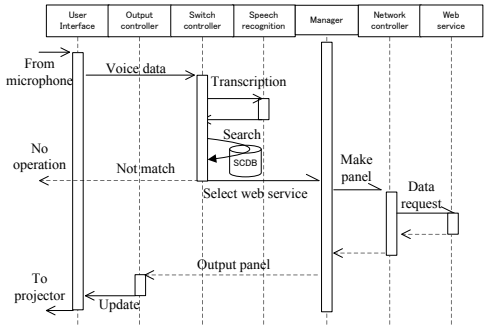

Fig. 3. A flow of the voice control

TABLE I. AN EXAMPLE OF A TABLE IN THE SCDB

Functions	Keywords (representative)		Web services
Calendar	calendar	schedule	Google Calendar
Twitter	twitter	tweet	Twitter
Route search	map	route	Google Maps
Watching movie	movie	youtube	YouTube
Telephone	chat	telephone	WebRTC

of function, service, and terms. Web service information is displayed, that is named "panel". In order to handle the panel, user's gesture is used.

Fig. 4 illustrates a flow of the hand gesture control. The system obtains the skeleton data from some depth data using a camera of Kinect and acquires hand status that is a pair of a hand gesture and a 3D position by gesture recognition module. The manager module carries out appropriately to require the web service based on the hand status through panel controller module. The processing result is sent to panel controller module, and it updates the panel. The system outputs the result for user by projector.

An image of running in prototype is shown in Fig. 5. We carry out a questionnaire to evaluate the system. The questionnaire is composed of several parts such as a score of interest, its reasons, a free description, and so on, and are carried out with a total of 54 people (22 men and 32 women). Fig. 6 (A) is shown the distribution of each age. Fig. 6 (B) is shown the distribution of interest.

IV. CONCLUSION AND FUTURE WORKS

In this paper, we propose a system to superpose information of web services into living space. The system is controlled by

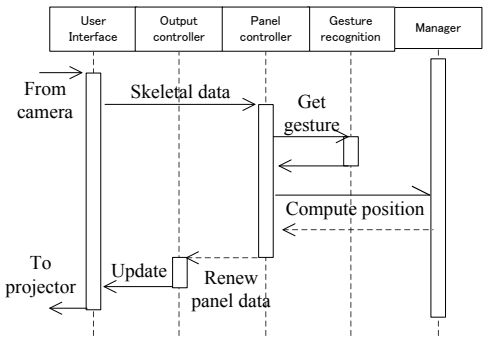

Fig. 4. A flow of the hand gesture control

Fig. 5. A prototype of the system

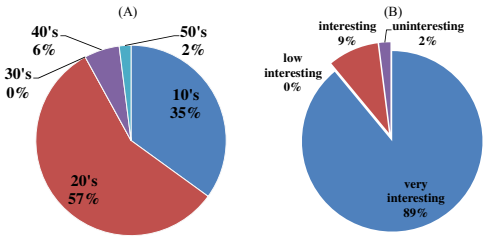

Fig. 6. The result of questionnaire

natural actions such as gestures and voice. We assemble the system by three input/output devices which are a microphone for the sound recognition, a camera for the gesture recognition, and a projector to display information. We described the entire process, the implementation and the evaluation of a prototype system. Then, we confirmed the usefulness of proposal system.

REFERENCES

[1] Cesar, Pablo and Geerts, David, *Past, present, and future of social TV: A categorization*: Consumer Communications and Networking Conference (CCNC), 2011 IEEE, pp. 347–351, 2011.

[2] Zivkov, Dusan and Davidovic, M and Vidovic, M and Zmukic, N and Andelkovic, A, *Integration of Skype client and performance analysts on television platform*: MIPRO, 2012 Proceedings of the 35th International Convention, pp. 479–482, 2012.

[3] Rivas-Costa, CARLOS and Gómez-Carballa, MIGUEL and Anido-Rifón, LUIS and Fernandez-Iglesias, MANUEL J and Valladares-Rodríguez, SONIA, *Controlling your Home from your TV*: Recent Advances in Electrical and Computer Engineering, pp. 124–130, 2013.

[4] Lee, Sang-Hyuk and Oh, Seung-Hyun, *A Kinect Sensor based Windows Control Interface*: International Journal of Control & Automation, Vol. 7, No. 3, 2014.

[5] Andics, Attila and McQueen, James M and Petersson, Karl Magnus and Gál, Viktor and Rudas, Gábor and Vidnyánszky, Zoltán, *Neural mechanisms for voice recognition*: Neuroimage, Vol. 52, No. 4, pp. 1528–1540, 2010.

[6] Zumalt, Joseph R, *Voice recognition technology: Has it come of age?*: Information technology and libraries, Vol. 24, No. 4, pp. 180–185, 2013.

[7] Schalkwyk, Johan and Beeferman, Doug and Beaufays, Françoise and Byrne, Bill and Chelba, Ciprian and Cohen, Mike and Kamvar, Maryam and Strope, Brian, *"Your Word is my Command": Google Search by Voice: A Case Study*: Advances in Speech Recognition, pp. 61–90, 2010.

[8] Halarnkar, Pallavi and Shah, Sahil and Shah, Harsh and Shah, Hardik and Shah, Jay, *Gesture Recognition Technology: A Review*: International Journal of Engineering Science & Technology, Vol. 4, No. 11, 2012.