Proceedings of KFIS Spring Conference 2006 Vol. 16, No. 1.

음성인식과 지문식별에 기초한 가상 상호작용

Virtual Interaction based on Speech Recognition and Fingerprint Verification

Sung-Ill Kim[1], Se-Jin Oh[2], Dong-Hun Kim[3], Sang-Yong Lee[4], Seung-Gook Hwang[5]

[1] Department of Electronic Engineering, Kyungnam University
E-mail: kimstar@kyungnam.ac.kr
[2] Radio Astronomy Division, Korea Astronomy and Space Science Institute
[3] Department of Electrical Engineering, Kyungnam University
[4]Division of Computer Science and Engineering, Kyungnam University
[5]Department of Industrial Engineering, Kyungnam University

Abstract

In this paper, we discuss the user-customized interaction for intelligent home environments. The interactive system is based upon the integrated techniques using speech recognition and fingerprint verification. For essential modules, the speech recognition and synthesis were basically used for a virtual interaction between the user and the proposed system. In experiments, particularly, the real-time speech recognizer based on the HM-Net(Hidden Markov Network) was incorporated into the integrated system. Besides, the fingerprint verification was adopted to customize home environments for a specific user. In evaluation, the results showed that the proposed system was easy to use for intelligent home environments, even though the performance of the speech recognizer was not better than the simulation results owing to the noisy environments

Key Words : Smart Home, Virtual Interaction, Speech Recognition, Fingerprint Verification

1. Introduction

Recently, the fingerprints have been one of the most widely used biometrics for personal verification as an important function of door locks. The system of fingerprint verification matches the given fingerprint with the one stored in its database and a degree of similarity, named score, is finally computed. If such score is higher than a certain acceptance threshold, then the person is classified as an inhabitant of the house. In the present paper, the fingerprint verification can be used for a user-customized interface in addition to a function of home security.

When talking about the intelligent home, it means different things to different people. The interactive system is integrated with the essential components, for example, such as speech recognition, speech synthesis, and fingerprint verification. When we get into our house through the verification process of fingerprint identification that is used for user-customization, the user-customized interaction is automatically formed so that the intelligent home can provide you with more convenient and comfortable living environments.

The basic idea is based on the fact that the place we spend most time at home is our living room, particularly in a sofa. The concept is started on the assumption that the interaction between user and system can be built when user sits in the sofa that is interconnected with the proposed system. The proposed system is designed to allow users to converse with their home based on the user-customized interaction where the system puts emphasis on an easy-to-use and user-friendly man-machine interface. As a consequence, the intelligent home, as an aim of this study, makes it possible to lead the living environments to the most suitable condition for users by integrating speech recognition with fingerprint verification. For fingerprint verification, in this study, it was used to customize the home environments for specific user. For speech recognition and synthesis, on the other hand, they were used for interaction between user and intelligent home.

2. The Interface for Speech Recognition Using HM-Net

HMM(Hidden Markov Model) is a mathematical model which has been widely used in speech recognition systems. In this study, we used HM-Net(Hidden Markov Network)[1,2] which is an efficient representation of context-dependent phonemes for speech recognition. The HM-Net, which has various state lengths and shares their states one another, is automatically generated by SSS(Successive State Splitting)[2,3]. The SSS is an iterative algorithm that progressively grows HM-Net, where each state in the network is associated with a 2-component Gaussian mixture.

In the algorithm, a state is selected to be split according to which has the largest divergence between its two mixtures. The state is then split on the contextual and temporal domains, and the one giving greater likelihood is chosen. The affected states are retrained using the Baum-Welch algorithm. The above procedure is iterated until getting to a pre-defined number of states.

The PDT-SSS(Phonetic Decision Tree-Successive State Splitting)[4] based on the SSS algorithm is a powerful technique to design topologies of tied-state models, and is possible to generate highly accurate HM-Net. Each state of HM-Net has the information such as state index, contextual class, lists of preceding and succeeding states, parameters of the output probability density distribution and the state transition probability. If contextual information is given, the model corresponding to the context can be determined by concatenating several associated states within the restriction of the preceding and succeeding state lists.

Fig. 1 shows an overall schematic of HM-Net speech recognition system. In case the speech signals are given to the HM-Net speech recognition system, the acoustic features are first picked out for pre-processing, and then given to the search module that uses both tree-structured lexicon and HM-Net Triphones. The final recognition results are then obtained by frame synchronous Viterbi beam search algorithm using word-pair grammar.

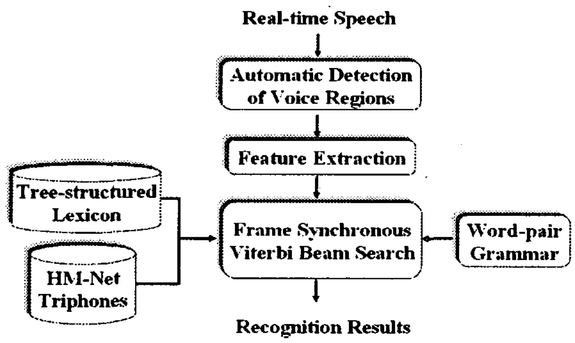

Fig. 1. Overall schematic of HM-Net speech recognition system

3. The Interface for Fingerprint Verification

The fingerprint verification has been widely used for granting or denying the access to certain typical resources such as computer, ATM, and such security areas. In this study, this technique was adopted for user-customized interaction.

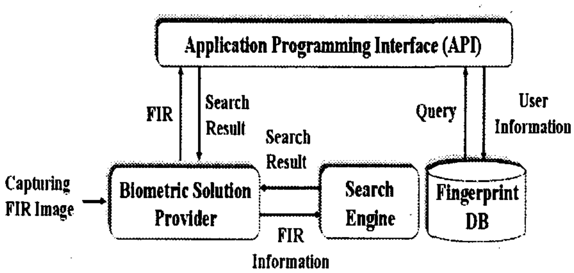

Fig. 2. System diagram for fingerprint verification. (FIR: fingerprint identification record)

Fig. 2 illustrates the overall diagram for procedure of fingerprint verification used in the proposed system. The module of fingerprint verification needs four kinds of different modules such as fingerprint device, device driver, fingerprint extraction and search engine module.

All these modules are controlled by an application programming interface. The BSP(Biometric Solution Provider) is the interface module between application programs and search engine. The BSP module is used to capture fingerprints, perform matching operation, and directly control the search engine. The search engine is the main module of the system to support large volume of 1:N matching operation. The engine registers, manages and searches fingerprint data in the memory resident fingerprint database.

The software interface for management consists of the functions for fingerprint registration and search. In experiments, we used the Fingkey Hamster of NITGEN for a fingerprint image capturing deviceas well as biometric service provider SDK for a software interface.

4. The Proposed Interactive System

The proposed system can be built by integrating two main module of both HM-Net speech recognition and fingerprint verification, mentioned above. Fig. 3 shows the flow diagram of the processing based on the proposed system, which is operated in real time. It indicate show to build the interaction between user and system. When user gets into the home through the verification of his or her fingerprint, for the first step, the system searches the pre-registered database and then presents a verification result with a synthesized voice. Simultaneously, the system makes a decision whether to activate speech recognition. In case speech recognition is activated, the system notifies user of necessary information such as important messages or schedules. In this way, the user-customized services are provided through the virtual interaction chiefly based on speech recognition and synthesis.

Fig. 4 shows the main frame of a user interface, which has been made by VC++, with the modules of speech recognition and fingerprint verification. The system provides several functions. First, it is possible for user to control multimedia application programs such as video, MP3 player, CD player etc. Besides, several kinds of electrical appliances such as electrical fans, lamps can be controlled by the power relay units of print-port interface.

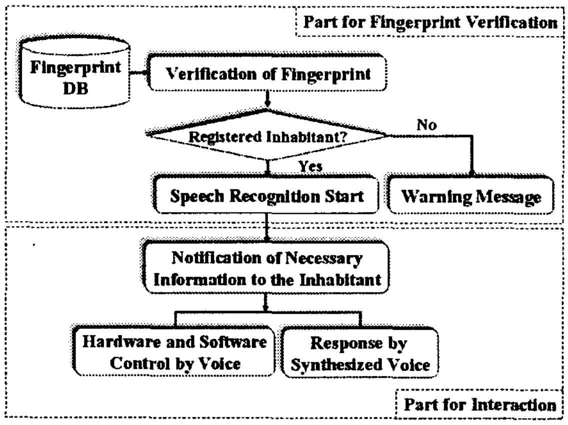

Fig. 3. The flow diagram of building interaction between user and system.

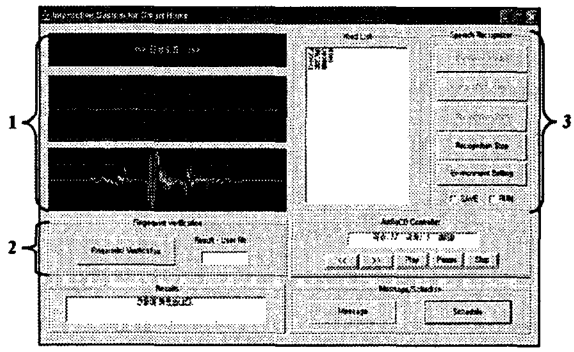

Fig. 4. Main frame of user interface window. (num.1,3: interface for speech recognition, num.2: interface for fingerprint verification)

5. Experimental Result

All speech data were sampled at 16kHz, quantized at 16 bits, pre-emphasized with a transfer function of $(1-0.97z^{-1})$, and processed to extract acoustic features using a 25ms Hamming window with a 10ms shift. The feature parameters consisted of total 39 order LPC Mel Cepstrum coefficients including normalized log-power, 1st and 2nd order delta coefficients.

The database used for speech recognition consists of two kinds of database, one of which is made by ETRI(The Electronics and Telecommunications Research Institute), and the other is made by KLE(Center for the Korean Language Engineering).

For experiments, total 40 male college students were participated in the evaluation of the system. For examining the human performance on the accuracies of the proposed system, we first showed them a demonstration of how to use and operate the system, and made them to use it for themselves.

As the evaluation using questionnaire, all participants marked ranks from 1- to 5-point about how easy and how useful they thought the system was to use. We could get the results as shown in Fig. 5 and 6 that the system was relatively easy to use and would be useful in real applications.

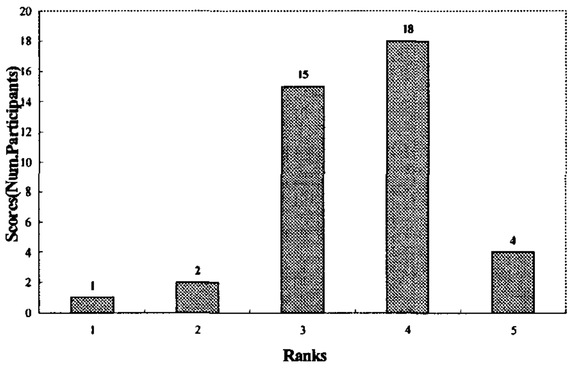

Fig. 5. Evaluation of the system using questionnaire in terms of how easy the system was to use.

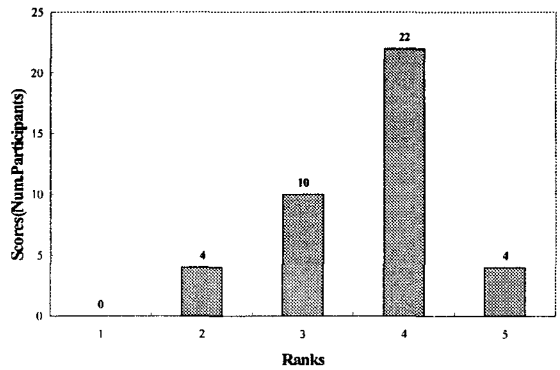

Fig. 6. Evaluation of the system using questionnaire in terms of how useful the system will be in real applications.

6. Conclusion

This study has described the user-customized interaction based on speech recognition and fingerprint verification for intelligent home environments. The results from the experimental evaluation have shown that the proposed system was relatively easy to use. This means a possibility for building a virtual interaction using the system that might give us much more convenient and comfortable living environments. However, the accuracy of speech recognition was unsatisfactory owing to the noisy environments, diverse speaking rates, and speaking styles of users. From the evaluation, we could obtain several ideas on the system as future works. Namely, the future works should be conducted on more natural methods of interaction so that the future systems would allow users to feel more natural in virtual interaction for intelligent home environments.

References

[1] M. Suzuki, S. Makino, A. Ito, H. Aso and H. Shimodaira, "A new HMnet construction algorithm requiring no contextual factors,"IEICE Trans. Inf. & Syst., Vol. E78-D, No. 6, pp. 662-669, 1995.

[2] M. Ostendoft and H. Singer, "HMM Topology design Using Maximum Likelihood Successive State Splitting," Computer Speech and Language Vol. 11, pp. 17-41, 1997.

[3] J. Takami and S. Sagayama, "A Successive State Splitting Algorithm for Efficient Allophone Modeling," Proc. of ICASSP'92, Vol. 1, pp. 573-576, 1992.

[4] T. Hori, M. Katoh, A. Ito and M. Kohda, "A Study on HM-Nets using Decision Tree-based Successive State Splitting," Proc. of ICSP'97, Vol.1, pp. 383-387, 1997.