WPS and Voice-XML-Based Multi-Modal Fusion Agent Using SNNR and Fuzzy Value

Jung-Hyun Kim and Kwang-Seok Hong

School of Information and Communication Engineering, Sungkyunkwan University, 300, Chunchun-dong, Jangan-gu, Suwon, KyungKi-do, 440-746, Korea
kjh0328@skku.edu and kshong@skku.ac.kr
http://hci.skku.ac.kr

Abstract. The traditional fusion methods of multiple sensing modalities are summarized with 1) data-level fusion, 2) feature-level fusion and 3) decision-level fusion. This paper suggests the decision-level fusion-oriented novel fusion and fission framework, and it implements WPS (Wearable Personal Station) and Voice-XML-Based Multi-Modal Fusion Agent (hereinafter, MMFA) using audio-gesture modalities. Because the MMFA provides different weight and a feed-back function in individual recognizer, according to SNNR(Signal Plus Noise to Noise Ratio) and fuzzy value, it may select an optimal instruction processing interface under a given situation or noisy environment, and can allow more interactive communication functions in noisy environment. In addition, the MMFA provides a wider range of personalized information more effectively as well as it not need complicated mathematical algorithm and computation costs that are concerned with multidimensional features and patterns (data) size, according as it use a WPS and distributed computing-based database and SQL-logic, for synchronization and fusion between modalities.

1 Introduction

A recent study on multimodal interaction describes the potential of multimodality in terms of increased adaptability, robustness and efficiency. There are some well-known approaches towards modeling and prototyping multimodal systems such as RDPM [1, 2], theoretical frameworks (e.g., modality theory [3, 4], TYCOON [5], CARE [6]) and others, which constitute a solid basis for multimodal system design, focusing on task specification, rapid prototyping and Wizard-of-Oz-testing.

However, the next generation HCI for more advanced and personalized PC system such as wearable computer and PDA based on wireless network and wearable computing, may require and allow new interfaces and interaction techniques such as tactile interfaces with haptic feedback methods, and gesture interfaces based on hand gestures, or mouse gestures sketched with a computer mouse or a stylus, to serve different kinds of users. In other words, for perceptual experience and behavior to benefit from the simultaneous stimulation of multiple sensory modalities that are concerned with human's (five) senses, fusion and fission technologies of the information from these modalities are very important and positively necessary.

T. Enokido, L. Barolli, and M. Takizawa (Eds.): NBiS 2007, LNCS 4658, pp. 523–532, 2007.

Consequently, we implement MMFA including synchronization between audio-gesture modalities by coupling the WPS-based embedded KSSL recognizer with a remote Voice-XML user, for improved multi-modal HCI in noisy environments, and suggest improved fusion and fission rules depending on SNNR (Signal Plus Noise to Noise Ratio) and fuzzy value, for simultaneous multi-modality.

This paper is organized as follows. In section 2, we describe an implementation of the WPS-based embedded KSSL recognizer for ubiquitous computing. Web-based speech recognition and synthesis system using Voice-XML is described briefly, in section 3. In section 4, we introduce SNNR and fuzzy value-based the MMFA architecture depending on improved fusion and fission rules, including a synchronization between audio-gesture modalities. In section 5, we evaluate and verify suggested the MMFA with experimental results, and finally, this study is summarized in section 6, together with an outline of challenges and future directions.

2 Embedded KSSL Recognizer

2.1 RDBMS-Based Feature Extraction and Instruction Recognition Models

We constructed 65 sentential and 165 word instruction models by coupling KSSL hand gestures with motion gestures that are referred to "Korean Standard Sign Language Tutor (KSSLT) [7]". In addition, for a clustering method to achieve efficient feature extraction and construction of recognition models based on distributed computing, we utilize and introduce an improved RDBMS(Relational DataBase Management System) clustering module [8], because statistical classification algorithms including K-means clustering, QT(Quality Threshold) clustering, and the Self-Organizing Map(SOM), have certain restrictions and problems, such as the necessity of complicated mathematical algorithm by multidimensional features, relativity of computation costs by pattern size, and minimization of memory swapping and assignment.

2.2 Pattern Recognition Using Fuzzy Max-Min Composition

As the fuzzy logic for KSSL recognition, we applied trapezoidal shaped membership functions for representation of fuzzy numbers-sets, and utilized the fuzzy max-min composition.

$$\begin{gathered} For\ (x, y) \in A \times B,\ (y, z) \in B \times C, \\ \mu_{S \bullet R}(x, z) = \underset{y}{Max}\ [Min\ (\mu_R(x, y), \mu_S(y, z))] \end{gathered} \tag{1}$$

Two fuzzy relations R and S are defined in sets A, B and C (we prescribed the accuracy of hand gestures and motion gestures, object KSSL recognition models as the sets of events that occur in KSSL recognition with the sets A, B and C). That is, $R \subseteq A \times B$, $S \subseteq B \times C$. The composition $S \cdot R = SR$ of two relations R and S is expressed by the relation from A to C, and this composition is defined in Eq. (1) [9], [10]. $S \cdot R$ from this elaboration is a subset of $A \times C$. That is, $S \cdot R \subseteq A \times C$. If the relations R and S are represented by matrices M_R and M_S, the matrix $M_{S \bullet R}$ corresponding to $S \cdot R$ is obtained from the product of M_R and M_S; $M_{S \bullet R} = M_R \cdot M_S$. The matrix $M_{S \bullet R}$ represents