Integrating Sketch and Speech Inputs using Spatial Information

Bee-Wah Lee
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Sarawak, Malaysia
bwlee@fit.unimas.my

Alvin W. Yeo
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
Sarawak, Malaysia
alvin@fit.unimas.my

ABSTRACT

Since the development of multimodal spatial query, the integration technique in determining the correct pair of multimodal inputs remains a problem in multimodal fusion. Although there exist integration techniques that have been proposed to resolve this problem, they are limited to the interaction with predefined speech and sketch commands. Furthermore, they are only designed to resolve the spatial query with single speech input and single sketch input. Therefore, when it comes to the introduction of multiple speech and sketch inputs in a single query, all the existing integration techniques are unable to resolve it. To date, no integration technique has been found that can resolve the Multiple Sentences and Sketch Objects Spatial Query. In this paper, the limitations of the existing integration techniques are discussed. A new integration technique in resolving this problem is described and compared with the widely used integration technique, Unification-based Integration Technique.

Categories and Subject Descriptors

H. [**Information Systems**]: H.1: Models and Principles: H.5 Information Interface and Presentation (I.7)

General Terms

Design, Experimentation, Human Factors, Measurement, Performance

Keywords

Multimodal interaction, multimodal spatial query, spatial query, multimodal spatial scene description, and Multiple Sentences and Sketch Objects Spatial Query

1. INTRODUCTION

Since the conventional Geographic Information System (GIS) applications are often difficult to use and take a long time to learn [2, 3], there is a need to make these applications easier to use especially by novice users. According to Schlaisich and Egenhofer [7], people often communicate about space by talking and simultaneously drawing freehand sketches. This natural human communication pattern can be adapted into the existing GIS applications to accommodate a wider range of users and for better ease-of-use. This natural human communication can be achieved by adding multimodal interactions such as the use of pen gestures and speech in the process of spatial query formulation. Currently, most of these multimodal spatial systems are only able to accept single speech input and single sketch input in the spatial query [6,7]. The most widely used multimodal integration technique in this context is Unification-based Multimodal Integration Techniques.

However, when responding to the pedestrian who asks for direction on the street, users' sketch and verbal descriptions tend to occur more frequently in continuous spoken stream and freehand sketch. In this situation, multiple speech sentences and multiple sketch objects are involved in the single query formulation. Unlike the single sentence and single sketch object spatial query, this multiple-input spatial query would result in more problems in identifying the correct pair of speech and sketch inputs, which refer to the same spatial object. Furthermore, to date no multimodal integration technique exists for this context. Thus, a user survey was conducted to determine the suitability of the existing integration techniques to resolve this type of Multiple Sentences and Multiple Sketch Objects Spatial Query. The existing techniques were found to be insufficient in resolving this type of spatial query. The limitations of the existing integration techniques are discussed in the next section.

2. LIMITATIONS OF THE UNIFICATION-BASED MULTIMODAL INTEGRATION TECHNIQUE

In Unification-based Integration Technique, a temporal constraint is used to obtain the correct pair of speech and sketch events for a spatial object. The temporal constraint states that the time of the speech input occurrence must either overlap with the time interval of sketch input or the onset of the speech input is within 4 seconds following the end of the sketch input [5,6]. Therefore, the time interval within the occurrences of the inputs is used as the integration parameter in this technique. The occurrences of specific keywords or command in speech are detected and time stamped accordingly based on the temporal constraint. The sketch inputs are also time stamped and matched with the predefined simple gesture and symbols in the database. Consequently, the users' inputs are restricted and users have to remember and use

ICMI'05, October 4–6, 2005, Trento, Italy.