Extraction of Visual Material and Spatial Information from Text Description for Scene Visualization

Xin Zeng, MLing Tan, and WJing Chen

School of Architecture and Art Central South University, China
zengxin@mail.csu.edu.cn

Abstract. The generation of 3D virtual scene could be greatly simplified by integration of natural language interface. This involves complex task of extracting of visual object, material and their spatial information from language description. The knowledge of visual perception, spatial language, natural language processing and graphic representation can be combined with one another to accomplish the task of generating a virtual scene. This paper first mentions a variety of research in developing language based scene application. This is followed by a brief introduction of natural language processing technology. Some related theories of human visual perception and spatial language are investigated in section 3. A representation formalism based on word-concept-visual information has been proposed by linking the levels of meaning and visualization in section 4. Finally, we draw conclusion and mentioned future work.

Keywords: Virtual Reality, CAD, Material and Spatial Representation, Text Visualization.

1 Introduction

The barriers to productive use of traditional menu-based virtual environment CAD tools have been recognized, and as interactive software tools of the future become more and more complex and knowledge-intensive, the possibilities for integrating a natural language interface and computer graphics by incorporating knowledge of human-to-human interaction has been considered in improved virtual environment user-interfaces. Different types of text to images conversion system have been attempted and considered by many researchers over the last few decades, such as text to static scene based applications [1,2,3,4,5], and text to animation based applications [6,7,8,9].

Fig. 1. Architecture of Language-based Graphic Application
All common approach of language-based graphic applications consists of three components as shown in Figure 1. The input sentence is analyzed by the language component, and then the semantic of sentence is mapped into low level parameterized data in the knowledge base component. The output is used by the graphic component and ultimately to construct a virtual scene. The work in this area focuses on how to extract the meaning of the description and represent in a computational virtual scene. This involves complex natural language processing and visualization, and the knowledge of visual perception, spatial language, natural language processing and graphic representation can be combined with one another to accomplish the task of generating a virtual environment.

2 Language Processing and Information Extraction

Natural language processing uses known data about words and predefined rules that show authorized word structures to give the meaning of descriptions to enable communication between people and computers [10]. The general process of natural language processing can be divided into following levels: morphological analysis, lexical tagging, syntactic parsing, and semantic analysis. During the process the knowledge of the real world information need be captured and encoded in databases for aiding language understanding. There are three types of knowledge which are used in natural language processing, namely syntactic knowledge, word-sense knowledge, and world knowledge [11]. When combined, the above three types of knowledge permit to construct a representation of the initial meaning of the sentence. After many years of development and research on natural language processing and understanding, the state-of-art natural language enabled technology is rapidly growing along with development of computer processing power. If we constrain our language interface, this may help us use exist natural language processing techniques to extract information from language description and produce reliable interpretations. In current, each language based graphic application has its own constrained input language to suit its particular objectives. The form of the language created restricts the concepts then can be expressed and, ultimately the scenes that can be generated.

3 Visual Information and Related Theories

When dealing with computational models for integrating natural language and graphic representation, it is useful to examine theories of human visual perception and spatial language to discover what kind of processes and representation lie in between. The integration of visual perception and language understanding is one of the research fields of artificial intelligence and cognitive science that concerned with machine-simulated intelligent [12, 13]. This involves object identification and placement, it also appears very important to areas of computer graphics such as generation of virtual environment and animation. In the research of visual perception, [14] investigate how to use visual information to construct an internal spatial representation of visible objects. An intermediate representation provides the link between language and Marr’s computational theory of vision. The object descriptions