The Implementation of Graphic Constraints for Automatic Text to Scene Conversion

Xin ZENG, Man-ling TAN and Shuai REN

School of Architecture and Art, Central South University, Changsha, Hunan, China
e-mail: zengxin@csu.edu.cn

Keywords: Graphic constrains, Visual semantic definition, Text to scene conversion

Abstract. The use of constraints in computer graphics has been widely used in computer graphic systems, the constraints can be specified that restrict the objects and determine the relations between them. We present an approach based on visual semantic definition by integration of related theory, advances in NLP and computer graphics technology to generate a 3D virtual environment. In the prototype system developed here, the graphic constraints refer to establishing the geometric relations among the 3D objects. The concept of rigidity of constraints (i.e. restrict the parameters of the position, orientation, etc.) was introduced to handle a set of the objects and their spatial representations.

Introduction

Interacting with the computer by using a natural language interface does not require substantial training or learning effort and it allows straightforward expression. This should make it relative easier and faster for non-professionals to create virtual environments compared to using traditional 3D graphic applications. Along with the increasing power of the computer, and the development of computer graphics and natural language processing technologies, there is a variety of applications adopting these methods to generate virtual scene by language description. Such as [1], [2], [3], [4]. These systems enable the users to generate still image or scenes from language descriptions with some character poses without real time interactivity. In this paper, based on our previous work [5], we present an approach based on visual semantic definition by integration of advances in computer graphics technology to generate a 3D virtual scene. The main challenges of this work are to encapsulate natural language understanding and computer graphic technology, incorporate theory of visual perception and spatial language, and finally generate appropriate virtual environment.

Semantic Analysis and Representation

The meaning of a sentence is most often represented as a formula in a logical language that represents the way we conceptually describe the world we perceive. A semantic representation frame depends upon the denotations of the phrase and is derived from the output of the parsed structure by semantic interpretation. Some basic knowledge of grammar and understanding of the structure of sentences can be added for semantic representation, to be translated into the following elements (e.g. location, object, attribute, action and position to correspond to nouns, adjectives, verbs and prepositions) to represent the various meanings and relations inherent from the input sentence. To illustrate the entire process of the language analysis, the data flow diagram of the language component is shown in Figure 1. The language module produces a dual representation during the language parsing process, with conceptual and visual levels. One is to convert the syntactic structure of sentence into a semantic representation to present the various meanings and relations inherent from the input sentences for further operation; the other is to allow the graphic component to begin to assemble the original virtual scene once the 3D objects (entities) associated noun phrases have been extracted.
Object and Spatial Semantic Definition

After the semantic representation, it is necessary to build an intermediate representation that bridges the gap between the language understanding and graphic visualization. The definition of visual semantic information of this prototype system has been most influenced by perception and spatial theory from [6], [7], [8]. It works as shown in Figure 2. Once the output of the XML semantic representation parse tree has been extracted through Knowledge base module, then the semantic representation is converted into a set of parameterized data by the Descriptionary. Consequently it interprets words that denote particular entities with the kind of the properties and spatial relations, etc. The spatial compound and inference module implements the implicit constraints and updates the parameterized data, and finally enables the graphic component to create the virtual environment to represent the various meanings and relations inherent from the input sentences.

In the current stage of development, the spatial related vocabulary has been semantically parameterized and classified into two XML files: Virtual Object Definition (VOD) describes nouns and Virtual Object Spatial-relation (VOS) associates prepositions. In general sense, the object type is defined explicitly to represent a physical object, each object in the environment corresponding to a related noun and associated with a geometry model in virtual object database. The VOD uses a computationally feasible representation to define the 3D objects along with a list of elements. The elements are defined with respect to implicit properties of the object, such as initial position, size, length, width and height, etc. that can be interpreted in graphic system. Inspired by previous practical work on graphical collision detection and the computer geometric theory of [7], [9], this semantic definition of 3D object associated noun uses a bounding box to simplify the complex object geometry.
The objects are idealized and simplified, for example by defining their default sizes and coordinates, the orientation constraints, and dimensions. The bounding-box as the boundary to all the objects, the definition of the spatial area of object is obtained by expending the idea of [10] from 2D to 3D.

The use of constraints in computer graphics has been widely used in computer graphic systems, the constraints can be specified that restrict the objects and determine the relations between them [11]. When we describe spatial relationships with natural language we often use spatial prepositions. The interpretation of spatial relations is often quite object-dependent [4, 8], objects that are not located in terms of absolute space, but always in terms of placed to a background. We distinguish between structural and spatial properties of objects. The structural properties stand to the spatial attributes within the object, meaning that each object has location, orientation, dimension, and shape as its internal and independent spatial attributes. Spatial properties of one object to another depend on geometric relations and it is the most common way to express object spatial relations. Herskovits’s geometrical idea indicates that spatial prepositions have ideal meanings (i.e. ideal geometric objects) together with a set of constraints associated with them in their lexical entries [7]. These geometric functions determine what the preposition contributes to the meaning of a particular situation. The information such as structural properties of the objects involved, initial position, orientation, axes of rotation etc. can be used to semantically parameterize the prepositions according the common situation.

The above XML frame shows that prepositions such as left of are handled by semantic functions that look at the left and right dependents of the preposition (i.e. subject and object) and construct a semantic spatial parameterization depending upon their properties. The element of the prepositions consists of three attributes, each attribute takes different values to indicate the position of the subject is determined by both subject and object associate objects’ structural properties; especially the position of object is key reference to the subject. The boundaries of objects are restricted as a bounding box with variables for their width, height and depth. The parameterization of a preposition in VOS is only defined to the most common situation. As the trick of visualizing spatial prepositions is that the visualization is not a one-to-one association, one preposition may stand for different geometric configuration in different objects. The implementation of these object-oriented geometric constraints and some extra spatial algorithms (i.e. object spatial compound, spatial reasoning) for spatial reconstruction are also programmed in both knowledge base and graphic module for handling the complexities of the spatial task of visualizing.

**Conclusion and Future work**

The originality of the approach described here lies in the generation of a 3D virtual environment based on visual semantic approach, which enables the user to manipulate the visual features of virtual scene in real-time. The use of word-concept-visual based graphic constraints assists in object definitions and provides visual information for spatial construction and spatial reasoning. In the prototype system developed here, the graphic constraints refer to establishing the geometric relations among the 3D objects. The concept of rigidity of constraints (i.e. restrict the parameters of the position, orientation, etc.) was introduced to handle a set of the objects and their spatial representations. Some screen shots of the system are presented in Figure 3.
Scene Generation: There is a swimming pool in front of a house. A tree is right of the house. A wooden bench is in front of the tree. A wall is left of the house. A lake picture is on the wall. A golden cow is in front of the wall.

Scene Manipulation: Change wall_1 to brick. Change bench_1 to red. Move swimming_pool_1 10 units left. Move tree_1 in front of house_1. Change cow_1 to blue

Future works will investigate new the placement algorithm according the specific shape in order to obtain more accurate and efficiency output.

References