IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Volume 14, PP 47-51

Automatic Visualization of Descriptive Text

Dharmit Viradia[1],Ajay Singh Thakurathi[2],Dhavalkumar Sonavaria[3], Apeksha Waghmare[4]

[123](Department of Computer Engineering, Atharva College of Engineering, India)
[4](Assistant Professor, Department of Computer Engineering, Atharva College of Engineering, India)

Abstract: *We analyze the case of putting forth a interface for generating 3D scenes from input texts. The component of natural language understanding is framed as a problem of semantic parsing, in which a structured representation of input texts is created and then linking it together for a meaningful representation of scenes in space. We incorporate a dataset of 3D objects that are recorded with natural language texts and learn how to generate 3d scenes from descriptive texts. This system renders 3D scenes for a correct user input by analyzing text parameters and also allows user to interact with the generated scene.*

I. Introduction

Attaining 3D scenes of our thoughts is very complex and difficult course. Common 3D scene generating softwares seem very complicated at first glance, also when user tries to create a scene, he has get familiar with the tools and interface, also creating every object from scratch will result into a long process[1]. Hence, a path based on natural language where in the virtual 3D environment are created directly through natural language not relying on special graphics tools which simplifies the process of 3D scene generation.

In this paper, we put forth a framework which accordingly generates 3D scenes from natural language. More accurately, just a simple descriptive input text serves a base for rendering a specific virtual 3D environment[1][3]. The description includes of the name of the objects appearing in the scene, as well as the spatial relationships between them. Our system concentrates on the key issues of Information Extraction (IE) as part of NLP(Natural language Processing), semantics, and graphical representation of a given text.[1]

II. Need

2.1 Motivation

There are many research going on Text to Scene conversion system. One of the examples of Text to Scene conversion system is PUT system . It identifies the spatial relationship from the text and arranges the objects as per this relationship. It depicts the natural language text in the form of PUT(X, Y, P), where X and Y are two objects and P is a spatial relationship between them. Carsim is a system which visualizes and animates car accidents from the description of accident. ScriptViz is a system which depicts the movie script into graphical form. And most recently WordsEye , which use large 3D model database and by combining this models it generates a 3D scene.We have surveyed many Text to Scene conversion systems to get the idea about how text to scene conversion takes place and visualizes a text in the form of readymade images which is not an appropriate method and depends on user selection of images, that should be avoided by using some ranking method. Other systems generate a 3D scene by combining different models, but they don't consider all aspects.They only focus on visualization of mechanics problem, It should consider other problems and it can be further extended for generating animation. WordsEye focuses on spatial relations of objects. It has a large database of 3D objects which also includes human and animal gestures. It is not able to depict actions of live objects properly; Whereas NLP Story Maker focuses on actions.

The aim of the Text to Scene conversion system is to intelligently convert the natural language text into its visual representation. At the front end, user enters the natural language text like story and the system will visualize it. This process of automatically generating a 3D scene from the text description involves Natural Language Processing (NLP) and 3D Computer Graphics. NLP is used to process the sentence in natural language and 3D Computer Graphics is used to visualize those sentences.

2.2 Basic Concept

The proposed system allows the arrangement of objects spatially in a given environment becomes tough due to changes in design this can be solved by developing a program which generates a scene of the given text. Our proposed system takes raw text as input identifies misspelled, incomplete and non visualizable words and suggests correction for such words to improve scene generation process. This system generates scene for the corrected input by identifying various scene parameters and also provides user interaction to render exact scene of user's choice.

III. Previous System

Already quite a few projects investigated the field of natural language input for creating virtual environments. In the following, we present some of the research that is most related to our system and motivated the development of our system.

The SHRDLU program as shown in Fig (1) was one of the earliest systems that was able to understand and evaluate natural language[4]. User interaction was allowed via simple English dialogs about a small blocks world shown on an early display screen. SHRDLU was primarily a language parser with the ability to use semantic information and context to interpret natural language input. It allows users to enter natural language commands, which are reissued to a virtual robot arm which moves blocks around in a small 'blocks world'[1]. block world means a virtual box filled with different blocks. However, the usable vocabulary for interaction was rather limited and the amount of referenced objects was restricted to a pre-existing environment.

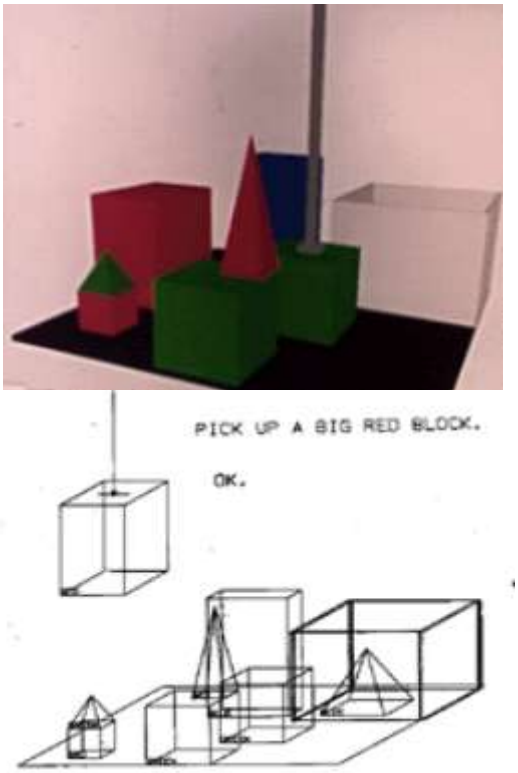


Fig 1 :SHRDLU Program

One of the most well-known projects in the field of language-based 3D scene generation is WordsEye as shown in Fig 2, created by B. Coyne and R. Sproat. It generates static scenes out of a user-given text. An entered text consists of simple sentences that describe positions of objects and their orientations, colors, textures, and sizes. Although it rationally anticipates natural language input, the translation of dimensional bonding does not succeed and the architecture of the input is rather limited. In order to obtain "natural" looking depictions, the user is also required to use constraints for arranging objects, which takes time and effort. Contrary, in our approach we disregard colors and textures because we want to focus on correctly interpreting spatial relations without the need of user interaction, and keep the input more flexible.

Fig 2 : Wordseye

The Text to Scene Generation as shown in Fig 3, project aims to explore how to automatically generate 3D scenes from a natural text description. When portraying a scene, individuals often skip relevant common sense knowledge about the placing of objects. This project pursues to determine such insight from a database of scenes and use the learned priors to conclude missing parameters when generating a scene[3].

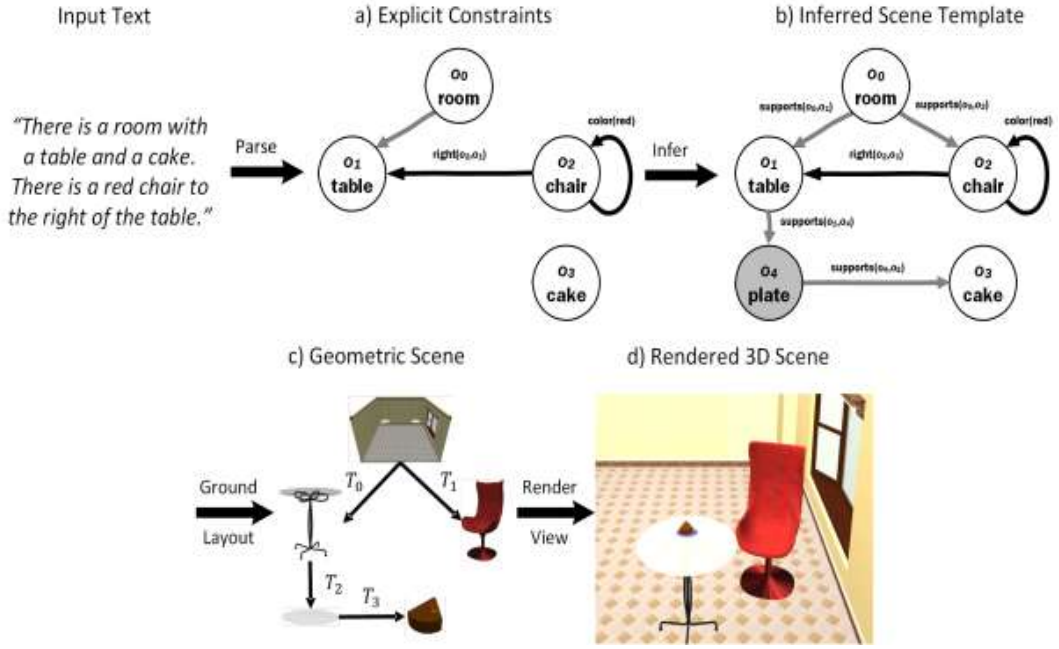


Fig 3: Text to Scene Generation

From all these previous work, we can conclude that automatic scene generation generally involves two parts, i.e. language engine and graphics engine. The language engine analyzes text descriptions about scene, and extracts the useful information about objects and their spatial relationships. The graphics engine creates the scenes based on the object layout determination according to spatial relationships and presents the visual features of the scene by text input. Most previous work paid more attention to the language engine, but less attention to the graphics engine which was just designed to present scene with poor visual effects. In the previous work, the existing excellent graphics engines of design applications were almost completely neglected to take advantage of to improve visual effects. A more important point is that those excellent graphics engines provide a series of tools to facilitate the subsequent operations such as rendering scene even creating animation in scene[2].

IV. Proposed System

In this system we are going to use following steps to convert text into 3D Scene

Input processing: Objects and its attributes must be identified from input query but some of the words may be misspelled and incomplete. Thus this input text must be processed to fine tune query.

Identify the words: Using WordNet, nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept.

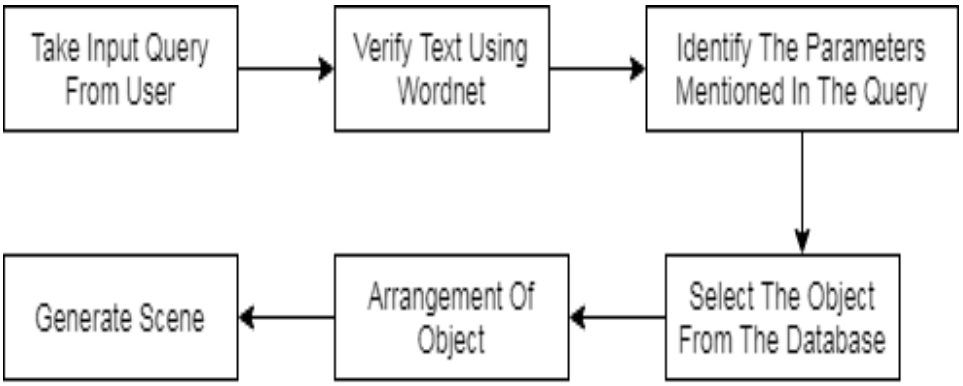
Identify explicit parameters: Once the query is processed, objects and attributes mentioned in query must be determined. In order to identify such explicit parameters system may use Part-Of-Speech(POS) tagger[4].

Identify implicit parameters: Some known facts are not mentioned in query and has to be inferred from explicit parameters. Scene type must also be inferred from objects mentioned in input text. Prior knowledge can be used to identify implicit parameters.

Identify Hierarchy: After identifying all the objects and attributes that must be present in scene we need to identify hierarchical arrangement of those objects. Spatial database can be used to identify such hierarchy.

Generate candidate scene: Once all the data has been collected about the scene it must be rendered by fetching appropriate model from 3D model database.

User interaction: User can give commands to change the view of scene as per his requirement. These choices can be saved for future use.



V. Result

Our system presents a unique approach to creating scenes and image. Our system allows a user to quickly generate virtual 3D environments by using natural language as input. Starting from a descriptive text, information about objects and spatial relationships are gathered and refined. The findings are used to link retrieved entities to appropriate models as well as deriving a directed graph representation of the text. With the aid of that digraph, spatial relations between objects are calculated. The resulting locations and models are finally assembled in an interactive virtual environment.

Our system will be an aid in many cases, providing interesting and surprising analysis. When users want to control a illustration more precisely, they can adjust their language to better specify the exact meaning and graphical constraints they conceive. We believe that the fast processing natural linguistic scene development systems will bring a natural and attracting way for day to day users to generate vivid imaginations and express themselves. In its current state, our system is only a first step toward our goal. There are several domains where the efficiency of the system needs to be improved, such as: Improvements in the coverage and robustness of the natural language processing, language input via automatic speech recognition rather than text; a larger inventory of objects, poses, atomic rules, and states of objects; mechanisms for depicting materials and textures; mechanisms for modifying geometric and surface properties of object parts; environments, activities, and common knowledge about them; shape deformation and natural phenomena.

VI. Conclusion

We studied many research papers and existing systems to determine the way how text can be converted into a scene. From this research papers, we have concluded that the system can be divided into two modules: NLP module and Graphics Module. From different tools available, we decide to use Stanford NLP as it contains all the text processing tools and For Graphics Generation we have used Java Monkey Engine and Blender models.

We have considered some basic spatial relationships like on, under, above, below etc. and created basic algorithm to convert the natural language text into scene. We will then use this algorithm to generate accurate spatially correct 3D images dynamically.

References

[1] Angel Chang , Will Monroe, Manolis Savva, Christopher Potts and Christopher D. Manning, Stanford University, Stanford, CA 94305." Text to 3D Scene Generation with Rich Lexical Grounding" at Stanford University.

[2] Sneha N. Dessai, Dhanaraj . "Creation of 3D Scene from Raw Text" IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India.

[3] Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright c 2016. All rights reserved. Draft of August 7, 2017.

[4] Will Monroe,"3D Scene Retrieval from Text with Semantic Parsing",Stanford University,Computer Science Department ,Stanford, CA 94305, USA.

.