ISSN 2093-0542

KOREAN SOCIETY FOR INTERNET INFORMATION

# The 16th International Conference on Internet

**(ICONI 2024)**

**Dec. 16-19, 2024 Taipei International Convention Center (TICC)**
**Taipei, Taiwan**
http://www.iconi.org

# *Proceedings of ICONI 2024*

**| Organized by |**
**Korean Society for Internet Information (KSII)**

**| Sponsored by |**
**INU Computational Research Center**

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

# Contents

KSII The 16th International Conference on Internet (ICONI) 2024.



# Automatic 3D Image Generation via UML Diagram based on Semantic Roles extracted with ChatGPT

**Hyuntae Kim[1], Kidu Kim[2], Jihoon Kong[1] and R. Young Chul Kim[1,*]**
[1] Software Engineering Laboratory, Graduate School, Hongik University
Sejong, South Korea
[2]AI Infrastructure Team, TTA
Seongnam, South Korea
[e-mail: hyuntaekim@g.hongik.ac.kr, kdkim@tta.or.kr, go400s@naver.com, bob@hongik.ac.kr]
*Corresponding author: R. Young Chul Kim

***Abstract***

With the development of Generative AI, interest in using it has increased. However, currently, Generative AI sometimes incorrectly interprets prompts entered by users when creating images and generates results different from their intentions. There are cases where images created according to the intention are also awkward. In addition, there is a problem that whenever an image is generated, objects or objects in the image are not reused and are continuously generated differently. To solve this problem, we propose automatic 3D image generation from natural language by combining linguistic techniques and software engineering techniques. Using linguistic theory, natural language sentences are accurately analyzed to generate sequence diagrams and images generated by diagrams. In addition, it is expected that accurate 3D images can be generated through natural language input by developing automation tools applying these mechanisms.

***Keywords***: Image Generation, Linguistic Theory, Sequence Diagram, LLM

## 1. Introduction

Recently, technology for creating 2D and 3D images or generating videos using Generative AI has been developed, attracting a lot of attention. However, there is a problem that objects and objects appearing in AI-generated images cannot be reused when the image is reproduced. In addition, there is a problem that AI misidentifies the prompts written by the user and generates images different from the intention [1].

To solve these problems, we propose a 3D image generation mechanism and tool through natural language analysis. We generate diagrams by analyzing natural language sentences. In addition, 3D images are generated using the generated diagrams.

Chapter 2 of this study describes the problems of Generative AI. Chapter 3 mentions our mechanisms, the tools to which they are applied, and application cases. The final chapter, 4, describes the conclusions.

## 2. The Problems of AI

Currently, generative AIs have advanced a lot so that they can generate not only general 2D images but also 3D images and videos. However, AI's natural language analysis capabilities are not perfect at present. The following **Fig. 1** is the

result of creating an image using ChatGPT, a representative Generative AI.

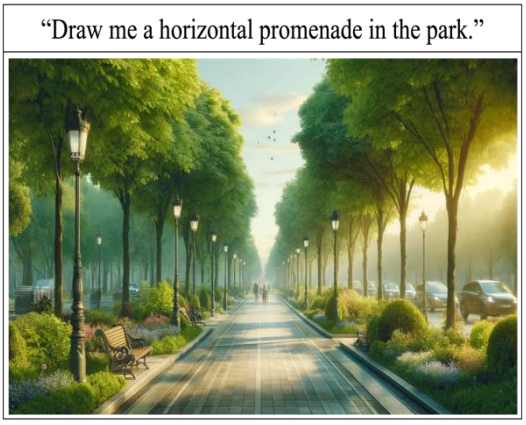


**Fig. 1.** Image Generated by GPT

**Fig. 1** did not analyze the requirements requested by the user well, resulting in a result different from the intention. Another problem with current AI is that it cannot create a perfect image. **Fig. 2** shows the result of creating images using several types of AI.

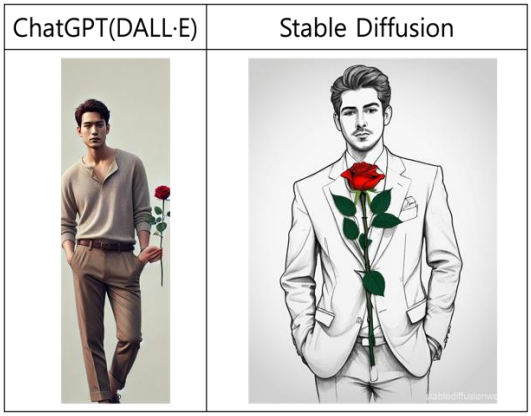


**Fig. 2.** Images Generated by Generative AIs

The prompt used to generate the images in the figure above is 'Draw a man standing with a red rose.' Both images have a standing man and a rose. However, in the image generated by ChatGPT, the man's hand is positioned abnormally. In the image generated by Stable Diffusion, the rose is floating in the air. Therefore, we try to solve these limitations using the mechanism we propose.

## 3. 3D Image Generation Approach

The approach of generating a 3D image from a natural language sentence proceeds in a total of three steps as shown in **Fig. 3** below.

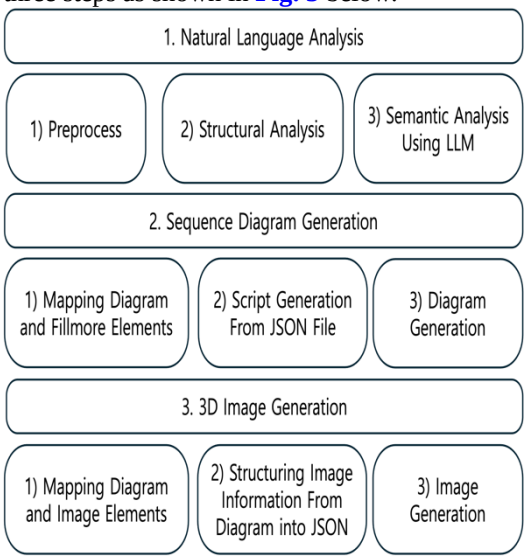


**Fig. 3.** Entire Process

### 3.1 Natural Language Preprocessing and Morpheme Analysis

The first step in the approach we propose is the natural language pre-processing step. In this step, the sentences are transformed so that a diagram can be generated from natural language sentences. The pre-processing process consists of 1) identification of clauses and phrases, 2) identification of subordinate clauses and phrases, 3) separation of clauses and phrases, and 4) adjustment of sentences. First, in step 1, sentences that are intact and can be alone are identified. In step 2, clauses and phrases with dependent information are identified and removed if unnecessary. In step 3, the clauses and phrases are separated into sentences. In step 4, the subject is made to be complete sentences, but some words are added, or words that are not necessary are deleted. The next step in pre-processing is the morpheme analysis step. For the pre-processed sentence, the morpheme of words in the sentence is analyzed using Stanford Parser[2]. Among the identified morpheme, tags representing pronouns are named PRP*, tags representing nouns are named NN*, and tags representing verbs are named VB*.

### 3.2 Semantic Analysis with LLM

Until previous studies, we redefined and manually identified Fillmore's case grammar [3]. The **Table 1** shows the redefined Case Grammar and explanation.

**Table 1.** Redefined Case Grammar

| Case | Description |
|---|---|
| Source | The entity that sent the message |
| Target | The entity that received the message |
| Instrument | The noun entity represents additional information of a verb, such as a tool or a place used for an event |
| Main Verb | The main verb of sentence |

To automate case identification, there was a limit to being rule-based, so this study automated the identification of redefined cases using ChatGPT's API in NodeJS. The following examples were included in the prompt to increase the accuracy of identification.

Part of Prompt

```
⋮
Q. Tom went to see Richard.
A.
{
source: Tom
target: Richard
instrument:
mainverb: went
}

Q. Tom told Richard to go to the canteen.
A.
{
source: Tom
target: Richard
instrument: canteen
mainverb: told
}
⋮
```

**Fig. 4.** Part of Prompt

After that, the case is identified using LLM, and the result is output in JSON file format, as shown in the above figure. The figure below is part of the result of case identification of the sentence 'Tom is going home. Tom looked sideways while walking. Tom saw Jason walking. Tom said hello to Jason.'

Part of JSON File

```
⋮
{
"sentence": "Tom looked sideways while walking.",
"source": "Tom",
"target": "Tom",
"instrument": "sideways",
"mainverb": "look"
},
{
"sentence": "Tom saw Jason walking.",
"source": "Tom",
"target": "Jason",
"instrument": "",
"mainverb": "saw"
},
⋮
```

**Fig. 5.** Part of JSON File

### 3.3 Sequence Diagram Generation

Our mechanism provides an intermediate output for the image generation process by representing the sentence analysis result as a sequence diagram before generating the image. Diagrams are generated in PlantUML by creating scripts based on the results of mapping the sequence diagram elements with Fillmore's case[4]. The script format used to generate the sequence diagram is as follows. 'Source -> Target: Verb(Instrument)' Finally, a diagram is created using JSON file information in the script. The figure below is the result of making a diagram with the stored JSON file using the tool.

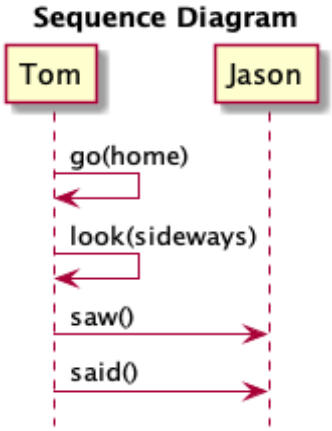


**Fig. 6.** Generated Sequence Diagram

### 3.4 3D Image Generation

To generate an image, information is extracted from a diagram to define the attributes of the

image. Based on the attributes of the image, a 3D image is generated using ThreeJS[5]. The figure below is the result of image generation for the sentence 'Tom saw Jason walking.'

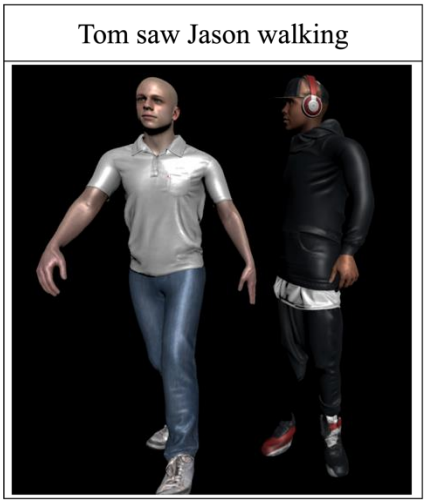


**Fig. 7.** Generated 3D Image

## 4. Conclusions

The mechanisms and tools we propose can accurately analyze the user's intention by using linguistic techniques. Among the linguistic techniques, LLM applied with Fillmore's theory was used to automatically identify the case of a noun without manually identifying it. Through this, we automated image generation from natural language. In addition, using the mechanism we propose, the same object can be reused, solving the problems of the Generative AI.

## ACKNOWLEDGMENT

This research was supported by Korea Creative Content Agency (KOCCA) grant funded by the Ministry of Culture, Sports and Tourism (MCST) in 2024 (Project Name: Artificial Intelligence-based User Interactive Storytelling 3D Scene Authoring Technology Development, Project Number: RS-2023-00227917307820870502O1) and National Research Foundation (NRF), Korea, under project BK21 Four.

## References

[1] H. Kim, J. Kim, J. Kong, K. Kim, R. Y. C. Kim, “3D Object Extraction Mechanism from Informal Natural Language Based Requirement Specifications,” *The Transactions of the Korea Information Processing Society*, vol.13, no.9, pp.453-459, 2024.

[2] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky, “The Stanford CoreNLP Natural Language Processing Toolkit,” in *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp.55-60, 2014.

[3] H. Kim, J. Kim, R. Y. C. Kim, “3D Object Extraction Mechanism via UML Sequence Models from Natural Language Requirements,” in *Proc. of the Annual Symposium of Korea Information Processing Society Conference (KIPS)*, vol.31, pp.490-493, 2024.

[4] PlantUML, https://plantuml.com/ko.

[5] Three.JS, https://threejs.org/.