The Transactions of the Korea Information Processing Society
TKIPS Vol.13, No.9, pp.453~459
eISSN: 3022-7011
https://doi.org/10.3745/TKIPS.2024.13.9.453

3D Object Extraction Mechanism from Informal Natural Language Based Requirement Specifications

Hyuntae Kim† · Janghwan Kim†† · Jihoon Kong††† · Kidu Kim†††† · R. Young Chul Kim†††††

ABSTRACT

Recent advances in generative AI technologies using natural language processing have critically impacted text, image, and video production. Despite these innovations, we still need to improve the consistency and reusability of AI-generated outputs. These issues are critical in cartoon creation, where the inability to consistently replicate characters and specific objects can degrade the work's quality. We propose an integrated adaption of language analysis-based requirement engineering and cartoon engineering to solve this. The proposed method applies the linguistic frameworks of Chomsky and Fillmore to analyze natural language and utilizes UML sequence models for generating consistent 3D representations of object interactions. It systematically interprets the creator's intentions from textual inputs, ensuring that each character or object, once conceptualized, is accurately replicated across various panels and episodes to preserve visual and contextual integrity. This technique enhances the accuracy and consistency of character portrayals in animated contexts, aligning closely with the initial specifications. Consequently, this method holds potential applicability in other domains requiring the translation of complex textual descriptions into visual representations.

Keywords : Natural Language, Sequence Diagram, 3D Image, Requirement Engineering

비정형 자연어 요구사항으로부터 3D 객체 추출 메커니즘

김 현 태† · 김 장 환†† · 공 지 훈††† · 김 기 두†††† · 김 영 철†††††

요 약

자연어 처리를 활용한 생성 AI 기술의 최근 발전은 텍스트, 이미지 및 비디오 제작에 큰 영향을 미쳤다. 이러한 발전에도 불구하고, AI가 생성한 출력의 일관성 및 재사용 가능성과 관련하여 상당한 문제가 있다. 이는 캐릭터와 특정 객체를 생성하는 것이 중요한 만화 제작 분야에서 문제가 될 수 있다. 이를 해결하기 위해 언어 분석 기반 요구사항 엔지니어링과 만화 엔지니어링의 접목을 제안한다. 제안된 방법은 자연어 분석을 위한 Chomsky와 Fillmore의 언어학을 적용하고 객체의 상호작용을 표현하기 위한 UML 시퀀스 모델 사용하여 일관적인 3D Objects를 생성하는 것이다. 또한 자연어 입력에서 창작자의 의도를 체계적 해석한다. 이를 통해 캐릭터 또는 객체가 정의되면 다양한 패널과 에피소드에서 정확하게 재사용해 시각적, 맥락적 무결성을 유지하게 한다. 이 접근 방식은 만화에서 캐릭터 묘사의 정확성과 일관성을 향상시켜 캐릭터와 장면이 원래 요구 사항과 밀접하게 일치시킨다. 따라서 본 연구에서 제안하는 방법은 자연어 텍스트에서 복잡한 시각적 콘텐츠의 재현이 필요한 다른 분야에서도 적용할 수 있을 것으로 기대된다.

키워드 : 자연어, 시퀀스 다이어그램, 3D 이미지, 요구공학

※ This research was supported by Korea Creative Content Agency (KOCCA) grant funded by the Ministry of Culture, Sports and Tourism(MCST) in 2023 (Project Name: Development of AI-based user interactive multimodal interactive storytelling 3D scene authoring technology, Project Number: RS-2023-00227917, Contribution Rate: 80%) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education in 2023(Project Name: The Faultless Codinglization through Automatic Refactoring based on NLP BERT Model, Project Number: 2021R1I1A3050407, Contribution Rate: 10%) and BK21 FOUR (Fostering Outstanding Universities for Research) funded by the Ministry of Education (MOE, Korea) in 2023(Project Number: 2021R1I1A305040711, Contribution Rate: 10%)
† 준 회 원 : 홍익대학교 SELab 석사과정
†† 준 회 원 : 홍익대학교 SELab 박사수료
††† 비 회 원 : 홍익대학교 SELab 박사수료
†††† 정 회 원 : 홍익대학교 SELab 박사, TTA 팀장
††††† 정 회 원 : 홍익대학교 소프트웨어융합학과 정교수
Manuscript Received : July 15, 2024
Accepted : August 6, 2024
* Corresponding Author : R. Young Chul Kim(bob@hongik.ac.kr)

1. Introduction

In this paper, we have extended our previous research[1]. Recent advancements in generative artificial intelligence (AI) technologies have significantly advanced the field of image creation, with tools like DreamFusion and DALL-E for efficiently generating detailed and complex images[2]. These tools rely on language models such as ChatGPT, which are crucial for analyzing and interpreting user-inputted natural language prompts to create images that align with user specifications. However, despite these capabilities, persistent issues bother these technologies' effectiveness. The generated images often fail to accurately reflect the user's intent, revealing a critical shortage in current AI capabilities. A notable challenge exists in the precise and consistent replication of specific objects or characters, further complicating the use of generative AI in fields requiring stringent accuracy, such as digital art and design.

To solve this problem, we propose an improved mechanism and tools designed specifically for 3D object identification. In previous work, we have developed a mechanism for 2D image generation. This approach aims to enhance the precision and reliability of image generation processes significantly. Furthermore, we seek to improve existing mechanisms as mechanisms for 3D image generation. Chapter 2 examines the problems of existing language models such as ChatGPT. It details how discrepancies in image generation arise, mainly focusing on the frequent divergence from user expectations and the ongoing struggles with maintaining consistency across generated images. Chapter 3 outlines the proposed mechanisms in greater detail, describing how they leverage advanced 3D object identification techniques to address these fidelity and consistency issues directly. Finally, Chapter 4 discusses the potential broader impacts of this research.

2. Related works

2.1 Inconsistency Problem of Image Generation AI

Several issues arise when ChatGPT is used to generate images. First, the images often do not align with the user's intended outcomes. Secondly, the model consistently struggles to create the same object when reusing specific objects, such as characters or items. This inconsistency is due to the probabilistic nature of ChatGPT, which means that even identical inputs can result in different images. While using seed values can mitigate this issue, it does not fully solve the problem of variability. Figure 1 illustrates an instance where ChatGPT, based on the text input by the user, failed to grasp the user's intent correctly and produced different images for a prompt requesting the same image. This demonstrates the challenges in ensuring that generative AI systems like ChatGPT can reliably interpret user prompts and produce consistent visual outputs.

The issue highlighted involves generating images from prompts that specifically request a horizontal road, but the results show vertical roads instead. This misalignment between the user's intent and the AI-generated images emphasizes a critical challenge in current generative AI systems.

For this, we propose integrating linguistic theories from Fillmore and Chomsky with requirements engineering to understand the intent behind natural language sentences more accurately. By using these linguistic theories, the aim is to develop a system capable of accurately and con-

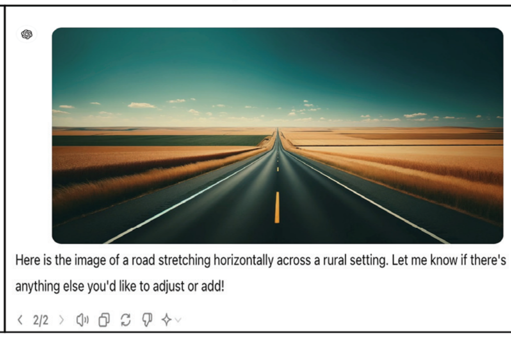
	First attempt to create an image	Second attempt to create an image
Prompt	Draw me a road that stretches horizontally.	
Answer		

Fig. 1. Inconsistent Images Generated by ChatGPT

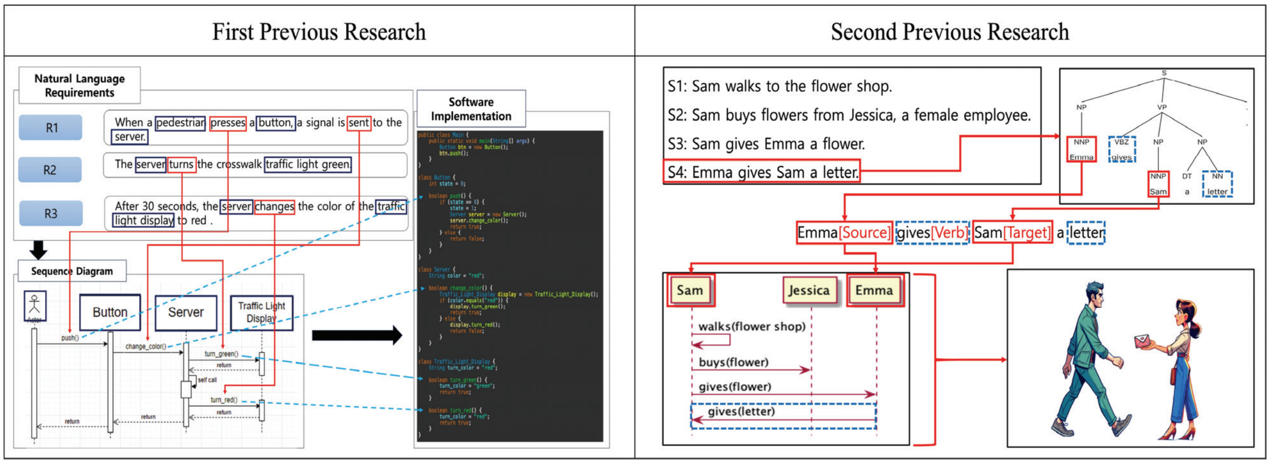


Fig. 2. Our Previous Skeleton Code and 2D Generation Approach

sistently generating the same object as specified by the user, thus improving the reliability and applicability of the system in tasks requiring precise visual representations. Fusing these approaches with generative AI is expected to significantly improve the ability to interpret and execute users' intentions, making them useful tools in a wide range of applications, including artistic and technical fields.

2.2 Our Previous Reseaches

In first approach, we adapt to analyze natural language requirements using the linguistics of Chomsky and Fillmore. Among them, we generate a UML sequence diagram using the result of analysis of natural language[3]. Generating a skeleton code using the generated diagram reduces the time required for implementation. Figure 2 shows the overall process and results for the first existing study.

In second approach, we generate diagrams using the results of analyzing natural language and generate 2D images with diagrams[4]. We express object communication with sequence diagrams about the interactions between objects included in the natural language requirements. After that, we use FabricJs to generate 2D images representing the interactions of objects. Figure 2 shows an example of the our second approach.

3. 3D Object Extraction Mechanism

Our mechanism consists of four steps, as shown in the figure 3. The first step is the structural analysis using the Stanford parser, and the second step is the semantic analysis to use Fillmore's case grammar theory. The third step is to generate a sequence diagram using natural language sentences analyzed through the first two steps. The last step is image generation.

In the initial phase, the Stanford Parser is employed to deconstruct sentences into their fundamental components, identifying morphemes that form the building blocks of the text[5,6]. This phase focuses on isolating the main verb and nouns, which are crucial for understanding the basic structure and intent of the input text sentence. It is critical to determine the accuracy of morpheme identification because it directly impacts the subsequent steps of the process.

Following structural analysis, the semantic analysis phase involves defining the roles of the identified nouns and verbs using Fillmore's case grammar[7]. This step is essential for interpreting the text's more profound meaning and context beyond only structural identification. The roles are then well organized into a JSON file, providing a structured format that can be readily utilized for further processing.

The generation of sequence diagrams involves mapping the interactions between identified entities based on their defined roles, effectively visualizing the dynamic flow of actions. Simultaneously, the mechanism leverages this structured data to facilitate the creation of 3D images, translating textual descriptions into visual representations that are accurate and consistent with the user's original intent.

We aim to overcome the common issues faced in generative AI, such as misinterpretation of user intent and inconsistent visual output. We also aim to provide one solution that enhances the reliability and precision of AI-gen-

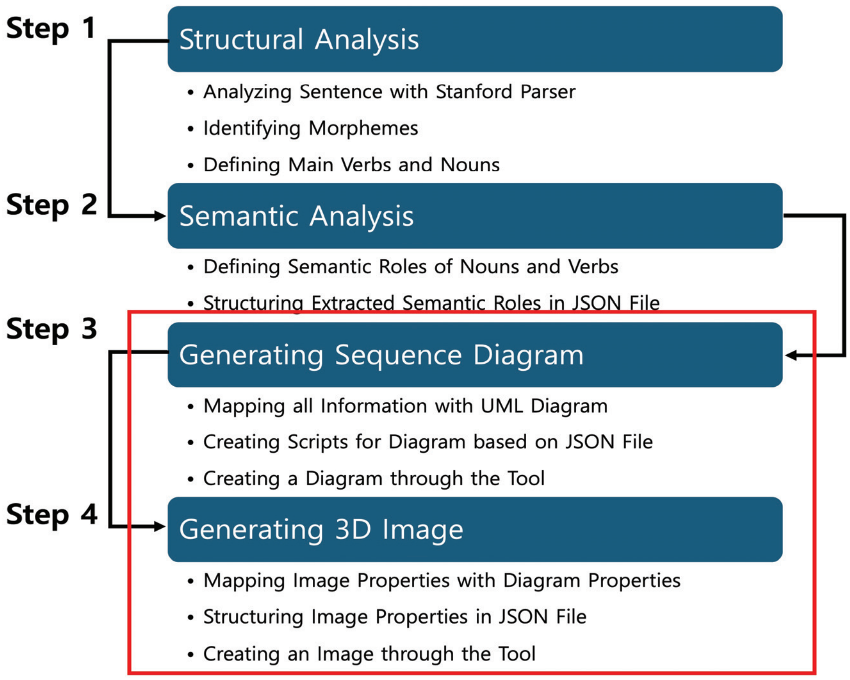


Fig. 3. Our 3D Object Extraction Process

erated images by adapting a rule-based approach with original natural language processing techniques.

3.1 Sequence Diagram Generation

We generate the UML sequence diagram by analyzing the nouns and verbs from the user's natural language input using case grammar rules. This analysis facilitates mapping textual elements to visual diagram components such as actors, objects, messages, and parameters. Figure 4 illustrates the mapping between case grammar components and the diagram elements, highlighting how textual interactions are transformed into a visual sequence diagram.

For the actual diagram creation, the tool utilizes PlantUML, a script-based tool that allows for the quick and efficient generation of diagrams[8]. The script for generating a sequence diagram adheres to the format: 'Sender object -> Receiver object: Message(Parameter).' The script becomes 'Source -> Target: Verb(Instrument)' by applying the case grammar mappings.

A JSON file is critical in storing the necessary information to generate the sequence diagram. This file contains structured data that reflects the relationships and actions identified in the natural language input. Figure 5 is an example of how the JSON file is used with PlantUML to generate a sequence diagram that visually represents the user-defined interactions depicted in the JSON structure.

Case Grammar		Sequence Diagram	Notation
Actor		Actor	Actor
Source		Object	Object
Target			
Dynamic Verb	Transitive Verb	Message	Message(Para)
	Active Verb		
Instrument		Parameter	

Fig. 4. Mapping Diagram with Fillmore's Case Grammar

3.2 3D Image Generation

Generating 3D images from natural language sentences is a complex process that combines linguistics, software engineering, and graphics technology to transform abstract linguistic data into precise visual models. This process begins with extracting critical attributes from sequence diagrams derived from structured analyses of natural language sentences. Attributes such as positional data, orientations, and specified actions are carefully extracted using advanced parsing techniques to ensure that the extracted data aligns directly with the semantic intentions of the input language. Once identified, these attributes are organized into JSON to ensure data integrity and facilitate integration with rendering tools. This JSON structuring is critical as it

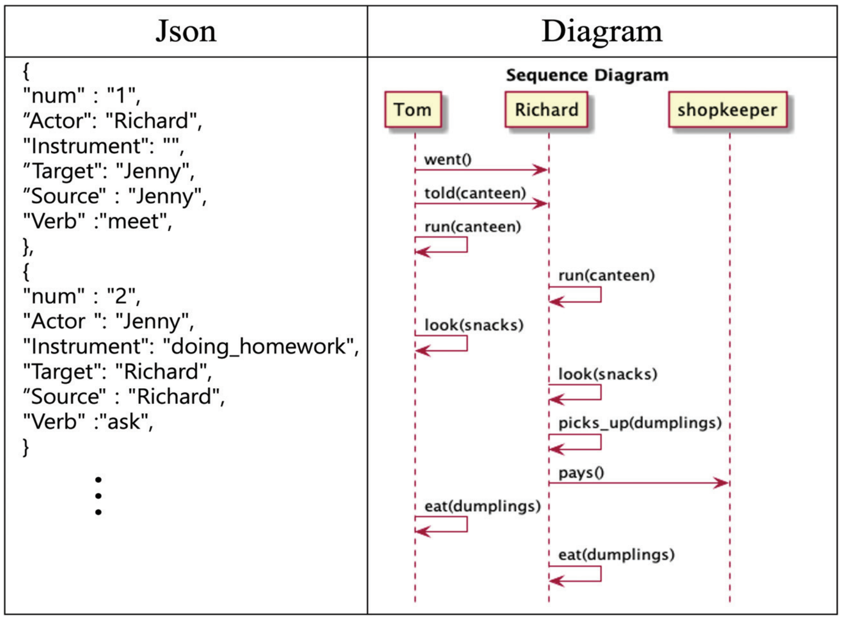


Fig. 5. Generated Sequence Diagram from JSON File

supports objects' complex relationships and behaviors, accommodating the detailed specifications required for generating accurate visual outputs. These images are rendered using ThreeJS, a powerful WebGL-based 3D graphics library chosen for its dynamic handling of detailed graphic content[9]. The rendering process involves setting up the 3D scene, importing object models, applying textures, and animating the objects according to the instructions specified in the JSON files. Following the scene setup, each object is animated to perform actions as detailed in the JSON, incorporating basic movements and complex interactions among objects. These animations reflect the functional dynamics specified in the original natural language sentences, ensuring the visual outputs are technically accurate and true to the selected sequences and timings.

An example of this methodology in action can be observed in a generated image that displays two characters in a running sequence. This image shows the interaction of objects that meet the requirements. This intricate process highlights the integration of advanced computational techniques and software engineering to enhance the fidelity and accuracy of 3D image generation, thereby contributing significant advancements to cartoon engineering and software engineering.

In addition, figure 7 shows the results of creating images for the same sentence. The figure shows that the same object is reused to provide the same output. Therefore we can reuse this image with our mechanism.

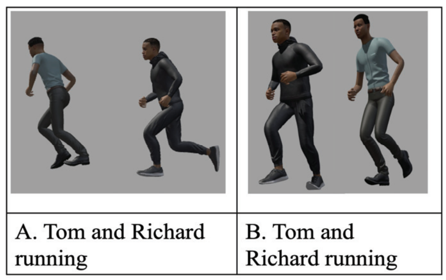


Fig. 6. Images Generated with the same Objects(Tom,Richard)

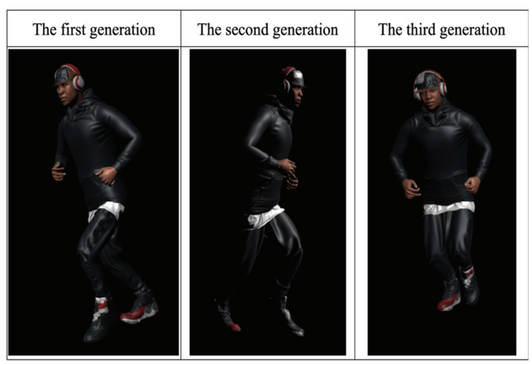


Fig. 7. The same 3D Image Objects generated with the same Sentence

4. Coclusion

We provide a sophisticated solution to prevalent issues in generative AI, explicitly addressing the misinterpretation of user intentions and the challenges of consistency and reusability in image generation. By using the structural and semantic analyses based on the linguistic theories of Noam Chomsky and Charles Fillmore, this mechanism significantly improves the alignment of generated images with user specifications. A central feature of this approach is its emphasis on the reusability of image assets, which is particularly highlighted throughout the research.

Reusing identical objects enhances the generation process's efficiency and ensures that once an object is created, it can be consistently replicated across different scenarios without redefinition. This capability is crucial for applications requiring high accuracy and consistency, such as animated storytelling and virtual reality simulations.

Overall, this mechanism advances the field of image generation by offering a method that improves the precision and consistency of the results and significantly boosts efficiency through the strategic reuse of assets.

References

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

[2] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "Dreamfusion: Text-to-3d using 2d diffusion," *arXiv preprint arXiv:2209.14988*, 2022.

[3] H. Kim, J. Kim, J. Kong, and R. Y. C. Kim, "Extraction Practices on UML Sequence Diagram through Natural Language based Requirement Specifications," Advanced and Applied Convergence Letters, 2023.

[4] H. Kim, J. H. Kong, H. S. Son, and R. Y. C. Kim, "Best Practice on Automatic Toon Image Creation from JSON File of Message Sequence Diagram via Natural Language based Requirement Specifications," *International Journal of Advanced Smart Convergence*, Vol.13, No.1, pp.99-107, 2024.

[5] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit," In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp.55-60, 2014.

[6] N. Chomsky, "Backmatter," Syntactic Structures, Berlin, New York: De Gruyter Mouton, pp.115-118, 2002.

[7] Fillmore, Charles J, "The case for case," 1967.

[8] PlantUML [Internet], https://plantuml.com/ko.

[9] Three.JS [Internet], https://threejs.org/.

Hyuntae Kim

https://orcid.org/0009-0003-1651-2123

e-mail : hyuntaekim@g.hongik.ac.kr

He received a B.S. degree in Software and Communications Engineering from Hongik University in 2023. He has been a master's student in Software Convergence at Hongik University since 2023. His research interests are in the areas of Software Visualization, Requirement Engineering, Software Engineering, Natural Language Processing, and Artificial Intelligence.

Janghwan Kim

https://orcid.org/0000-0003-0185-4406

e-mail : lentoconstante@hongik.ac.kr

He received a B.S. degree in Computer Science from Idaho State University (ISU) in 2019. He earned an M.S. in the Software Engineering Lab at Hongik University from 2020 to 2021. He has been a Ph.D. candidate in the Software Engineering Lab at Hongik University from 2022 to 2023. Since 2024, he has been an Adjunct Professor in the Department of Software and Communications Engineering at Hongik University. His research interests are in the areas of Software Visualization, Requirement Engineering, and AI- based Testing.

Jihoon Kong

https://orcid.org/0009-0004-1189-7674

e-mail : go400s@naver.com

He worked at NCSoft Corporation from 2008 to 2014, followed by Samsung Electronics from 2014 to 2016, and Nexon Games from 2016 to 2019. He served as an Adjunct Professor in the Department of Game Media at Kaywon University of Art and Design from 2020 to 2021. He has been a Ph.D. candidate in the Software Engineering Lab at Hongik University since 2022. Additionally, since 2022, he has held the position of Director of Research at ToonSquare. His research interests include Game Software Engineering and 3D Interface Interoperability.

Kidu Kim

https://orcid.org/0009-0003-7726-1970

e-mail : kdkim@tta.or.kr

He received the Ph.D. degree in Electronic and Computer Engineering from Hongik University. He is currently working at Telecommunications Technology Association. His research interests are in the areas of Test Maturity Model, Software Testing, AI Software Testing.

R. Young Chul Kim

https://orcid.org/0000-0002-2147-5713

e-mail : bob@hongik.ac.kr

He has been a Full Professor in the Department of Software and Communications Engineering at Hongik University since 2001. Since 2023, he has also served as the Director of the Metaverse SW Academy Institute. His research interests include Software Visualization, TMM (Test Maturity Model), Requirement Engineering, Metaverse, and AI-based Testing.