Evaluating the usage of Text to 3D scene generation methods in Game-Based Learning

Alexandru Gradinaru
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
alexandru.gradinaru@upb.ro

Mihnea Vrejoiu
National Institute for Research and Development in Informatics
Bucharest, Romania
mihnea.vrejoiu@ici.ro

Florica Moldoveanu
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
florica.moldoveanu@upb.ro

Maria-Anca Balutoiu
Faculty of Automatic Control and Computer Science
University Politehnica of Bucharest
Bucharest, Romania
maria_anca.balutoiu@upb.ro

***Abstract*—Computer games as an educational tool significantly improved information retention, student involvement, and motivation. Despite the growing popularity and the demonstrated positive impact, little progress is made to let educators create games without learning game programming or investing massive amounts of time, money, and human resources. This paper considers the usage of procedural content generation in simplifying the generation of game environments. Given that most educational material is in text format, we evaluate the possibility of generating 3D scenes starting from natural language text. A comprehensive survey of the current methods is presented, analyzing whether they can be successfully used in game-based learning. The study discusses several current limitations and possible improvements, highlighting a need for better progress in the domain.**

***Keywords*—text to 3d scene, natural language processing, text to scene conversion, text to graphics, visual representation, game-based learning**

I. Introduction

The use of augmented reality (AR), game-based learning (GBL), and gamification are among the most notable technological trends in education [1]–[4], especially in the context of an emerging metaverse concept. Gamification is a different concept than GBL: it represents the process by which game elements are used in different contexts, so as to generate fun experiences with a specific, usually hidden, purpose [5], [6]. An example of gamification can be adding a reward system based on the results obtained on a quiz. It provides a purpose, clear rules, and ways to receive the reward, recurring elements in game design. Previous studies have shown that both game-based education and gamification in education significantly improve the educational process by increasing motivation and student involvement but also the degree of information retention and learning, generally related to immersion and fun aspect [3], [5]–[7].

AR, gamification, and GBL have been long studied and applied in computer science and engineering education, with at least 156 studies published between 1976 and 2016 [8], and much more after 2016. Research cover topics like introductory programming [9]–[11], math [12], [13], physics [14], science [15], [16], software engineering [8], [17], mechanical engineering and many other, targeting both higher education and primary education.

However, despite the growing interest and usage of these technologies, there are considerable limitations in their development and use, the most notable being the high production costs, in the case of GBL and AR, alongside the difficulty of designing interactive educational content in such a way that both correct instruction and student evaluation can be achieved [18], [19].

Procedural content generation (PCG) can be one solution to simplify the creation of virtual scenes, reduce development costs, and enable users without extensive game development knowledge (i.e., primary school teachers, and engineering teachers without programming experience) to design educational or AR games. Although remarkable progress has been made in generative fields (e.g., text to 3D object generation [20], [21], image to 3D object generation [22], [23], text to image [24]–[27], sketch to image [26], [28], 3D indoor room layout synthesis [29], 3D reconstruction [30], [31], and others), the generation of 3D scenes starting from a text in natural language does not seem to have shown much improvement lately.

Considering most of the current educational material is in text format, text to 3D scene generation methods could be the most promising way to generate 3D virtual scenes, that can be used in AR or GBL.

II. Methodology

A systematic literature review was made to identify text to 3D scene generation methods. The research methodology consisted of searching articles after keywords in multiple databases (e.g. Google Scholar, Clarivate, ArXiv, Elsevier, Semantic Scholar), then looking for references or citations of all related articles. The research was initially focused on 3D scene generation methods that use large natural language corpus bodies (i.e., book chapters, novels), but there were few relevant results targeting the identification of scenes, characters, or actions, with a few studies using annotated text (i.e., film scenario, structured text) to generate 3D scenes and other large corpus or literary research divided generally into several categories: sentiment analysis, book classification (according to the genre, emotion, action type and others), text summarization, vocabulary analysis, and other text mining methods.

We extended the research to cover all text to scene generation methods. This way, we surveyed more than 700 articles, identifying multiple approaches: text to images, text to scene graph, text to 3D scene, text to 3D model, text to texture, text to animation, text to video, text games, extracting spatial ontology from text and indoor scene synthesis.

Finally, 111 articles were selected from 1992 to 2023, after removing duplicates, including original work on new methods for text-to-3D scene conversion, thesis, surveys, evaluation, or case studies.

The results presented in this article have been funded by the Ministry of Investments and European Projects through the Human Capital Sectoral Operational Program 2014-2020, Contract no. 62461/03.06.2022, SMIS code 153735.

979-8-3503-1339-0/23/$31.00 ©2023 IEEE
DOI 10.1109/CSCS59211.2023.00105

Several qualitative criteria were defined to evaluate the possibility of using current text to 3D scene methods in generating content for game-based learning starting from existing text material. Criteria, objectives, and evaluation methods are presented in Table 1.

TABLE I. EVALUATION CRITERIA

Criteria	Objective	Evaluation method
Source code	Testing the method	Searched after mentions inside the article, using google search, github or paperswithcode
Result accuracy	Accuracy of the resulting text to scene conversion	Analyze paper results or direct testing if source code available
Scene quality	Graphics quality of the resulting scenes	Analyze paper results or direct testing if source code available
Scene complexity	Scene complexity level	Analyze paper results or direct testing if source code available
Text length	The maximum length of text	Analyze paper results or direct testing if source code available
Text peculiarity	How descriptive or specific should the text be? Works on real existing text or needs text written especially for the method	Analyze paper results or direct testing if source code available
Language variety	What kind of words can be used. Support for special terminology (e.g., medical, engineering)	Analyze paper results or direct testing if source code available
Multiple scene support	Text having multiple scenes (e.g., literary novels)	Analyze paper results or direct testing if source code available
Export	Capability to export results in a popular format (e.g., GLTF, USD)	Analyze paper results or direct testing if source code available

III. TEXT TO 3D SCENE GENERATION METHODS

Yamada et al. (1992) [32] is one of the earliest studies researching the reconstruction of the geometric model of a scene, starting from natural language descriptions. The authors implemented a prototype called SPINT (Spatial Representation INTerpreter) that can extract spatial constraints from text and represent them in a primitive, low-quality 3-dimensional world. The method can interpret multiple sentences in the natural language, only in the Japanese language.

WordsEye [33], although published in 2001, is among the most well-known systems for generating scenes based on text. The model uses linguistic analysis, predefined patterns, and actions to interpret short sentences. The work is referenced in almost all research papers on the topic of the text to 3D scene generation. Further research on improving the WordsEye system by the authors has been presented in [34]–[40].

WordsEye has a large database of 3D objects that can be rendered, more than 12000, but does not target real-world text, but rather imaginary scenes. If the identified objects are not in the dataset, the word is represented in 3D. Since there could be multiple ambiguities in the sentences, the resulting 3D scene could be something unexpected [41].

WordsEye is able to create only static scenes and does not treat the temporal factor, where information is presented over time [42].

Fig. 1. Generated scenes with WordsEye [33]

A more recent analysis of WordsEye's capabilities is carried out in 2018 by Ulinski et al. [43], done from the perspective of generating scenes starting from short sentences vs. Google search results. The results rate WordsEye as superior for imaginative sentences, and Google for realistic sentences. Some important limitations are highlighted, resulting from semantic interpretation, missing or unrepresentative 3D objects, syntax or punctuation problems, and others.

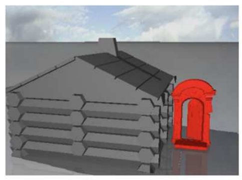

Fig. 2. A generated scene for "A house with a red door" in WordsEye [43]

Furthermore, Ulinski introduced in her doctoral thesis WELT (the WordsEye Linguistics Tools) and SpatialNet, which provides a representation for expressing spatial relations, and demonstrated its use for spatial language elicitation through crowdsourcing annotation, thus increasing the accuracy of the resulting scenes [44].

Wordseye can be tested online on their website, but registrations have been closed since 2022, indicating a newer improved version is soon to be launched.

CarSim [45], initially published in 2001, is another highly influential system, that describes a system developed to reconstruct, visualize and animate scenes from car accident reports, written in natural language. The prototype can identify objects, extract events, and model trajectories and collisions from paragraphs containing several sentences. The system is limited to text regarding car accidents or trajectory modeling. The first version was developed only for the French language, but there are several research papers that describe prototypes adapted for English [46] and Swedish [47] [48]. The system was tested on real-world text corpus, showing good accuracy.

Zeng et al. describe in 2003 [49] and 2005 [50] 3DSV (3D Story Visualizer), another approach that enables non-professionals to generate 3D virtual environments (3DVE) using Java, VRML, and XML technologies. The method is offering similar features and quality as WordsEye and CarSim. It can process more than 100 individual incremental sentences with simplified descriptive text but can also accept editing input that can manipulate the properties of existing 3D virtual objects. The vocabulary is in English, but it is limited.

Liu and Leung [51] published ScriptViz (2006), a system developed in Java and OpenGL that can create story visualization for screenplays. The story can be described using simple sentences in natural language, structured as screenplay scripts. Each sentence is processed one at a time. The user can control the scene background, environment settings, and camera position. Each scene is treated and represented separately. The scenes are limited to a few avatar models and objects, and the scripts model the position and animation generation in real time. The results are accurate, but the quality of the graphics is outdated.

There is also other research that uses structured natural language (e.g., movie scripts or annotated systems) to generate 3D scenes, images, sequences of images, or animations. Among them, we can mention CARDINAL - a system for the interactive creation of film scenarios with 2D and 3D visual effects [52], ShowRunner – a system for playing a well-defined and annotated scenario [53], TakeToons – a system for creating facial/character animations based on annotated text [54], generating animations specified in a text [55], SceneMaker – a system that aims to facilitate film production, with particular attention to detecting emotional aspects in text [56], [57], another study entitled SceneMaker - proposing an architecture for producing multimodal animated scene from film and drama scripts [58], and others [59], [60].

In 2006, Ma proposed CONFUCIUS [61], an automatic conversion system from natural language to 3D animation. CONFUCIUS is handling mostly scene animations but has features for scene composition and control. It can handle only simple single sentences at a time, which contain action verbs. It has around 50 3D models that can be used for visual representation and animation. The accuracy of the representation is good most of the time, but the graphics quality is outdated.

Severesky and Yin [62] present a system that targets automatic 3D scene generation from voice and text descriptions, studying approaches to object placement and spatial relationships, based on voxelated objects. Only a few relationships are supported, and the paper does not give many details about the number of objects that can be used. Graphics quality is low, and objects are not textured. Text is processed sentence by sentence and needs to be specific.

Glass and Bangay [63] proposed in 2017 a conversion system that targets fiction text. The system has a model database. It only represents 3D avatars in a predefined scene and models their movement in time. Glass, in his doctoral thesis [64], presented the first automatic system that is able to convert natural language sourced from fiction books to multi-modal 3D animated virtual environments. It uses procedural generation for background scenery, audio representations of text, and time-based animations. The fiction text is annotated using a pattern-based machine learning approach using manually and automatically generated annotations. Still, scenes need to be processed individually with objects from a 3D model database, and the processing is limited by the range of the annotation categories. Berkland and Bangay [65] created a fully-playable adventure game using XML configurations in the Glest game engine. Although limited, the study manages to convert a small custom-written story into a 3D interactive game environment with events and dialogue.

Simpler NLP-based 3D generation systems are presented as an aid for children with autism or mental retardation converting Turkish sentences into visual 3D representation [66], interactive spatial visual representations [67], automating 3d scene generation in Autodesk Maya [68], [69], generating scene graphs for Indonesian text [70], aiding children's spatial knowledge understanding [71], using multi-agents to visualize text descriptions by animating objects in the Alice editor [72], generating static scenes in the Alice editor from full natural language sentences [73], visualization of mechanics problems [74], 3D visualization and animation of verbal concepts [75], scene layout from Chinese text [76], automated simulation creation of military operations scenes [77], immersive visualizations of 3D disaster scenes [78], authoring augmented reality experiences through text [79], using virtual reality as an aid as cognitive learning environment [80], help dyslexic students to comprehend text [81], traditional Chinese art preservation [82], and other [83]–[94]. Most of them can handle only sentences parsed individually or in a short text, have a simple vocabulary, and present limited results.

AVDT (Automatic Visualization of Descriptive Text) [95] is a system with similar capabilities as WordsEye, allowing for a more flexible linguistic input but with a much smaller 3D model database and no textures or colors.

There is also related research in the problem of extracting [96] and representing visual or spatial information from natural text, with methods using XML [97], [98], graph grammars [99], or space vectors [100].

Chang presents some important research in the domain, with several papers [101]–[104] and a *Text to 3D Scene Generation* doctoral thesis published at Stanford as a book in 2015 [105]. The system has a database of 12490 3D models divided into 270 categories. The results are accurate, and the graphics are acceptable. Furthermore, a large dataset of 3D scenes annotated with natural language descriptions is being published to enable machine learning techniques. However, the approach is applied to small sentence text which must include absolutely all the details of the scene and is limited to indoor environments. The latest research paper introduces SceneSeer, an improved system that addresses some of the limitations and is adding a scene inference module. A similar approach based on SceneSeer is presented by Hossain and Salam in 2017 [106].

Fig. 3. "There is a sandwich on a plate" generated with SceneSeer [107]

There is a lot of research dealing specifically with generating interior spaces and especially with generating interior layouts, starting from text descriptions, but these are too specific and generally use templates or datasets trained only on interior rooms, being tangentially relevant through the lens of textual analysis and inferring additional context information for the environment [29], [108]–[117]. For example, if certain details are missing from the text description, they can be filled in, or if the action is taking place in a room, it can be automatically generated. There are also approaches that can generate 3D virtual worlds starting from descriptions and floorplans [118].

Mirzaei et al. [119] developed a story creation platform, rendered in Unity 3D, that extracts characters and represents them as avatars in a predefined environment, deducing their position in the scene, identifying actions, and applying proper animation. Furthermore, it uses speech, avatar gestures, and text descriptions regarding the feelings, emotions, or thoughts of the characters. The system allows viewing the scene from multiple perspectives and can handle multiple storylines. The geometry is represented in a low poly approach, but the graphics quality is good. Still, the system is limited to representing character interaction.

There are also some reviews that cover the topic of converting text into 3D scenes: Rouhizadeh (2011) [120], Hassani and Lee (2017) [121], and Ghoraba and Lakhfif (2022) [42]. All the reviews make an overview of text generation systems in graphics and make a survey of frameworks or methods of generating 2D and 3D scenes starting from text, with some examples and conclusions. All of them analyze and compare mainly the popular CONFUCIUS, CarSim, WordsEye, AVDT, Stanford Text to Scene (SceneSeer), and just a few other solutions, WordsEye being evaluated as the best in terms of graphic quality and representation but lacking animation or interactivity elements. However, all the presented systems are quite old, noticing an absence of newer research on this topic.

AnimaChaotic [122], an article recently published in 2022, takes up the subject from a somewhat newer perspective and proposes the transformation of children's short stories into small animated videos using 3D scene representations. Although more advanced than WordsEye, the results are limited to several previously defined scenes, inferred from the input text using naïve Bayes on the objects. It cannot render text covering multiple scenes or scene transitions. A comparison is also made with similar systems, especially WordsEye, highlighting better scene choice accuracy, animation capabilities, and emotion extraction. However, details on the object database are not mentioned, and other inferring techniques or limitations are not clearly presented.

IV. Discussion

The current study highlights several limitations regarding the current text to 3D scene methods. Although old, the most reliable system found in literature is still WordsEye, with other notable mentions being AVDT, SceneSeer, and maybe AnimaChaotic.

Research using large text corpus to generate complex 3D scenes is absent, and there are just a few example studies that use existing fictional text or text taken from unprocessed, natural language (e.g., from books). Only a few systems have export capabilities, being developed directly in Autodesk Maya or using Unity as a rendering engine.

A good qualitative analysis of the existing methods could not be done, as source code or other testing methods are absent, WordsEye being the only system that allowed online testing until 2022, directly through their website. Furthermore, research on the topic is old, and the graphics quality is outdated, so the defined qualitative metrics are not relevant by today's standards, with none of the analyzed solutions being viable solutions to generate game environments for learning.

Another limitation that stands out is the need to use an unnatural text, overloaded with spatial-temporal details of placing and relating objects. This makes the definition of spaces based on simple natural language or already extensive texts (e.g., from teaching manuals, or exercise collections) almost impossible.

All in all, generating 3D scenes from natural text is still a technical challenge. In [123] the difficulty of generating spaces is explored, highlighting many elements that should be extracted from natural language, divided into several categories: perceptional (e.g., lightness, color, sound, taste, smell, temperature, touch), spatial (e.g., dimensional, directional), human-related (e.g., appearance, mood, behavior) or material-related (e.g., material, consistency, weather). Moreover, there are still challenges like natural language understanding, knowledge representation, common-sense knowledge, implicit knowledge, action learning [121], placement (explicit, implicit, contact) constraints, co-reference, occlusion, relative size and shape of the object in a 3D space [124].

The main problem we identified is that many important elements are often not found in the natural text, so they must be inferred by other methods. For example, if we input the sentence *John was driving to the store*, it is assumable that John is inside a vehicle, driving on a road somewhere in a town. Sproat et al. [125] proposes a solution by using concordance lines for rooms, seasons, and times of the day. Other proposed solutions involve using artificial intelligence to represent the common sense of the real world [124], combine rule-based with data-driven paradigms to design adaptive and interactive systems that utilize available resources (e.g. lexicons and semantic networks) and advanced learning techniques (e.g. active learning, shallow semantic analysis, deep learning, deep reinforcement learning) [121], using imagistic modeling and understanding [126] [127], using annotation systems to create semantic databases [128] or inferring objects using Wikipedia and WordNet [129].

For indoor scenes, a possible solution could be to build models (neural or semantic) that can replace the lack of information from descriptive texts that can learn various configurations of interior scenes (e.g., rooms, bedrooms, kitchens), and generate an optimal result based on the text description, or based on the idea of interior arrangements, approaches covered in multiple research articles.

Another limitation of current methods is the inability to handle multiple scenes and large text corpus. Reference [130] presents a method to identify and divide scenes in a multi-scene text through analysis of scene features. Text summarization techniques could also help identify scenes and extract general information [131]–[134].

V. Conclusions

In this paper we presented a comprehensive survey of the current text to 3D scene methods, analyzing whether they can be successfully used in game-based learning. Multiple evaluation criteria were defined and verified in the literature, highlighting little progress in the domain.

There are multiple limitations discussed, mainly due to the complexity of the natural language, missing spatial information, or common-sense knowledge, demonstrating that text to 3D scene generation is still a challenge, and is not yet suitable to generate 3D virtual environments mature enough to be used in game-based learning.

Possible improvements involve a better usage of modern learning techniques and data-fusion or multimodal approaches, combining information from multiple sources (e.g. text, images, semantic networks) to better infer missing information.

Acknowledgment

The results presented in this article have been funded by the Ministry of Investments and European Projects through the Human Capital Sectoral Operational Program 2014-2020, Contract no. 62461/03.06.2022, SMIS code 153735.

References

[1] D. Alalgawi and S. B. Sadkhan, "Gamification Trends in E-learning – A Survey," *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, pp. 193–198, May 2022, doi: 10.1109/IICETA54559.2022.9888570.

[2] A. Khaldi, R. Bouzidi, and F. Nader, "Gamification of e-learning in higher education: a systematic literature review," *Smart Learn. Environ.*, vol. 10, no. 1, p. 10, Jan. 2023, doi: 10.1186/s40561-023-00227-z.

[3] P. Gupta and P. Goyal, "Is game-based pedagogy just a fad? A self-determination theory approach to gamification in higher education," *IJEM*, vol. 36, no. 3, pp. 341–356, Mar. 2022, doi: 10.1108/IJEM-04-2021-0126.

[4] R. Wazirali, "Aligning Education with Vision 2030 Using Augmented Reality," *Computer Systems Science and Engineering*, vol. 36, no. 2, pp. 339–351, 2021, doi: 10.32604/csse.2021.014301.

[5] R. al azawi, M. Bulshi, and F. farsi, "Educational Gamification Vs. Game Based Learning: Comparative Study," *International Journal of Innovation, Management and Technology (IJIMT)*, vol. 7, pp. 131–136, Aug. 2016, doi: 10.18178/ijimt.2016.7.4.659.

[6] J. Majuri, J. Koivisto, and J. Hamari, *Gamification of Education and Learning: A Review of Empirical Literature.* 2018.

[7] M.-M. Fernandez-Antolin, J. M. del Río, and R.-A. Gonzalez-Lezcano, "The use of gamification in higher technical education: perception of university students on innovative teaching materials," *Int J Technol Des Educ*, vol. 31, no. 5, pp. 1019–1038, Nov. 2021, doi: 10.1007/s10798-020-09583-0.

[8] M. R. de A. Souza, L. Veado, R. T. Moreira, E. Figueiredo, and H. Costa, "A systematic mapping study on game-related methods for software engineering education," *Information and Software Technology*, vol. 95, pp. 201–218, Mar. 2018, doi: 10.1016/j.infsof.2017.09.014.

[9] L. Rodrigues *et al.*, "Are They Learning or Playing? Moderator Conditions of Gamification's Success in Programming Classrooms," *ACM Trans. Comput. Educ.*, vol. 22, no. 3, pp. 1–27, Sep. 2022, doi: 10.1145/3485732.

[10] M. Holenko Dlab and N. Hoic-Bozic, "Effectiveness of game development-based learning for acquiring programming skills in lower secondary education in Croatia," *Educ Inf Technol*, vol. 26, no. 4, pp. 4433–4456, Jul. 2021, doi: 10.1007/s10639-021-10471-w.

[11] A. K. Al-Makhzoomy, K. Zhang, and T. Spannaus, "Game Development-Based Learning: A New Paradigm for Teaching Computer and Object-Oriented Programming," 2020, pp. 244–259. doi: 10.4018/978-1-7998-0249-5.ch012.

[12] E. Demitriadou, K.-E. Stavroulia, and A. Lanitis, "Comparative evaluation of virtual and augmented reality for teaching mathematics in primary education," *Educ Inf Technol*, vol. 25, no. 1, pp. 381–401, Jan. 2020, doi: 10.1007/s10639-019-09973-5.

[13] I. Kazanidis and N. Pellas, "Developing and Assessing Augmented Reality Applications for Mathematics with Trainee Instructional Media Designers: An Exploratory Study on User Experience," *J. Univers. Comput. Sci.*, 2019, Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Developing-and-Assessing-Augmented-Reality-for-with-Kazanidis-Pellas/d0d874a3f6a77932553d8f4b5655f8ba82e67bb4

[14] A. Dünser, L. Walker, H. Horner, and D. Bentall, "Creating interactive physics education books with augmented reality," in *Proceedings of the 24th Australian Computer-Human Interaction Conference*, Melbourne Australia, Nov. 2012, pp. 107–114. doi: 10.1145/2414536.2414554.

[15] I. Silva, A. Wong, B. Auria, D. Zambrano, and V. Echeverria, "Gamification in Engineering Education: Exploring Students' Performance, Motivation, and Engagement," *2022 IEEE Sixth Ecuador Technical Chapters Meeting (ETCM)*, pp. 1–6, Oct. 2022, doi: 10.1109/ETCM56276.2022.9935729.

[16] J. Díaz-Ramírez, "Gamification in engineering education – An empirical assessment on learning and game performance," *Heliyon*, vol. 6, no. 9, p. e04972, Sep. 2020, doi: 10.1016/j.heliyon.2020.e04972.

[17] M. M. Alhammad and A. M. Moreno, "Gamification in software engineering education: A systematic mapping," *Journal of Systems and Software*, vol. 141, pp. 131–150, Jul. 2018, doi: 10.1016/j.jss.2018.03.065.

[18] C. Alonso-Fernández, I. Pérez-Colado, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Improving Serious Games Analyzing Learning Analytics Data: Lessons Learned," in *Games and Learning Alliance*, Cham, 2019, pp. 287–296. doi: 10.1007/978-3-030-11548-7_27.

[19] J. Hamari, J. Koivisto, and H. Sarsa, "Does Gamification Work? – A Literature Review of Empirical Studies on Gamification," in *2014 47th Hawaii International Conference on System Sciences*, Jan. 2014, pp. 3025–3034. doi: 10.1109/HICSS.2014.377.

[20] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, "DreamFusion: Text-to-3D using 2D Diffusion." arXiv, Sep. 29, 2022. doi: 10.48550/arXiv.2209.14988.

[21] C.-H. Lin *et al.*, "Magic3D: High-Resolution Text-to-3D Content Creation." arXiv, Nov. 18, 2022. doi: 10.48550/arXiv.2211.10440.

[22] W. Kuo, A. Angelova, T.-Y. Lin, and A. Dai, "Mask2CAD: 3D Shape Prediction by Learning to Segment and Retrieve." arXiv, Jul. 25, 2020. doi: 10.48550/arXiv.2007.13034.

[23] G. Gkioxari, J. Malik, and J. Johnson, "Mesh R-CNN." arXiv, Jan. 25, 2020. doi: 10.48550/arXiv.1906.02739.

[24] A. Ramesh *et al.*, "Zero-Shot Text-to-Image Generation." arXiv, Feb. 26, 2021. doi: 10.48550/arXiv.2102.12092.

[25] A. Jain, B. Mildenhall, J. T. Barron, P. Abbeel, and B. Poole, "Zero-Shot Text-Guided Object Generation With Dream Fields," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 867–876. Accessed: Jan. 30, 2023. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Jain_Zero-Shot_Text-Guided_Object_Generation_With_Dream_Fields_CVPR_2022_paper.html

[26] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic Image Synthesis with Spatially-Adaptive Normalization." arXiv, Nov. 05, 2019. doi: 10.48550/arXiv.1903.07291.

[27] K. Hu, W. Liao, M. Y. Yang, and B. Rosenhahn, "Text to Image Generation with Semantic-Spatial Aware GAN." arXiv, Mar. 24, 2022. doi: 10.48550/arXiv.2104.00567.

[28] A. Voynov, K. Aberman, and D. Cohen-Or, "Sketch-Guided Text-to-Image Diffusion Models." arXiv, Nov. 24, 2022. doi: 10.48550/arXiv.2211.13752.

[29] S.-H. Zhang, S.-K. Zhang, Y. Liang, and P. Hall, "A Survey of 3D Indoor Scene Synthesis," *Journal of Computer Science and Technology*, vol. 34, pp. 594–608, May 2019, doi: 10.1007/s11390-019-1929-5.

[30] M. A. Bautista *et al.*, "GAUDI: A Neural Architect for Immersive 3D Scene Generation." arXiv, Jul. 27, 2022. doi: 10.48550/arXiv.2207.13751.

[31] C. Yu and Y. Wang, "3D-Scene-GAN: Three-dimensional Scene Reconstruction with Generative Adversarial Networks," Mar. 2018.

[32] A. Yamada, T. Yamamoto, H. Ikeda, T. Nishida, and S. Doshita, "Reconstructing spatial image from natural language texts," in *Proceedings of the 14th conference on Computational linguistics -*, Nantes, France, 1992, vol. 4, p. 1279. doi: 10.3115/992424.992482.

[33] B. Coyne and R. Sproat, "WordsEye: an automatic text-to-scene conversion system," *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 487–496, Aug. 2001, doi: 10.1145/383259.383316.

[34] R. E. Coyne, A. Klapheke, M. Rouhizadeh, R. Sproat, and D. Bauer, "Annotation Tools and Knowledge Representation for a Text-To-Scene System," 2012. doi: 10.7916/D84F205R.

[35] D. Hutchison *et al.*, "Frame Semantics in Text-to-Scene Generation," Berlin, Heidelberg, 2010, vol. 6279, pp. 375–384. doi: 10.1007/978-3-642-15384-6_40.

[36] M. Rouhizadeh, M. Bowler, R. Sproat, and B. Coyne, "Data collection and normalization for building the Scenario-Based Lexical Knowledge Resource of a text-to-scene conversion system," *2010 Fifth International Workshop Semantic Media Adaptation and Personalization*, pp. 25–30, Dec. 2010, doi: 10.1109/SMAP.2010.5706851.

[37] R. Sproat, R. E. Coyne, and J. B. Hirschberg, "Spatial Relations in Text-to-Scene Conversion," 2010. doi: 10.7916/D8902C78.

[38] R. E. Coyne, C. Schudel, M. Bitz, and J. B. Hirschberg, "Evaluating a Text-to-Scene Generation System as an Aid to Literacy," 2011. doi: 10.7916/D8FR0532.

[39] M. Rouhizadeh, D. Bauer, R. E. Coyne, O. C. Rambow, and R. Sproat, "Collecting Spatial Information for Locations in a Text-to-Scene Conversion System," 2011. doi: 10.7916/D8KH0WT0.

[40] M. Rouhizadeh, B. Coyne, and R. Sproat, "Collecting Semantic Information for Locations in the Scenario-Based Lexical Knowledge Resource of a Text-to-Scene Conversion System," Berlin, Heidelberg, 2011, vol. 6884, pp. 378–387. doi: 10.1007/978-3-642-23866-6_40.

[41] S. N. Dessai, "Text to 3 D Scene Generation," 2016. Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Text-to-3-D-Scene-Generation-Dessai/ef0a9b191d9ddd67293018b36d31296a103ac89d

[42] M. A. Ghorab and A. Lakhfif, "Text to 3D,2D scene generation systems, Frameworks and approaches: a survey," in *2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS)*, Oct. 2022, pp. 1–6. doi: 10.1109/PAIS56586.2022.9946876.

[43] M. Ulinski, B. Coyne, and J. Hirschberg, "Evaluating the WordsEye Text-to-Scene System: Imaginative and Realistic Sentences," presented at the International Conference on Language Resources and Evaluation, May 2018. Accessed: Feb. 07, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Evaluating-the-WordsEye-Text-to-Scene-System%3A-and-Ulinski-Coyne/a2a032f907c19256863e1e011b834eb1f8057a4c

[44] M. E. Ulinski, "Leveraging Text-to-Scene Generation for Language Elicitation and Documentation," 2019. doi: 10.7916/D8-2WMG-2230.

[45] S. Dupuy, A. Egges, V. Legendre, and P. Nugues, "Generating a 3D simulation of a car accident from a written description in natural language: the CarSim system," *Proceedings of the workshop on Temporal and spatial information processing* -, vol. 13, pp. 1–8, 2001, doi: 10.3115/1118238.1118239.

[46] O. Åkerberg, H. Svensson, B. Schulz, and P. Nugues, "CarSim: an automatic 3D text-to-scene conversion system applied to road accident reports," in *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - EACL '03*, Budapest, Hungary, 2003, vol. 2, p. 191. doi: 10.3115/1067737.1067782.

[47] R. Johansson, A. Berglund, M. Danielsson, and P. Nugues, "Automatic Text-to-Scene Conversion in the Traffic Accident Domain," presented at the International Joint Conference on Artificial Intelligence, Jul. 2005. Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Automatic-Text-to-Scene-Conversion-in-the-Traffic-Johansson-Berglund/e8d54b1cb99af8bb44ff9e0c9bdc95dee23a3bde?sort=pub-date

[48] A. Berglund, R. Johansson, and P. Nugues, "A Machine Learning Approach to Extract Temporal Information from Texts in Swedish and Generate Animated 3 D Scenes," 2006. Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/A-Machine-Learning-Approach-to-Extract-Temporal-in-Berglund-Johansson/fe19d7ddc53112ac7f116b1a9895ea7fabb0ca02

[49] X. Zeng, Q. H. Mehdi, and N. E. Gough, "Shape of the story: story visualization techniques," *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.*, pp. 144–149, 2003, doi: 10.1109/IV.2003.1217971.

[50] X. Zeng, Q. Mehdi, and N. Gough, "3D Scene Creation Using Story-Based Descriptions," 2005. Accessed: Mar. 04, 2023. [Online]. Available: https://www.semanticscholar.org/paper/3D-Scene-Creation-Using-Story-Based-Descriptions-Zeng-Mehdi/7dd4a8c190144a7be7e8f54bdd82d8568599d698

[51] Z.-Q. Liu and K.-M. Leung, "Script visualization (ScriptViz): a smart system that makes writing fun," *Soft Comput*, vol. 10, no. 1, pp. 34–40, Jan. 2006, doi: 10.1007/s00500-005-0461-4.

[52] M. Marti *et al.*, "CARDINAL: Computer Assisted Authoring of Movie Scripts," *23rd International Conference on Intelligent User Interfaces*, pp. 509–519, Mar. 2018, doi: 10.1145/3172944.3172972.

[53] R. Sanghrajka, R. M. Young, B. Salisbury, and E. W. Lang, "ShowRunner: A Tool for Storyline Execution/Visualization in 3D Game Environments," Cham, 2019, vol. 11869, pp. 323–327. doi: 10.1007/978-3-030-33894-7_32.

[54] H. Subramonyam, W. Li, E. Adar, and M. Dontcheva, "TakeToons: Script-driven Performance Animation," in *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, New York, NY, USA, Oct. 2018, pp. 663–674. doi: 10.1145/3242587.3242618.

[55] K. Sathe, "Generating Animations from Instructional Text," *IJATCSE*, vol. 9, no. 3, pp. 3023–3027, Jun. 2020, doi: 10.30534/ijatcse/2020/81932020.

[56] E. Hanser, P. Mc Kevitt, T. Lunney, and J. Condell, "SceneMaker: Intelligent Multimodal Visualisation of Natural Language Scripts," Berlin, Heidelberg, 2010, vol. 6206, pp. 144–153. doi: 10.1007/978-3-642-17080-5_17.

[57] E. Hanser, P. Mc Kevitt, T. Lunney, and J. Condell, "SceneMaker: Automatic Visualisation of Screenplays," Berlin, Heidelberg, 2009, vol. 5803, pp. 265–272. doi: 10.1007/978-3-642-04617-9_34.

[58] M. Akser *et al.*, "SceneMaker: Creative Technology for Digital StoryTelling," Cham, 2017, vol. 196, pp. 29–38. doi: 10.1007/978-3-319-55834-9_4.

[59] R. Velonoromanalintantely, T. Andriamarozakaina, C. Sanza, V. Gaildrat, M. Martinez-Thomas, and P. Matthieu, "Theatrical Text to 3D Virtual Scenography," Berlin, Heidelberg, 2013, vol. 441, pp. 23–40. doi: 10.1007/978-3-642-31745-3_2.

[60] M. Kapadia, J. Falk, F. Zünd, M. Marti, R. W. Sumner, and M. Gross, "Computer-assisted authoring of interactive narratives," in *Proceedings of the 19th Symposium on Interactive 3D Graphics and Games*, New York, NY, USA, Feb. 2015, pp. 85–92. doi: 10.1145/2699276.2699279.

[61] M. Ma, "Automatic Conversion of Natural Language to 3D Animation," Jul. 2006. Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Automatic-Conversion-of-Natural-Language-to-3D-Ma/55a2a4d9e585bb2829a316e96e76f6a62bf5cf3b

[62] L. Seversky and L. Yin, "Real-time automatic 3D scene generation from natural language voice and text descriptions," Oct. 2006, pp. 61–64. doi: 10.1145/1180639.1180660.

[63] K. Glass and S. Bangay, "Constraint-based conversion of fiction text to a time-based graphical representation," in *Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, Port Elizabeth South Africa, Oct. 2007, pp. 19–28. doi: 10.1145/1292491.1292494.

[64] K. R. Glass, "Automating the conversion of natural language fiction to multi-modal 3D animated virtual environments," 2009. Accessed: Mar. 04, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Automating-the-conversion-of-natural-language-to-3D-Glass/2f3052458ffc7b4cb68fbf8f3032fecce8a35e43

[65] R. Berkland and S. Bangay, "Identifying annotations for adventure game generation from fiction text," in *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*, Bela Bela South Africa, Oct. 2010, pp. 19–28. doi: 10.1145/1899503.1899506.

[66] Y. Kılıçaslan, Ö. Uçar, and E. S. Güner, "An NLP-Based 3D Scene Generation System for Children with Autism or Mental Retardation," Berlin, Heidelberg, 2008, vol. 5097, pp. 929–938. doi: 10.1007/978-3-540-69731-2_88.

[67] D. Tappan, "Knowledge-Based Spatial Reasoning for Scene Generation from Text Descriptions," presented at the AAAI Conference on Artificial Intelligence, Jul. 2008. Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Knowledge-Based-Spatial-Reasoning-for-Scene-from-Tappan/0c999e177ebdf17f742a7f88a7189b8b52964166

[68] Jiajie Lu, Canlin Li, Chao Yin, and Lizhuang Ma, “A new framework for automatic 3D scene construction from text description,” *2010 IEEE International Conference on Progress in Informatics and Computing*, pp. 964–968, Dec. 2010, doi: 10.1109/PIC.2010.5687870.

[69] Canlin Li, Chao Yin, Jiajie Lu, and Lizhuang Ma, “Automatic 3D scene generation based on Maya,” *2009 IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design*, pp. 981–985, Nov. 2009, doi: 10.1109/CAIDCD.2009.5375246.

[70] M. A. Helfiandri, F. Zakhralativa Ruskanda, and M. L. Khodra, “Generating Scene Descriptor from Indonesian Narrative Text,” in *2020 International Conference on ICT for Smart Society (ICISS)*, Nov. 2020, pp. 1–6. doi: 10.1109/ICISS50791.2020.9307536.

[71] X. Zeng and M. Tan, “The Development of Text to Scene Tool for Aiding Children’s Spatial Knowledge Understanding and Language Learning,” *2010 International Conference on Internet Technology and Applications*, pp. 1–4, Aug. 2010, doi: 10.1109/ITAPP.2010.5566174.

[72] E. Bolaño-Rodríguez, J. C. González-Moreno, D. Ramos-Valcarcel, and L. Vázquez-López, “Using Multi-Agent Systems to Visualize Text Descriptions,” Berlin, Heidelberg, 2011, vol. 88, pp. 39–45. doi: 10.1007/978-3-642-19875-5_5.

[73] E. Ünal, “A Language Visualization System,” 2014. Accessed: Mar. 12, 2023. [Online]. Available: https://www.semanticscholar.org/paper/A-Language-Visualization-System-%C3%9Cnal/a527fe2b32e1abc63c907ade0d5b7876a2d56f96

[74] A. Nayak, B. Geleda, A. Sakhapara, A. Singh, and N. Acharya, “Visualization of Mechanics Problems based on Natural Language Processing,” *IJCA*, vol. 116, no. 14, pp. 34–37, Apr. 2015, doi: 10.5120/20408-2766.

[75] C. Claude-Lachenaud, E. Charton, B. Ozell, and M. Gagnon, “A multimodal interpreter for 3D visualization and animation of verbal concepts,” presented at the International Conference on Language Resources and Evaluation, May 2014. Accessed: Mar. 12, 2023. [Online]. Available: https://www.semanticscholar.org/paper/A-multimodal-interpreter-for-3D-visualization-and-Claude-Lachenaud-Charton/3f87b87345a79862d9e4f0c458d42ffbc0d01d14

[76] F. Yang, Y. Zhou, and X. Luo, “Scene layout in text-to-scene conversion,” *The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014)*, pp. 891–895, Nov. 2014, doi: 10.1109/ICSAI.2014.7009411.

[77] J. Balint, J. Allbeck, M. Hieb, and G. Mason, “Automated Simulation Creation from Military Operations Documents,” 2015. Accessed: Mar. 04, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Automated-Simulation-Creation-from-Military-Balint-Allbeck/a136c984169c3423a6f0bc7a1f50e419d75298a7

[78] Y. Hu *et al.*, “Construction and Optimization of Three-Dimensional Disaster Scenes within Mobile Virtual Reality,” *IJGI*, vol. 7, no. 6, p. 215, Jun. 2018, doi: 10.3390/ijgi7060215.

[79] S. Shekhar, P. Maheshwari, M. J, A. Singhal, K. K. Singh, and K. Krishna, “ARComposer: Authoring Augmented Reality Experiences through Text,” *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 63–65, Oct. 2019, doi: 10.1145/3332167.3357116.

[80] P. Jain *et al.*, “Virtual reality: an aid as cognitive learning environment—a case study of Hindi language,” *Virtual Reality*, vol. 24, no. 4, pp. 771–781, Dec. 2020, doi: 10.1007/s10055-020-00426-w.

[81] G. Vyas, P. Kadam, and M. Thaker, “Enabl: Visual Aid for Dyslexic based on Natural language procesing,” *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 52–57, Sep. 2020, doi: 10.1109/ICOSEC49089.2020.9215268.

[82] H. Liang, F. Bao, Y. Sun, C. Ge, and J. Chang, “Virtual Scenes Construction Promotes Traditional Chinese Art Preservation,” Cham, 2021, vol. 13002, pp. 621–632. doi: 10.1007/978-3-030-89029-2_46.

[83] P. Jain *et al.*, “Evaluation of Automatic Text Visualization Systems: A Case Study,” Singapore, 2021, vol. 1141, pp. 25–37. doi: 10.1007/978-981-15-3383-9_3.

[84] G. A. Krikos, N. N. Karanikolas, G. Miaoulis, and A. Voulodimos, “Greek language object representation scene system: a text interface for designing X3D objects and X3D scenes,” *Proceedings of the 23rd Pan-Hellenic Conference on Informatics*, pp. 164–167, Nov. 2019, doi: 10.1145/3368640.3368666.

[85] R. A. Kadir, A. R. M. Hashim, R. Wirza, and A. Mustapha, “3D Visualization of Simple Natural Language Statement Using Semantic Description,” Berlin, Heidelberg, 2011, vol. 7066, pp. 36–44. doi: 10.1007/978-3-642-25191-7_5.

[86] F. Yang, J. Sun, and Z. Huang, “Preliminary Implementation of Text-to-scene System:,” presented at the First International Conference on Information Sciences, Machinery, Materials and Energy, Chongqing, China, 2015. doi: 10.2991/icismme-15.2015.278.

[87] R. Delair, H. Sojitra, K. Patel, K. Bhatt, K. Gharia, and R. Mahajan, “Text to Scene conversion for smart learning,” *2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*, pp. 76–81, Dec. 2016, doi: 10.1109/ICAECCT.2016.7942559.

[88] S. N. Dessai and R. Dhanaraj, “Text to 3D Scene Generation”.

[89] S. N. Dessai and R. Dhanaraj, “Creation of 3D scene from raw text,” *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pp. 1466–1469, May 2016, doi: 10.1109/RTEICT.2016.7808075.

[90] P. Jain, H. Darbari, and V. C. Bhavsar, “Spatial intelligence from hindi language text for scene generation,” *2017 2nd International Conference for Convergence in Technology (I2CT)*, pp. 132–138, Apr. 2017, doi: 10.1109/I2CT.2017.8226108.

[91] X. Zeng, M. Tan, and S. Ren, “The Implementation of Graphic Constraints for Automatic Text to Scene Conversion,” *dtcse*, no. aics, Apr. 2017, doi: 10.12783/dtcse/aics2016/8224.

[92] P. Jain, R. Bhavsar, A. Kumar, B. V. Pawar, H. Darbari, and V. C. Bhavsar, “Tree Adjoining Grammar Based Parser for a Hindi Text-to-Scene Conversion System,” *2018 3rd International Conference for Convergence in Technology (I2CT)*, pp. 1–7, Apr. 2018, doi: 10.1109/I2CT.2018.8529491.

[93] D. Viradia, A. S. Thakurathi, D. Sonavaria, and A. Waghmare, “Automatic Visualization of Descriptive Text,” 2018. Accessed: Mar. 04, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Automatic-Visualization-of-Descriptive-Text-Viradia-Thakurathi/009f95d51cf51ab9f83ab2e7df11b8b7b9f39315

[94] S. Deepa and D. P. Kumar, “INTELLIGENT USER INTERACTIVE MODEL FOR REAL TIME TEXT-GRAPHIC GENERATION.” Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/INTELLIGENT-USER-INTERACTIVE-MODEL-FOR-REAL-TIME-Deepa-Kumar/a808dbe9641909fbcf6fbdba9f8bdd09405d524b

[95] C. Spika, K. Schwarz, H. Dammertz, and H. P. A. Lensch, “AVDT - Automatic Visualization of Descriptive Texts,” in *Vision*, 2011, vol. Modeling, p. 8 pages. doi: 10.2312/PE/VMV/VMV11/129-136.

[96] V. Bova and E. Cardillo, “Descriptive Sentence Extraction for Text to 3D Scene Generation,” Apr. 2018. Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Descriptive-Sentence-Extraction-for-Text-to-3D-Bova-Cardillo/2ef6ed1506660fdd0b81f91655e1772f902b0ed4

[97] X. Zeng, “Visual semantic approach for virtual scene generation,” *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pp. 553–556, Dec. 2011, doi: 10.1145/2087756.2087862.

[98] X. Zeng, Ml. Tan, and Wj. Chen, “Extraction of Visual Material and Spatial Information from Text Description for Scene Visualization,” Berlin, Heidelberg, 2011, vol. 104, pp. 735–741. doi: 10.1007/978-3-642-23777-5_118.

[99] D. Bauer, “Understanding Descriptions of Visual Scenes Using Graph Grammars,” *AAAI*, vol. 27, no. 1, pp. 1656–1657, Jun. 2013, doi: 10.1609/aaai.v27i1.8498.

[100] C. Wang, B. Tang, and S. Cao, “Space Vector Generation for 3D Scenes from Text Descriptions,” in *Proceedings of the 2019 3rd International Conference on Artificial Intelligence and Virtual Reality*, Singapore Singapore, Jul. 2019, pp. 8–11. doi: 10.1145/3348488.3348497.

[101] W. Monroe, “3D Scene Retrieval from Text with Semantic Parsing,” 2014. Accessed: Mar. 04, 2023. [Online]. Available: https://www.semanticscholar.org/paper/3D-Scene-Retrieval-

from-Text-with-Semantic-Parsing-Monroe/d07290c564cf27f8da8f92be8f120fddbf4e6af0

[102] A. Chang, M. Savva, and C. Manning, "Semantic Parsing for Text to 3D Scene Generation," in *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, Baltimore, MD, 2014, pp. 17–21. doi: 10.3115/v1/W14-2404.

[103] A. Chang, M. Savva, and C. Manning, "Interactive Learning of Spatial Knowledge for Text to 3D Scene Generation," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Baltimore, Maryland, USA, Jun. 2014, pp. 14–21. doi: 10.3115/v1/W14-3102.

[104] A. Chang, W. Monroe, M. Savva, C. Potts, and C. D. Manning, "Text to 3D Scene Generation with Rich Lexical Grounding." arXiv, Jun. 04, 2015. doi: 10.48550/arXiv.1505.06289.

[105] A. X. Chang, *Text to 3D Scene Generation*. Stanford University, 2015.

[106] M. Hossain and A. Salam, "Text-to-3D-Scene Generation using Semantic Parsing and Spatial Knowledge with Rule Based System," *International Journal of Computer Science Issues*, vol. 14, Sep. 2017, doi: 10.20943/01201705.3741.

[107] A. X. Chang, M. Eric, M. Savva, and C. D. Manning, "SceneSeer: 3D Scene Design with Natural Language," *ArXiv*, Feb. 2017, Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/SceneSeer%3A-3D-Scene-Design-with-Natural-Language-Chang-Eric/dd3cd261320bb3f6dcc455a918422821c5f14ade

[108] L. Gao, J.-M. Sun, K. Mo, Y.-K. Lai, L. J. Guibas, and J. Yang, "SceneHGN: Hierarchical Graph Networks for 3D Indoor Scene Generation with Fine-Grained Geometry," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–18, 2023, doi: 10.1109/TPAMI.2023.3237577.

[109] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan, "Example-based synthesis of 3D object arrangements," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 1–11, Nov. 2012, doi: 10.1145/2366145.2366154.

[110] Y. Liang, F. Xu, S.-H. Zhang, Y. Lai, and T. Mu, "Knowledge graph construction with structure and parameter learning for indoor scene design," *Computational Visual Media*, 2018, Accessed: Feb. 07, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Knowledge-graph-construction-with-structure-and-for-Liang-Xu/92603835b00190a92d1bf0b324193bee08dcd79c

[111] X. Yang, F. Hu, and L. Ye, "Text to Scene: A System of Configurable 3D Indoor Scene Synthesis," in *Proceedings of the 29th ACM International Conference on Multimedia*, New York, NY, USA, Oct. 2021, pp. 2819–2821. doi: 10.1145/3474085.3478570.

[112] G. Radevski, G. Collell, M.-F. Moens, and T. Tuytelaars, "Decoding Language Spatial Relations to 2D Spatial Arrangements," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Online, 2020, pp. 4549–4560. doi: 10.18653/v1/2020.findings-emnlp.408.

[113] E. Karan, S. Asgari, and A. Rashidi, "A Markov Decision Process Workflow for Automating Interior Design," *KSCE J Civ Eng*, vol. 25, no. 9, pp. 3199–3212, Sep. 2021, doi: 10.1007/s12205-021-1272-6.

[114] J. Kang *et al.*, "Fuzzy Speech Driven Indoor Modeling System," *2019 International Conference on Virtual Reality and Visualization (ICVRV)*, pp. 41–49, Nov. 2019, doi: 10.1109/ICVRV47840.2019.00016.

[115] J. T. Balint and R. Bidarra, "A generalized semantic representation for procedural generation of rooms," *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pp. 1–8, Aug. 2019, doi: 10.1145/3337722.3341848.

[116] M. Keshavarzi, "Indoor Scene Augmentation via Scene Graph Priors," 2022. Accessed: Feb. 07, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Indoor-Scene-Augmentation-via-Scene-Graph-Priors-Keshavarzi/007b39325422d5660b22a81b37038c46688bc47b

[117] Y. Liang, S.-H. Zhang, and R. Martin, "Learning guidelines for automatic indoor scene design," *Multimedia Tools and Applications*, 2018, Accessed: Feb. 07, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Learning-guidelines-for-automatic-indoor-scene-Liang-Zhang/53a620f102a99716bea89b8a618c4209d85ecaec

[118] T. Trescak, M. Esteva, and I. Rodriguez, "A Virtual World Grammar for automatic generation of virtual worlds," *Vis Comput*, vol. 26, no. 6–8, pp. 521–531, Jun. 2010, doi: 10.1007/s00371-010-0473-7.

[119] F. Meunier *et al.*, "Collaborative learning through story envisioning in virtual reality," Dec. 2019, 1st ed., pp. 297–303. doi: 10.14705/rpnet.2019.38.1026.

[120] M. Rouhizadeh, *An Overview of Text-to-Scene Conversion Systems*. 2011. doi: 10.13140/RG.2.2.12488.96008/2.

[121] K. Hassani and W.-S. Lee, "Visualizing Natural Language Descriptions: A Survey," *ACM Comput. Surv.*, vol. 49, no. 1, pp. 1–34, Mar. 2017, doi: 10.1145/2932710.

[122] R. Abdel-Salam, R. Gody, M. Maher, H. Hosny, and A. S. Kaseb, "AnimaChaotic: AI-based Automatic Conversion of Children's Stories to Animated 3D Videos:," in *Proceedings of the 14th International Conference on Agents and Artificial Intelligence*, Online Streaming, --- Select a Country ---, 2022, pp. 226–235. doi: 10.5220/0010815100003116.

[123] X. Zeng and T. Mling, "A Review of Scene Visualization Based on Language Descriptions," *2009 Sixth International Conference on Computer Graphics, Imaging and Visualization*, pp. 429–433, Aug. 2009, doi: 10.1109/CGIV.2009.9.

[124] D. B. S. Shylaja, "A Survey on Challenges in Text to Scene Generation," 2020. Accessed: Mar. 04, 2023. [Online]. Available: https://www.semanticscholar.org/paper/A-Survey-on-Challenges-in-Text-to-Scene-Generation-Shylaja/d889eaa3154e43b944ba216544266e0a90aa37c2

[125] R. Sproat, "Inferring the environment in a text-to-scene conversion system," in *Proceedings of the 1st international conference on Knowledge capture*, Victoria British Columbia Canada, Oct. 2001, pp. 147–154. doi: 10.1145/500737.500761.

[126] E. Bigelow, D. Scarafoni, L. Schubert, and A. Wilson, "On the need for imagistic modeling in story understanding," in *Biologically Inspired Cognitive Architectures*, Jan. 2015, vol. 11, pp. 22–28. doi: 10.1016/j.bica.2014.11.012.

[127] A. Li, J. Sun, J. Y.-H. Ng, R. Yu, V. I. Morariu, and L. S. Davis, "Generating Holistic 3D Scene Abstractions for Text-Based Image Retrieval," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1942–1950, Jul. 2017, doi: 10.1109/CVPR.2017.210.

[128] G. Abrami, A. Henlein, A. Kett, and A. Mehler, "Text2SceneVR: Generating Hypertexts with VAnnotatoR as a Pre-processing Step for Text2Scene Systems," *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, pp. 177–186, Jul. 2020, doi: 10.1145/3372923.3404791.

[129] A. Cropper, "Identifying and inferring objects from textual descriptions of scenes from books," 2014, p. 8 pages. doi: 10.4230/OASICS.ICCSW.2014.19.

[130] F. Yang, Z. Yuan, and X. Cheng, "A Scene Division Method Based on Theme:," *Proceedings of the International Conference on Advances in Computer Technology, Information Science and Communications*, pp. 44–48, 2019, doi: 10.5220/0008098000440048.

[131] R. Vijaya Saraswathi, R. V. Chunchu, S. Kunchala, M. Varun, T. Begari, and S. Bodduru, "A LSTM based Deep Learning Model for Text Summarization," *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, pp. 1063–1068, Dec. 2022, doi: 10.1109/ICECA55336.2022.10009541.

[132] N. Ibrahim Altmami and M. El Bachir Menai, "Automatic summarization of scientific articles: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1011–1028, Apr. 2022, doi: 10.1016/j.jksuci.2020.04.020.

[133] A. S. Uban and C. Caragea, "Generating Summaries for Scientific Paper Review," *ArXiv*, Sep. 2021, Accessed: Mar. 05, 2023. [Online]. Available: https://www.semanticscholar.org/paper/Generating-Summaries-for-Scientific-Paper-Review-Uban-Caragea/b9e2d8721c0c33e75ecedf05492b56a632b53bbb

[134] D. Yadav, J. Desai, and A. K. Yadav, "Automatic Text Summarization Methods: A Comprehensive Review," 2022, doi: 10.48550/ARXIV.2204.01849.