ARComposer: Authoring Augmented Reality Experiences through Text

Sumit Shekhar
Adobe Research
sushekha@adobe.com

Paridhi Maheshwari
Adobe Research
parimahe@adobe.com

Monisha J
IIT Madras
monishaj.65@gmail.com

Amrit Singhal
IIT Kanpur
amritsinghal97@gmail.com

Kush Kumar Singh
IIT Roorkee
singhkush114@gmail.com

Kundan Krishna
Carnegie Mellon University
kundank@andrew.cmu.edu

ABSTRACT
Augmented Reality (AR) is rapidly gaining popularity, enhancing human perception of the real world by augmenting digital experiences. Existing tools for authoring AR scenes are either template based or require domain knowledge from experts, and are therefore restrictive. ARComposer is a novel interface that enables easy authoring of AR experiences from free-form text describing the scene. Our proposed interface allows creators to compose varied scenes comprising of multiple objects with diverse relationships to each other as well as human models with animations, starting merely with a textual description. A qualitative evaluation shows that ARComposer provides a good flow experience to its users. Furthermore, a crowd-sourced experiment evaluating various aspects of the rendered AR scenes indicates the viability of the proposed approach.

Author Keywords
Augmented Reality, AR Rendering, Natural Language, 3D models

CCS Concepts
•**Human-centered computing** → **Mixed / augmented reality; Visualization systems and tools;** *Empirical studies in HCI;* •**Computing methodologies** → *Machine learning;*

INTRODUCTION
Augmented reality (AR) is the integration of digitally created content into the real world to provide a visually enriched experience to its users. AR blends the lines between digital and physical worlds and offers an interactive way of engaging with the surroundings. Owing to its ability to immerse digital content, various industries are exploring the possibilities offered by AR to elevate the user experience. Some prominent applications of AR include educational tools [12], navigation and path planning [5], medical visualization [11] and gaming [15].

With increase in the applications of AR, there has been significant technological advancement in tools for developing AR content. Google ARCore[1] and Apple ARKit[2] particularly have made AR development accessible to developers around the globe. AR has also been harvested well by apps like IKEA Place[3] and Pokemon Go[4] to enrich the user experience. For example, IKEA Place allows users to place catalogue furniture in the real world setting and examine furniture compatibility.

While such apps allow the user to design motionless scenes by placing models at specific locations, there is little flexibility for customizing these AR scenes. There have been efforts in enabling re-coloring of the scene [16] or custom selection of models for AR display [6]. However, the creation of AR scenes involving model interaction has not been studied extensively. While DART [10] and ComposeAR [14] provide a schema-based approach to create AR experiences, they require the user to describe the scene using a defined template which hinders the creation experience of the user. MagicBook [1] and ARTutor [9] use marker technology to trigger pre-built augmentations. There are also firms such as Carlton books[5] and Alive Studios[6], which engage in converting text-based content like kids' stories into AR experiences. But these are typically performed by creative professionals and take considerable time for creation. This reliance on domain experts to develop rich AR scenes renders these processes non-scalable. Hence, there is a technological need for enabling novice users to create personalized AR scenes.

To this end, we propose a novel framework - ARComposer, which allows users to compose experiences in AR using free-form text descriptions. Natural language input has been shown to considerably improve the user experience while using conversational interfaces [7] and it also allows rich descriptions

UIST '19 Adjunct, October 20-23, 2019, New Orleans, LA, USA

ACM ISBN 978-1-4503-6817-9/19/10.
https://doi.org/10.1145/3332167.3357116

[1] https://developers.google.com/ar/
[2] https://developer.apple.com/arkit/
[3] https://highlights.ikea.com/2017/ikea-place/
[4] https://www.pokemongo.com/en-us/
[5] https://www.carltonbooks.co.uk/
[6] https://alivestudiosco.com/

from users. We leverage progress in machine learning techniques to accomplish an easy-to-use interface for users. Specifically, we build upon the work in text-to-visual conversion [13, 3, 2] for text parsing and machine learning models for predicting spatial arrangements. ARComposer works with text describing both static and dynamic scenes, and also handles implicit mentions. Further, a background image is added to describe parts of the scene which cannot be described sufficiently through 3D models, such as 'day' and 'night'.

The augmented reality output is realized using Google's ARCore library through the Unity3D [7] application. The objects required to render the scene in AR are obtained from a curated model repository. Human models and animations are from the publicly available Mixamo[8] collection, while models for the common objects are fetched from ShapeNet [4], a large corpus of 3D objects with rich annotations.

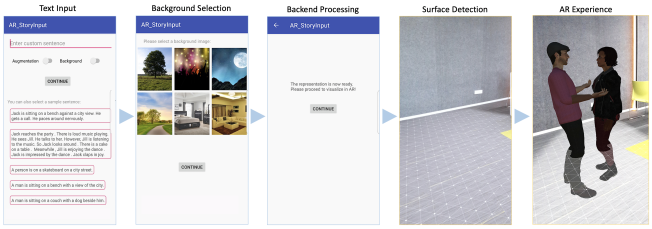

Figure 1. ARComposer allows users to create an AR rendering through a simple process. The user enters free-form text input. Once processed, they proceed to visualize the scene in AR.

INTERFACE

ARComposer follows a simple interface design to allow users to focus on the creation of AR experiences using free-form text. Figure 1 illustrates the interface design, which includes a text box to input the scene description and two toggle buttons for scene augmentation and background inclusion.

The user first describes the AR scene that they wish to design. ARComposer allows the user to include various commonplace objects and human models with animations exemplified in Figure 2. The user can describe their relations (e.g., "*A chair next to the bed*") and can include humans or animals (e.g., "*A man is sitting on a sofa with a dog*"). There is a **Scene Augmentation** option to infer implicit information in the text using common-world knowledge. For example, the object *chair* can be augmented for the input *"Tom is sitting"*. The interface also provides an option to choose a suitable background image to the AR scene.

Objects				Animations	
Ball	Bench	Frisbee	Skateboard	Clap	Look
Bat	Plate	Cabinet	Television	Run	Slap
Girl	Sofa	Camera	Umbrella	Sit	Dance
Man	Table	Laptop	Vegetables	Talk	Yawn

Figure 2. Examples from ARComposer model repository.

[7] https://unity3d.com/
[8] www.mixamo.com

On clicking the **Continue** button, ARComposer processes the text as shown in Figure 3. Once the output is ready, the user can choose any flat surface to render the scene in AR, and view it from different viewpoints.

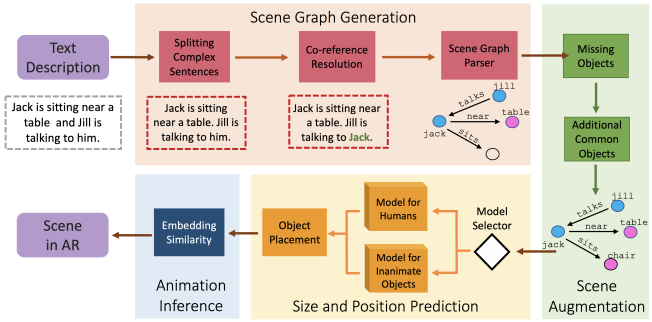

Figure 3. Different processing steps at the backend of ARComposer.

EVALUATING ARCOMPOSER

ARComposer has been evaluated using a comprehensive in-person survey ($N = 20$, 7 females, age $M = 24.2$, $SD = 4.92$) as well as a large-scale crowd-sourced experiment ($N = 278$). For the in-person survey, participants freely explored the app for ~ 30 minutes and constructed scenes of their choice. Participants with diverse backgrounds (young undergraduates, researchers, design experts, IT professionals) were chosen to provide us with varied perspectives and creative needs to designing AR scenes. Finally, through a semi-structured questionnaire, they provided qualitative feedback about the interface, as well as rated the generated AR scenes for correctness. The rating was done on a 5-point Likert scale with the items *Very poor*, *Poor*, *Neutral*, *Good*, *Very good*, results of which have been tabulated in Table 1. Additionally, an extensive survey was carried out on Amazon Mechanical Turk (MTurk) to test the scene correctness. Workers were shown scenes corresponding to different captions from MS COCO [8] and asked to rate as before. Ratings show that ARComposer performs well on different aspects of scene generation and also the overall coherence of the scene. Some feedback from the in-person survey:

"*The app can be used to create first draft of ideas that creative personas may have in their mind and reduces the cognitive load of starting a work from scratch.*"

"*I liked that the size of the objects was relatively more realistic in the scene. It was easy to visualize and control.*"

Aspect	In-Person ($n = 20$)	MTurk ($n = 278$)
Position	3.70 ± 1.06	3.38 ± 1.34
Size	3.78 ± 1.30	3.33 ± 1.26
Augmentation	3.84 ± 0.94	3.32 ± 1.23
Background	3.62 ± 1.11	3.16 ± 1.28
Human Actions	4.12 ± 1.09	3.45 ± 1.15
Overall	3.65 ± 1.17	3.20 ± 1.22

Table 1. Mean and standard deviation of ratings (scale 1-5) for different aspects in the In-Person and MTurk surveys.

REFERENCES

[1] Mark Billinghurst, Hirkazu Kato, and Ivan Poupyrev. 2001. The MagicBook: 1Moving Seamlessly Between Reality and Virtuality. *IEEE Comput. Graph. Appl.* 21, 3 (May 2001), 6–8. DOI: http://dx.doi.org/10.1109/38.920621

[2] Angel Chang, Will Monroe, Manolis Savva, Christopher Potts, and Christopher D Manning. 2015b. Text to 3d scene generation with rich lexical grounding. *arXiv preprint arXiv:1505.06289* (2015).

[3] Angel Chang, Manolis Savva, and Christopher Manning. 2014. Semantic parsing for text to 3d scene generation. In *Proceedings of the ACL 2014 Workshop on Semantic Parsing*. 17–21.

[4] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. 2015a. *ShapeNet: An Information-Rich 3D Model Repository*. Technical Report arXiv:1512.03012 [cs.GR]. Stanford University — Princeton University — Toyota Technological Institute at Chicago.

[5] Frank J Delgado, Scott Altman, Michael F Abernathy, and Janis White. 2000. Virtual cockpit window for the X-38 crew return vehicle. In *Enhanced and Synthetic Vision 2000*, Vol. 4023. International Society for Optics and Photonics, 63–71.

[6] Andreas Dünser, Lawrence Walker, Heather Horner, and Daniel Bentall. 2012. Creating interactive physics education books with augmented reality. In *Proceedings of the 24th Australian computer-human interaction conference*. ACM, 107–114.

[7] Lorenz Cuno Klopfenstein, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 555–565.

[8] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[9] Chris Lytridis, Avgoustos Tsinakos, and Ioannis Kazanidis. 2018. ARTutor-An Augmented Reality Platform for Interactive Distance Learning. *Education Sciences* 8, 1 (2018), 6.

[10] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. 2004. DART: a toolkit for rapid design exploration of augmented reality experiences. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 197–206.

[11] Nicolas Loy Rodas and Nicolas Padoy. 2014. 3D global estimation and augmented reality visualization of intra-operative X-ray dose. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 415–422.

[12] Marc Ericson C Santos, Angie Chen, Takafumi Taketomi, Goshiro Yamamoto, Jun Miyazaki, and Hirokazu Kato. 2014. Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on learning technologies* 7, 1 (2014), 38–56.

[13] Sebastian Schuster, Ranjay Krishna, Angel Chang, Li Fei-Fei, and Christopher D. Manning. 2015. Generating Semantically Precise Scene Graphs from Textual Descriptions for Improved Image Retrieval. In *Workshop on Vision and Language (VL15)*. Association for Computational Linguistics, Lisbon, Portugal.

[14] Hartmut Seichter, Julian Looser, and Mark Billinghurst. 2008. ComposAR: An Intuitive Tool for Authoring AR Applications. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR '08)*. IEEE Computer Society, Washington, DC, USA, 177–178. DOI: http://dx.doi.org/10.1109/ISMAR.2008.4637354

[15] Rachel Swatman. 2016. Pokémon Go catches five new world records. *línea]: http://www. guinnessworldrecords. com/news/2016/8/pokemon-go-catches-five-world-records-439327 (consultado el 20 de diciembre de 2016)* (2016).

[16] Fabio Zünd, Mattia Ryffel, Stéphane Magnenat, Alessia Marra, Maurizio Nitti, Mubbasir Kapadia, Gioacchino Noris, Kenny Mitchell, Markus Gross, and Robert W Sumner. 2015. Augmented creativity: Bridging the real and virtual worlds to enhance creative play. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications*. ACM, 21.