Interactive Augmented Reality Storytelling Guided by Scene Semantics (Supplementary Material)

CHANGYANG LI, George Mason University, USA
WANWAN LI, George Mason University and University of South Florida, USA
HAIKUN HUANG, George Mason University, USA
LAP-FAI YU, George Mason University, USA

CCS Concepts: • **Computing methodologies → Graphics systems and interfaces**; **Augmented reality**.

Additional Key Words and Phrases: augmented reality, storytelling, character animation

ACM Reference Format:
Changyang Li, WanWan Li, Haikun Huang, and Lap-Fai Yu. 2022. Interactive Augmented Reality Storytelling Guided by Scene Semantics (Supplementary Material). *ACM Trans. Graph.* 41, 4, Article 91 (July 2022), 7 pages. https://doi.org/10.1145/3528223.3530061

1 COLLECTING INDOOR ACTIVITY PRIORS

Instead of capturing images or videos from the real world and retrieving activity priors from them like [Savva et al. 2014, 2016], we used a more direct method to collect indoor activity priors by asking specific people room and activity-related preferences. Concretely, we asked people, for a given activity in a type of room, if the furniture objects as shown in the scene are available, which option they would prefer. For example, we asked people, for the activity *watch TV* in a *living room*, if they preferred to sit on a *sofa* or a *chair*. Comprehensive room-activity-option conditions in our survey are shown in Figure 3. We also provide reference pictures to depict such scenarios. Some examples are shown in Figure 4. Note that those pictures only contain necessary characters and objects in a solid color background, and are used for illustrating the abstract contexts only. Therefore, we informed the users that the pictures did not refer to specific scenarios. By doing so, we tried to avoid introducing potential scene-related (e.g. decoration style) biases such that people's judgements were not affected by such biases.

Our survey was sent out on Amazon Mechanical Turk. In total, we collected data from 200 participants, each providing answers to all 33 questions. So we collected a dataset containing 6,600 answers in total. Based on this dataset, we calculated the priors for choosing different possible cases (e.g., object to sit on when watching TV) given activity-room conditions, and built spatial graph branches on event graphs. In an event that depicts activities of multiple characters, we derive different possible cases for each of them using the activity priors, and thus can create spatial graph variations corresponding to combinations of all those cases. Note that, for some straightforward scenarios, we also include some deterministic rules in addition to the collected priors. For example, for activities in a *kitchen*, we define that the activity *cook* must be associated with a *stove* and the activity *wash dishes* must be associated with a *sink*.

In our experiments, we generalize the usage of activity priors in similar scenarios. For example, we use the collected priors of a *conference room* for scenarios including *meeting area* and *discussion area* in stories described in Table 2.

2 ADDITIONAL EXPERIMENTS

2.1 Augmenting Layouts for AR Stories

Our approach assumes that the physical furniture and appliances described in an input story exist in the real scene. However, in case certain objects are missing, we could manually complement those essential objects using virtual 3D models. Different from virtual items in our framework, which are small and movable, the placement of such objects is usually constrained by the room layouts and structures. Instead of augmenting those objects manually, like what we do in the previous experiment, we extend our approach with an automatic furniture placement method [Yu et al. 2011] to optimize the placement of furniture and appliance objects. This method constrains object placements with pre-defined layout-related cost functions, thus is highly compatible with our approach.

To augment a general scene with specific furniture objects missing, we first traverse the event graph and check what furniture objects should be augmented. Following that, we apply a layout optimization, which takes the existing furniture objects into consideration while setting them as static. Our AR storytelling approach can then be employed to generate animations in the synthesized room layout (containing both real and added virtual objects).

We conducted an experiment to demonstrate such an extension. Figure 1 shows the setup and results. In this experiment, the input real scene only contained a desk, a chair and a standing lamp. We show examples of augmenting the scene to fit two events, one where the AR player and two characters watched TV; and another where the player and characters were chatting with each other. Missing but required objects, including a TV and some objects that provided two sit slots (complemented as a sofa) in the *watch TV* event, and some objects that provided two seat slots (complemented as two chairs) in the *chat* event, were automatically placed. We also included some decorations for the Christmas theme in the automatic layout augmenting process, including Christmas trees, carpets, fireplaces and gift boxes. In our results, the AR player sat

Authors' addresses: Changyang Li, George Mason University, USA, cli25@gmu.edu; WanWan Li, George Mason University and University of South Florida, USA, wli17@gmu.edu; Haikun Huang, George Mason University, USA, hhuang25@gmu.edu; Lap-Fai Yu, George Mason University, USA, craigyu@gmu.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2022 Copyright held by the owner/author(s).
0730-0301/2022/7-ART91
https://doi.org/10.1145/3528223.3530061

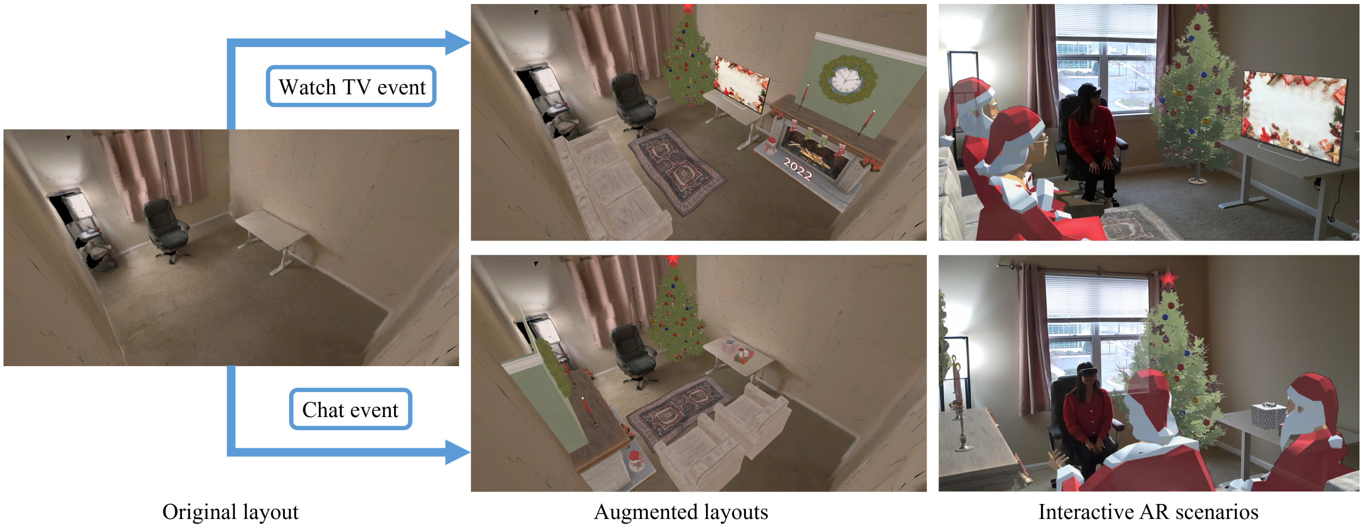

Fig. 1. Augmenting a layout with a virtual TV and couch (left) and virtual chairs (right) for AR storytelling. Other virtual decoration objects are also automatically populated to fit with the theme.

on the physical chair and the virtual characters sat on the virtual sofa/chairs to complete the event according to the AR story plot.

2.2 Retargeting AR Stories to Different Scenes

We also evaluate our approach's ability to retarget the same story to different scenes of a specific type. Particularly, we choose four apartment scenes from the Matterport3D dataset [Chang et al. 2017] to play the *apartment* story same as the previous experiments. Note that in this experiment, sampled stories are fully virtual since we did not have access to the corresponding physical environments for AR experiences. Figure 2 shows the results. Since some of the 3D scene scans do not contain a coffee maker, which are necessary in our story, we manually placed a virtual coffee maker model to complement those scenes. The results validate that our approach can retarget the same story to different scenes, which is a key strength of our approach. Such a functionality lets players instantiate AR stories in their own environments and participate via AR devices.

3 USER STUDY

3.1 Visualization of Story Trials

For all story trials either manually created by designers or automatically assembled by our approach, we show visualizations of selected events, where the AR player and multiple virtual characters coexisted, in Figure 5. For all conditions *I*, *II* and *III*, while the selections of character slots for the AR player were fixed throughout the stories, other virtual characters were posed in diverse ways. Note that, unlike other results in the main paper that approximated the AR player's poses for the blue avatar, we only placed the avatar idly at the character slots that the designers assigned. The reason is because Figure 5 does not show replays of real story trajectories, but the original human designs together with the corresponding automatically assembled stories.

Table 1. Accumulated trajectory lengths (in meters) for characters [m], [a], and [b] in the *office* story. I^o, II^o and III^o were synthesized by our approach. I^m, II^m and III^m were created by designers.

	I^m	I^o	II^m	II^o	III^m	III^o
character [m]	57.27	**55.88**	58.97	**52.75**	58.97	**55.88**
character [a]	59.94	**52.86**	54.66	**53.14**	56.20	**53.51**
character [b]	57.78	**55.96**	**54.18**	54.80	52.49	**51.53**

3.2 Additional Evaluation Metrics

In addition to metrics about plausibility concerning individual activities, group activities and item placements, we also asked the following questions and collected answers in a 5-point Likert scale: (1) How understandable the AR story is; (2) How enjoyable the AR storytelling is; (3) How is the AR storytelling is; (4) How interactive the AR storytelling is; (5) How comfortable the AR storytelling is; (6) How convenient the system is. Numerical results in Table 3 indicate that there were no significant differences between stories assembled by our approach and those created by human designers with respect to these additional metrics.

We also evaluated the accumulated trajectory length for each virtual character. Note that since the characters' positions are determined in all story trials, the trajectory of each character is estimated by calculating shortest paths between event transitions. According to the results in Table 1, stories created by our approach reduce the overall accumulated trajectory lengths for all characters, hence reducing the wait time for the AR player. The largest difference is as much as 7.08m for character [a] in I^m and I^o.

REFERENCES

Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2017. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158* (2017).

Fig. 2. Retargeting the *apartment* story (refer to Table 2) to four different apartment scenes from the Matterport3D dataset [Chang et al. 2017]. Note that results in this experiment are fully virtual without player's participation in AR, thus [P] in the story is replaced by another virtual character [c].

Manolis Savva, Angel X Chang, Pat Hanrahan, Matthew Fisher, and Matthias Nießner. 2014. SceneGrok: Inferring action maps in 3D environments. *ACM transactions on graphics (TOG)* 33, 6 (2014), 1–10.

Manolis Savva, Angel X Chang, Pat Hanrahan, Matthew Fisher, and Matthias Nießner. 2016. Pigraphs: learning interaction snapshots from observations. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–12.

Lap Fai Yu, Sai Kit Yeung, Chi Keung Tang, Demetri Terzopoulos, Tony F Chan, and Stanley J Osher. 2011. Make it home: automatic optimization of furniture arrangement. *ACM Transactions on Graphics (TOG)-Proceedings of ACM SIGGRAPH 2011, v. 30,(4), July 2011, article no. 86* 30, 4 (2011).

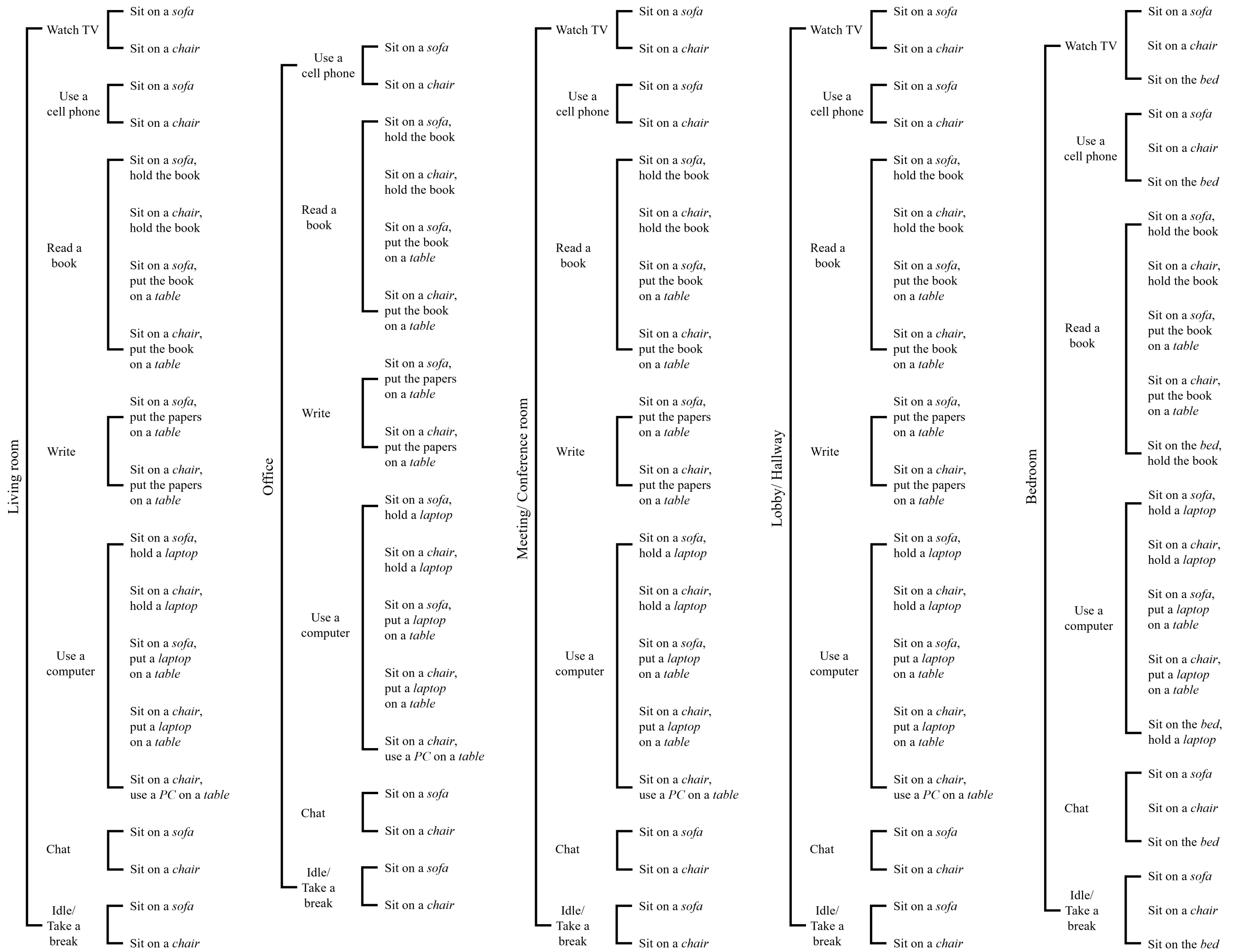

Fig. 3. All of the room-activity-option conditions included in our survey for collecting indoor activity priors. For each hierarchy, the first-level includes the room, the second-level includes activities in that room, and the third-level includes options given the room-activity condition. Possible furniture objects and items are shown in italic.

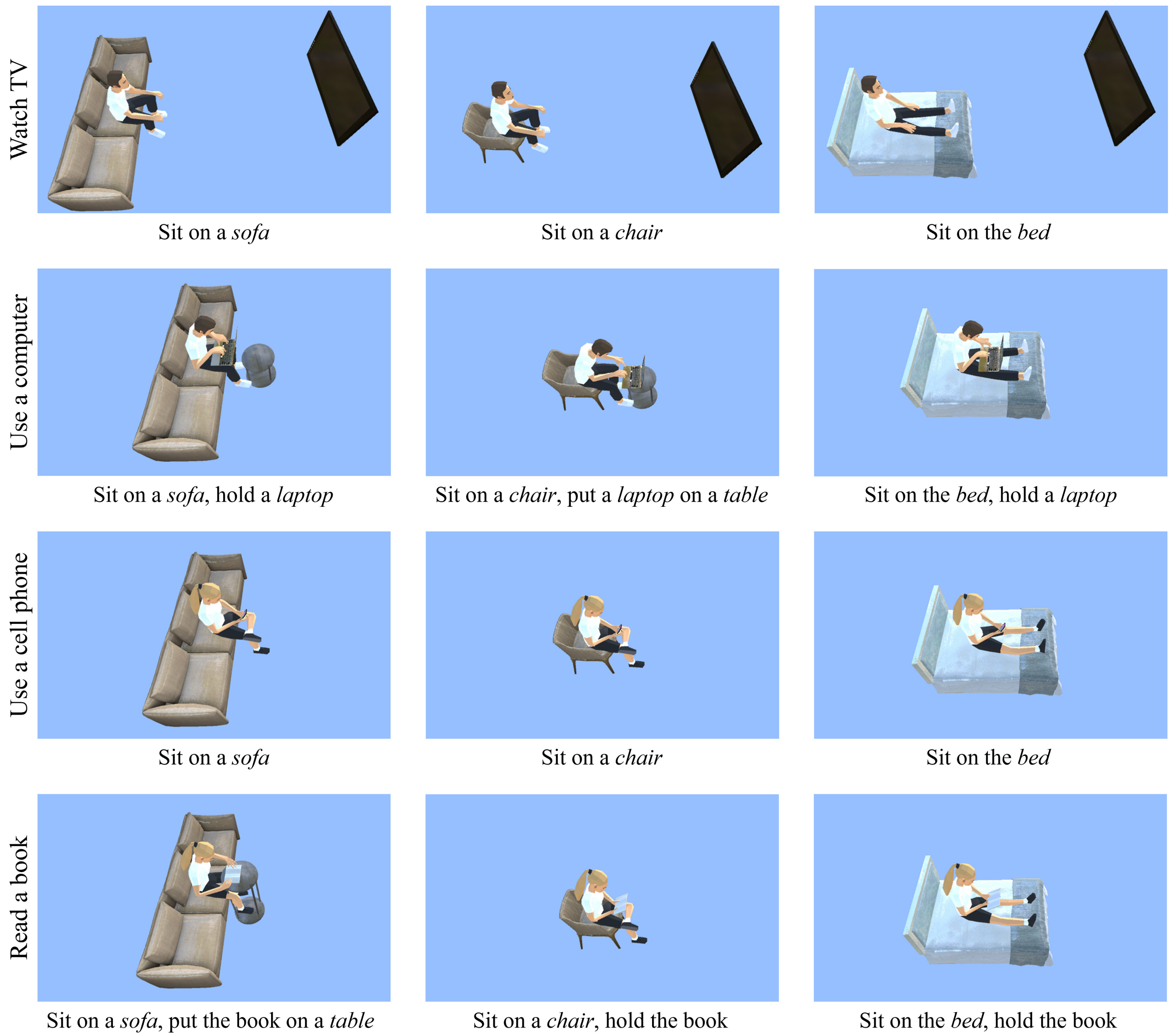

Fig. 4. Examples of reference pictures in our survey for collecting the indoor activity priors. The pictures are for illustration and only depict abstract contexts in given scenarios.

Table 2. Descriptions of events of the four stories in different types of indoor scenes. [P] denotes the AR player, [m] denotes the manager, [prof] denotes the professor, and others are general characters without specific roles.

Apartment	Office	Teaching building	Makerspace
Event 1: [P] uses a cell phone in the dining area, [a] cooks in the kitchen, [b] uses a computer in the living room	Event 1: [m] hosts a meeting in the meeting area, [P,a,b] listen to [m]	Event 1: [P] reads a book, [a] texts [prof] to make an appointment, in the meeting area	Event 1: [P,a,b] discuss in the hallway
Event 2: [P,a] carry food in the kitchen	Event 2: [P,a,b] discuss in the meeting area, [m] uses a computer in the office	Event 2: [P,a] buy drinks at vending machines in the hallway	Event 2: [P] uploads printing tasks in the media room, [a,b] chat in the hallway
Event 3: [P,a] deliver food in the dining area	Event 3: [P,a,b] use computers in the workspace	Event 3: [P,a] chat in the hallway	Event 3: [P] prints design drawings in the hallway
Event 4: [P,a,b] eat in the dining area	Event 4: [P] prints files in the printing room	Event 4: [P,prof,a] discuss in the discussion area	Event 4: [P] presents designs, [a,b] listen to [P], in the media room
Event 5: [P,b] carry dishes in the dining area, [a] makes coffee in the kitchen	Event 5: [P,m] discuss in the manager's office	Event 5: [P,a] discuss in the discussion area, [prof] leaves	Event 5: [P,a,b] edit digital designs on computers in the media room
Event 6: [P,b] wash dishes in the kitchen	Event 6: [P] presents in the manager's office, [m] listens to [P], [a,b] carry snacks in the pantry	Event 6: [P] prints files in the hallway, [a] sets up a computer in the meeting area	Event 6: [b] prints the 3D model using a 3D printer in the makerspace
Event 7: [P,b] carry coffee in the kitchen	Event 7: [a,b] deliver snacks in the meeting area	Event 7: [P] delivers files to [a] in the meeting area	Event 7: [P,a] watch the 3D printing in the makerspace
Event 8: [P,a,b] chat in the living room	Event 8: [P,m,a,b] chat in the meeting area	Event 8: [P] presents the project to [prof,a,b,c] in the meeting area	Event 8: [P,a,b] chat in the makerspace

Table 3. Quantitative results of participants' ratings on (1) How understandable the AR story is (understandability); (2) How enjoyable the AR storytelling is (enjoyment); (3) How immersive the AR storytelling is (immersiveness); (4) How interactive the AR storytelling is (interactiveness); (5) How comfortable the AR storytelling is (comfortableness); and (6) How convenient the system is (convenience). The average (avg.), standard deviation (std.) and p values are shown. I^o, II^o and III^o were produced by our approach, and I^m, II^m and III^m were created by designers.

	Understandability			Enjoyment			Immersiveness			Interactiveness			Comfortableness			Convenience		
	avg.	std.	p	avg.	std.	p	avg.	std.	p	avg.	std.	p	avg.	std.	p	avg.	std.	p
I^m	4.2	0.75	1.00	4.6	0.49	0.80	3.9	0.70	1.00	3.7	0.90	1.00	4.4	0.66	0.75	4.3	0.78	1.00
I^o	4.2	0.75		4.6	0.66		3.9	0.70		3.7	0.90		4.5	0.67		4.3	0.78	
II^m	4.8	0.40	0.63	4.4	0.80	0.81	4.2	0.75	0.79	3.6	0.80	0.79	4.6	0.49	1.00	4.3	0.46	0.66
II^o	4.7	0.46		4.5	0.92		4.1	0.83		3.7	0.78		4.6	0.49		4.4	0.49	
III^m	4.6	0.66	1.00	4.3	0.90	1.00	4.2	0.87	0.80	3.7	0.90	1.00	4.4	0.80	0.77	4.4	0.80	1.00
III^o	4.6	0.66		4.3	0.90		4.3	0.78		3.7	0.90		4.5	0.67		4.4	0.80	

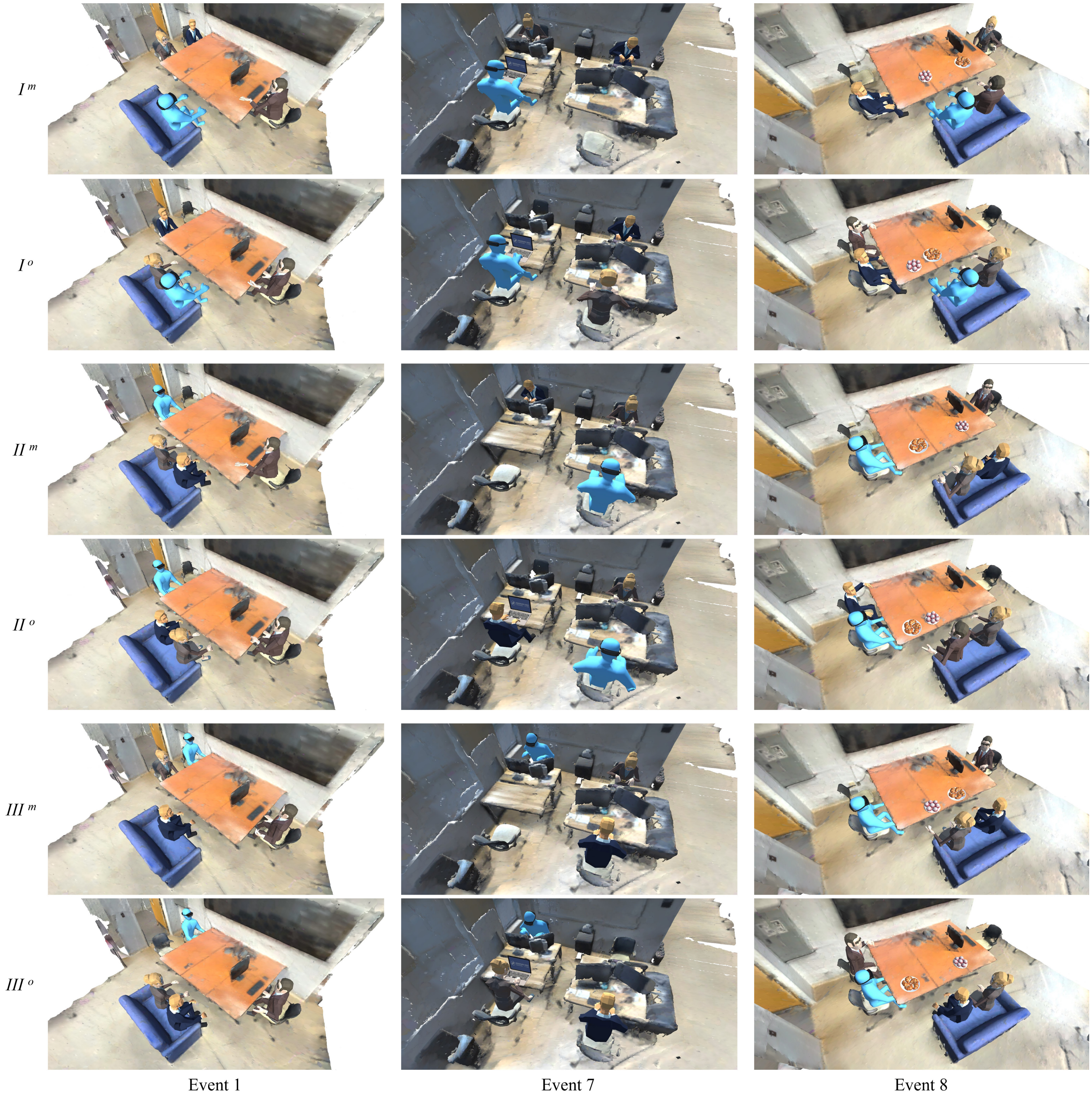

Fig. 5. Visualization of selected events, which are described in Table 2, of all story trials in our user study.