

# Synthesizing Scene-Aware Virtual Reality Teleport Graphs (Supplementary Material)

CHANGYANG LI, George Mason University, USA  
HAIKUN HUANG, George Mason University, USA  
JYH-MING LIEN, George Mason University, USA  
LAP-FAI YU, George Mason University, USA

## ACM Reference Format:

Changyang Li, Haikun Huang, Jyh-Ming Lien, and Lap-Fai Yu. 2021. Synthesizing Scene-Aware Virtual Reality Teleport Graphs (Supplementary Material). *ACM Trans. Graph.* 40, 6, Article 229 (December 2021), 10 pages. <https://doi.org/10.1145/3478513.3480478>

## 1 SCENE PERCEPTION PREDICTION

### 1.1 Comparing SP Graphs with HOI Graphs

While our definition of SP graphs are inspired by the notion of HOI graphs, we discuss two major differences in this section.

First, in the HOI detection task, an image shows how people physically interact with objects from a third-person view. In an HOI graph, depending on the number of detected objects, there could be multiple human nodes and object nodes; and edges could be densely generated such that the whole graph is a complete graph. In contrast, in an SP graph, a panoramic image corresponds to the first-person view of an observer who does not explicitly show up in the image. Instead of modeling physical human-object contacts, we model the observer's visual interaction with objects by perceiving his surroundings, which fits with our notion of scene perception. Therefore, edges are only generated for human-object pairs.

Second, HOI graphs relies on visual features (e.g., visual representations of human gestures and object shapes) extracted from RGB images using convolutional neural networks to classify the types of interactive activities between people and objects. While visual features could capture low-level visual details, in SP graphs, we propose to use high-level semantic features, which allow our method to work in more general cases. In our implementation, we use the category features and depth features for general and robust representations of SP graphs.

### 1.2 Flattening SP Graphs into Sequences

We discuss the equivalence of an SP graph  $G$  and its flattened sequence  $s(G)$ . First, to transform  $G$  into  $s(G)$ , while the human node

---

Authors' addresses: Changyang Li, George Mason University, USA, [cli25@gmu.edu](mailto:cli25@gmu.edu); Haikun Huang, George Mason University, USA, [hhuang25@gmu.edu](mailto:hhuang25@gmu.edu); Jyh-Ming Lien, George Mason University, USA, [jmlien@gmu.edu](mailto:jmlien@gmu.edu); Lap-Fai Yu, George Mason University, USA, [craigyu@gmu.edu](mailto:craigyu@gmu.edu).

---

Permission to make digital or hard copies of all or part 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 components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2021 Association for Computing Machinery.

0730-0301/2021/12-ART229 \$15.00

<https://doi.org/10.1145/3478513.3480478>

and edges are ignored, features of all other nodes corresponding to the detected objects are reserved. In other words,  $s(G)$  can be simply regarded as an array of object nodes sorted by the features. Second, transforming  $s(G)$  back to  $G$  only requires adding the human node back and building edges between every human-object node pairs. Since  $G$  and  $s(G)$  are equivalent, the process of flattening  $G$  does not lose any information.

### 1.3 Pilot Study for Rational Choice Assumption

We ran a two-alternative forced choice (2AFC) pilot study to briefly validate the rational choice assumption that the positions in the scenes where the authors of the Gibson Environment [Xia et al. 2018] chose to capture equirectangular panoramic images are desirable. In our study, participants visited a pair of virtual positions in an immersive manner through VR devices, and they were asked to choose the position with a more desirable view based on the surrounding objects (e.g., furniture, decoration). In each comparison pair, one position is randomly selected from the assumed optimal examples, and another position is randomly sampled from the same scene of the picked optimal example. We tried to avoid cases where the views of the random samples and the optimal examples are too similar to each other (i.e., where the random samples are close to the optimal examples). In order to guarantee their differences, we discard random samples whose similarity value to the optimal set is higher than 0.8 (calculated by the graph kernel  $K(\cdot)$ ).

We recruited 30 participants in the pilot study. Each participant was asked to make 20 comparisons. The null hypothesis  $H_0$  was that users perceive no significant differences in the desirability of panoramic views at optimal examples and random samples. The alternative hypothesis  $H_1$  was that users did perceive significant differences. According to the results, 440 out of all the 600 selections are the optimal examples.

We first adopted the Chi-square nonparametric analysis technique. The obtained optimal/random frequencies were compared to an expected 300/300 result to ascertain whether this difference is significant. The analysis suggests that participants favor optimal examples with a 99% a confidence level ( $p = 2.93e-30$ ). Second, we adopted a Bayesian analysis to determine whether the number of participants who selected the assumed optimal examples was what would be expected by chance, or if there was a preference pattern. We assumed that the participant had a probability  $P$  of picking the assumed optima, and used a binomial distribution to model the results. We computed the odds  $O_{10}$  on  $H_1$  over  $H_0$ , and the results ( $O_{10} = 1.45e28$ ) showed "very strong evidence" favoring  $H_1$  according to [Rouder et al. 2009].

## 2 TELEPORT GRAPH OPTIMIZATION

### 2.1 Balancing Action Acceptance Probability

In practice, the teleport graphs are fairly sparse. The selected positions in a teleport graph only account for a small portion of all candidate positions in a scene and the ratio  $\frac{|\mathcal{V}|}{|\mathcal{P}|}$  is usually low. As a result, in most cases the term  $\frac{|\mathcal{P}-\mathcal{V}|}{|\mathcal{V}^*|}$  is large, thus  $A_a$  is close to 1 and  $A_b$  is close to 0. To address this issue, we regularize the number of positions to stay close to an expected range. Intuitively, the number of positions  $|\mathcal{V}|$  in a teleport graph should increase with the area of free space, which can be approximated by the number of candidate positions  $|\mathcal{P}|$  as the candidate positions are densely sampled. Recall that as we extract data from the Gibson Environment [Xia et al. 2018], we fit a polynomial function  $\lambda(|\mathcal{P}|)$  to estimate the number of selected positions based on the number of candidate positions  $|\mathcal{P}|$ . Therefore we can balance  $A_a$  and  $A_b$  by penalizing the solution when the number of positions exceeds  $\lambda(|\mathcal{P}|)$ , which is encoded as the regularization cost in Section 5.2 of our main paper.

## 3 RESULTS

### 3.1 Scene Perception Score Prediction

To complement results in the main paper, we show the visualization of SP score maps in Figure 3 obtained using our SP predictor (GAT) and the baseline (CNN) for *Scene 1* to *Scene 4*, which are used in the main paper experiments. By comparing the score maps produced by the two methods, we observe that the SP predictor yields results similar to the ground truth. In particular, in some local areas (e.g., *Scene 1*'s bedroom, *Scene 2*'s bathroom, *Scene 3*'s master bedroom, *Scene 4*'s corridor), the score variation patterns are well predicted. In comparison, the score maps by the baseline appear overly smoothed; sharp SP score variations in some local areas are not well captured.

We also show quantitative results of SP score prediction for *Scene 5* to *Scene 16* in the 16-scene dataset in Table 1, which further validate that our GAT-based SP predictor is more reliable for handling diverse cases.

### 3.2 Additional Teleport Graph Synthesis for Indoor Scenes

We demonstrate additional teleport graph syntheses for *Scene 5* to *Scene 16*. Eight of the results in Figure 4 and Figure 5 are in synthetic scenes, and the other four in Figure 6 are in 3D reconstructed scenes extracted from the Matterport3D dataset [Chang et al. 2017]. The results indicate that our approach can be widely adopted in various virtual indoor environments of different sizes, structures and styles.

### 3.3 Teleport Graphs with and without Scene Perception

When synthesizing teleport graphs, the SP score cost serves as a soft constraint that encourages graphs to include positions with higher SP scores, which indicate higher visual desirability. According to our experiment observations, while the connectivity constraint is usually more easily satisfied, an intuitive consequence of omitting the SP score consideration is that it stimulates teleport graphs to cover a larger area of the virtual environment and avoid redundant coverage, no matter how the views look like.

We make two comparisons in Figure 7 to qualitatively show how views at selected positions are influenced without the SP score

Table 1. Quantitative results of SP score prediction for our SP predictor (GAT) and the baseline (CNN) in the experiment indoor scenes. For AE  $p$ -value of *Ours/Baseline*, bold indicate significant differences with 95% confidence.

Scene	Pos.#	RMSE		AE Variance		AE
		Ours	Baseline	Ours	Baseline	$p$ -value
5	2,279	<b>0.062</b>	0.099	0.0050	<b>0.0038</b>	<b>2.39e-61</b>
6	2,651	<b>0.074</b>	0.149	<b>0.0102</b>	0.0163	<b>9.31e-26</b>
7	3,409	<b>0.064</b>	0.117	<b>0.0079</b>	0.0096	<b>2.31e-102</b>
8	2,226	<b>0.062</b>	0.086	<b>0.0020</b>	0.0029	<b>1.86e-21</b>
9	2,619	<b>0.049</b>	0.072	<b>0.0017</b>	0.0021	<b>5.04e-11</b>
10	2,715	<b>0.057</b>	0.111	<b>0.0039</b>	0.0185	<b>5.75e-43</b>
11	3,431	<b>0.057</b>	0.106	<b>0.0022</b>	0.0050	<b>6.71e-55</b>
12	2,223	<b>0.068</b>	0.096	<b>0.0014</b>	0.0019	<b>1.61e-2</b>
13	9,985	<b>0.056</b>	0.085	<b>0.0024</b>	0.0032	<b>1.30e-9</b>
14	20,595	<b>0.078</b>	0.108	<b>0.0029</b>	0.0051	<b>2.60e-207</b>
15	4,670	<b>0.057</b>	0.107	<b>0.0031</b>	0.0051	<b>7.24e-24</b>
16	13,735	<b>0.072</b>	0.118	<b>0.0030</b>	0.0065	<b>3.13e-196</b>

consideration. Similar to results in Figure 7 in the main paper, for comparison and illustration purposes, we use results synthesized with all costs as the initialization and omit the SP score cost to synthesize new results. In each comparison group, we select a node with a small move but a large SP score variation to show the change of its first-person view. In *Scene 1*, from the all costs result to the no SP score cost result, the selected node moves away from the wall above to cover a larger area. While connectivity to the top-left room is still maintained, the view at the selected node is severely occluded, and we have difficulty recognizing the table from this view. In another case in *Scene 2*, for the no SP score cost result, one node is set free from the rightmost room to cover another area inside the scene. In order to maintain connectivity, the selected node is moved to an extreme position very close to the wall. Similarly, without considering visual information, the view at the selected node is severely occluded.

The results suggest that a tradeoff between scene perception and other constraints is necessary, and the SP cost consideration is indispensable for synthesizing visually desirable teleport graphs.

### 3.4 Other Environments

*Canyon.* We apply our approach to synthesize a teleport graph for another outdoor environment, a canyon, as shown in Figure 8. Akin to the natural park example in the main paper, we define a teleport position's SP score based on whether a flag is visible from this position within 100m. The synthesized teleport graph shows that most (88.89%) of the teleport positions have a visible flag nearby even though the user's visibility could be occluded by rocks often. Meanwhile the teleport graph maintains connectivity in this complex environment.

*Outlet.* We also applied our approach for a multi-floor outlet scene. Unlike the outdoor environments, we modified the definition of SP score cost in this example to generate a teleport graph that fits with a user's shopping preferences in visiting a virtual outlet in VR. Akin to some real-world shopping experiences, we assume that the user is only interested in certain categories of shops. The

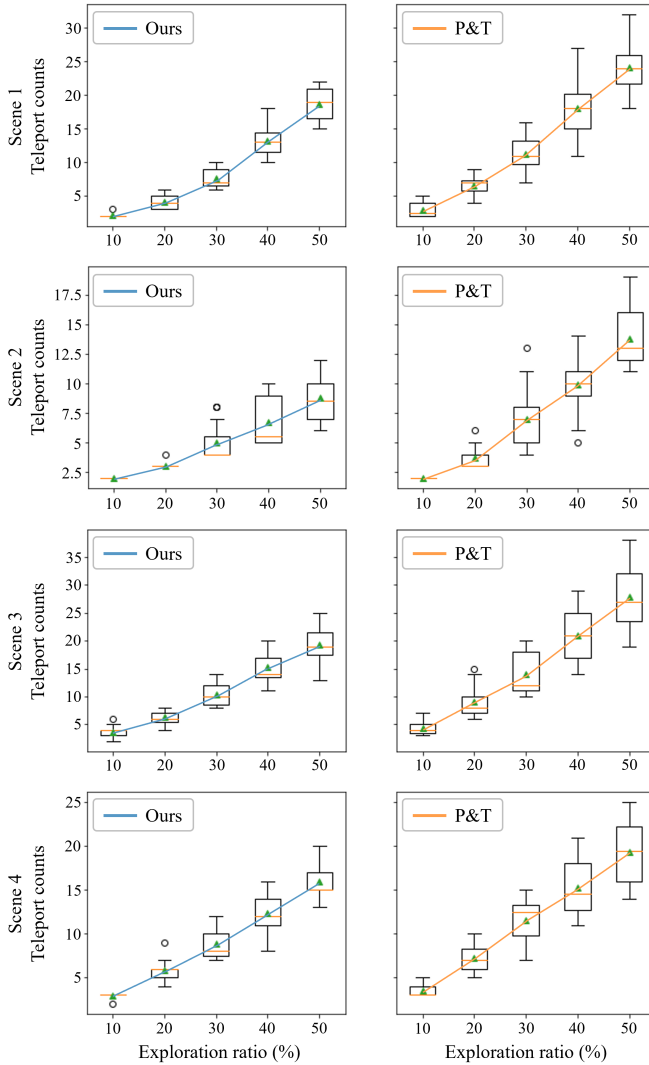


Fig. 1. Boxplots of the number of teleport counts versus exploration ratio for all conditions. The dots, triangles and orange lines indicate outliers, means, and medians, respectively. The whiskers indicate the min/max values. The box indicate the first/third quartiles.

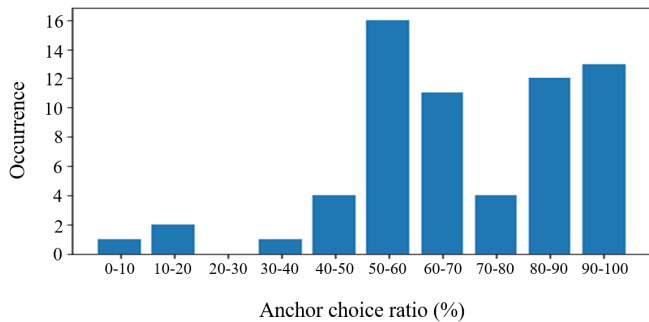


Fig. 2. Distribution of participants' anchor choice ratios.

SP score here is defined based on the proximity and visibility of the nearest store of the user's interested category.

If a teleport position is near the front door (i.e. within 5m) of a store of the user's interested category, the position has an SP score of 1.0. If a teleport position allows the user to see a store of the user's interested category within a certain distance (i.e. within 50m), the position has an SP score of 0.5. Otherwise, the position has an SP score of 0.0.

Figure 9 shows a teleport graph synthesized with this SP score definition. We set shoes stores as the stores of interest. As this scene consists of multiple floors, we fix nodes at the ends of the stairs at each floor to keep the whole graph connected. According to the result, with the SP score cost, nodes are more densely sampled near the shoe stores. In contrast, without using the SP score cost, the nodes of the synthesized teleport graph are more evenly distributed in the free space.

## 4 USER STUDY

### 4.1 Navigation Results

We demonstrate more details of the navigation results by showing boxplots for all conditions in Figure 1. While the results support our conclusion that *ours* generally reduces the teleport counts, the fact that the results of *ours* show less deviation overall also suggests that participants using *ours* had more stable teleport operations when navigating in virtual environments.

### 4.2 Anchor Choice Ratio

Since participants using *ours* also had the freedom to manually choose destinations other than the anchor positions of the graph, we cannot draw conclusions about *ours* from records where the anchor choice ratios are low. We examine the distribution of participants' anchor choice ratios and present the results in Figure 2. According to the results, participants show generally high anchor choice ratios when using *ours*.

## 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. *International Conference on 3D Vision (3DV)* (2017).
- Jeffrey N Rouder, Paul L Speckman, Dongchu Sun, Richard D Morey, and Geoffrey Iverson. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review* 16, 2 (2009), 225–237.
- Fei Xia, Amir R. Zamir, Zhi-Yang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. 2018. Gibson env: real-world perception for embodied agents. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

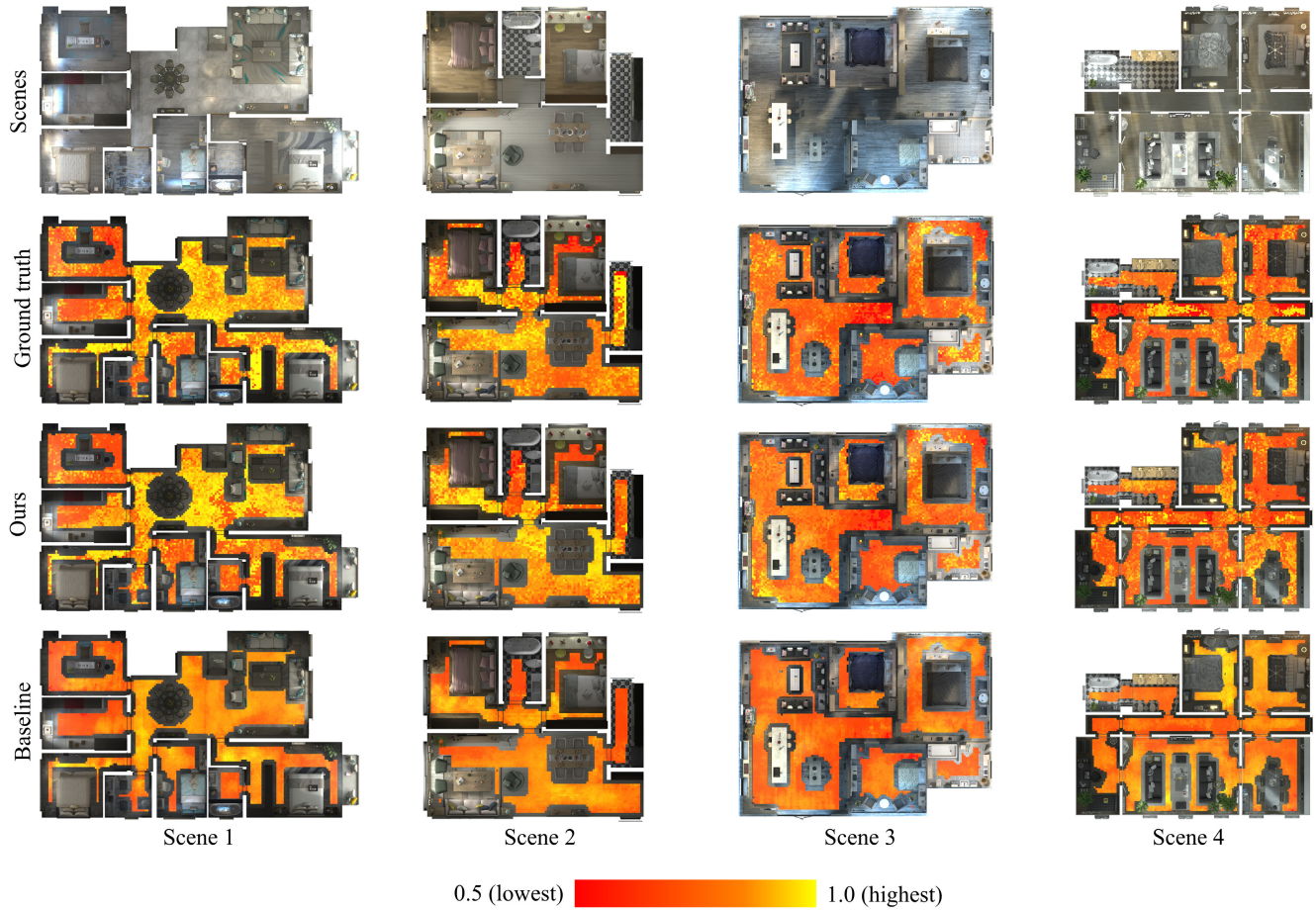


Fig. 3. A comparison of SP score maps produced by our SP predictor (GAT) and the baseline (CNN) for *Scene 1* to *Scene 4*.



Fig. 4. Teleport graph syntheses and predicted SP score maps for *Scene 5* to *Scene 8* (synthetic scenes).



Fig. 5. Teleport graph syntheses and predicted SP score maps for *Scene 9* to *Scene 12* (synthetic scenes).

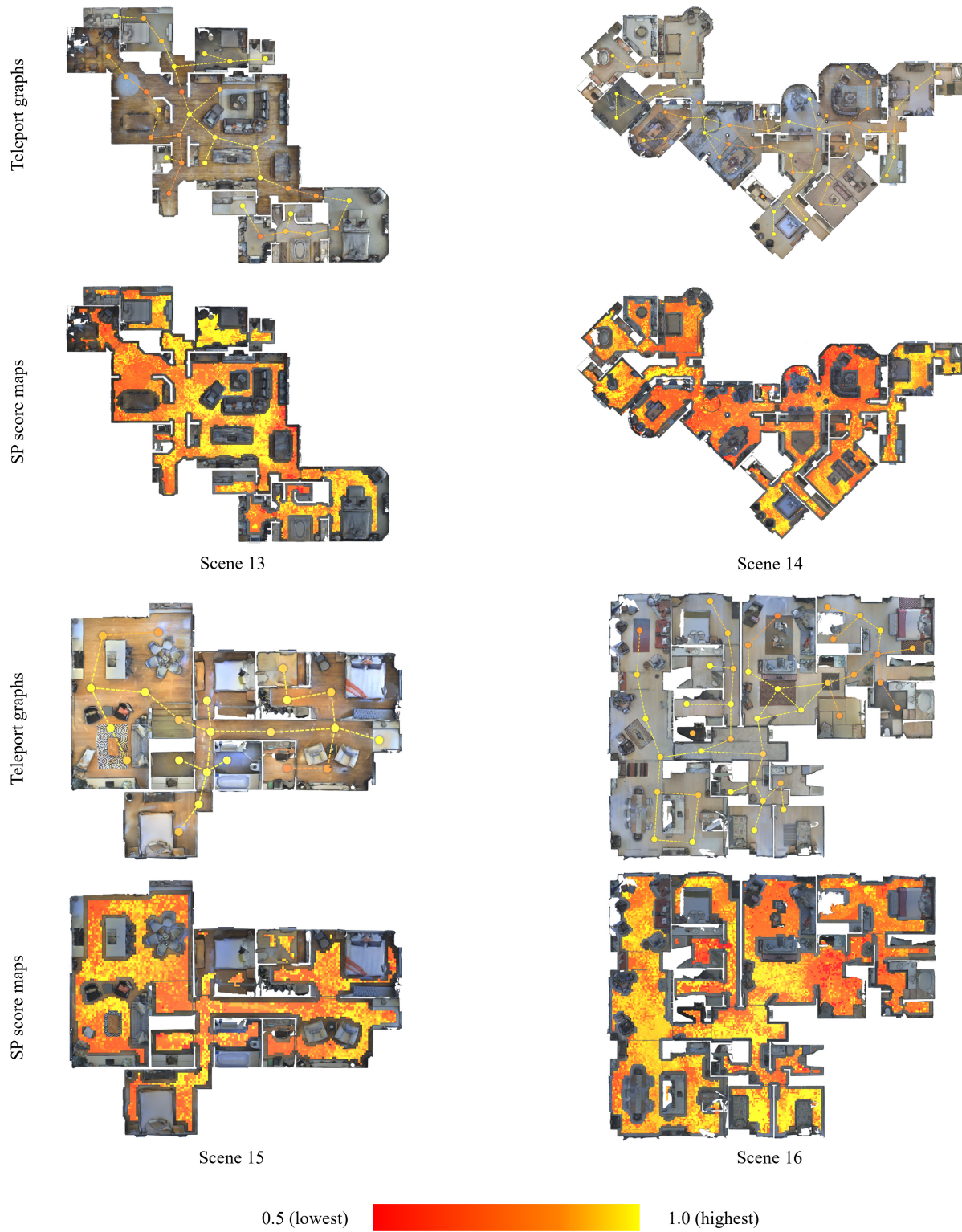


Fig. 6. Teleport graph syntheses and predicted SP score maps for *Scene 13* to *Scene 16* (3D scans from the Matterport3D dataset [Chang et al. 2017]).

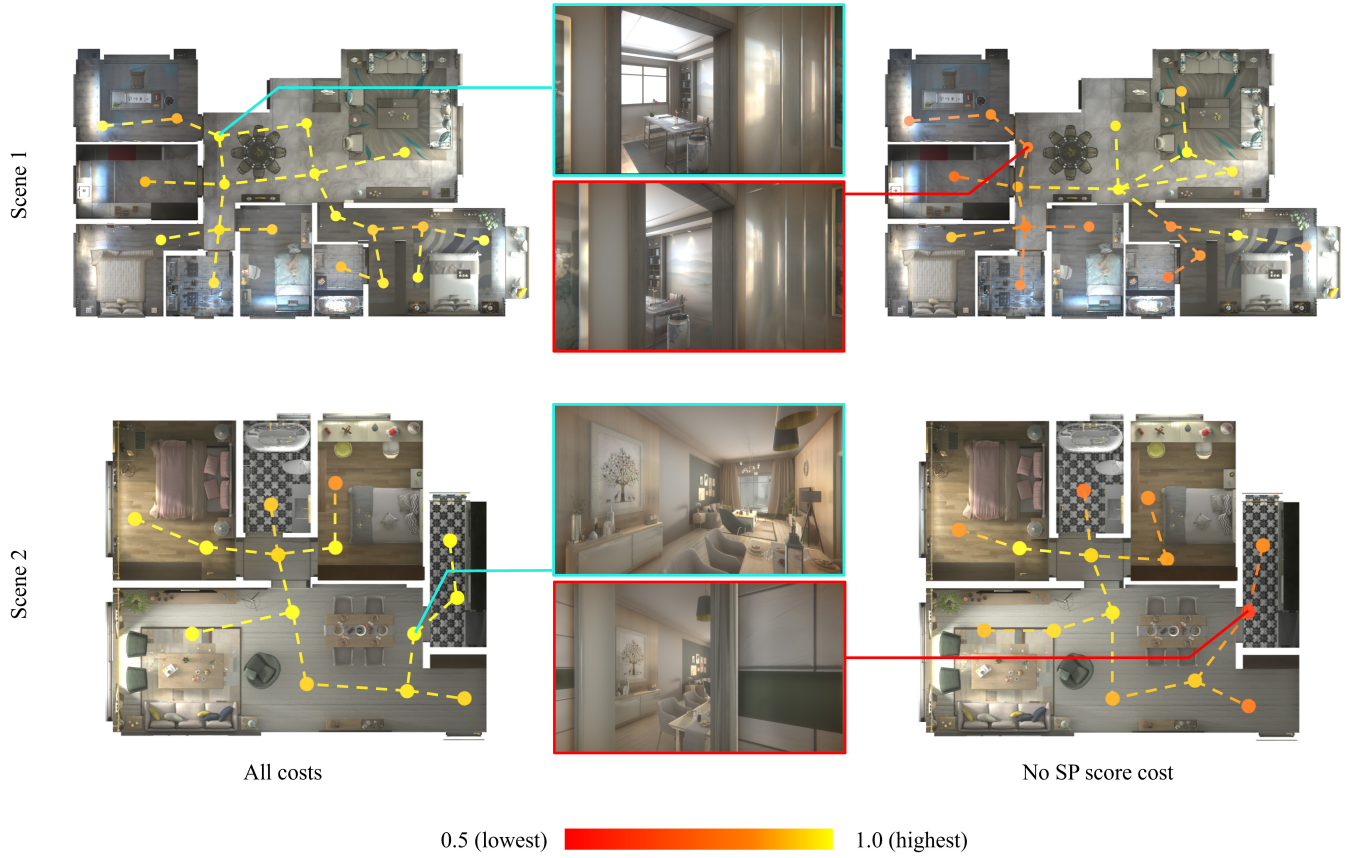
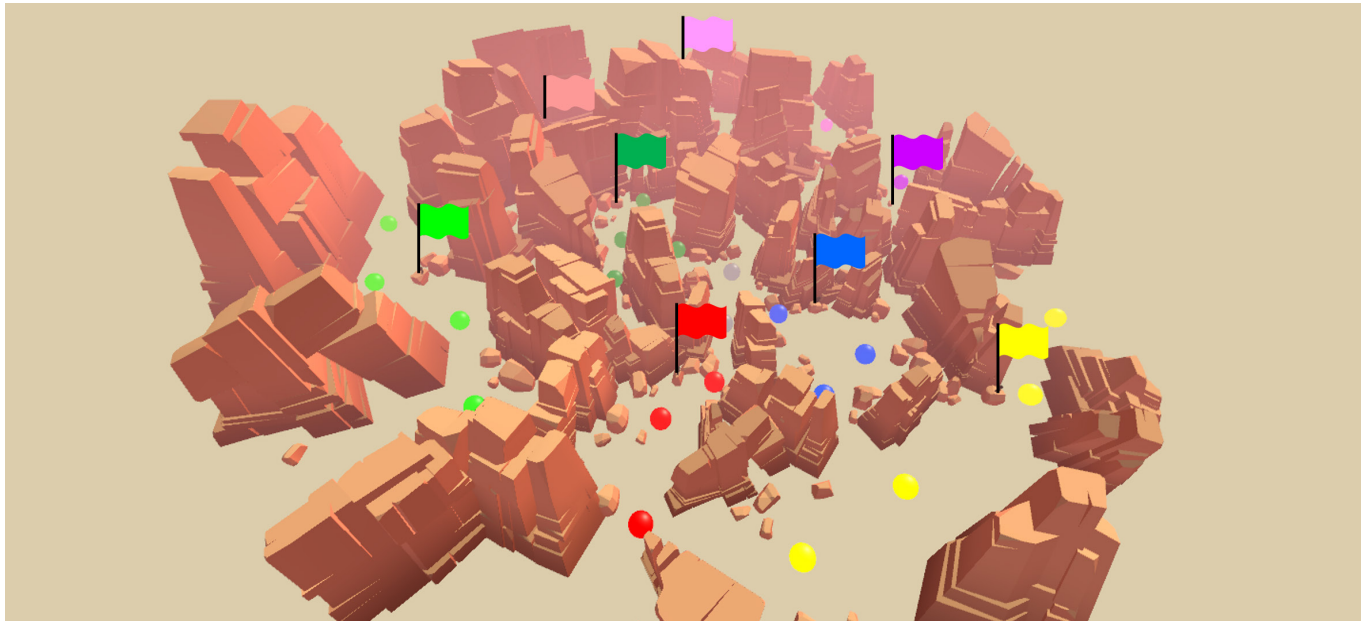
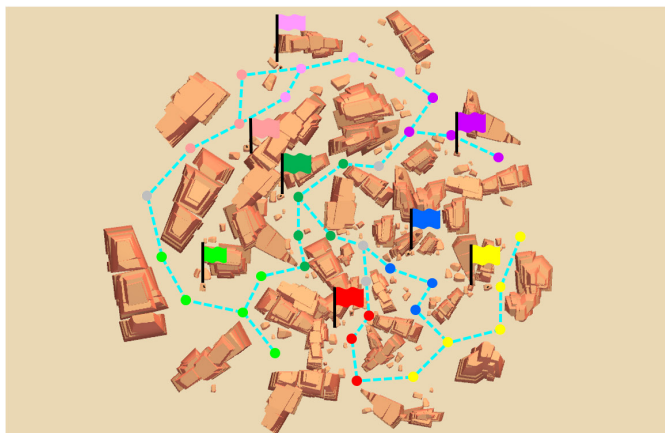


Fig. 7. Qualitative comparisons between synthesized graphs with and without the SP score consideration in *Scene 1* and *Scene 2*. First-person views are shown at selected nodes.

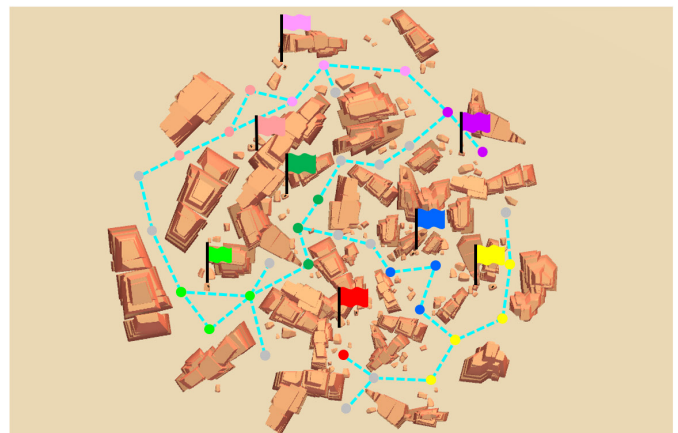




All costs (perspective view)



All costs (top view)



No SP score cost (top view)

Fig. 8. Teleport graphs synthesized for a canyon. Gray nodes denote positions where no flag is visible. For the graph synthesized with all costs, 88.89% of the positions have a visible flag in their proximity. For the graph synthesized without the SP score cost, only 62.86% of the positions have a visible flag.

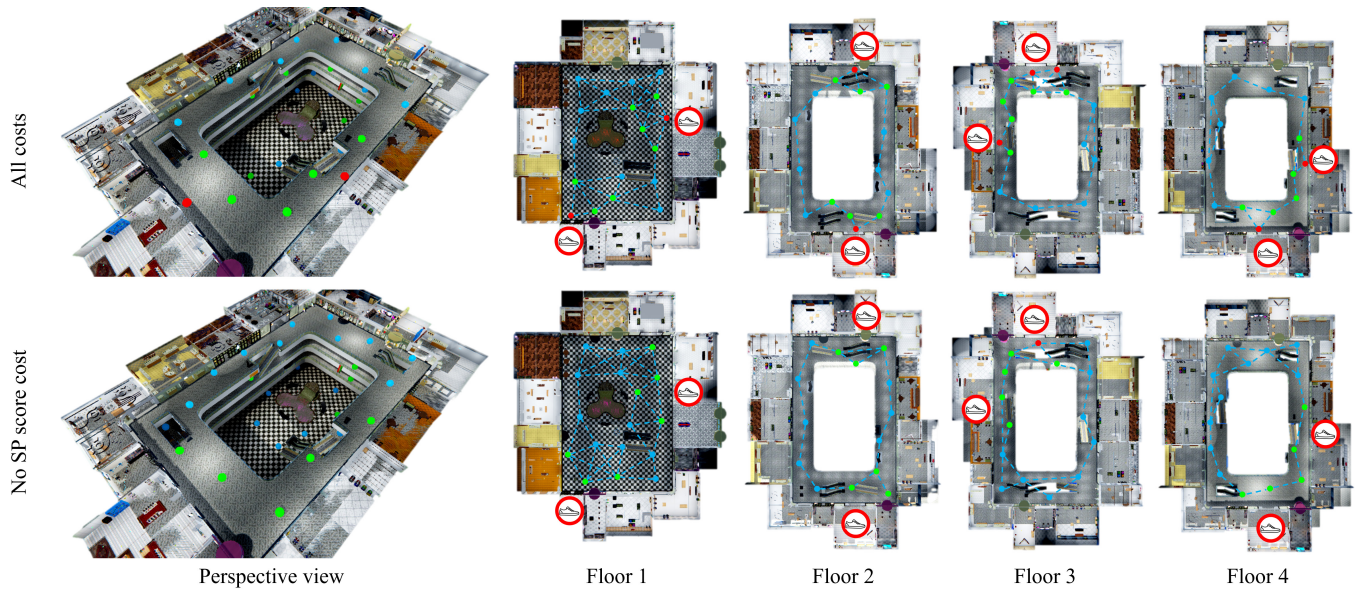


Fig. 9. Teleport graphs synthesized for an outlet scene consisting of four floors. The SP score is defined based on the proximity and visibility of shoe stores. Red nodes refer to positions within 5m from a shoe store. Green nodes refer to positions where a shoe store is visible within 50m. Cyan nodes refer to the remaining positions. With the SP score cost, more nodes in front of a shoe store are sampled to form the teleport graph.