JTS

The Journal of Transdisciplinary Studies, 4(1): 93-110, 2020
https://doi.org/10.22685/JTS.2020.4.1.093

Spatial Similarity Measurement in Dissimilar Indoor Environments Using Semantic Features for Virtual Space Services*

Sukjoo Hong[1], Mi Chang[2], Urvashi Dwivedi[3], Valentin Porcellini[4], Chae-Seok Lee[5], Ji-Hyun Lee[6]

[1]Master Student, Graduate School of Culture Technology, KAIST, Korea (sukjoo.hong@kaist.ac.kr), First Author
[2]Doctoral Student, Graduate School of Culture Technology, KAIST, Korea (rosechang@kaist.ac.kr)
[3]Master Student, Graduate School of Culture Technology, KAIST, Korea (urv892@gmail.com)
[4]Master Student, Dept. of Information Science, Cornell University, Ithaca NY (vp272@cornell.edu)
[5]Postdoctoral researcher, Institute for Information Technology Convergence, KAIST, Korea (quarry@kaist.ac.kr)
[6]Professor, Graduate School of Culture Technology, KAIST, Korea (jihyunlee@kaist.ac.kr), Corresponding Author

ARTICLE INFO

Article history:
Received 24 June 2020
Revised 25 June 2020
Accepted 27 June 2020

Keywords:
Indoor Representation Method,
Semantic Feature,
Space Comparison Criteria,
Spatial Similarity Measurement

ABSTRACT

In order to provide intelligent services in an indoor environment at furniture level, it is important to understand not only the physical attributes but also the semantic features of an interior space. Spatial similarity measurement is a comparison method conducted on top of a hierarchical graph representation to calculate numeric values considering quantified spatial features. Three consecutive tasks are required: indoor space modeling, user study acquiring spatial context, and spatial similarity calculation. To evaluate the effectiveness of the method, a comparison was made with the existing random sampling method. The findings surpassed the random sampling results in four experimental cases. Spatial similarity measurement is therefore capable of analyzing the spatial context with proposed similarity criteria and equations, and it also suggests the probability of local similarity of a functional space.

1. Introduction

In recent years, user experience-based technologies such as AR and VR have been developed, and new forms of telecommunication have emerged to replace traditional video-conferencing (Alexiou et al., 2004; Heun, Kasahara, & Maes, 2013; Rácz & Zilizi, 2018). Telepresence, which is a holographic communication solution that allows remote participants to feel like they are all together in the same room, is one such technology. Participants can interact with each other's avatars in a virtual space, communicate in real-time, and share content (Dreshaj, 2015). To achieve this, it is essential

* This work was supported by Samsung Research Funding Center of Samsung Electronics under Project Number SRFC-IT1701-00.

to provide an appropriate placement of remote partner avatars through the analysis of different indoor spatial information. However, daily life-based avatar placement is challenging because avatar placement takes into account the factors that preserve spatial and interactional behavior, unlike more traditional video teleconferencing through a monitor screen. It varies from day to day and from person to person depending on people's living habits or indoor environment. In a daily life-based situation, the proper placement of an avatar is based on a user's context in space, and the different location elements between spaces affect the reality of the user avatar (Richard et al., 2018). Therefore, to effectively show the reality of the avatar in telepresence, it is necessary to latch onto users' exact context and the meaning of an object in indoor space information (Nevelsteen, 2015). It is also essential to systematically analyze object characteristics of dissimilar spaces with the same criteria in order to convey a sense of reality while communicating with an avatar.

However, there are three problems with the current situation. First, most quantitative measurements of indoor spaces are too simple. For example, when describing living rooms, they are treated as spaces of the same character, even though there might be two living rooms of different layout and with different arrangement of objects. Additionally, depending on the different purpose and arrangement of objects, users' behavior patterns or elements of the space may differ. In order to provide an optimized user experience through the understanding of context, it is important to analyze and define the configuration of objects in each space in detail. Second, to accurately analyze the location of an avatar, a method that understands the actual space semantically, is needed. Based on the existing FBS framework (Gero & Kannengiesser, 2014), we have attempted to solve this problem by developing a mapping method for a real space. We named the spatial representation method of furniture Function-Behavior-Structure Map (FBSMAP), using semantic features. Third, in order to find the proper placement of an avatar when two avatars are in a single space, a comparison standard of similarity for the entire indoor space is required. Measuring spatial similarity is complicated because it must reflect the properties of numerous objects that comprise the indoor space (Narengerel et al., 2018). Precise criteria that can reflect the properties of various objects need to be established for measuring the similarity of two different spaces.

Therefore, the goal of this work is to: 1) analyze object configuration in an indoor space for understanding user context, 2) develop a method for applying to the actual space, and 3) compare the spatial similarity of dissimilar spaces through this semantic method.

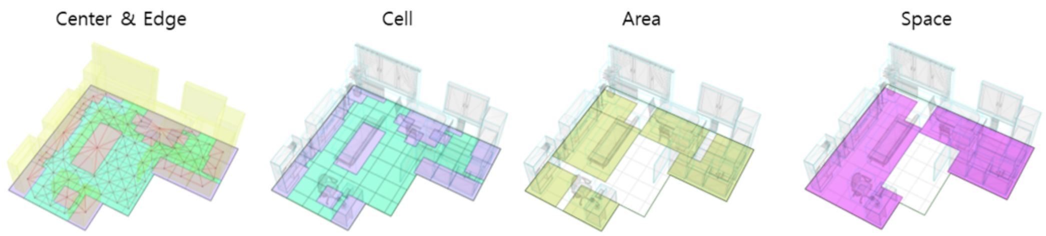
The four main stages of semantic spatial comparison are discussed. The first stage is to quantitatively implement the indoor space modeling through FBSMAP, and the second stage acquires the properties of the modeled space through questionnaires. The third stage is to establish spatial similarity evaluation criteria to compare the acquired space, and fourth, spatial similarity evaluation is performed to provide the optimal virtual space service.

This study presents the criteria for the comparison of dissimilar indoor spaces in detail through more specific methods than those of previous studies. This will offer an effective way to place a user's avatar in the correct location during telepresence. It will also provide an optimization of the user-based experience through the understanding and analyzing of the context in the user's indoor space.

2. Related Work

2.1 Semantic features of an indoor environment

Comprehensive space analyses for the purpose of defining semantic features are found in previous literature. For example, Norberg-Schulz (1971) introduced an existential space theory that categorized the environment into geography, landscape, urban environment, house, and the "thing." Each categorical layer has its own three basic elements derived from the way humans utilize and perceive space: center, path, area. In contrast, the FBS framework (Gero & Kannengiesser, 2014) investigates designed things to define and specify their characteristics. It articulates a designed object into three ontological categories: "Function" is the purpose of the designed object, and "behavior" implies the attributes that can be derived from it, and "Structure" indicates the components of the designed object. The FBS framework is applicable to architecture and furniture design. Finally, to manage building components and indoor facilities, the Open Geospatial Consortium provides the IndoorGML framework (www.indoorgml.net) to standardize indoor spatial information. In order to represent geometric and semantic indoor space, it uses "cellular space," (also called "cell"), which is the smallest structural unit having a singular architectural significance. Research that applies the FBS framework to analyze indoor environments is presented in Narangerel et al. (2018). They took into consideration not only physical properties but also semantic features. Function-Behavior-Structure Map (FBSMAP) is the spatial representation method at furniture level. Since FBSMAP covers detailed components and semantic features of an indoor environment, we decided to use the term FBSMAP. Figure 1 compares spatial terminologies considered in this research.



Existential space		Path	Center	Area	
IndoorGML	Centroid	Edge	Cell	Space	
FBSMAP	Center	Edge	Cell	Area	Space
Definition	The centroid or central point of cell	The path information derived from the relation between each cell	The initial space where a human can be situated in or make use of	The territory where a human can perform similar actions near his or her vicinity	The combination of multiple area having similar spatial features

Fig. 1. A comparison of terminologies of Existential space, IndoorGML, and FBSMAP

2.2 Spatial similarity measurement

Li and Fonseca (2006) proposed essential elements for a spatial measurement process that allows for the preparation of similarity assessment between spatial scenes. As shown in Table 1, they considered scene level, which measures relationships between two different scenes, and object level, comparing the attributes of objects in a scene one by one. In the current study, the topological, directional, metric distance, distribution, and attributes relationships are included in the model to measure commonality and contrast.

Table 1. Basic elements in the spatial measurement process (Li & Fonseca, 2006)

Level of comparison	Types of similarity measured		
Scene	Relationships	Spatial	Topological
			Direction
			Metric distance
			Distribution
		Non-spatial	Attribute distance
Object	Attributes	Geometric	Types of objects
		Thematic	Attribute comparison

Spatial similarity measurement is guided by a scene and layer-level comparison that draws a parallel between objects in the scene and the overall relationship of each scene. In order to analyze the characteristics of a spatial configuration, Park (2016) proposed a method of representing a residential space through nodes and edges in a graph, as shown in Fig.2. The topological compositions are classified by their types and adjacent relationships between subspaces. Diakité and Zlatanova (2017) introduced the Flexible Space Subdivision framework that allows for the identification of a space that can be used for indoor navigation purposes. The type of indoor objects was considered based on their classification and functions to define the distribution and attributes relationship. As mentioned in the study, however, the semantic information available in the building model was limited, so that information about the mobility of furniture was only possible to consider. Both approaches exploited one or two criteria for spatial configuration and comparison. To measure spatial similarity, a method that reflects the context of the environment is required.

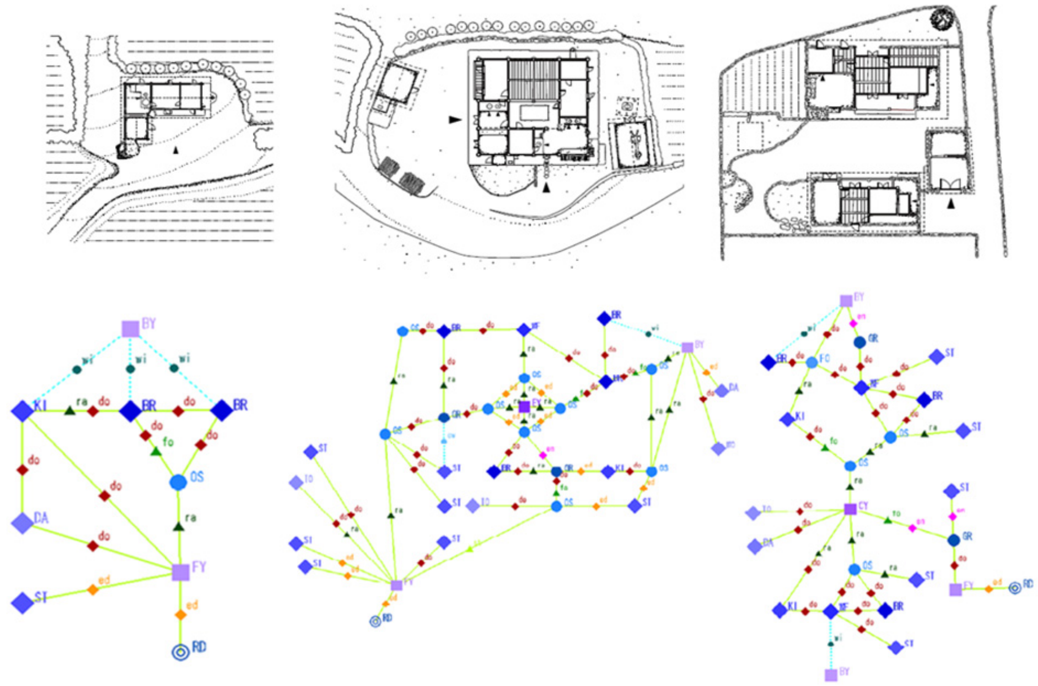


Fig. 2. Adjacent graphs of three traditional Korean dwellings, taken from Park (2016)

3. Methodology

3.1 A hierarchical graph representation for indoor spaces

To understand an indoor environment, space representation should be examined in respect to physical attributes and semantic features. The concept that regards space as multi-layered subspaces represented as nodes or vertices in accordance with the hierarchical representation is discussed in before. However, neither the representation of furniture nor the semantic features of a spatial component are considered in IndoorGML (www.indoorgml.net). According to the level of detail that captures building components, a node represents a room or furniture. A room should be subdivided into the level of an object with functional information so as to understand the interaction between a human and the indoor space. Furthermore, each component sharing similar semantic features needs to be combined into higher concepts such as “living room” or “bedroom.” Using a computer-aided design program, it is possible to find the physical attribute of a single item of furniture by calculating the bounding box of each object. As was pointed out by some of the authors of this paper (Narengerel et al., 2018), however, acquiring semantic features remains challenging. In this study, a crowdsourcing method is conducted to gather information about the context of an indoor environment. Even though a FBSMAP visualization of different houses explicitly characterizes an indoor environment, as illustrated in Fig.3, the difficulty lies in comparing one space with another. To measure similarity and difference, establishing a standard needs to come first.

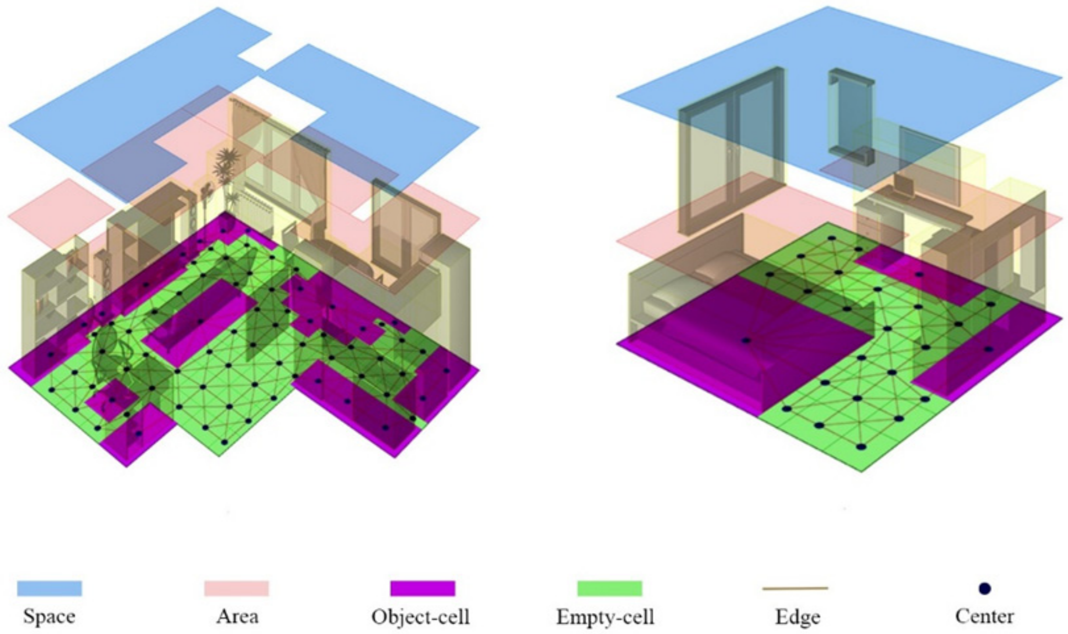


Fig. 3. FBSMAP visualization of mid and compact-sized houses (Narengerel et al., 2018)

To describe an indoor space in detail, furniture should be of the minimum unit in an indoor representation, which corresponds to a *cell* in FBSMAP. Each *cell* is classified into an *object-cell* and *empty-cell* depending on whether it contains furniture or not. In the *cell* layer, a *cell* is represented as a unique node in the graph, and an *edge* connects two cells together. As shown in Fig. 3., the *object* and *empty-cells* contain one node and are separately classified into green and purple colors. Furthermore, adjacent *cells* sharing similar behavior are merged into a single *area*, which is also described as one node (including multiple cells). For example, various objects that share similar functionality such as a dining table and cooking area provide the notion of one functional space.

For depicting the topology of a house and the location of a human (Fig. 4.), a hierarchical node-edge graph is created, which is further transformed into a binary tree. To calculate the similarity measurement between two houses, manipulation cost of the graph is thus calculated. We consider a binary tree, which is initialized by a root node; it is the *cell* where a person is located. Child nodes are determined by the connectivity of the node-edge graph. For showing the hierarchy among several functional areas, we try to precisely depict the node connectivity in the expansion order from the root node. The activity spaces that are reachable through the shortest distance act as the child node for the root node. The nodes that are additionally connected or closer to the activity space nodes behave as the children nodes. In a *cell* layer, the nodes represent *cells*, and the graph indicates the relationship among several functional areas and furniture properties.

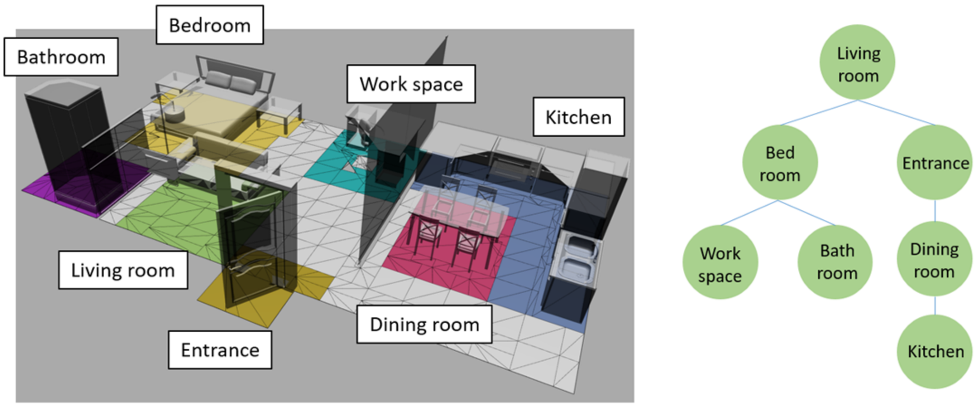


Fig. 4. Graphical (tree) representation of a sample home's topology

It is possible to generate an *area* layer from the *cell layer*, where a node represents the cluster of *cells*. These clusters of cells contain several furniture items that act as a single unit to form a functional area. For example, the "living room" node in the *area* layer comprises nodes in the *cell* layer, such as "sofa," "coffee table," and "TV" nodes. In other words, when we form several *area layers* from *cell layers*, a single indoor environment can be represented as multiple hierarchical graphs.

3.2 Survey for semantic features using crowdsourcing

A FBSMAP needs to generate not only physical attributes such as size, the position in a room, or distances between objects, but also semantic features such as how a user can behave with a single object, or how pieces of furniture can be clustered as a higher level of space. The semantic feature in terms of *object-cell* is the spatial feature that is determined by a performance or activity of the user with particular furniture. For example, one person might say the behaviors of a chair are "sit," "recline," and/or "stand on." The name of the object, "sofa," may imply the behavior of it. However, a list of behavior gathered from several respondents makes it possible to analyze similarity quantitatively. Secondly, the semantic feature for the boundary of the clustered *cells* is variable, depending on the context of the indoor environment. The *area* is the combination of *cells* that contains multiple furniture items by their behavioral similarity, which is driven by their list of behavior. Even though a coffee table, a sofa, and a chair might be placed near to one another, a chair cannot be clustered in the same *area* if it does not share a similar behavior list with the other items, and if someone does not judge it as one cluster. Lastly, the semantic feature of *area* is the function of FBSMAP. The function of *area* is the spatial feature that a person can interact with the set of furniture. It is a more abstract level of the semantic feature than what the *object-cell* contains. It is often expressed as a functional component such as a living room, dining area, kitchen, and so on.

To generalize the semantic feature mentioned above, we conducted an online survey using a web application on Amazon Mechanical Turk. Respondents were asked to provide responses on three aspects: the behavior of *object-cell*, the cluster of *cells* sharing similar behavior, and the function of *area*. We prepared 3D model home from the SUNCG (Song et al., 2017) dataset, using varied house models with different size, usage, and room number, for example. Before it is provided, house models are already processed with the *cell* division phase, which means generating a bounding box of furniture and *empty-cell* on empty space. From that, it was possible to identify specific furniture and to select the indoor object based on its *cell*. Five pieces of furniture, which were considered the most essential objects having a decisive effect on the characteristics of the indoor environment, were highlighted. Respondents were required to type the name of behavior at least 3. In sequence, the *cells* were clicked depending on their spatial functionality and were differentiated from the colors. Once *areas* were all made, respondents wrote down the name of the function at least 3.

3.3 Spatial similarity criteria and equation

Scene-level estimation inquires about the overall relationship of each scene, which comprises of a topological, directional, metric distance, distribution, and attribute (non-spatial) relationship. In this comparison we only considered similarity measures in 2D space, not 3D space. We compute a similarity percentage for each cell of the activity space. The cell obtaining the highest percentage would then be considered as an avatar's final destination or location. For observing the cell for an avatar that is closest to at human's location, we made use of different similarity features, as listed in Table 2. The table introduces the similarity measures that we have taken into consideration for the calculation of the finest cell.

Table 2. The criteria for indoor space similarity measurement (based on Li and Fonseca (2006))

Similarity Measure	Spatial Information	In-Room Environment
Topological Relationship	Objects or stimuli present in the enclosed areas and path.	Doors, study areas, etc.
Direction	The direction of the path or route.	Number and types of turns involved to reach from source to destination.
Metric Distance	Length of the path or route.	Amount of distance covered in order to reach a specific point from the start.
Distribution	This includes an activity center or activity space.	Distribution of the activities performed based on the kind of space and objects.
Attribute Distance (Non-Spatial)	This also includes an activity space.	Landscape features near the space where activity by the person will be performed.

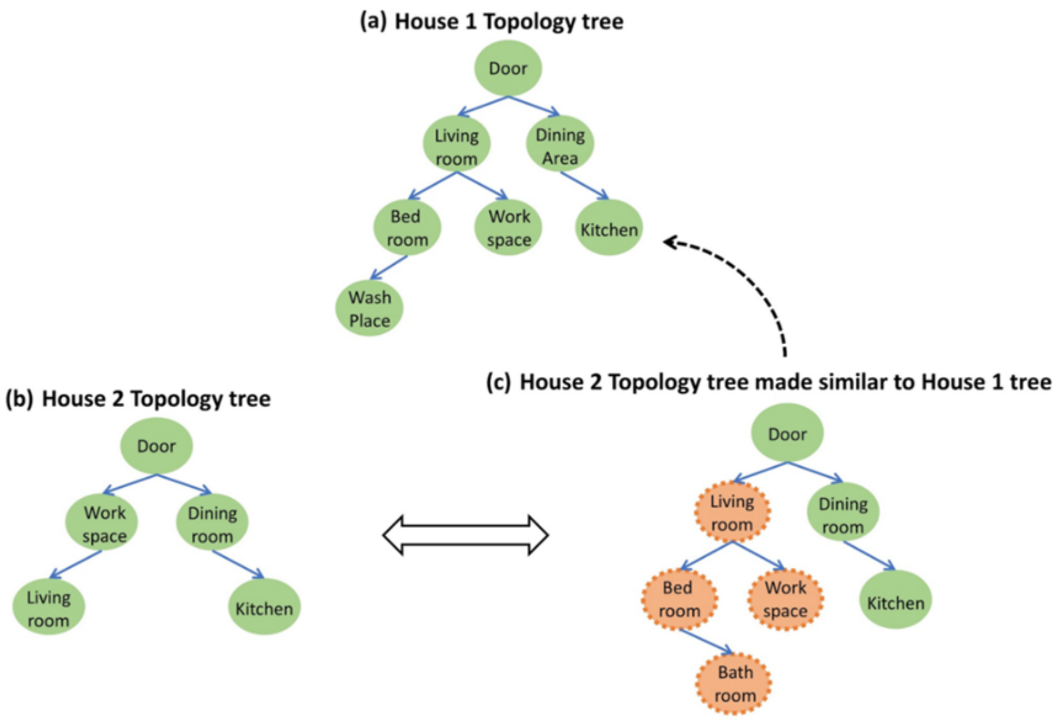
3.3.1 Topological relationship

The foremost spatial similarity feature is the topological comparison between different scenes. The objective is to augment the missing nodes graphically (activity spaces or objects). This requires us to 1) make a thorough transformation of the topology of one house (avatar space) similar to another house (where a person is located), and 2) determine the similarity or dissimilarity between two spatial scenes (in our case, we considered indoor layouts of model homes). To make the latter speculation workable, we created a simple representation of numerous activity spaces present in the specimen house models. We generated straightforward graphs in the order of increasing cell distance from the door, that is, the root node. After obtaining the graphs, we performed the graph edit distance operation to make two graphs' topologies equivalent. Graph edit distance measures the distance between two graphs, g1 and g2, by the amount of manipulation that is required to transform g1 into g2. The similarity calculation can be executed using the formula:

$$S_{topology} = 1 - \frac{l}{n(x)}$$

Where, l is the edit distance, that is, the number of edit operations required (deletion, insertion, substitution) to make two graphs similar, and n(x) is the number of nodes in the compared topology graph.

In this work, the node connectivity in increasing distance order (here cell distance) from the root node was considered, and no other form of connectivity was contemplated. The directly accessible activity spaces will serve as the children nodes to the root node or the main node. These children nodes are further linked to the activity space nodes that are directly connected with them or adjacent to them. After obtaining the graphs for each house model to be compared, the edit distance was applied to calculate the similarity between the house models. Figure 5 shows an example of graph edit distance application on a house tree to make the topology similar.



(a) House 1 topology; (b) House 2 topology; (c) To achieve augmented representation of House 1's topology (for the avatar), we performed node insertion and substitution using the edit distance operation in House 2's graph (Dwivedi et al., 2019).

Fig. 5. Graphical (tree) representation of two house topologies

3.3.2 Direction and metric distance

The next two spatial similarities considered are direction and metric distance, and these were calculated simultaneously. We considered four different angles (0°: the avatar is looking forward; 90°: the avatar is looking right; 180°: the avatar is looking backward; 270°: the avatar is looking left). For determining the similarity of distance among two spaces, the distance between objects that fall within the avatar's field of view and the avatar's cell is calculated. This helped us to establish several objects that are inside the field of view of both the avatar and the human. The distance of those objects from the avatar's location and how similar they are located by human perception in the compared space were then deduced. After obtaining this information, we then found the best angle from which an avatar sees the same objects as seen from the humans' view. We derived the following formula to calculate the mean percentage of all the attribute or object distances in an avatar's field of view:

$$S_{dist\&dire}(x) = \begin{cases} \sum_{i=1}^{k} \frac{d_{x_i}}{d_{z_i}}, & \forall d_{x_i} \le 2d_{z_i} \\ k, & Otherwise \end{cases}$$

Where dxi is the distance between the cell x and an attribute i (in the case of an avatar), and dzi is the distance between the human cell and the attribute. To avoid complexity and to offer a perfect range, we considered a 180° field of view for an avatar even though a human being's field of view is typically only 114°.

3.3.3 Distribution

Similarity evaluation for scenes with many spatial objects is complex. Therefore, it is important to measure the distribution of all the activities that can be performed around a set of attributes. Apart from measuring the topological and directional similarity and, distance, an activity distribution assessment seemed inevitable. Scenes with a discrete set of objects and dissimilar spatial distribution were considered. Usually a human performs set of activities when surrounded by specific objects and this activity area is determined by the dominance of particular objects for e.g. TV and sofa in the living room. Distribution similarity measurement makes use of a list of several activities that can be performed in an activity area in order to provide a percentage. For instance, the objects surrounding someone and the activity that he/she performs with those objects and around them. We looked for similar activities in the same area but in two different house models (one where a human is present and the other where an avatar is placed) to infer the scenario for the avatar placement. The similarity percentage between two house models for distribution assessment can be identified by performing the list comparison. This can be obtained through the following cross product:

$$S_{list}(x) = \frac{m(x)}{l}$$

Where, l is the size of the similarity measurement list in House 1 (human space) and m(x) is the number of activities or items (in the case of attribute, discussed later) common in both houses.

3.3.4 Attributes

The similarity measurement for attributes is similar to the measurement approach in Section 3.3.3. The only difference is that instead of an activity list, we considered the simple distribution of several objects in space. For instance, if the attributes for the current human position in House 1 are a desk and a chair, we considered a similarity of 100% for a cell in House 2 that possesses the same attributes or more (e.g., a desk, a chair, and a table lamp). It should be taken into account, that House 2 is the space for avatar placement. Since the similarity percentage for the attribute list comparison is a simple list differentiation, it can also be obtained through the cross product mentioned in Section 3.3.3.

3.4 User study to acquire spatial context

The feature we proposed in Section 3.1. allowed us to rank cells from the highest similarity percentage to the lowest. We conducted a user study to observe whether users would follow a common pattern (i.e., they would choose cells that would be close to each other) for the avatar placement. For example, if the cells chosen by users are very close to each other, we may be able to conclude that there would be a common area where the avatar has the best placement. In future work, we will develop the weights associated with our similarity criteria so that they reflect the users' choices. We conducted a user study with n=W participants, and we ran a browser-based 3D application with three.js in which we loaded house models from the SUNCG (Song et al., 2017) dataset, processed with Rhino and Grasshopper, to generate distinctive empty cells and furniture cells. We displayed an initial house that represented where a human would be standing. This house was displayed with a top view and a first-person view (i.e., the human's field of view), with a height of 160 cm. The survey respondents had to select the cell and the field of view that best matches, according to them, the human's position and field of view. In this study, we represented only one human in one house so that we can focus on the semantic aspects of the house. Other criteria could be discussed but are not taken into account, which could, for example, involve the presence of two people, such as the importance of the gaze of the avatar when the two people are having a conversation. Hence, we have chosen to focus on how the users would place an avatar according to the surrounding objects in a house. We ran five comparisons (one house model compared with another one), and within these comparisons, we have proposed 15 different locations where a human could be per comparison. Consequently, the respondents were given 75 unique questions and cases. In Fig. 6, the left red cell represents where a human is positioned, and the right red cell represents where the user chose to place the avatar. The yellow arrow for both the top views represents where the avatar is looking. We gave the survey respondents the option to navigate through the house and rotate the avatar with their keyboard's arrow keys, to provide a greater sense of realism than just clicking on the best cell on the top view as if they were walking in the house (the latter was also possible if they preferred). The users could also zoom in/out and rotate the top views. We chose to load house textures for the first-person view to give a better feeling of realism and immersion when the user moved through the house.

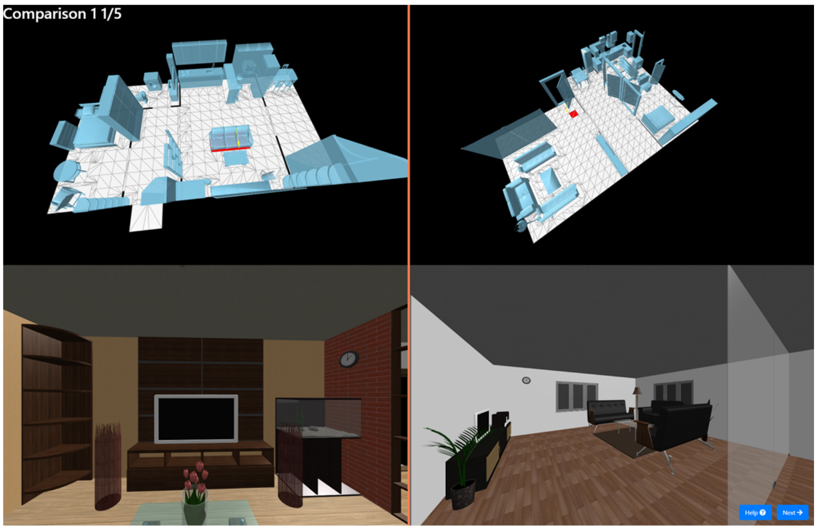


Fig. 6. Images that were used in the user study. Based on the position in the reference house (left), a user selects the most similar position in the target house (right).

4. Implementation

4.1 Overview of the research

To propose the spatial similarity measurement for the indoor environment, we performed three steps, as illustrated in Fig. 7. First, we processed the indoor space modeling with example 3D house models. Second, we proposed spatial similarity criteria and equations. Third, we calculated similarity metrics and evaluate them based on a user study.

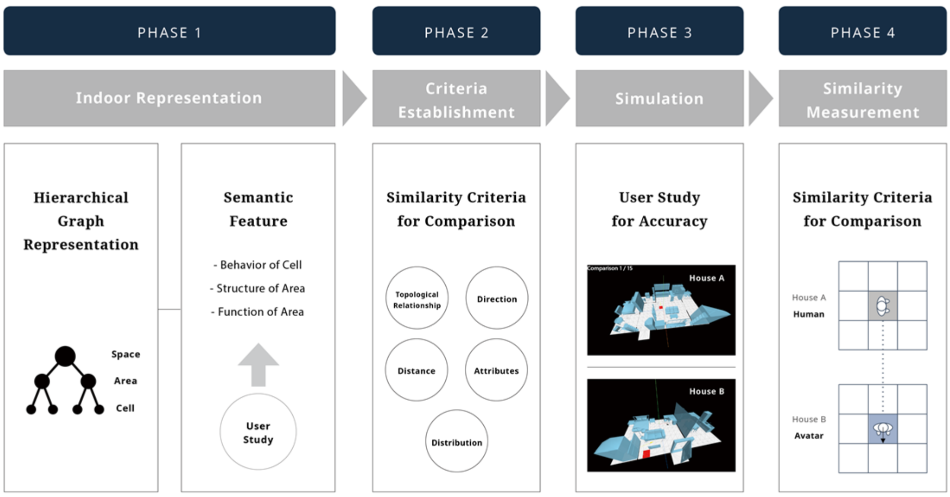


Fig. 7. Space modeling and similarity measurement process diagram

4.2 Indoor space modeling

In order to calculate the similarity between indoor environments, each of them should be represented as a space model. Since the methodology, which was suggested by Narangerel et al. (Narengerel et al., 2018), quantifies physical and semantic characteristics, we decided to make use of FBSMAP as a modeling method. To test and verify our proposed similarity measurement methodology, four house models were selected from the SUNCG (Song et al., 2017) dataset based on the size and the composition of rooms. The floor plan of those four models is shown in Fig.8. This figure also depicts the alignment of cells by segmented lines on the floor and the generated areas by clustered cells filled with different colors. According to the procedure in FBSMAP methodology, the cell layer was processed with the 3D modeling software, Rhinoceros (https://www.rhino3d.com), and the algorithmic computation program, Grasshopper (https://www.grasshopper3d.com). In this step, the physical property of indoor objects and the labeled data for furniture were addressed. In sequence, the information gathered from the survey conducted in Section 3.2.2. was used to add up semantic features that are the behavior of *object-cell* and the boundary and function of each area. It was possible to implement FBSMAP on the selected 3D models successfully. The implemented space models will play a critical role in measuring space similarity.

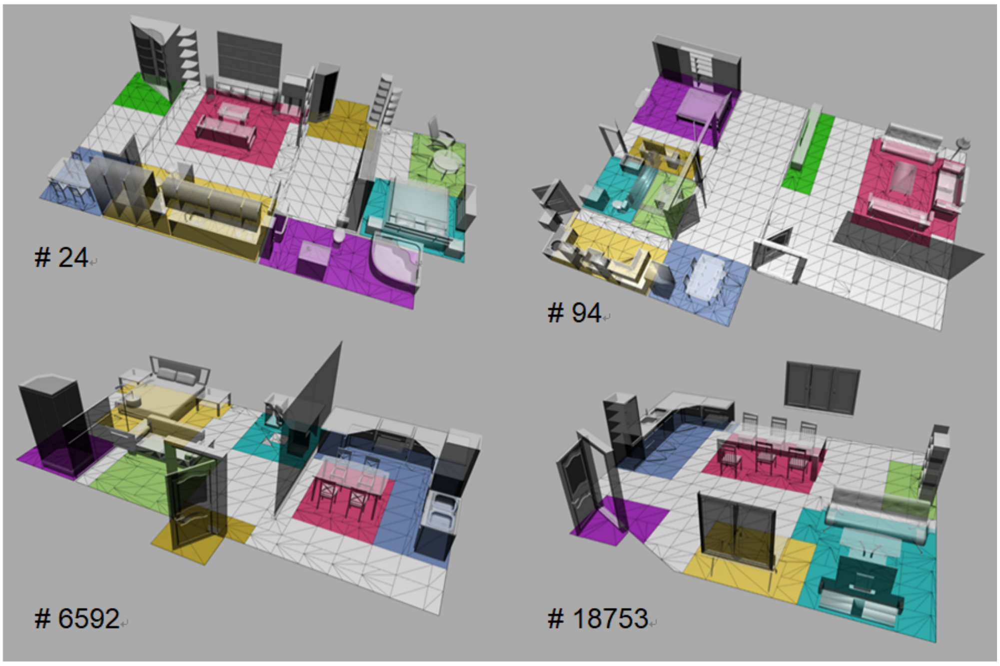


Fig. 8. Floor plans of the sample house models

4.3 Calculating spatial similarity metrics

Spatial similarity measurement was performed to place an avatar on the target house from the comparison house based on the methodology described in Section 3. The proposed method suggests the position in the target house based on the calculated metrics considering spatial features on the position in the comparison house. Since there are various types of homes in terms of size or arrangement, four models were chosen from the SUNCG dataset; two (#24 and #94) for large size and holding multiple furniture, and two (#6592 and #18753) for small and fewer indoor objects, as depicted in Fig. 8. The FBSMAP of those houses was already generated from the indoor space modeling. The physical information, which includes cell division, furniture labeling, and indoor object arrangement, is calculated using Rhino3d and Grasshopper. The semantic features, including the behavior with furniture, area shape, area function, were obtained from the user study. There were four comparison sets following the types of houses: from big to big (#24→#94), from small to small (#6592→#18753), from big to small (#24→#6592), and from small to big (#18753→#94). Each comparison set had five subcases, which were five different positions in the comparison house. For example, Fig. 9 illustrates the case that, compares from house #24 to #6592. Based on the cell colored in green in the left house (#24), spatial similarity measurement was performed to predict the best candidate cells in the right house (#6592). Red-colored cells indicate high spatial similarity, and blue ones do not. The spatial similarity criteria and equations mentioned in Section 3.3 are implemented with Python programming language. Python libraries such as NumPy, Scikit-learn, and Pandas, which are for mathematical function of the numeric value and matrix, were also used for implementing algorithm and data processing. Each calculated value was scaled from 0 to 1 so as not to make only one value dilute the others, and then all the values were summed into one, which means that the relations between spatial similarity features were not considered at that point.

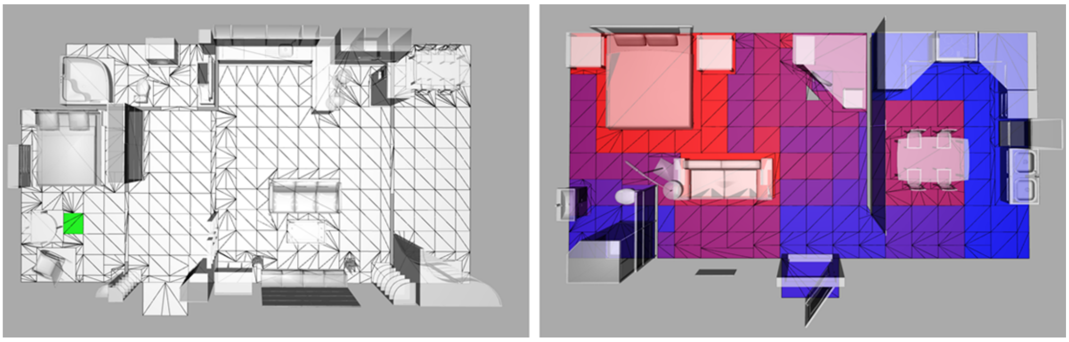


Fig. 9. Calculated spatial similarity measurement of each cell (right), based on the green-colored cell in the reference house (left).

4.4 Evaluation of spatial similarity measurement

In order to evaluate our proposed method, the performance of the spatial similarity metrics needs to be measured in a quantitative way and this value should be compared with another method. We calculated the unique value of each cell, considering the five similarity criteria given in Section 3.3. The cells scoring the highest value were taken to be the candidates for the most similar space. Since the cell itself only has a physical value, which is the coordinate of the floor, the candidates were converted into area information as per FBSMAP. With the suggested candidate area set, we measured the balanced accuracy, as proposed by Brodersen et al. (2010), on the basis of the human result as a ground truth. It was considered that the positive and negative answer is not only the true condition but also the predicted condition. For example, the comparison between every cell in house #18753 and #94 for cell #73, the ground truth from the user study suggests that the answer is four of a dining room, three of a TV space, two of a living room, and one of a library. On the contrary, our spatial similarity measurement method gives the result of seven of a dining room, one of a kitchen, one of a library, and one of a bathroom. So the true positive rate is eight over thirteen, and the true negative rate is five over seven. Therefore the balanced accuracy is 0.664, which is the mean of those two values. To compare how our method performs, we generated 10,000 random sample data under the same condition. The counted lists calculated the balanced accuracy and the average. The result was 0.498, and it usually converged to 0.5. This means that without knowing the spatial similarity, the result should be fifty-fifty. Table 3 shows the result of all comparison sets for every five subcases.

Table 3. The balanced accuracy calculated for four comparison sets

Comparison 1	#24→#94	Our method	Random sampling
Case 1	Furniture_5_Cell	0.85185185	0.69242671
Case 2	Empty_12_Cell	0.64102564	0.5177742
Case 3	Empty_36_Cell	0.79411765	0.58324792
Case 4	Empty_146_Cell	0.88461538	0.43908764
Case 5	Empty_175_Cell	0.46640316	0.63862503
		0.727602736	0.5742323
Comparison 2	#6592→#18753	Our method	Random sampling
Case 1	Empty_21_Cell	0.85185185	0.65189376
Case 2	Empty_35_Cell	0.90909091	0.43861898
Case 3	Empty_3_Cell	0	0.39868583
Case 4	Empty_7_Cell	0.83333333	0.5441535
Case 5	Empty_27_Cell	0.83333333	0.41391031
		0.685521884	0.489452476
Comparison 3	#24→#6592	Our method	Random sampling
Case 1	Furniture_5_Cell	0.8	0.51976908
Case 2	Empty_12_Cell	0	0.4695891
Case 3	Empty_36_Cell	0.14545455	0.5577087
Case 4	Empty_146_Cell	0.76666667	0.42023599
Case 5	Empty_175_Cell	0.8	0.62812628
		0.502424244	0.51908583
Comparison 4	#18753→#94	Our method	Random sampling
Case 1	Empty_73_Cell	0.66483516	0.49894066
Case 2	Empty_43_Cell	0.82352941	0.5640466
Case 3	Furniture_0_Cell	0.76470588	0.52322557
Case 4	Empty_39_Cell	0.76923077	0.47146323
Case 5	Empty_34_Cell	0.83333333	0.45486179
		0.77112691	0.50250757

5. Discussion

In our suggested spatial similarity measurement, the context of the indoor floor plan and furniture arrangement was considered. In order to achieve the semantic comparison of different spaces, the idea of taking a physical structure into a hierarchical node-edge graph was taken into account. The topological representation model has a key part in the spatial concepts of a room and building components. As depicted in Fig. 4., not only *object-cells* but also *empty-cells* are assigned with the code, which is called “function” in the “area” layer. In other words, the semantic spatial concept became a node in the indoor topology. On the basis of an individual person's position, the relationship between nodes is represented as edges for connectivity. The solution to analyze the space with semantic features makes our method robust.

Spatial similarity measurement aims to calculate the similarity of indoor spaces with a quantitative method. The spatial similarity criteria come from Li and Fonseca (2006), and each criterion was revised so that it could be applied to the indoor comparison. “Direction,” “metric distance,” and “distribution” were scaled down according to the furniture unit. “Topological relationship” utilized a hierarchical node-edge graph, and “attribute distance” measured the arrangement of furniture concerning semantic features defined as “behavior” and “function.” The fundamental intention of spatial similarity equations is to calculate the cost of becoming equal to the standard house based on spatial context. With those five criteria and equations, it is possible to compare different indoor spaces quantitatively and comprehensively.

The result of the four comparison sets shows a difference according to the size of the indoor space. Houses #24 and #94 are comparatively bigger than houses #6592 and #18753, which hold a greater number of furniture items and support complex activity. The comparisons of similar houses in terms of the property size present better results than the others. On the other hand, our method shows a worse result than random sampling in “comparison 3,” which compares from a big house to a small one. It can be interpreted that selecting spatial functionality in limited choices with multiple options is a challenging task for the suggested spatial similarity measurement.

Even though our method surpassed the random sampling method, there is still room for improvement. Our method considers five criteria equal, which means each criterion calculates its value, and the final metric sums up five values. It is reasonable that those five criteria should be considered differently depending on the context of the comparison house. If the weights for the criteria could be appropriately set, the result might be adjusted to be better than the current method. It is possible to obtain weights using multivariate linear regression under sufficient data from a user study. Each house model has unique spatial characteristics, and the task is finding approximate values for five criteria, which is, after all, the optimization problem.

6. Conclusion

In order to convey the reality of an avatar in telepresence, a systematic analysis of indoor space information and the user’s context is required. We have presented a method to analyze the object configuration of an indoor space and to apply the analysis to an actual space for understanding the user context. We implemented indoor space modeling quantitatively through FBSMAP, and we analyzed the behavior of an *object-cell* and the boundary and function of each area by conducting questionnaires with quantified indoor space modeling. In addition, the proposed methodology can compare spatial similarity in dissimilar spaces. It consists of five criteria, which are: Topological Relationship, Direction, Metric Distance, Distribution, and Attribute Distance. With that, five established similarity measures have been used for virtual telepresence. Although the proposed method outperformed the random sampling method in evaluation, it is possible to raise its performance to a higher level. The greater the number of responses from a user study with various types of floorplan, the more generalized calculation. Considering the similarity relationship assigning weights to each criterion on the basis of importance also makes our method robust for atypical/unconventional cases.

References

Alexiou, A., Bouras, C., Giannaka, E., Kapoulas, V., Nani, M., & Tsiatsos, T. (2004). *Using VR technology to support e-learning: the 3D virtual radiopharmacy laboratory, 24th International Conference on Distributed Computing Systems Workshops*, *2004*. Proceedings. DOI: 10.1109/ICDCSW.2004.1284042

Heun, V., Kasahara, S., & Maes, P. (2013). *Smarter Objects: Using AR technology to Program Physical Objects and their Interactions*, CHI 2013: Changing Perspectives, Paris, France.

Rácz, A., & Zilizi, G. (2018). *VR aided architecture and interior design, 2018 International Conference on Advances in Computing and Communication Engineering*, Paris, France.

Dreshaj, E. (2015). *Holosuite: an exploration into interactive holographic telepresence* (PhD thesis). Massachusetts Institute of Technology, Boston, Massachusetts.

Richard, P., Burkhardt, J-M., Lubart, T., Bourgeois-Bougrine, S., & Barré, J. (2018). *The Effect of Virtual Environment and User/Designer Collaboration on the Creative Co-design Process*, Congress of the International Ergonomics Association.

Nevelsteen, K. (2015). Virtual World, Defined from a Technological Perspective, and Applied to Video Games, Mixed Reality and the Metaverse. *International Journal of Computer Games Technology, 29*, 1-22.

Narengerel, A., Hong, S., Lee, C., & Lee, J. (2018). *FBSMAP: The Spatial Representation Method for Intelligent Semantic Service in Indoor Environment, Proceedings of the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia 2018*, Volume 2, 587-596.

Norberg-Schulz, C. (1971). *Existence, space & architecture*. Studio Vista, London, England.

"IndoorGML OGC." IndoorGML OGC, www.indoorgml.net/. (accessed 12th March, 2019)

Gero, J. & Kannengiesser, U. (2014). The function-behaviour-structure ontology of design. *An anthology of theories and models of design*. Springer, London, 263-283.

Li, B., & Fonseca, F. (2006). TDD: A comprehensive model for qualitative spatial similarity assessment. *Spatial Cognition and Computation, 6*(1), 31-62.

Park, J. (2016). Spatial Configuration Analysis on Three Korean Traditional Dwellings Using the Graph Theory. *Journal of Asian Architecture and Building Engineering, 15*(1), 103-110.

Diakité, A. A., & Zlatanova, S. (2017). Spatial subdivision of complex indoor environments for 3D indoor navigation. *International Journal of Geographical Information Science, 32*(2), 213-235.

"Rhinoceros." Robert McNeel & Associates, https://www.rhino3d.com/. (accessed 25th July, 2019)

"Grasshopper - Algorithmic modeling for rhino." Scott Davidson, https://www.grasshopper3d.com/. (accessed 25th July, 2019)

Song, S., Yu, F., Zeng, A., Chang, A. X., Savva, M., & Funkhouser, T. (2017). Semantic scene completion from a single depth image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1746-1754.

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann. J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, 3121-3124. IEEE.

Dwivedi, U., Porcellini, V., Hong, S., Chang, Z., & Lee, J. (2019). Computing Spatial Features to Allocate Collision-free Motion-Paths for Tele-presence Avatars. In *Intelligent & Informed*, 245-254. CAADRIA.