OPTIMAL NUMBER OF CUE OBJECTS FOR PHOTO-BASED INDOOR LOCALIZATION

Youngsun Chung, Daeyoung Gil & Ghang Lee
Department of Architecture and Architectural Engineering, Yonsei University, South Korea

ABSTRACT: *Building information modeling (BIM) is widely used to generate indoor images for indoor localization. However, changes in camera angles and indoor conditions mean that photos are much more changeable than BIM images. This makes any attempt at localization based on the similarity between real photos and BIM images challenging. To overcome this limitation, we propose a reasoning-based approach for determining the location of a photo by detecting the cue objects in the photo and the relationships between them. The aim of this preliminary study was to determine the optimal number of cue objects required for an indoor image. If there are too few cue objects in an indoor image, it results in an excessive number of location candidates. Conversely, if there are too many cue objects, the accuracy of object detection in an image decreases. Theoretically, a larger number of cue objects would improve the reasoning process; however, too many cue objects could lead to declining object detection performance. The experimental results demonstrated that of two to five cue objects, three cue objects is most likely to yield optimal performance.*

KEYWORDS: *indoor location determination, BIM, reasoning*

1. INTRODUCTION

Photos are commonly used as a medium to support building maintenance and defect management (Kim et al., 2014). These photos are sometimes taken by experienced field workers, but most are captured by unskilled workers or individuals with a limited understanding of the building, such as occupants (Kang et al., 2019). In existing maintenance systems, users are typically required to manually tag the locations where a photo was taken to utilize the system effectively. In particular, when occupants report defects, the specific locations and conditions of the defects are often described in unstructured text, making management even more challenging. Various methods have been used to accurately determine the locations where photos were taken. Several image-based indoor positioning methods have been explored, including approaches which search for the most similar BIM screenshot image to a target photo (Ha et al., 2018) or regress the camera position using deep learning algorithms (Acharya et al., 2019). Nevertheless, image-based methods typically rely on extensive image training and are sensitive to changes in indoor conditions (Kim & Kim, 2023). This sensitivity becomes especially problematic in buildings that have multiple varying factors, including interior fittings and lighting.

To overcome these limitations, especially the sensitivity to changes in indoor conditions, we propose an indoor localization method based on reasoning-based localization method that uses cue objects and their spatial relationships. Unlike furniture, cue objects, such as doors and windows, can serve as stable reference points because they rarely change. To achieve this goal, the first step is to determine the optimal number of cue objects required. If there are too few cue objects, they may not provide sufficient information for localization. However, if there are too many cue objects, the accumulated accuracy of cue-object detection decreases. The aim of this preliminary study was to validate our proposed method by determining the optimal number of cue objects required in an image to accurately locate the positions of indoor photos. To select the optimal number, we developed a prototype localization method based on the spatial relationships among cue objects, which involved comparing the similarities between cue objects and their spatial relationships in a target indoor image with those in a BIM model. We evaluated the performance of the proposed method by varying the number of cue objects in an indoor image, using the mean probability to accurately determine the location where the image was taken. To validate the proposed method, we measured the performance using the mean probability of localization and varied the number of objects within the photos.

This paper consists of five sections. Following this introduction, the second section discusses previous studies related to the research. The third section describes the research methodology and explains the details of the experiments. The fourth section presents the analysis and results of the experiments, and the final section concludes the paper by discussing the main findings, contributions, and limitations of the research.

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Youngsun Chung, Daeyoung Gil, Ghang Lee, *Optimal Number of Cue Objects for Photo-Based Indoor Localization,* pp. 977-987, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.98

2. BACKGROUND

2.1 Indoor Localization Using Images

Recent developments in computer vision have led to many attempts at indoor localization. Ha et al. (2018) suggested an indoor localization approach using BIM and a visual geometry group (VGG) model (Simonyan & Zisserman, 2015). They used the proposed model to retrieve the most similar BIM screenshot image to a given photo and to determine where the photo was taken. Alam et al. (2022) conducted a similar investigation based on recurrent neural networks (RNNs) to find the correct position of an indoor camera. These methods were intuitive and moderately effective but sensitive to variations in indoor decorations and lighting conditions arising due to their reliance on identifying the most similar screenshot images.

To enhance robustness, developers have attempted to utilize image datasets with camera trajectory data. In the context of BIM-PoseNet (Acharya et al., 2019) and related studies, various researchers have trained models based on deep convolutional neural networks (DCNNs) and large datasets of indoor BIM screenshot images to determine the positions and angles of the cameras used to capture photos. Two such studies were based on RNNs (Acharya et al., 2020) and channel-wise transformer localization (CT-Loc; Kim & Kim, 2023). Although BIM-PoseNet and CT-Loc applications are robust under varying lighting conditions due to an edge extraction method, they still cannot adapt to changes in furniture arrangements or interior decorations. Additionally, the studies were limited to the fixed linear paths of the cameras and excluded the simultaneous handling of close-range and wide-angle images. However, close-up photos are typically taken to capture the appearance of small-sized defects clearly, but for effective building defect management, both wide- and close-range images are required (i.e., photos need to be taken from a distance to address defects that cover a wide area or where the spatial context of the defects is crucial).

2.2 Indoor Localization Using Objects

To overcome the problem of condition changes in images, several researchers have proposed methods that utilize objects within images for indoor localization. Bay et al. (2006) investigated image-based indoor localization using speed-up robust features (SURF; Guan et al., 2016) and unique landmarks, such as posters or logos. Similarly, Li et al. (2022) used multiple visual landmarks and incorporated smartphone compass readings to improve performance. However, using posters as references is not practical because posters may change frequently, causing difficulties in keeping indoor landmark databases up to date.

To overcome these limitations, our author team proposed a method that used semantic segmentation and pose estimation for the positions of cue objects in indoor photos. They aimed to identify the indoor location where the cue objects in photos and conducted a proof-of-concept study (Kim, 2022). However, the method was only tested on objects photographed at relatively short distances.

In summary, previous localization methods based on images have revealed weaknesses in analyzing images under varying conditions. While the methods that employed edge-rendered images and semantic segmentation proved helpful in increasing the robustness of localization under different lighting conditions, they were still ineffective in capturing changes in interior items or furniture. As a solution, we previously proposed a method that focused on cue objects that rarely changed over time, such as light switches and fire extinguishers, and conducted a preliminary study (Kim, 2022). To further develop the method, in this study, we conducted a set of experiments to determine the optimal number of cue objects required for the method.

3. RESEARCH METHOD

The research flowchart for this study is depicted in Fig. 1. The indoor localization method first obtains information about cue objects and their spatial relationships using computer vision technology. We trained and validated the object detection model on an object detection training dataset using the object types and spatial order of the bounding boxes detected to reason indoor locations by comparing them with the spatial relationships among BIM cue objects. This method is based on left–right relationships of cue objects. Location reasoning may result in one or more specific sets of candidate cue objects. Each candidate represents a potential location depicted in the image. We evaluated the performance of the reasoning model based on the probability of accurately determining the location where the photo was taken, considering the number of candidates and the object detection accuracy. We tested the method with varying numbers of cue objects and found the optimal number when the model achieved the highest performance.

In the experimental phase, to which this paper relates, we began by determining the types of objects that would be used as cue objects and establishing the range of objects present in the image. We then created a sample for the BIM model of a housing unit, shown as a BIM DB in Fig. 1, which incorporated 10 different types of cue objects across 11 rooms. We generated 12,861 BIM screenshot images and used 12,671 images to train and validate the object detection model. We employed the remaining 190 images, each of which included 2–5 cue objects, to evaluate indoor localization performance and find the optimal number of cue objects.

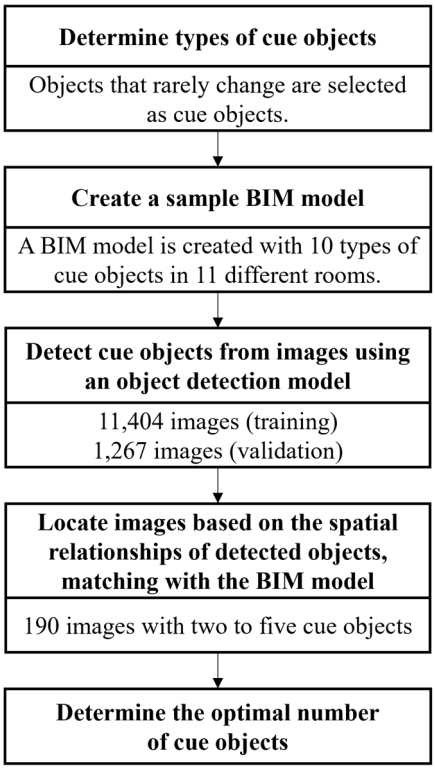

Fig. 1: Research flowchart

3.1 Selection of Cue Objects

For the experiments, we set the following criteria for cue objects:

1) The appearances or positions of cue objects should rarely change; thus, transient items, such as tables and posters, should not be considered cue objects.

2) Ideally, objects should be unique to a certain space and representative of the space. However, since a few objects remain constant over time and are exclusive to a particular space, non-unique objects, such as light switches or doors, could also be considered cue objects.

To determine positions using the spatial relationships among objects that remained relatively constant, we selected objects that fulfilled the above criteria, which resulted in the 10 cue objects listed in Table 1, including three types of doors, a window, a power socket, a light switch, a sink, a toilet bowl, a showerhead, and a kitchen cabinet, being chosen for the experiments.

3.2 BIM Model Creation

We created a sample BIM, we created a model of a housing unit for the experiment. Fig. 2 presents the axonometric view and the plan of the unit model. The model consisted of 11 rooms (the living-dining-kitchen [LDK] space, bedroom 1, bedroom 2, bedroom 3, bedroom 4, bathroom 1, bathroom 2, pantry, closet, balcony, and entrance), all of which had boundary walls, except for the LDK space and the entrance. Although some rooms were significantly different from one another (e.g., bedroom 1 and 2), there were also similarities between certain rooms (e.g., bedrooms 2 and 3). Table 1 provides detailed room information for the housing unit, including the room number, name, and list of cue objects present inside each room, along with their respective quantities. We placed the cue

objects in plausible locations and varied their numbers and placements. We classified cue objects with differing appearances within the same category as distinct types. For example, we categorized doors into three different types. Doors A and B were both indoor wooden doors, but door B differed from door A by having two panels. Meanwhile, door C was a steel front door.

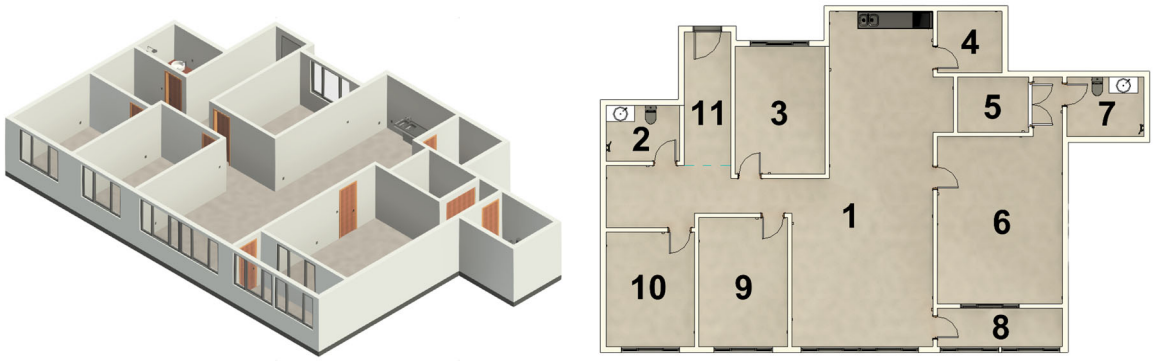

Fig. 2: Axonometric view of the housing unit model (left) and plan of the unit (right)

Table 1: Room information (number, name, and cue objects) for the BIM model of housing unit

Room Number	Room Name	Object Name and Quantity
1	LDK	Door A, 7; Window, 2; Power socket, 8; Light switch, 5; Kitchen cabinet, 1
2	Bathroom 1	Door A, 1; Power socket, 1; Sink, 1; Toilet, 1; Showerhead, 1
3	Bedroom4	Door A, 1; Window, 1; Power socket, 1; Light switch, 1
4	Pantry	Door A, 1; Power socket, 1
5	Closet	Door B, 1; Power socket, 1; Light switch, 1
6	Bedroom 1	Door A, 2; Door B, 1; Window, 1; Power socket, 4; Light switch, 2
7	Bathroom 2	Door A, 1; Power socket, 1; Sink, 1; Toilet, 1; Showerhead, 1
8	Balcony	Door A, 1; Window, 2; Power socket, 1
9	Bedroom 3	Door A, 1; Window, 1; Power socket, 1; Light switch, 1
10	Bedroom 2	Door A, 1; Window, 1; Power socket, 1; Switch, 1
11	Entrance	Door C, 1

3.3 Image Dataset Preparation

To train the model and evaluate the performance of the indoor localization method, we generated 12,861 BIM screenshot images, of which 12,671 were used for object detection and 190 for indoor localization method evaluation. Specifically, we used 11,404 images (roughly 90% of the object detection dataset) for training and 1,267 images for model validation. The dataset with 190 images for the indoor localization method was labeled differently from the previous dataset. The dataset for object detection was labeled according to bounded boxes, whereas the dataset for localization was annotated according to the information for the target room. Table 2 shows the major characteristics of the two image datasets. We created the dataset for object detection using a script that automatically captured the appearance of objects within the BIM model and labeled them accordingly. However, we manually created the dataset to validate the indoor localization method by directly capturing BIM model views.

Table 2: Characteristics of each image dataset

	Dataset for object detection	Dataset for indoor localization method evaluation
Purpose	To train and validate the object detection model	To validate the indoor localization method
Labeling	Cue objects with bounded boxes	Rooms and associated cue objects
Data size	11,404 (training)/1,267 (validation)	190 (validation)
Creation method	Automatically captured BIM model views	Manually captured BIM model views
Image size	785 × 785	1024 × 767

The images used for training and validating the object detection model were square, measured 785 pixels on each side, and were rendered in a realistic style. The dataset creation method is depicted in Fig. 3. We employed visual scripting to automatically generate these images. The viewing point from the camera's location and the target point where it is directed are both required to capture BIM screenshots. We began the viewing point and target point acquisition process by extracting room boundaries from the model and calculating the midpoint of each boundary side. We established viewing points by vertically elevating the midpoint of each boundary 150 cm from the floor to position the camera at average eye level. Viewing points were positioned along room boundaries rather than at the room centroids to capture images from the maximum distance within a room and thereby capture a greater number of cue objects. The scale of the captured objects was similar to that of the objects captured in the indoor localization method evaluation dataset when the viewing points were positioned along room boundaries. We then set target points by vertically elevating the midpoint of each boundary, spanning 40–200 cm, at intervals of 20 cm from the floor. To capture the desired views, we positioned a camera with a field of view (FOV) of 50°, which is the base angle of a normal lens, at the viewing point and directed it toward the target points. We repeated this process for each room in the BIM housing unit model. We produced the initial BIM images using Dynamo. Each was 1,047 × 785 pixels and was subsequently cropped into left, center, and right portions to create 785 × 785-pixel images. We removed redundant images that did not contain any cue objects. In total, we generated 12,671 images and used them to train and validate the object detection model.

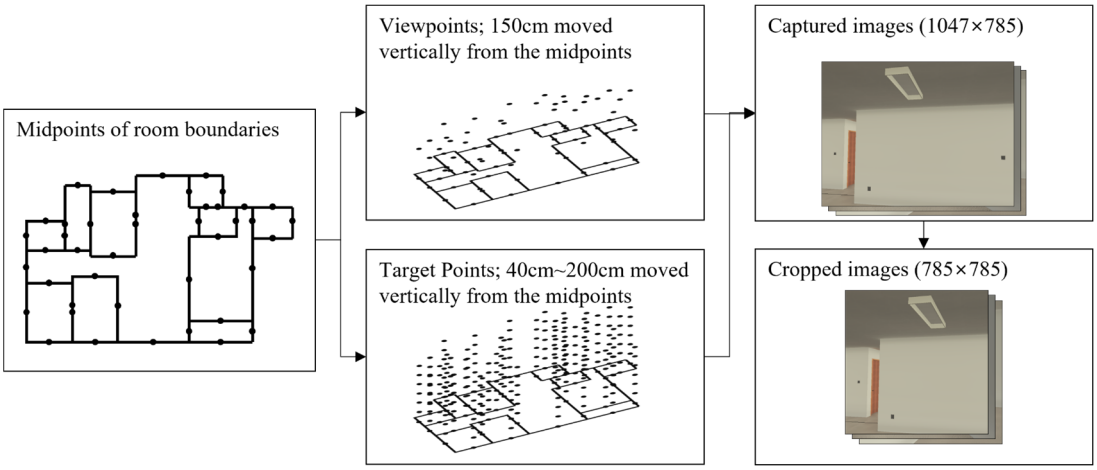

Fig. 3: Dataset creation for the object detection model

Additionally, we manually generated 190 images with varying numbers of cue objects, ranging from two to five, to determine the optimal number of cue objects for an image. The images were divided based on the number of cue objects (n) present in each image. We determined the range of n based on the following rationale: To deduce the location based on the spatial relationship between objects, at least two cue objects were minimally required. For the maximum number of cue objects, assuming that all cue objects were accurately detected, having more cue objects in the image made it easier to accurately determine the location. However, it was very unlikely for an image to include more than five cue objects. Moreover, as the number of cue objects necessary for the proposed method increased, the cumulative object detection error rate increased accordingly. Therefore, we set the maximum n-value to 5. Table 3 provides the distribution of images across the rooms. For scenarios with 2, 3, or 4 cue objects, 50 images were captured for each scenario. For scenarios with 5 cue objects, 40 images were captured, due to the limited existence of views that met the criteria. We determined the number of images for each room based on the

availability of the desired view and the size of the room. Initially, we assessed whether each room could provide a view with a certain number of cue objects within the FOV of 50°. Then, considering the available rooms, we allocated the number of images for each room proportionally based on their respective areas. As shown in Table 3, the number of available rooms decreased significantly as the number of cue objects in the image increased. To fulfill the research objective, the image needed to include all contiguous cue objects. Further details regarding this condition will be discussed in Section 3.4.

Table 3: Number of images taken for each room

	Number	1	2	3	4	5	6	7	8	9	10	11	Total
Room information	**Area (m^2)**	60	5	14	5	5	27	5	6	13	12	8	160
	Cue object quantity	23	5	5	2	3	10	5	5	5	5	1	69
Number of images	***n* = 2**	20	2	5	0	2	9	2	2	4	4	0	50
	***n* = 3**	20	2	5	0	2	9	2	2	4	4	0	50
	***n* = 4**	29	2	0	0	0	13	3	3	0	0	0	50
	***n* = 5**	34	0	0	0	0	6	0	0	0	0	0	40

3.4 Object Detection

We used You Only Look Once (YOLO) (Redmon et al., 2016) for this study, which is one of the most widely used networks for object detection. We trained the model using 11,404 images and set aside 1,267 images for model validation. To rationalize the indoor location and the spatial relationships among cue objects within an image, we applied the trained object detection model to the indoor localization method evaluation image dataset, which varied the number of cue objects contained in each image.

3.5 Indoor Location Reasoning

The goal of indoor location reasoning is to determine the locations at which the positional relationships between cue objects obtained through object detection in the images align with the positional relationships the cue objects have within the BIM model. This involves analyzing the X and Y coordinates of the bounding boxes of cue objects in the images to infer whether one object is to the left or right of another object, or above or below it. However, in this experiment, we specifically focused on the left–right relationships between cue objects, as they tended to have fewer variations and provided greater accuracy. Based on the object detection results, we created a cue object list by arranging the objects in ascending order according to their X-coordinate values.

To identify the locations where the BIM information matched the information from the image, we considered how the model's information would manifest in the image. To determine which objects could be observed to the left or right of a specific object when taking a photo, we employed clockwise ordering of the objects present in each room. First, we extracted the positions of the cue objects and the room boundaries, which enabled us to determine the locations and relationships of cue objects within each room. Based on the extracted information, we sequentially listed all the objects in a clockwise direction along the boundaries of each room. Subsequently, to ensure that the list represented the relationships between objects, regardless of the starting point, we copied the elements from the front of the list, counting one less than the number of objects found in the image, and added them to the end of the list. If the quantity of elements added to the list exceeded the count of objects identified in the picture minus one, it could potentially result in duplicates during localization. Conversely, if the number of added elements was less than the count of objects identified in the picture minus one, it could lead to potential omissions during localization. Fig. 4 depicts an example. If the cue objects in a room were arranged clockwise as a, b, c, d, and e, the original list was [a, b, c, d, e]. If the image contained three cue objects, additional elements of the list 'a, b', which was one less than the total number of cue objects in the image, were appended at the end of the list, resulting in [a, b, c, d, e, a, b]. Matching parts were then sought between the cue object list created for each room and the list of cue objects present in the image. The matched cue objects were considered candidates for the location from which the photo was taken. For instance, if the cue object list for room A is [a, b, c, d, e, a, b], and the cue object list for the image is [a, b, c], there is one matching object arrangement. Therefore, [a, b, c] inside room A ([**a, b, c**, d, e, a, b])

becomes a candidate location. There could be multiple candidates for each image, or no candidates if the object detection result was incorrect.

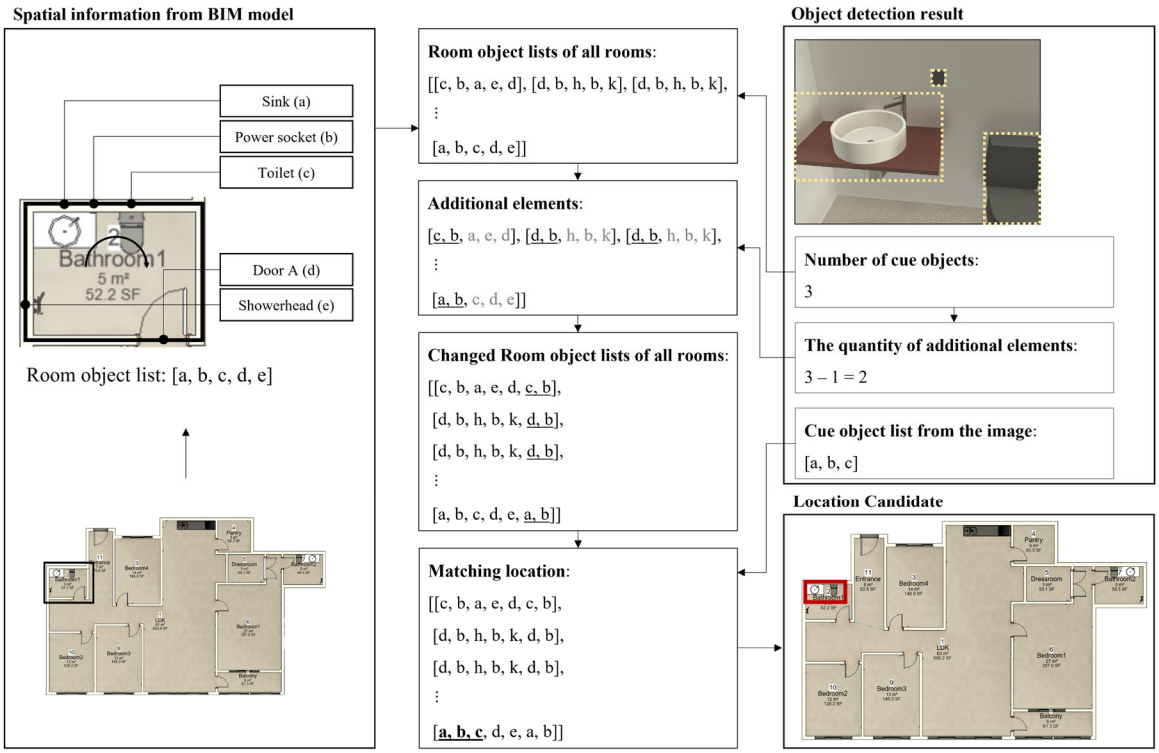

Fig. 4: The process of object location matching

4. RESULTS

4.1 Object Detection

Table 4 presents the model's performance metrics. The object detection model used in the study achieved $mAP_{0.5}$ of 0.9403 and F1 score of 0.9595 (Table 4). Accurate object detection within images certainly resulted in improved performance in subsequent indoor localization tasks because our proposed method relies on the results of object detection; however, the performance tended to degrade exponentially as the number of objects within the images increased as long as the object detection performance reached 100%. The results in Table 4 show that the overall performance decreased, theoretically, by about 6% on average with the addition of each cue object, considering the $mAP_{0.5}$ performance. The proposed localization method also relies on location reasoning, which is positively influenced by increases in cue objects. Thus, the number of cue objects, whether too large or too small, can be detrimental to performance, emphasizing the importance of optimizing the number of cue objects. In addition, Fig. 5 illustrates that the method performed better for objects with less skewness and larger sizes. Due to the presence of only one door B in the BIM model with significant skewness in the image, the accuracy was low for this object. The accuracy for light switches (the smallest of the objects) was slightly lower than for the other objects. Based on the results, it appears that choosing larger cue objects for the localization method would probably have resulted in improved performance.

Table 4: Performance of the object detection model based on the validation dataset

Precision	Recall	F1 score	mAP_0.5
0.9914	0.9295	0.9595	0.9403

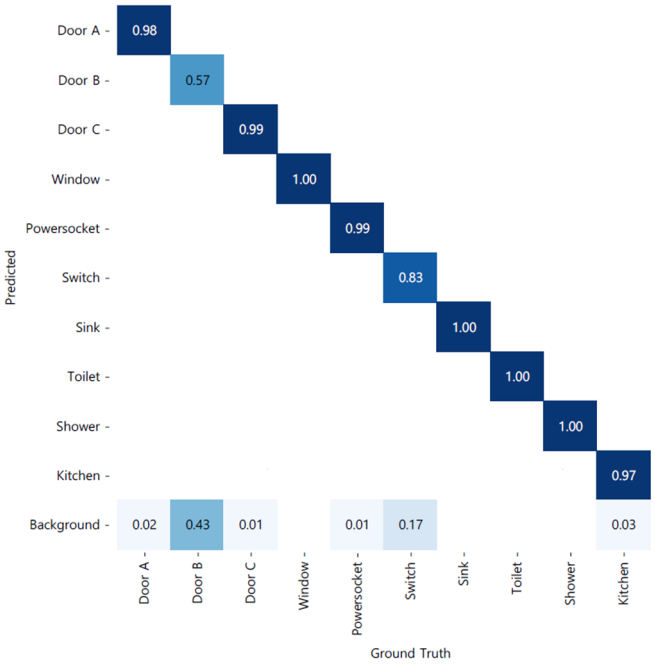

Fig. 5: Confusion matrix for the object detection results

4.2 Optimal Number of Cue Objects

Table 5 shows the evaluation results for the indoor localization method based on object detection and location reasoning for each number of cue objects in the images. Since the first step of the proposed method is to correctly detect every cue object in an image, only cases of every cue object in an image being predicted correctly are considered correct cases. As the number of objects in the image increased, object detection accuracy tended to decrease. This decrease in accuracy was minimal when the number of cue objects (n) changed from 2 (0.88) to 3 (0.80). The magnitude of the decrease was exponential, resulting in an accuracy of 0.20 when $n = 5$. The magnitude of the decrease was more significant than the theoretically estimated 6% decrease for the addition of an object, which was based on the object detection model performance for an individual object.

To identify the correct location among the candidate locations, the object detection results must be accurate. If object detection yields incorrect results, there may be no correct candidates or no candidates present. However, if the object detection results are correct, then at least one of the generated candidates is guaranteed to be correct. Therefore, after performing object detection, we conducted location reasoning for cases where there was a correct candidate and counted the number of generated candidates. Assuming that there was a correct candidate among the generated candidates, we calculated the probability of finding the correct location. However, in cases where the object detection results are wrong, there may be no correct candidate. Therefore, we multiplied the probability of finding the correct location when the correct location was present among the candidates by using object detection accuracy to calculate the probability of locating the correct position using the given method.

We used three metrics to evaluate the performance of the indoor localization method: (A) the mean number of location candidates when the correct location was present among the candidates, (B) the mean probability of finding the correct location when the correct location was present among the candidates, and (C) the mean probability of finding the correct location for both cases when the correct location was present among the candidates and when it was not. The three metrics provided answers to three research questions: (A) How many location candidates will be generated, depending on the number of cue objects? (B) What is the probability of finding the correct location from among the location candidates if the object detection is conducted correctly? (C) What is the probability of finding the correct location with the proposed method, considering both object detection accuracy and the performance of location reasoning. Metric (C) was the primary metric for assessing the model's performance. To compare the performance of the model with and without the influence of object detection, we considered the results based on predicted and actual object information. (A) decreased as the number of cue objects

in the image decreased and reached 1.000 when $n = 5$, and (B) could be calculated as the mean value of the inverse of (A), with a higher value indicating better performance. Since (C) considered both the object detection accuracy and the localization reasoning performance, it could be calculated by multiplying the object detection accuracy with (B); hence, a higher value of (C) indicated superior overall model performance.

When cases with correct object detection were considered, the localization performance increased as n increased. However, localization performance did not increase proportionally to n because the accuracy of object detection significantly decreased as n increased. Therefore, when using the predicted object information, the highest performance was observed at $n = 3$, with a probability of 0.283 for finding the correct location. At $n = 4$, the probability of finding the correct location was 0.276, which represented a slight decrease but showed a similar performance to that at $n = 3$. Therefore, even a slight improvement in object detection accuracy for $n = 4$ had the potential to yield a better score than the case at $n = 3$. This result shows that the presence of three to four cue objects in the image yielded optimal results within the given framework.

Table 5: Model evaluation results

Number of cue objects in the image (n)	Object detection accuracy	Localization performance based on predicted object information			Localization performance based on actual object information		
		(A)	(B)	(C)	(A)	(B)	(C)
2	0.88	3.841	0.260	0.229	3.800	0.263	0.263
3	0.80	2.825	0.354	0.283	2.816	0.355	0.355
4	0.52	1.885	0.531	0.276	1.700	0.588	0.588
5	0.20	1.000	1.000	0.200	1.000	1.000	1.000

(A) The mean number of location candidates when the correct location was present among the candidates

(B) The mean probability of finding the correct location when the correct location was included among the candidates.

(C) The mean probability of finding the correct location for both cases when the correct location was present within the candidates and when it was not.

5. CONCLUSION

Many previous studies on indoor localization have been based on the similarities between photos and BIM images. However, these approaches may exhibit weaknesses under varied lighting conditions and with different wallpapers and furniture locations. To overcome these limitations, we propose a reasoning-based approach based on cue objects in photos. With this approach, it is essential to optimize the number of cue objects for detection since it significantly influences performance. Hence, the aim of this preliminary study was to find the optimal number of cue objects in a photo that yielded the best localization performance. The proposed localization method uses spatial information on cue objects detected by a computer vision algorithm to locate shots by analyzing the spatial relationships among the objects found in an image and comparing them with those in the BIM model. We evaluated the method's performance by assessing the probability of accurately determining the location from which a photo was taken, varying the number of cue objects in each photo from two to five.

The experimental results indicated that the model showed the best performance when three cue objects were present in an image. When the number of cue objects increased to four, the probability of accurately determining the exact location decreased slightly compared to the case with three cue objects, mainly due to the dramatic decrease in object detection accuracy. As the number of cue objects captured from the image increased, the number of small and skewed objects also tended to increase, which led to a decrease in overall accuracy. Having a higher number of cue objects can make it easier to deduce the location of a shot accurately, but it may decrease object detection accuracy.

The major contribution of this study lies in suggesting the optimal number of cue objects that should be present in images to determine the locations of photos shot in indoor spaces. This finding highlights the importance of

balancing detection performance and reasoning capability in object detection-based indoor localization and considering the number of detected objects. The experimental results provide insights into areas for improvement in future research. First, the method did not perform well when cue-object arrangements in rooms were similar. This limitation could be addressed by considering the size of cue objects and incorporating more diverse spatial information. Second, the error rate accumulated at each stage: the cue-object detection stage, the spatial relationship detection stage, and the location deduction stage. Further research is expected to improve the proposed method to a practically applicable level. The results of this research will be integrated into construction management and maintenance software, enabling the automatic tagging of locations in provided images.

6. ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C3008209).

REFERENCES

Acharya, D., Khoshelham, K., & Winter, S. (2019). BIM-PoseNet: Indoor camera localisation using a 3D indoor model and deep learning from synthetic images. ISPRS Journal of Photogrammetry and Remote Sensing, 150, 245–258. https://doi.org/10.1016/j.isprsjprs.2019.02.020

Acharya, D., Roy, S., Khoshelham, K., & Winter, S. (2020). A recurrent deep network for estimating the pose of real indoor images from synthetic image sequences. Sensors, 20(19), 1–20. https://doi.org/10.3390/s20195492

Alam, M., Hossain, A. K. M., & Mohamed, F. (2022). Performance evaluation of recurrent neural networks applied to indoor camera localization. International Journal of Emerging Technology and Advanced Engineering, 12(8), 116–124. https://doi.org/10.46338/ijetae0822_15

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), Computer Vision – ECCV 2006 (pp. 404–417). Springer. https://doi.org/10.1007/11744023_32

Guan, K., Ma, L., Tan, X., & Guo, S. (2016). Vision-based indoor localization approach based on SURF and landmark. 2016 International Wireless Communications and Mobile Computing Conference (IWCMC), (pp. 655–659). https://doi.org/10.1109/IWCMC.2016.7577134

Ha, I., Kim, H., Park, S., & Kim, H. (2018). Image retrieval using BIM and features from a pretrained VGG network for indoor localization. Building and Environment, 140, 23–31. https://doi.org/10.1016/j.buildenv.2018.05.026

Kang, H., Park, Y., & Kim, Y. (2019). Improvement model of defect information management system for apartment buildings. Korean Journal of Construction Engineering and Management, 20(4), 13–21. https://doi.org/10.6106/KJCEM.2019.20.4.013

Kim, D., & Kim, J. (2023). CT-Loc: Cross-domain visual localization with a channel-wise transformer. Neural Networks, 158, 369–383. https://doi.org/10.1016/j.neunet.2022.11.014

Kim, J. (2022). Identifying indoor locations of close-up photos using deep learning and building information modeling objects. Yonsei University. http://www.riss.kr/link?id=T16372630

Kim, K.-T., Lim, M.-G., & Kim, G.-T. (2014). History management technology of building construction and maintenance using vector photo information and BIM. Journal of the Korea Institute of Building Construction, 14(6), 605–613. https://doi.org/10.5345/JKIBC.2014.14.6.605

Li, Y., Kambhamettu, R. H., Hu, Y., & Zhang, R. (2022). ImPos: An image-based indoor positioning system. 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC), (pp. 144–150). https://doi.org/10.1109/CCNC49033.2022.9700699

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Computer Vision and Pattern Recognition Conference (CVPR), (pp. 779-788). https://openaccess.thecvf.com/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. arXiv. https://doi.org/10.48550/arXiv.1409.1556