Image-based Indoor Localization using BIM and Features of CNN

Inhae Ha[a], Hongjo Kim[a], Somin Park[a], and Hyoungkwan Kim[a]

[a]School of Civil and Environmental Engineering, Yonsei University, Seoul, South Korea
E-mail: haine919@yonsei.ac.kr, hongjo@yonsei.ac.kr, somin109@yonsei.ac.kr, hyoungkwan@yonsei.ac.kr

Abstract –

This study suggests an indoor localization method to estimate the location of a user of a mobile device with imaging capability. The proposed method uses a matching approach between an actual photograph and a rendered BIM (building information modeling) image. A pre-trained VGG 16 network is used for feature extraction. Experimental results show that the best image matching performance can be obtained when using features from pooling layer 4 of VGG16. The proposed method allows for indoor localization only by image matching without additional sensing information.

Keywords –

Indoor Localization; Cross-Domain Image Matching; Convolutional Neural Network; Feature Extraction;

1 Introduction

A technique of acquiring the position and orientation information of a person in indoor environment is essential for presenting the information that the person needs at the location. The construction industry utilizes indoor localization technology for maintenance of facilities using augmented reality [4], evacuation route guidance in case of disasters [3], and understanding work situation [1,7].

Vision-based indoor localization estimates the user's location based on visual information obtained from the indoor image. Image-based indoor localization methods can utilize a pre-built image dataset that contains indoor photographs. Since the images in the dataset have the basic information (position and orientation) necessary for localization, the person's indoor position can be estimated by searching the image in the dataset, most similar to the photograph taken indoors. However, it is time-consuming and labor intensive to build datasets of indoor environments for localization.

This study utilizes the BIM (building information modelling) model of a building for its indoor localization, to construct the image dataset. BIM has become increasingly utilized as it is proving its versatile utilities in the construction industry. Since BIM contains a range of information of a building and it is easy to extract information at the location of interest, BIM is also used for localization with sensors [6-8].

This paper proposes a deep learning-based method to estimate the indoor position of a mobile device user, using an image dataset constructed from a BIM model. A deep learning network is an advanced form of traditional neural network, strengthened by the ability to learn important features without relying on human intervention. Image features extracted from a deep learning network are used to compare similarity between photograph and BIM-based images. That is, the visual characteristics of BIM is used for image-based indoor localization, without using additional sensing information.

2 Methodology

The image dataset is constructed by rendering indoor BIM views to images. The rendered images are visually similar to the actual indoor photographs, but there is a difference in style because the domains are different from each other, as shown in Figure 1. The evaluation of the similarity between images for image retrieval is mainly done by comparing their features. Conventional feature extraction methods, such as SIFT (scale-invariant feature transform), have been widely used for the same domain comparison in previous studies.

Figure 1. BIM image (left) and photograph (right) taken from the same location and orientation

This study proposes using the CNN (convolutional neural network) for feature extraction to compare indoor photographs and BIM images. CNN is a kind of deep neural network, appropriate for image processing. It consists of stacked layers and learns to extract meaningful features of images. In a convolutional layer, features in adjacent parts in image are extracted in a form of two dimensional array (feature map) and semantically related features are merged in a pooling layer [5].

Feature maps that passed through each layer of pre-trained CNNs are used as features of images for the cross-domain image retrieval. The CNN network trained for object classification with ImageNet dataset [2], a large image dataset, shows excellent performance in feature extraction even when applied to other datasets [9].

A pre-trained CNN is used to match indoor photographs with BIM images in the dataset based on the similarity between images. The evaluation of the similarity between the cross-domain images for image matching is mainly done by comparing their features with cosine distance. Figure 2 shows the proposed method.

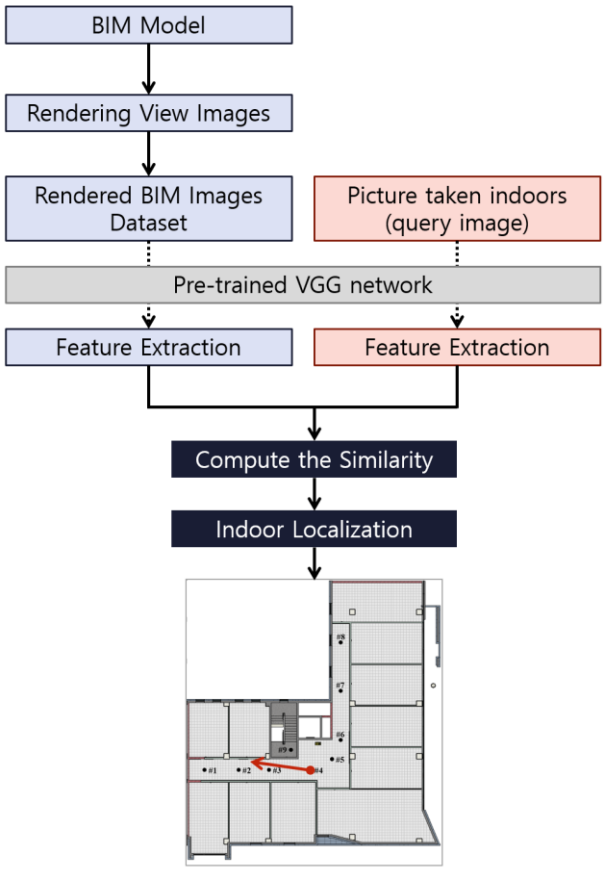

Figure 2. Image matching for indoor localization using BIM and CNN features.

3 Experiment

Experiments were carried out at the North Wing of the 1st Engineering Building of Yonsei University in Korea. As shown in Figure 3, the BIM image dataset for the experiment consists of images rendered in various directions at nine locations.

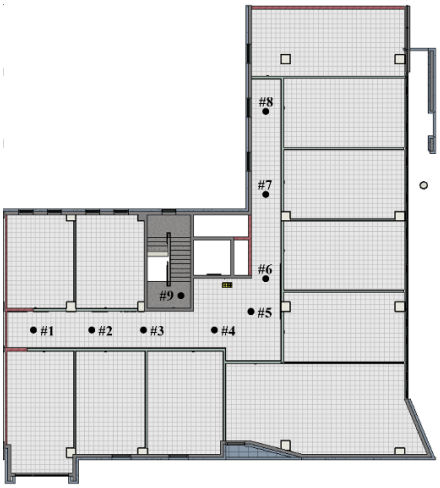

Figure 3. Floor plan of the building where the experiment was performed and the locations selected for view extraction.

Two datasets were created to evaluate the performance of the method. Level 1 dataset contained images with relatively fewer overlapping views as shown in Figure 4(a), and Level 2 dataset contained views in all directions, which resulted in more overlapping views as shown in Figure 4(b). The indoor photographs were taken in the same location and direction as the BIM images of the dataset, and the similarity with the dataset images was evaluated.

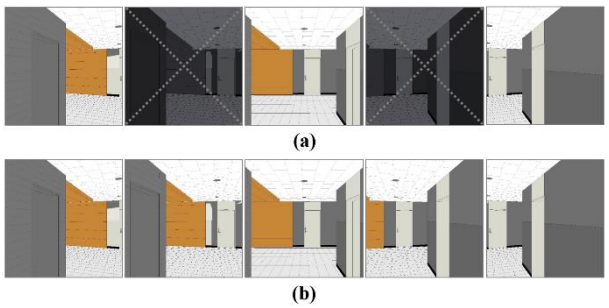

Figure 4. Examples of dataset images; a) Level 1: fewer overlapping views; b) Level 2: more overlapping views.

The VGG 16 network [10] was used as a CNN in the experiment. There are five pooling layers in the VGG

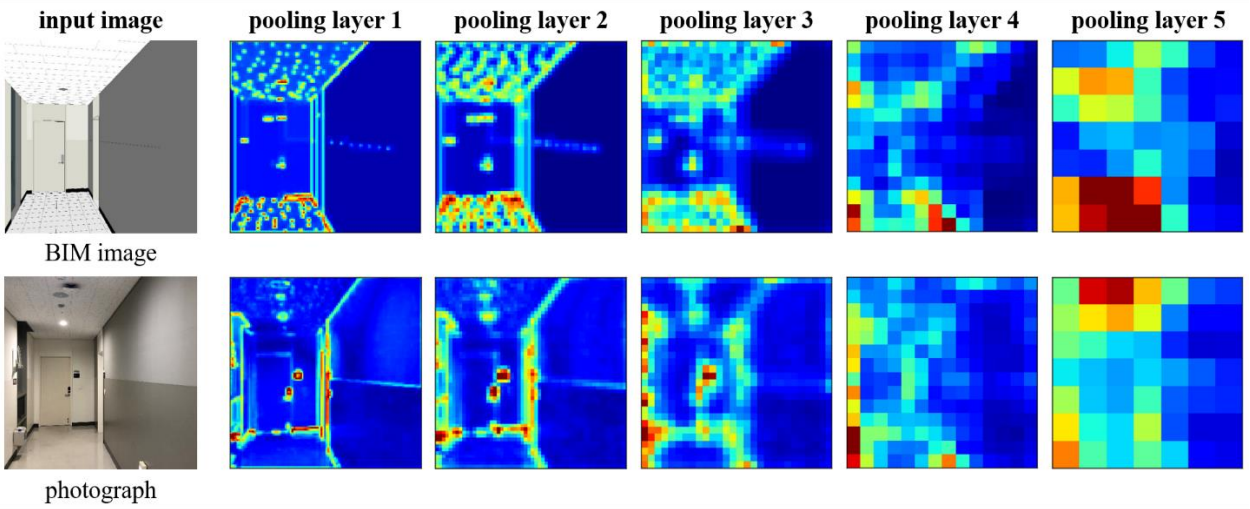

Figure 5. Visualized feature maps of BIM image and indoor photograph

network. The VGG network in the experiment was pre-trained with ImageNet dataset and the outputs of five pooling layers of the pre-trained VGG 16 network were used as features of the image. To evaluate the similarity between the cross-domain images, features from the pooling layer of the same order were employed. For example, the second layer output of the network for the actual photograph was compared with the second layer outputs of the network for the BIM images.

Table 1 shows the results of image matching based on features obtained from each pooling layer of the pre-trained VGG network for the two datasets. When using the features obtained from pooling layer 4, the image matching accuracy is over 90% in both datasets.

Table 1. Experimental Result

Dataset (number of images)	Level 1 (54)	Level 2 (86)
	Accuracy (%)	
Pooling layer 1	64.81	54.65
Pooling layer 2	64.81	52.33
Pooling layer 3	79.63	77.91
Pooling layer 4	**92.59**	**90.70**
Pooling layer 5	74.07	75.58

4 Discussion

Experimental results verified the proposed image-based indoor localization method with the high matching accuracy. The BIM image that was selected as the most similar image to the indoor photograph had the information about indoor location and orientation. In other words, when the matching was correctly done, the position at which the indoor photograph was taken could be estimated by the proposed method.

This study revealed which layer of VGG 16 network extracts the most proper features for matching the indoor photograph and BIM image. Figure 5 exhibits visualized feature maps from each pooling layer when paired BIM image and indoor photograph pass through the VGG 16 network. As the network deepens, it can be observed that features corresponding to large parts of the image are gradually extracted. Through the features obtained from pooling layer 4, which shows the best performance, it can be inferred that a global descriptor representing the structural information of the indoor is suitable for the cross-domain image matching.

As shown in Figure 5, a noticeable difference in the arrangement of colors can be identified in feature maps extracted from the front of the network. In other words, when the domains between images are different, the features that stand out in the images can be different and the difference is more apparent for features extracted from a local area. Therefore, it can be confirmed that the global descriptor is required for cross-domain image matching, and the experimental result verified that the features extracted from pooling layer 4 have the best image matching capability.

The proposed method performed well both in Level 1 and Level 2 datasets. In Level 2, there were images with high degree of view overlap between images and the number of images was about 60% larger as compared with Level 1. However, the accuracy of image matching of Level 2 was not affected much–decrease by 2% compared with Level 1. These experimental results indicate that the proposed method can be applied well in larger indoor environments.

5 Conclusions

This paper proposes an image-based indoor localization method to identify a mobile device user's location and orientation by matching the indoor photograph acquired by the user to the image rendered from the BIM model. Features for evaluating the similarity between the indoor photograph and the rendered BIM image were extracted from the pre-trained VGG 16 network. A field experiment, involving a floor of an actual building, was conducted to verify the proposed method.

This study has three major contributions. First, since the existing BIM model is used, the labor required to construct the dataset is reduced compared to the existing methods. Secondly, the feature extraction layer most suitable for the cross-domain image comparison was identified in the VGG network. Finally, since the pre-trained CNN is utilized, the proposed method can be applied in another place without modifying the network.

As a future study, it is necessary to verify the proposed method with a dataset configured in a more diverse environment. In addition, since using only a single image for localization may have limitation, it can be supplemented by using multiple images for more accurate localization. The proposed method, with the additional research, is expected to improve the traditional facility management process.

Acknowledgements

This work was supported by a grant (18CTAP-C133290-02) from Infrastructure and transportation technology promotion research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

[1] A.M. Costin, J. Teizer, B. Schoner, RFID and BIM-enabled worker location tracking to support real-time building protocol and data visualization, Journal of Information Technology in Construction (ITcon) 20 (29) (2015) 495-517.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE, 2009, pp. 248-255.

[3] R. Giuliano, F. Mazzenga, M. Petracca, M. Vari, Indoor localization system for first responders in emergency scenario, Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International, IEEE, 2013, pp. 1821-1826.

[4] C. Koch, M. Neges, M. König, M. Abramovici, Natural markers for augmented reality-based indoor navigation and facility maintenance, Automation in Construction 48 (2014) 18-30.

[5] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436-444.

[6] N. Li, B. Becerik-Gerber, B. Krishnamachari, L. Soibelman, A BIM centered indoor localization algorithm to support building fire emergency response operations, Automation in Construction 42 (2014) 78-89.

[7] J. Park, J. Chen, Y.K. Cho, Self-corrective knowledge-based hybrid tracking system using BIM and multimodal sensors, Advanced Engineering Informatics 32 (2017) 126-138.

[8] J. Park, Y.K. Cho, D. Martinez, A BIM and UWB integrated mobile robot navigation system for indoor position tracking applications, J. Constr. Eng. Project Manage 6 (2) (2016) 30-39.

[9] A. Sharif Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2014, pp. 806-813.

[10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014).