Building and Environment 140 (2018) 23–31

Contents lists available at ScienceDirect

Building and Environment

journal homepage: www.elsevier.com/locate/buildenv

Image retrieval using BIM and features from pretrained VGG network for indoor localization

Inhae Ha, Hongjo Kim, Somin Park, Hyoungkwan Kim*

School of Civil and Environmental Engineering, Yonsei University, South Korea

A R T I C L E I N F O

Keywords:
Image-based indoor localization
Cross-domain image retrieval
BIM
Feature extraction
Global descriptor

A B S T R A C T

Various devices that are used indoors require information regarding the user's position and orientation. This information enables the devices to offer the user customized and more relevant information. This study presents a new image-based indoor localization method using building information modeling (BIM) and convolutional neural networks (CNNs). This method constructs a dataset with rendered BIM images and searches the dataset for images most similar to indoor photographs, thereby estimating the indoor position and orientation of the photograph. A pretrained CNN (the VGG network) is used for image feature extraction for the similarity evaluation of two different types of images (BIM rendered and real images). Experiments were performed in real buildings to verify the method, and the matching accuracy is 91.61% for a total of 143 images. The results also confirm that pooling layer 4 in the VGG network is best suited for feature selection.

1. Introduction

Indoor localization is the process of understanding the location and orientation of an object in an indoor environment. Indoor localization methods are mainly used for augmenting visual information [1], indoor navigation [2], and tracking the indoor location of a person [3]. In the architectural, engineering, and construction (AEC) industries, indoor localization technology is used for various purposes, including facility maintenance using augmented reality [4], evacuation route guidance in emergency situations [5], and the tracking of workers or equipment [6–9].

In indoor environments, it is difficult to receive global navigation satellite system (GNSS) signals, which are mainly used for outdoor localization. Therefore, indoor localization methods using various sensors are being studied. Radio-based indoor localization is a method of estimating the position by transmitting and receiving signals. Radio-based indoor localization methods include radio-frequency identification (RFID) [7,10–13], Bluetooth [8,14–16], wireless local area networks (WLANs) [6,17–20], and ultra-wideband (UWB) technology [21–24]. However, these methods require preliminary tasks such as attaching a sensor in advance to the object whose indoor position is to be determined or making a map of the signal strength by location. In addition, radio-based methods are not very suitable for indicating the direction of an object because the position is determined by the signal [25].

Vision-based indoor localization has attracted attention as a cost-effective method [26]. Image-based indoor localization can utilize an image dataset for the prebuilt indoor environment and indoor photographs taken with a camera at the user's location. As the dataset has the prior information (position and orientation) necessary for localization, a person's indoor position is determined by extracting the image most similar to the photograph taken indoors from the dataset. The types of images in the dataset are monocular images [2,25,27–31] and omnidirectional images [32–34]. The use of an omnidirectional image is more advantageous for estimating the trajectory and orientation than the use of a monocular image, but it requires more cameras than a monocular image, which increases the cost and makes photographing more difficult.

In order to determine the indoor location information, a map of the indoor environment is required. The map is usually obtained from a three-dimensional (3D) model, which requires a number of indoor images or point clouds obtained from equipment such as laser scanners. Indoor images can be used for the retrieval of similar images in indoor localization and are also used in 3D modeling through structure from motion (SfM) algorithms [28,30,31,35]. These methods require a separate process to take photographs or point clouds to build a model including every location where indoor localization is desired.

Another way to use images for indoor localization is to construct a dataset by rendering a view in the form of a two-dimensional (2D) image from the 3D model of a building and to compare it with a photograph taken in a real indoor environment [27,33,36]. In Refs. [27,33,36], linear features are extracted from images to identify the

* Corresponding author.
E-mail addresses: haine919@yonsei.ac.kr (I. Ha), hongjo@yonsei.ac.kr (H. Kim), somin109@yonsei.ac.kr (S. Park), hyoungkwan@yonsei.ac.kr (H. Kim).

https://doi.org/10.1016/j.buildenv.2018.05.026
Received 13 February 2018; Received in revised form 11 May 2018; Accepted 12 May 2018
Available online 14 May 2018
0360-1323/

locations related to the features. The lines of an image representing an indoor scene are related to the structural characteristics of each location; therefore, they can be useful for determining the location. Although these studies have significantly advanced the body of knowledge in image-based indoor localization, there remains room for improvement. Higher levels of performance may be possible with additional information such as the color or texture of an image. Simpler or lighter equipment settings may provide the user with a more cost-effective solution for indoor localization.

Building information modeling (BIM) is increasingly being used in a wide variety of areas in the construction industry owing to its effectiveness [37–39]. Once a BIM model is developed, it is easy to obtain a desired map from any position; therefore, there have been attempts to utilize BIM in applications that use localization [7,8,24,29,40]. For localization, BIM is mostly used with sensors such as RFID [7], Bluetooth [8], radio-frequency beacons [40], and UWB [24]. Application-wise, BIM is used to provide geometric information [8,40], robot path extraction [24], and the integrated management and visualization of the information provided by the sensors [7]. The study presented in Ref. [29] proposes a method for recognizing the corner of a tile in an indoor floor photograph and identifying the indoor location using the coordinate information of the tile in BIM. On the basis of the utility of BIM, this paper presents an image-based indoor localization method that retrieves rendered BIM images most similar to the photographs taken from the user's location.

As shown in Fig. 1, the visual characteristics of an image rendered from the 3D model differ from the photograph obtained in the actual indoor environment. It is therefore difficult to perform image retrieval by conducting a cross-domain comparison between images. Previous studies that aim to estimate the indoor location by image retrieval use the photographs acquired in the indoor environment for both the query image and the images in the dataset; scale-invariant feature transform (SIFT) [25,30,34] or speeded-up robust features (SURF) [32] have been used for feature extraction in the same domain for comparison purposes. Fig. 2 shows that SIFT works well for images of the same domain but can have problems if the image types are different.

In this study, a convolutional neural network (CNN) is used to extract features for cross-domain image retrieval. CNNs are deep learning networks suitable for image processing. Unlike existing handcrafted feature extraction methods, CNNs have exhibited excellent performance in various fields of computer vision, such as classification, segmentation [41], and image-based localization [42,43], using the ability to learn visual features by itself. CNNs trained for image classification have proven to be a powerful tool for extracting the generic features from an image [44]. In fact, the intermediate output of each layer can be used as a feature map in the image retrieval process [45,46]. Thus, the proposed method utilizes a CNN trained for classification. The CNN is pretrained by ImageNet [47], a large labeled image dataset for feature extraction, and the extracted features are used for image retrieval.

2. Methodology

Fig. 3 shows the proposed method of image-based indoor localization. The core idea is that images are processed through the CNN to produce their multiscale features. The features from the rendered BIM images and the indoor photograph are vectorized so that they can be compared with each other. Once the image most similar to the user's photograph is retrieved from the dataset, the user's indoor position is estimated.

CNNs consist of multiple layers of neurons that learn to extract the meaningful features of an image to perform a desired task. The types of layers that make up a CNN are typically divided into convolutional,

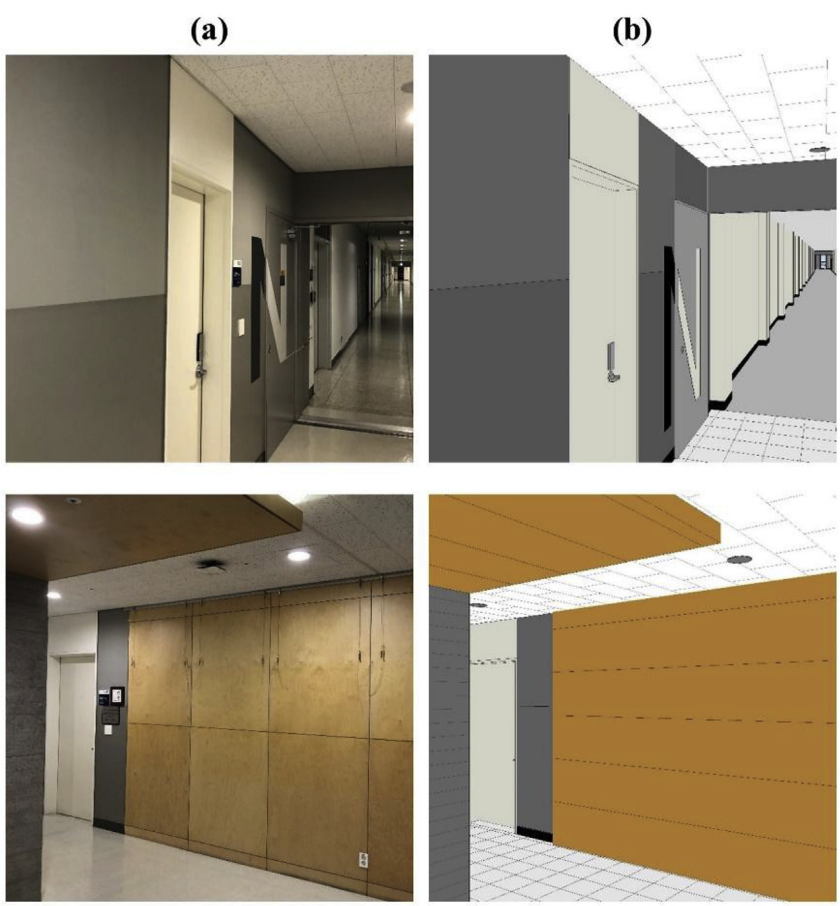

Fig. 1. Different image types: (a) indoor photographs and (b) images rendered by BIM.