# Place Recognition Using Multiple Wearable Cameras

Kyungmin Min, Seonghun Lee, Kee-Eung Kim, and Jin Hyung Kim

Korea Advanced Institute of Science and Technology,
373-1 Guseong-dong, Yuseong-gu, Daejeon, Korea
{kmmin, leesh, kekim, jkim}@ai.kaist.ac.kr

**Abstract.** Recognizing a user's location is the most challenging problem for providing intelligent location-based services. In this paper, we presented a real-time camera-based system for the place recognition problem. This system takes streams of scene images of a learned environment from user-worn cameras and produces the class label of the current place as an output. Multiple cameras are used to collect multi-directional scene images because utilizing multiple images yields better and robust recognition than a single image. For more robust recognition, we utilized spatial relationships between the places. In addition that, a temporal reasoning is incorporated with a Markov model to reflect typical staying time at each place. Recognition experiments, which were conducted in a real environment in a university campus, showed that the proposed method yields a very promising result.

**Keywords:** context recognition, place recognition, image understanding, wearable computing, hidden Markov models.

## 1 Introduction

Recognizing the situation of a user is an important problem for context aware intelligent services. Depending on the service to be provided, a user's context can be defined in various ways, such as location, current activity, physical state, and so on,. Among these, place information (a labeled location such as classroom, lobby, corridor, etc.) can be useful to provide various services such as mobility aids for the visually impaired [1], spatially-based notes and memory aids [2]. It can also be used as a basic feature to recognize high level contexts such as the user's current activity. For example, if we know a user is in a classroom, we may assume with high likelihood that he is attending a lecture.

Approaches for place recognition can be grouped into three categories; using a Global Positioning System (GPS), using Radio Frequency Identification (RFID) tags pre-attached on target places, and using sensors worn by the user. However, GPS has limited precision and only indoor availability. The RFID is also limited in the sense that we need to deal with the high setup cost of attaching a large number of tags in various places. Therefore, the approach based on user-mounted sensors is being actively pursued. Many kinds of sensors, such as microphones [3], accelerometers [4], and cameras [5] [6], can be used in this purpose. However, the image data from cameras is the most informative.

H. Ichikawa et al. (Eds.): UCS 2007, LNCS 4836, pp. 266–273, 2007.

In this paper, we focus on developing a camera-based place recognition system. The goal of our system is to recognize the user's current place in real time when the user navigates in a known environment with attached cameras.

Even though images contain rich information, it is hard to utilize the information efficiently because the images are usually degraded by motion blur, change of illumination, and several other factors. Previous approaches used only one camera capturing images in the front direction [5] [6]. However, these approaches are subject to incorrect recognition whenever the front-directional image (i.e, the single source of information) is degraded by noise or does not contain unique features of a certain place.

To make the inference process more robust, temporal reasoning was adopted in a number of previous studies. It uses sequences of visited places up to the previous time to recognize the current place. It was modeled often with the $1^{st}$ order Markov assumption [5] [6]. This modeling makes the computation simple by assuming only the first preceding place affects the current place. However, this assumption is not usually correct in the problem of place recognition.

In this paper, we propose two additional features to make up for the weak points of the previous camera-based approaches. The basic idea of each approach is as follows. The first one is to use multiple directional scene images obtained by using multiple cameras instead of single. With this, our system is able to recognize places more correctly. The second feature is to use information of staying time at a certain place for better temporal reasoning. In addition to the relationship between previous and current places, this temporal reasoning approach allows us to utilize better the map knowledge of target environment and, therefore, yields more robust recognition result.

The paper is organized as follows. Section 2 briefly describes the overview of the place recognition system and Section 3 explains the method of constructing the transition model. We validate in Section 4 the proposed method through the experiments conducted in the real environment and conclude in Section 5.

## 2 Place Recognition System Overview

The goal of the place recognition system is to determine the most likely current place from given image streams up to the current time. Our recognition system consists of learning module and recognition module. The learning module learns the transition probability and observation probability from the environment at the system development stage. The transition probability is the probability that the user moves from one place to another which is based on a given sequence of visited places up to the previous time step. The observation probability is the probability of a certain image to be observed in a given place. These two probabilities are combined as Hidden Markov Model (HMM) where the hidden nodes represent the user's place, the observation variables represent the captured image, and links represent the transition probabilities between nodes.

The recognition module performs real-time classification based on the models constructed by the learning module. The recognition module consists of a wavelet-based feature extractor and HMM-based classifier. The feature extractor uses the wavelet image decomposition method known as steerable pyramid [7]. By applying the steerable pyramid on the input image frame with four orientations and four scales, we got 16 decomposed images. To capture global image properties, we divided each decomposed image into 4x4 cells and take the mean of the feature values in the cell. Therefore we obtained 16 features per decomposed image and a total of 16x16=256 features per input image frame. By the principal component analysis, we finally obtained 80-dimensional feature vectors.

The observation probability for each directional image, $p(z_d|Q)$, is estimated with Parzen window method, a well-known non-parametric probability density estimation method. We used a Gaussian as the Parzen window function and the window size is obtained from several trials. With an assumption that every direction image is independent to each other, we obtain the observation probability of a set four images by multiplying the individual probabilities.

Then our recognition problem can be formulated as finding the maximal a posteriori probability place given image sequences with the trained HMM. Let $Q_t$ denote the place label at $t$ and $z_{1:t}$ denote image up to time $t$. Then, a posteriori probability is formulated as follows:

$$\begin{aligned} & P(Q_t = q \mid z_{1:t}) \propto p(z_t \mid Q_t = q) P(Q_t = q \mid z_{1:t-1}) \\ & and \\ & P(Q_t = q \mid z_{1:t-1}) = \sum P(Q_t = q \mid Q_{1:t-1}) P(Q_{1:t-1} \mid z_{1:t-1}) \end{aligned} \tag{1}$$

where $p(z_t|Q_t=q)$ represents the observation probability of image $z$ at place $q$ at $t$ and $P(Q_t=q|Q_{1:t-1})$ represents the transition probability to reach $q$. Since we already learned the observation probabilities with Parzen window method as described above, we only required to model the transition probability in order to solve this equation and obtain the posteriori probability.

## 3 Transition Modeling

In many problems using hidden Markov models, transition probability is learned from sample data because system designers usually do not have enough prior knowledge to construct the correct transition probability table. However, in the place recognition problem, the prior knowledge of the location of each place allows to clearly decide whether it is possible to move from current place to another. Therefore, we can build a robust model by determining transition probability along the prior knowledge rather than learning from sample data.

The probability of transition from the current place to the next place is determined by the possibility of the transition. The probability is set to zero when the next place is impossible to move from the current place. We assume that transitions to places reachable from the current place are equally likely.

To decide the validity of transition, we consider two kinds of constraints. First one states that a place transition can be occurred only between adjacent places. We call this constraint ‘spatial constraint’. The second constraint states that to move to another place from current place, one must stay in current place at least some time periods. We call this constraint ‘staying time constraint’.

In the next chapters, we will explain the concept of the constraints and discuss how to apply these constraints for solving this problem.

### 3.1 Spatial Constraint

In the real world, it is impossible for a user to be in a classroom just after appearing on the roof of the building. This is because there is a spatial constraint in the place transition. This means that within one time step, the user can either stay in the current place or move to adjacent place, but cannot move to non-adjacent places. To make the transition probability include this constraint, the neighborhood information between every pair of places is required. This information can be represented efficiently by using a graph structure, which is called place transition graph. It is defined as an undirected graph such that its nodes correspond to places and edges exist only between adjacent places.

Figure 1 shows an example of inferring the user's place using the spatial constraint. The map composed of lobby, corridor, lab and bathroom (Figure 1 (a)) is represented as the place transition graph (Figure 1 (b)). Then, even in the case that the current image features does not give enough information to decide whether the user is located in the corridor or the bathroom, the system could correctly decide the current place as the corridor by using the spatial constraint. (If the user was in the lab in the previous time step, he cannot be in the bathroom now because the lab and the bathroom is not connected) (Figure 1 (c)).

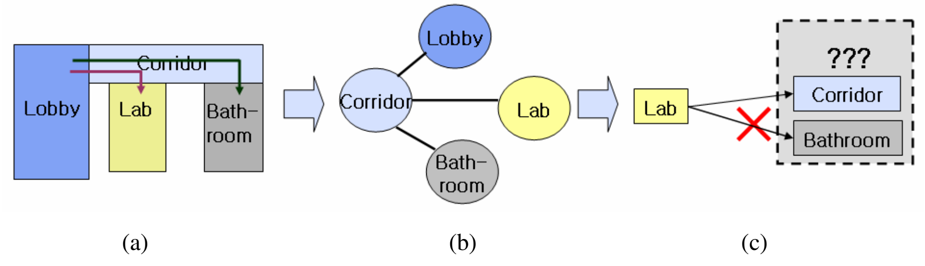

**Fig. 1.** Example of inference based on spatial constraint

### 3.2 Staying Time Constraint

As it is impossible to move between non-adjacent places directly, it is also impossible to move to a far place in a short time. For example, we can think of a situation that the user appears at 1st floor, stairs, and 2nd floor in 3 continuous time steps. If we only focus on the moving path, the path of '1st floor → stairs → 2nd floor' has no problem. However, in the real world, the situation is impossible because the staying time in stairs is too short to move from 1st floor to 2nd floor.

To summarize, it is impossible to move to a place and immediately move to another place again. In other words, it is possible to move to another place only after staying more than certain time periods in the current place. We will call this minimum staying time. After constructing a table of minimum staying time, we can decide invalid moving paths by comparing the staying time in a moving path to the minimum staying time.

Figure 2 shows an example of inference considering the staying time constraint. The table of minimum staying time (Figure 2 (b)) was constructed from the map of places (Figure 2 (a)). In this example, it is hard to distinguish lab and bathroom only with current image data and spatial constraint. However, by considering the staying time that the user came from lobby to corridor and stayed in 3 time steps, we reject the bathroom because the staying time is shorter than the minimum staying time in 'lobby → corridor → bathroom' path. And because the minimum staying time is longer than the minimum staying time of 'lobby → corridor → lab' path, we can select lab as the current place (Figure 2 (c)).

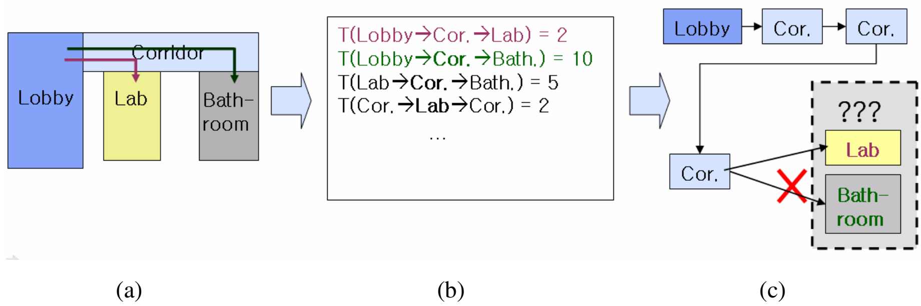

**Fig. 2.** Example of inference based on staying time constraint

## 4 Experiment

We performed experiment with the proposed method in the real environment. As target place, 11 places were chosen in the department of CS building at KAIST. The locations of the places are shown in Figure 3. Subjects were to wear cameras and explore places in free order. The image streams are used for training and testing.

The wearable test bed was composed of four web-cams, a mini PC and a vest. The four web-cams were attached to the shoulders of the vest and subjects were to move wearing the vest. This system allowed us to acquire images under realistic conditions while the user navigates the environment. In this way, 12,606 images for each direction, a total of 50,424 images were collected from six different subjects. 2~3 images per second were captured, and the size of each image was 320x240. We collected six image sequences from six subjects. We trained and evaluated the system by using 6-fold cross validation.

To analyze the effect of using multiple cameras, we evaluated the system twice; with using only one camera and with using four cameras. Similarly, to analyze the effect of considering two types of constraints in temporal reasoning, we performed evaluation for three cases; using no temporal reasoning, considering spatial constraint only, and considering both constraints. Figure 4 shows the recognition rate of the six experiments which differs in number of cameras used and types of constraints considered.

As we expected, using four cameras gave higher recognition rate than using only one. We could also find that considering more constraints in temporal reasoning gives higher recognition result. The proposed approach with four cameras considering both of the constraints recognized 90.91% of images correctly.

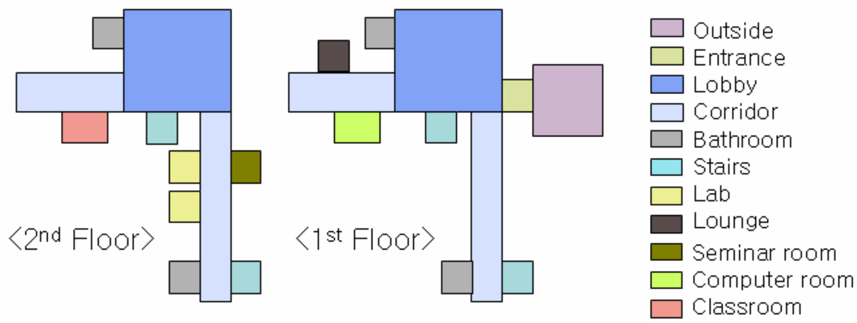

**Fig. 3.** Target places

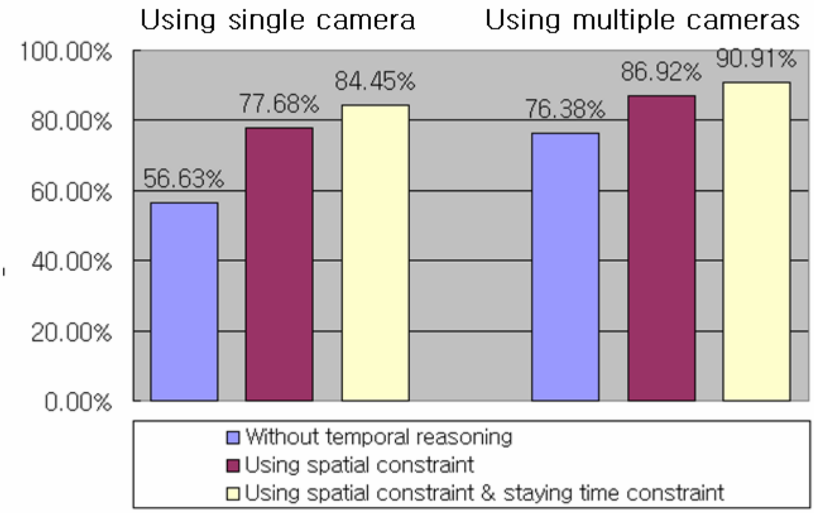

**Fig. 4.** Experimental result

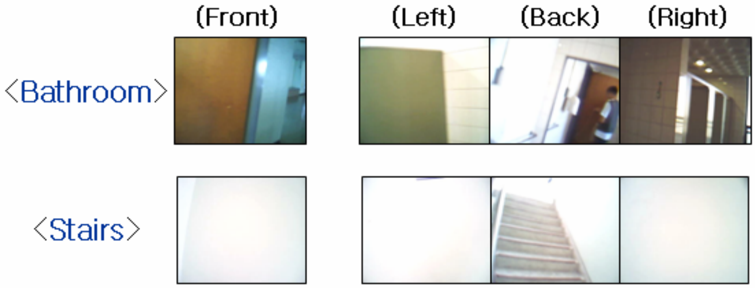

**Fig. 5.** Examples of corrected error by using multiple cameras

We confirmed the effect of four cameras by taking some examples from the data set. Figure 5 shows the examples of corrected misrecognition by using four direction cameras. In the case of bathroom images, when only the front image was used, the system recognized it as a seminar room instead of a bathroom. This is because the front direction scene image is common scene which can be observed in seminar room as well as bathroom. However, the scene images of other directions contain unique features of a bathroom, so the likelihood of bathroom is the larger than any other places. By comparing the sum of likelihoods of each image, bathroom got the highest likelihood and recognition was corrected. Images of stairs showed similar result. With only the front image, the system misrecognized it as outside. However, the result was corrected as stairs by considering the four direction images.

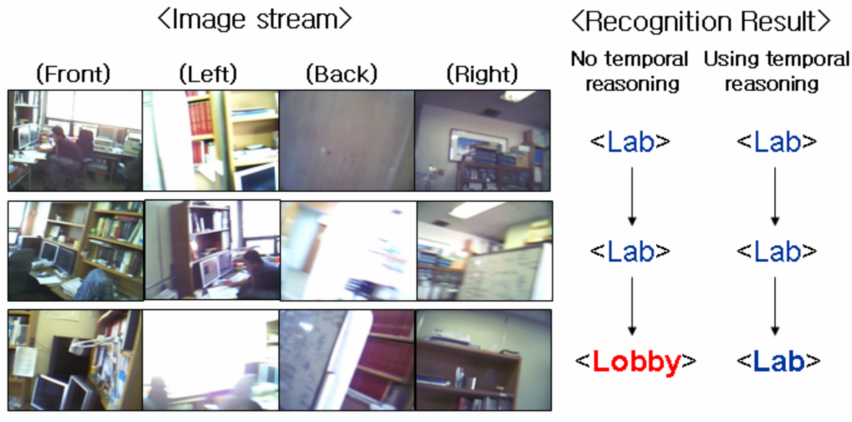

**Fig. 6.** Example of corrected error by using temporal reasoning

Figure 6 shows an example of continuously captured images and the recognition results with and without temporal reasoning. Without temporal reasoning, recognition result was incorrect in the third time step, though the results in the first and second time step were correct. However, with temporal reasoning, the system recognized all three images correctly because the results of previous time steps gave high probability of being in the lab to the inference in the third time step.

## 5 Conclusion

The problem of recognizing a user's location is the most crucial for providing intelligent location-based services. The camera-based approach has been actively pursued for the place recognition problem to exploit the rich information contained in the images. However, because the images are usually degraded several kinds of noise, the problem is still considered as a difficult problem.

In this work, we propose efficient methods for camera-based place recognition. In the proposed method, multiple cameras are used to collect multi-direction scene images. By using multi-direction images instead of single image, the system can recognize the place even if some images have insufficient information.

For more robust recognition, we utilized spatial relationships between the places. In addition that, a temporal reasoning is incorporated with a Markov model to reflect typical staying time at each place

Recognition experiments were conducted in 11 places in the real environment with both the previous method and the proposed method. The proposed approach recognized 90.91% of images correctly.

## Acknowledgements

This research is supported by Ubiquitous Computing and Network (UCN) Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

## References

1. Loomis, J.M.R., Golledge, R.G., Klatzky, R.L., Speigle, J.M., Tietz, J.: Personal guidance system for the visually impaired. In: 1st annual ACM conference on Assistive Technologies, pp. 85–90. ACM Press, New York (1994)
2. Rhodes, B., Starner, T.: Remembrance agent: a continuously running automated information retrieval system. In: 1st International Conference on the Practical Application of Intelligent Agents and Multi Agent Technology, pp. 487–495 (1996)
3. Clarkson, B., Mase, K., Pentland, A.: Recognizing User Context via Wearable Sensors. In: 4th IEEE International Symposium on Wearable Computers, pp. 69–74. IEEE Press, Los Alamitos (2000)
4. Lee, S., Mase, K.: Activity and Location Recognition Using Wearable Sensors. Pervasive Computing 1(3), 24–32 (2002)
5. Torralba, A., Murphy, K.P., Freeman, W.T., Rubin, M.A.: Context-based vision system for place and object recognition. In: 9th IEEE Int'l Conf. on Computer Vision, vol. 1, pp. 273–280. IEEE Press, Los Alamitos (2003)
6. Li, F., Kosecka, J.: Probabilistic Location Recognition using Reduced Feature Set. In: IEEE Int. Conf. on Robotics and Automation, pp. 3405–3410. IEEE Press, Los Alamitos (2006)
7. Simoncelli, E.P., Freeman, W.T.: The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: IEEE Int. Conf. on Image Processing, vol. 3, pp. 444–447. IEEE Press, Los Alamitos (1995)