International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.1 (2014), pp.77-88
http://dx.doi.org/10.14257/ijmue.2014.9.1.08

Automatic Landmark Selection Using Clustering for Robot SLAM

Chuho Yi[1], Yongmin Shin[2] and Jungwon Cho[3,*]

[1] *Future IT R&D Lab, LG Electronics, Korea*
[2] *Home Appliance R&D Lab, LG Electronics, Korea*
[3] *Department of Computer Education, Jeju National University, Korea*

chuho.yi@gmail.com, ymsin4@hotmail.com, jwcho@jejunu.ac.kr

Abstract

We discuss the current technology behind automatic selection of landmarks by simultaneous localization and mapping (SLAM), using a single camera in an unfamiliar indoor environment, and we propose an improved method. As currently implemented, automatic landmark selection by vision-based SLAM results in many useless landmarks, because features of the image are distinguished from the surrounding environment and are detected repeatedly. These useless landmarks create a serious problem for the SLAM system because they complicate data association. To solve this problem, we propose a method in which a robot initially collects landmarks through automatic detection while traversing the entire area where the robot performs SLAM and then, through clustering, selects only those landmarks that exhibit high rarity. This enhances system performance. Experimental results showed that this method of automatic landmark selection results in a high-rarity landmark being selected. Our method improves the performance of SLAM compared to conventional methods, an increases the accuracy of data associations.

Keywords: *SLAM (Simultaneous Localization and Mapping), Automatic landmark selection, Visual attention, Clustering*

1. Introduction

Simultaneous localization and mapping (SLAM) is a method in which the work required by a robot to create a map of its environment is executed at the same time as its position is estimated through the use of sensors while the robot navigates an unfamiliar environment [1]. A camera sensor is usually used for SLAM because a great deal of information can be readily obtained that can be used to recognize an object and a particular place and because the cost is low, compared to other sensors. In general, because of the difficulty of data association, vision-based SLAM uses a multi-dimensional feature descriptor or information about the landmark known in advance. However, advance knowledge may be impossible if the robot needs to perform in an unfamiliar environment, and a multi-dimensional feature descriptor limits operation. Therefore, a technique to automatically select a meaningful landmark and to limit the number of landmarks is needed. Many vision-based SLAMs use a local image feature descriptor to define a landmark. Examples of this are shift invariant feature transform (SIFT) [2], speeded-up robust features (SURF) [3], and Harris corner [4]. The process of matching local feature points entails a massive computation cost due to the inclusion of multi-dimensional data. If SLAM is performed in a way that generates feature points of the entire image followed by matching of the entire set of feature points, then time and computational

[*] *Corresponding author.*

ISSN: 1975-0080 IJMUE

complexity are significantly increased. Therefore, determining a region of interest (ROI) within the image early on in the process enables SLAM to be performed efficiently using observation data from selected areas.

Itti and Siagian [5] proposed applying a visual attention system to robot localization. Figure 1 illustrates the use of visual attention to estimate a robot's position. A green rectangle is a first-matched area in the map, using GIST features [6]. After the first-matched regions are used to estimate the robot position, a SIFT feature descriptor defines a node of the map, as shown on the right side of the figure. However, real-time implementation of this method entails a high overhead because of the vast number of operations required to match SIFT and GIST features simultaneously.

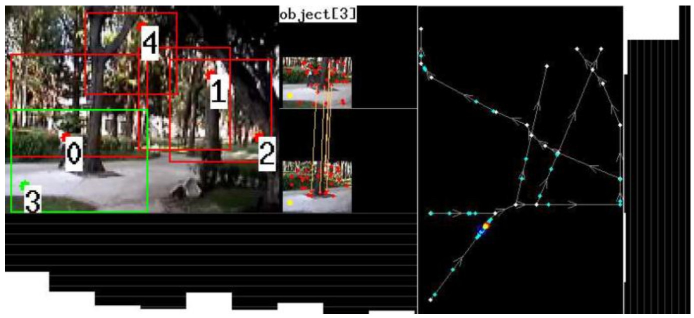


Figure 1. Robot localization using visual attention [5]

Frintrop [7] proposed a method of landmark selection using VOCUS and subsequently applied the method to vision-based SLAM [8]. His method for selection and registration of a landmark uses prior target-oriented learning to improve the results of matching. However, VOCUS is inadequate for use in an unfamiliar environment because it requires that the landmark selection be trained in advance.

In this paper, to automatically select an ROI in the image and to decide its usefulness as a landmark, we use a visual attention system that mimics the process of human visual perception. We previously proposed a modified visual attention system that uses corner features instead of a Gabor filter value [9]. In this paper, we propose using clustering in the unknown and globally similar environments to automatically select a landmark and to reliably perform vision-based SLAM. Our proposed landmark selection method is based on the rarity of detected regions as an indicator of the usefulness of a landmark in the environment where the robot is active. This improves both data association and the performance of SLAM. To automatically select a useful landmark, all regions detected during a pre-traveling stage are clustered in the visual feature space. Then each clustered group is projected onto a map and only the observation data that exhibit a narrow distribution are determined to be landmarks. We evaluated the proposed method by comparing the performance of traditional vision-based SLAM to that using only useful, high-rarity landmarks obtained by clustering.

2. The Proposed Method

We previously proposed a modified visual attention system that could automatically detect a characteristic region as a landmark and showed that it improved the performance of vision-based SLAM [9]. However, when the visual characteristics of the environment are globally similar, as they are in an indoor environment (*e.g.*, Figure 2), various problems arise. Figure 2 shows a corridor environment in the IT/BT building at Hanyang University, Seoul, Korea. There are numerous brown bars on the walls, as well as a painted band near the floor and ceiling, as part of the interior design. Other than these, which form the most visible region in the image, there are few distinctive landmarks.

(a) detected region in image with visual attention (only one)

(b) detected region in a map

Figure 2. Example of similar indoor environments, globally

The automatically detected regions in the environment, using the visual attention system, are shown as red squares in Figure 2(a). Because the detected brown bars contained only a simple feature and were located all over the environment, the data association often led to an incorrect match. That is, because all brown bars were observed in the entire map, as shown in Figure 2(b), the data association was unreliable because the same was observed at all positions. To overcome this difficulty, we propose a method for analyzing the characteristic value of specific landmarks in the current environment and using clustering to select useful landmarks.

2. 1. Clustering with the visual feature value

An ideal landmark is one that is observed in only one region of the map and that has high rarity; such a landmark is ideal for data association. If there is a landmark that uniquely represents one location on the map, then vision-based SLAM is likely to work well. However, when it is not possible to use any prior information about the environment (*i.e.*, it is unfamiliar), then every landmark with a characteristic image feature is detected, and there is no rarity. Thus, all detected regions are determined to be landmarks, and these are used by SLAM as observation data. Using numerous detected regions as landmarks is very inefficient in terms of system efficiency and computation time. Furthermore, the lack of rarity increases the probability of errors, which have an adverse effect on the overall performance of the system.

To overcome this problem, we developed a method for selectively determining useful landmarks through clustering during a pre-traveling stage. We assume that it is desirable to remove landmarks that have similar image features (*e.g.*, intensity, color, and corners). In an environment such as that shown in Figure 2, in which a common feature (*i.e.*, the brown bars attached to the walls) occurs throughout the environment as part of an interior design, there are many similar encounters in the indoor space during robot travel. Such detected regions are narrowly distributed in a feature vector space because the features of color and intensity are similar between images, but it will have a wide distribution when it is projected on the map. Conversely, a projected distribution of useful landmarks on a map has a narrow distribution, such as the red circle shown in Figure 3.

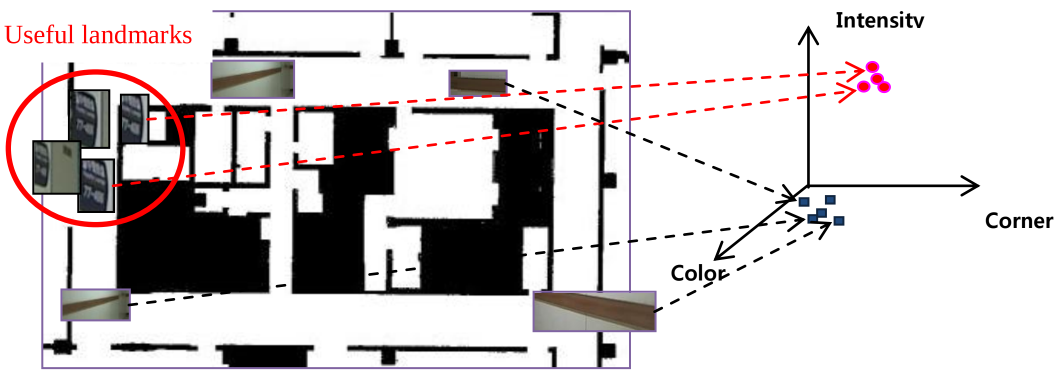


Figure 3. Example of landmarks expressed in 3D feature space and on a 2D map

To verify the relationship between a position on the map and a position in the feature space, we used the mean-shift algorithm to cluster the feature values of the detected region [10]. Mean-shift is an algorithm for clustering similar data while moving the center of gravity towards the average of the peripheral data that exists around the data. Figure 4 shows the results of mean-shift clustering in the 2D feature space of all regions detected during the pre-traveling stage.

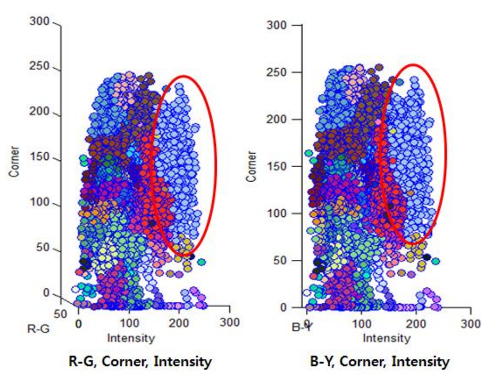


Figure 4. Results of mean-shift clustering

Figure 5 shows images that were clustered together, with the detected regions corresponding to the circles in the mean-shift results of Figure 4. The images in Figure 5 were clustered into groups corresponding to (a) the brown bars on the walls, (b) light reflection areas, and (c) useful landmarks. As a result of the clustering shown in Figures 4 and 5, similar detected regions with similar visual characteristics were grouped together.

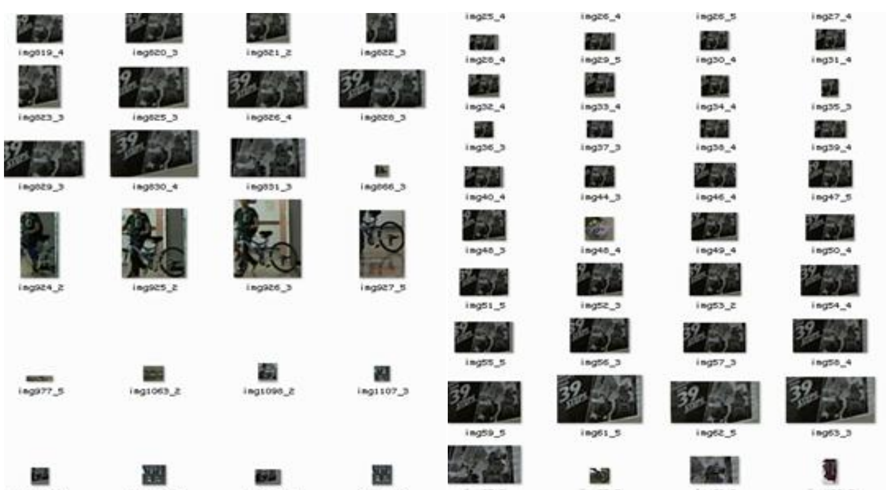
(a) clustered landmarks 1 (b) clustered landmarks 2 (c) clustered landmarks 3

Figure 5. An example of clustered landmarks

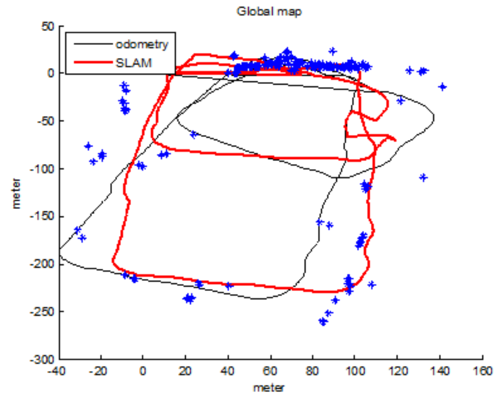
2. 2. Landmark selection using clustering

In this section, we explain how the entropy of the position distribution of a landmark group is calculated from the clustered group projections on the map and how this entropy value is used to analyze and selectively determine a useful landmark.

Figure 6 shows detected landmarks that were clustered together and the results of projecting these landmarks onto a map. The landmarks are quite similar to each other, and are narrowly distributed when projected onto a map. In this paper, we assume that these landmarks are useful.

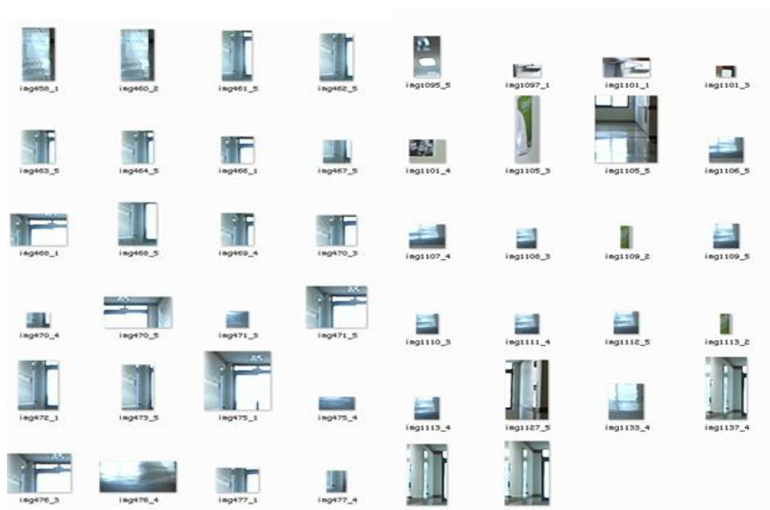


(a) landmarks clustered into a single group (samples have high rarity)

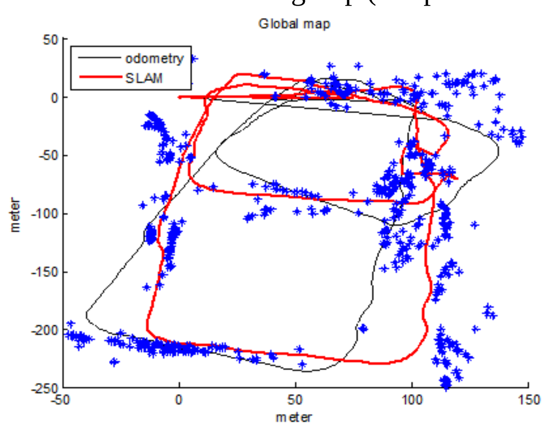


(b) results of projecting the landmarks in (a) onto a map

Figure 6. Thumbnail images of clustered landmarks and a result of projections onto map I



(a) landmarks clustered into one group (samples have low rarity)



(b) results of projecting the landmarks in (a) onto a map

Figure 7. Thumbnail images of clustered landmarks and a result of projections onto map II

By comparison, the detected landmarks shown in Figure 7(a) were clustered using the mean-shift method into groups with similar image features but low rarity, and therefore are not useful. When projected onto a map (Figure 7b), these landmarks exhibit a wide distribution of position. As illustrated by the results of Figures 6 and 7, it is possible to determine which group of clustered landmarks is useful in the current environment by calculating the entropy of the position distribution after projection onto a map. In other words, the landmarks in a group with a wide distribution on the map have low rarity, and if used would likely result in incorrect data associations. Therefore, through clustering and projection, landmarks detected during pre-traveling are evaluated for their usefulness, and only the useful landmarks are applied to vision-based SLAM. As a result, it is possible to significantly reduce the number of computations and incorrect data associations.

To analyze the rarity of landmarks in the current environment, Eq. (1) is used to calculate the complexity of the position distribution of a cluster after projection of the cluster onto a quantized map (Figure 8). The quantization of the map determines the x-axis and y-axis and the complexity of the position distribution is calculated separately.

$$\text{distribution complexity} = -\sum_{i=0}^{q} P_{L_i} log P_{L_i} \qquad (1)$$

In Eq. (1), q is the total number of quantizations for the map, L is a quantized region, and P_{L_i} is the probability that a landmark exists in the ith region. The value of the entropy value is obtained from P_{L_i} for the frequency of occurrence of the separated landmarks in the quantized region.

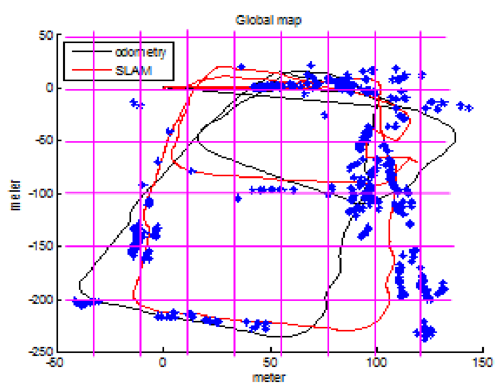
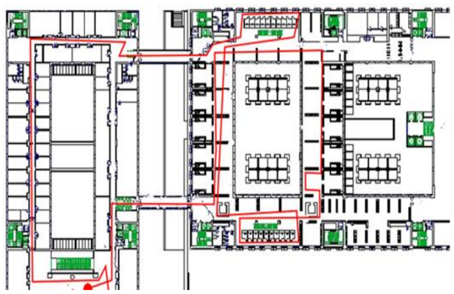


Figure 8. An example of a quantized map

In our experiment, the environment consisted of a 150 × 300 m space, and the complexity of the position distribution was calculated using quantized units of 25 m. The number of landmarks and observations was reduced as a result of automatic landmark selection using visual attention. However, because landmarks with low rarity in the observation region were not used, the reliability of the data association was significantly improved.

3. Experimental Results

We evaluated our proposed method using the experimental data provided by RAWSEEDS for objective evaluation [11]. RAWSEEDS is a public data set that includes images from a single camera, odometer data of robots, and corrected robot trajectories (ground-truth); thus, it is suitable for evaluating the method proposed in this paper. The red line in Figure 10(a) is a ground truth for the robot trajectory in an indoor university environment, and Figure 10(b) shows captured images.

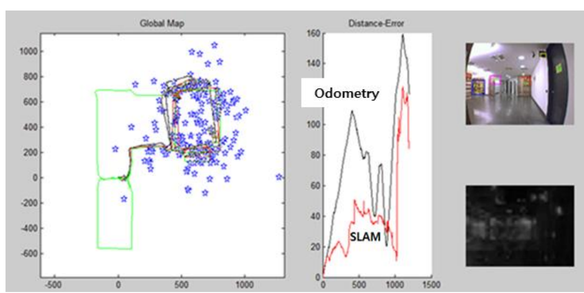


(a) the ground truth for the robot trajectory

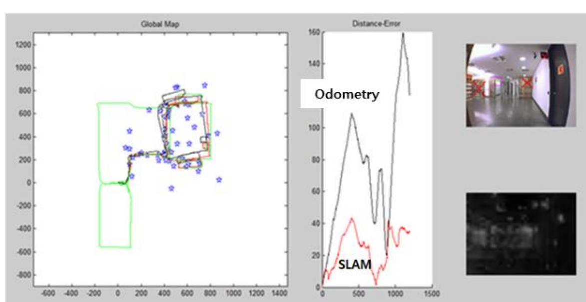
(b) captured images in indoor experiments

Figure 9. Robot trajectory and the experimental indoor environments

As shown in Figure 9(a), the data set provided by RAWSEEDS is suitable for evaluation of a vision-based SLAM because there is a closed-loop. Figure 10(a) is a result of using the observation data for all landmarks, and the left side of Figure 10 is a result of mapping. The red line in the middle graph is an error value obtained by subtracting ground-truth from the estimated robot position, and the black line is an error vale obtained by subtracting ground-truth from the odometer data. The upper right image in the Figure 10 is the result of selected landmarks in the current image, and the bottom right figure is a result of visual attention.



(a) SLAM result using all landmarks as the observation data



(b) SLAM result using our proposed method

Figure 10. SLAM results using all landmarks and selective selection by clustering

The conventional method produced significantly more landmarks, which means that the mapping had been performed incorrectly, because it includes a wider area than the robot path. In contract, in our method (Figure 10b), only useful landmarks were registered along the robot path, drastically decreasing the variance in the closed-loop section in which approximately 1000 images were analyzed.

Table 1. Comparison of errors between the proposed method and the conventional method

Method	Average distance for error (*m*)	Variance for error
Odometer (raw data)	7.558	14.8144
All landmarks as the observation data	4.024	8.8201
Proposed method	2.959	4.4394

Table 1 compares the average error between the proposed method and the conventional method. The average error was reduced by 1.07m using the proposed method. This represents a decrease of approximately 27%. Our proposed method attempts to solve the problem of generating numerous similar landmarks in a real environment when using an image feature descriptor from automatic landmarks selection. If only useful (*i.e.*, high rarity) landmarks are selected, then the probability of correct data association increases. Our proposed method selects useful landmarks in the current environment, determined by clustering the accumulated feature values observed through visual attention during a pre-traveling stage, thereby improving the performance of vision-based SLAM.

4. Conclusions

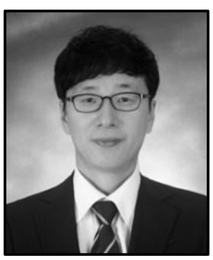
Our method for determining automatic landmarks selectively by clustering the feature values obtained by visual attention during a pre-traveling stage represents an improvement to vision-based SLAM. We used the visual attention system to automatically detect the landmarks in an unfamiliar environment. Useful landmarks with high rarity were selected by this method, and then vision-based SLAM was performed with only useful landmarks. Our proposed method improved the performance of robot localization and mapping in an unfamiliar environment.

In a future study, we intend to research an online automatic-selective landmarks selection method for SLAM that does not require pre-traveling. In addition, we plan to proceed with research on how to use information other than local image features to improve the reliability of data association in a large, essentially homogeneous environment.

References

[1] S. Thrun, W. Burgard and D. Fox, "Probabilistic ROBOTICS", MIT Press, **(2006)**.
[2] D. Lowe, "Object recognition from local scale-invariant features", Proceedings of the International Conference on Computer Vision, vol. 2, **(1999)**, pp. 1150–1157.
[3] H. Bay, A, Ess, T. Tuytelaars and L. Gool, "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding, vol. 110, **(2008)**, pp. 346-359.
[4] C. Harris and M. Stephens, "A combined corner and edge detector", Proceedings of the 4th Alvey Vision Conference, **(1988)**, pp. 147-151.
[5] C. Siagian and L. Itti, "Biologically Inspired Mobile Robot Vision Localization", IEEE Transactions on Robotics, vol. 25, no. 4, **(2009)**, pp. 861-873.
[6] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention", IEEE Trans. Pattern Anal. Mach. Intell., **(2007)**, pp. 300–312.
[7] S. Frintrop and P. Jensfelt, "Attentional Landmarks and Active Gaze Control for Visual SLAM", IEEE Transactions on Robotics, Special Issue on Visual SLAM, vol. 24, no. 5, **(2008)**.
[8] S. Frintrop, P. Jensfelt and H. Christensen, "Attentional Landmark Selection for Visual SLAM", Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, **(2006)**.
[9] Y. Shin, C. Yi, I. H. Suh and B. U. Choi, "Visual-Attention Using Corner Feature Based SLAM in Indoor Environment", The Magazine of IEEK, vol. 49, **(2012)**, pp.90-101.
[10] D. Comaniciu and P. Meer, "Mean Shift: A robust approach toward feature space analysis", IEEE Trans. Pattern Anal. Machine Intell., vol. 24, **(2002)**, pp. 603-619.
[11] RAWSEEDS: Robotics Advancement through Web-publishing of Sensorial and Elaborated Extensive Data Sets, Bicocca_2009-02-25b, http://www.rawseeds.org/rs/capture_sessions/view/5, **(2009)**.

Authors



Chuho Yi

He received a B.S. in the School of Electrical and Computer Engineering from the University of Seoul, Seoul, Korea in 2000 and received his MS and Ph.D. in the Department of Electronics and Computer Engineering from Hanyang University, Seoul, Korea, in 2002 and 2012, respectively. From 2003 to 2006, he worked as a senior researcher of the Production Engineering Research Institute at LG Electronics, Korea. He is currently a chief research engineer at the Future IT R&D Laboratory, LG Electronics, Seoul, Korea. He is an author of over 10 papers in refereed international journals and conference proceedings. His research interests include SLAM, semantic map-building, active localization, navigation, Bayesian models, and robot vision.

Yongmin Shin

He received a B.S. in the Information and Communication Department of Engineering from the Dongguk University, Korea in 2010 and received his MS in the Department of Intelligent Robot Engineering from Hanyang University, Seoul, Korea, in 2012. He is currently a research engineer at the Home Appliance R&D Lab, LG Electronics, Seoul, Korea. His research interests include robot vision, SLAM, and image processing.

Jungwon Cho

He received a B.S. degree in Information & Telecommunication Engineering from the University of Incheon, Incheon, S.Korea at 1996, and earned M.S. and Ph.D. degrees in Electronic Communication Engineering from Hanyang University, Seoul, S.Korea in 1998 and 2004, respectively. In 2004, he joined Jeju National University, Jeju, S.Korea, as a Professor at the Department of Computer Education. He was Vice-dean at the College of Education in 2011 and 2012. He also conducted research at Purdue University as a Visiting Scholar in 2007-2008. He is an author of over 35 papers in refereed international journals and conference proceedings. His research interests include computer education, information ethics, smart learning, and multimedia information retrieval. He is a member of the IEEE and the IEICE.