실내 환경에서의 로봇 자율주행을 위한 천장영상으로부터의 이종 특징점을 이용한 단일비전 기반 자기 위치 추정 시스템

Monocular Vision Based Localization System using Hybrid Features from Ceiling Images for Robot Navigation in an Indoor Environment

강 정 원[1], 방 석 원†, 크리스토퍼 쥐 애키슨[2],
홍 영 진[3], 서 진 호[3], 이 정 우[3], 정 명 진[1]

Jungwon Kang[1], Seok Won Bang†, Christopher G. Atkeson[2]
Youngjin Hong[3], Jinho Suh[3], Jungwoo Lee[3], Myung Jin Chung[1]

Abstract This paper presents a localization system using ceiling images in a large indoor environment. For a system with low cost and complexity, we propose a single camera based system that utilizes ceiling images acquired from a camera installed to point upwards. For reliable operation, we propose a method using hybrid features which include natural landmarks in a natural scene and artificial landmarks observable in an infrared ray domain. Compared with previous works utilizing only infrared based features, our method reduces the required number of artificial features as we exploit both natural and artificial features. In addition, compared with previous works using only natural scene, our method has an advantage in the convergence speed and robustness as an observation of an artificial feature provides a crucial clue for robot pose estimation. In an experiment with challenging situations in a real environment, our method was performed impressively in terms of the robustness and accuracy. To our knowledge, our method is the first ceiling vision based localization method using features from both visible and infrared rays domains. Our system can be easily utilized with a variety of service robot applications in a large indoor environment.

Keywords: Pose Estimation, Localization, Ceiling Vision

1. Introduction

Accurate and robust localization in a large indoor environment is a challenging problem for the autonomous navigation of a mobile robot. As a service robot, such as a cleaning robot has been introduced in the commercial market, the system cost has become an important factor. Among sensors for localization, a vision sensor can be the most suitable solution as its price has decreased gradually so that it can be applied to commercial appliances.

Solving localization and SLAM (Simultaneous Localization And Mapping) with vision sensors has been a central research topic in the robotics and computer vision community for several decades. Stereo vision has been widely used in this area [1], as it directly provides the depth information of the scene. On the other hand, some works have proposed the use of only a single vision system [2][3][4][5][6]. The single vision based approach is especially interesting as it requires minimal system resources.

※ This work was supported by Pohang Institute of Intelligent Robotics. In addition, this research was supported in part by the MKE(The Ministry of Knowledge Economy), Korea, under the Human Resources Development Program for Convergence Robot Specialists support program supervised by the NIPA(National IT Industry Promotion Agency) (NIPA-2011-C7000-1001-0007).

1 Department of Electrical Engineering, KAIST
2 The Robotics Institute, Carnegie Mellon University
3 Pohang Institute of Intelligent Robotics
† Corresponding author: Scholar at the Robotics Institute, CMU

It is also adequate for practical applications with low system cost.

Our work begins the development of a practical localization system that achieves both robustness and accuracy in a large indoor environment. For a system with low cost and complexity, we propose a single camera based system. Our system utilizes ceiling images acquired from a camera installed to point toward the ceiling. Many previous single vision based works have used frontal views. On the other hand, with regard to an indoor environment, the other available choice is to use ceiling images, and some previous works using ceiling images have shown noticeable results [4][5][6]. It is advantageous to use ceiling images for localization and mapping in local feature tracking as well as wide baseline feature matching, as there are only rotation and affine transformations without scale changes in feature appearances.

For reliable operation in a large scale environment, we propose a method using hybrid features which include natural landmarks in a natural scene and artificial landmarks observable in an infrared ray domain. Corner points and blobs in a natural scene, observable in visible ray domains, are utilized as natural features. Artificial features are landmarks made of infrared retroreflective materials, only visible in an infrared ray domain. As we assign a unique code to the physical body of the artificial feature, the feature becomes a highly distinctive landmark, which helps to solve challenging problems such as initial robot pose estimation and the kidnapped problem. Moreover, the artificial features are robustly detectable under variation of ambient visible rays as they react to infrared rays. Due to this advantage the infrared based features have been used in several works for localization [8][9][10]. However, as they have utilized only infrared based features, they require a large number of artificial features to handle a large environment. Compared with previous works utilizing only infrared based features, our method reduces the required number of artificial features as we exploit both natural and artificial features. In addition, compared with previous works using natural scene [4][5][6][7], our method has an advantage in the convergence speed and robustness as an observation of an artificial feature provides a crucial clue for robot pose estimation. In an experiment with challenging situations in a real environment, our method was performed impressively in terms of the robustness and accuracy.

To our knowledge, our method is the first ceiling vision based localization method using features from both visible and infrared rays domains. Our system can be easily utilized for a variety of service robot applications such as cleaning robots and guide robots in large indoor environments.

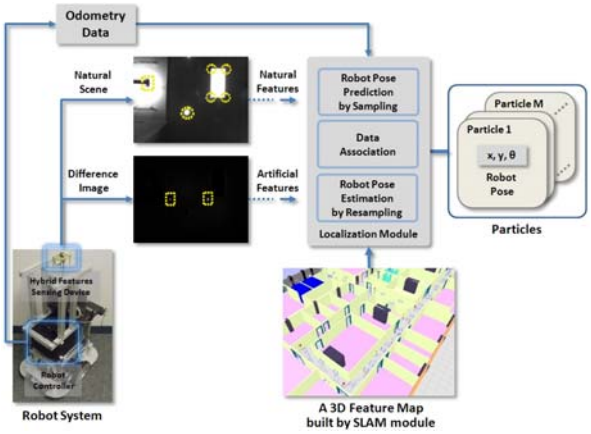


Fig. 1. Overview of our system for localization

This paper is organized as follows. In the upcoming section, an overview of our system is presented. In Section 3, we describe feature extraction methods with the proposed hybrid feature sensing device. The methods for building a map and localization are described in Sections 4 and 5, respectively. The experimental results are shown in Section 6, and our work is concluded in Section 7.

2. Overview of Our System

An overview of our system is described in Fig. 1. The proposed hybrid feature sensing device provides a natural scene image and a difference image at the same time. Natural features include corner points and blobs in a natural scene. Meanwhile, artificial features are extracted from the difference image.

Having an accurate map is a prerequisite for accurate localization. Here, creating a map involves the 3D position estimation of each feature using measurements of the features in a 2D image space. In order to make a map, a particle filter based SLAM technique is employed. The output of the SLAM is a map that is composed of a set of 3D positions of each feature with its description. As a description of a feature, the appearance of each corner point and the area of each blob extracted from a natural scene are stored in the map. In the case of artificial features, the analyzed code is saved. The map that the maximum weighted particle possesses is used for localization in navigation.

Once the map is built, a robot is capable of estimating its pose in the environment during navigation. Our localization method is also based on a particle filter in which each particle is a hypothesis of the robot pose. Each particle is weighted according to the difference between the measurement and estimation of the feature locations in an image space. Whenever

an artificial feature is observed, relocation is performed to correct the robot pose, which prevents the divergence of localization and secures the robustness in localization.

3. Hybrid Feature Extraction

3.1 Hybrid Feature Sensing Device

The hybrid feature sensing device was developed to provide images for detecting both natural and artificial features. The designed artificial features are made of retroreflective materials which react to infrared rays. They can be detected from a difference image, which is acquired from a pair of an image with IR LEDs on and an image with IR LEDs off in two consecutive frames. Such a difference image based system for extracting infrared features has been widely used for localization [8][9], as well as human eye detection and tracking [11][12]. Moreover, the difference image based system allows us to utilize both natural and artificial features. Natural features can be detected from images with IR LEDs off. Artificial features can be extracted from difference images acquired by the absolute difference between images with IR LEDs on and off. Therefore, our device was developed to capture images with IR LEDs on and off alternately.

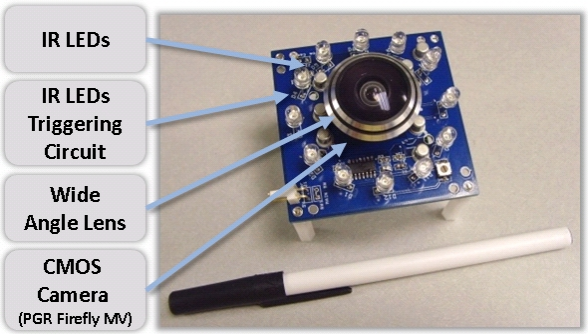


Fig. 2. The developed hybrid feature sensing device

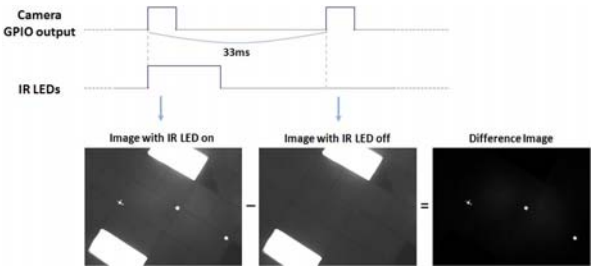


Fig. 3. Timing diagram of the hybrid feature sensing device

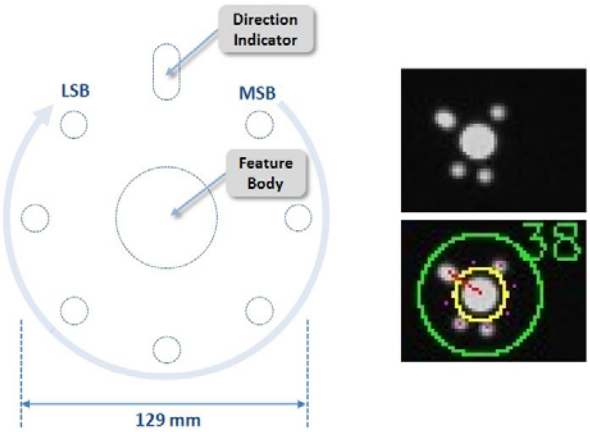


Fig. 4. Design of an artificial feature (left), its appearances in a difference image and a code decoding result (right)

Fig. 5. An artificial feature on the ceiling with flash off (left) and on (right)

Fig. 2 shows the developed hybrid feature sensing device. The device is composed of a cmos camera, a wide angle lens, 12 infrared LEDs with wavelengths of 735 nm around the camera and an LED triggering circuit. The LED triggering circuit consists of a D flipflop for making an LED control signal and an amplifier for driving LEDs.

When an image is captured, i.e. the camera shutter opens, the rising edge signal comes out from the camera general purpose input/output port. With this rising edge signal, an LED triggering circuit generates rectangular pulses whose frequency is half that of the rising edge signal. This half frequency rectangular pulse triggers the IR LEDs, and we get an image with IR LEDs on when the rectangular pulse turns on IR LEDs and vice-versa. Consequently, we get an image with IR LEDs on and an image with IR LEDs off in two consecutive frames as in Fig. 3. The device provides images whose pixel size is 640 x 480, up to 30 Hz.

3.2 Design of Artificial Features

In order to make artificial features, we chose infrared retroreflective materials to satisfy the following conditions. First, the material should be insensitive to the illumination change. Second, it should be passive. It should not need external power. Finally, it should be as invisible to human eyes as possible to harmonize with the surroundings.

Our aim was to make robustly detectable artificial features with each unique code as an identification tag. Through experiments using the retroreflective materials, we found that the intensity in the image was gradually changed between the feature area and the background. This behavior was not appropriate for extracting contour or shape by detecting the abrupt intensity change, as in the previous works using artificial markers in the visible rays domain [13]. Therefore, we designed the artificial features that consist of circles so that the center of mass of each circle area can be used to encode the feature information. Fig. 4 shows the designed artificial feature, its appearances in a difference image and a code decoding result. Fig. 5 shows the feature installed on the ceiling. The designed feature consists of a feature body, a direction indicator, and small circles for codes as binary numbers. The feature can encode 128 different codes, as it has 7 small circles area for codes.

3.3 Natural Features Extraction

In recent research, the features named SIFT [14] and SURF [15] have been widely used because they are robustly detected and matched regardless of the scaling and rotation of the features in images. However, since we use ceiling images where there is no scaling change of the features, we simply detect corner points as natural features instead of extracting those computationally expensive features.

After applying Gaussian smoothing to an image with IR LEDs off to make the image insensitive to noise, we extract corner points that satisfy the condition mentioned in [16].

The other natural feature is a blob in a natural scene. A covariance based shape analysis method is used to extract it from an image. After applying an intensity threshold to a given image, a noisy area is removed by morphological operation, and then connected components are extracted. After applying an area threshold to each component in order to remove too large or too small components, the shape of each candidate area is analyzed using the covariance matrix C defined as

$$D = \begin{bmatrix} \mathbf{x} - \hat{\mathbf{x}} & \mathbf{y} - \hat{\mathbf{y}} \end{bmatrix} \tag{1}$$

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = \frac{1}{N} D^T D \tag{2}$$

Here $\mathbf{x}$ and $\mathbf{y}$ are column vectors with pixel coordinate values in each component, $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are column vectors with average pixel coordinate values in each component, and N is the total number of pixels in each component.

A component that satisfies the conditions $|c_{11}| \simeq |c_{22}|$ and $|c_{12}| \simeq 0$ is regarded as a blob feature.

3.4 Artificial Features Extraction

The artificial features can be extracted from a difference image by the following three steps: feature body extraction, direction indicator extraction, and code analysis.

The first step is to extract a circular shaped feature body. This is done with the same method for a blob extraction in a natural scene, except here we use a difference image. The second step is to find a direction indicator around the feature body. After setting a search region to include area of a direction indicator and code components around the feature body, the intensity threshold is applied to the search region. After performing connected components labeling in the search region, the largest component is regarded as a direction indicator. The final step is to read the code of the feature, which is simply done by reading intensity values from MSB to LSB. The *i*th bit α_i equals 1 if $p_i \geq T$, where p_i is the intensity value at the point of *i*th bit in the image and T is the intensity threshold value used in the second step. Otherwise the α_i equals zero. The code is given as $\sum_{i=0}^{7} \alpha_i \cdot 2^i$.

4. Map Building

4.1 SLAM Framework for Building a Map

The Rao-Blackwellized particle filter based method is employed as a SLAM framework [17]. The key idea of the SLAM method is to maintain a set of particles. Each particle contains an estimated robot pose x_t^k, and a map in which each feature is represented by local Gaussian. The map includes a set of Kalman filters with mean $\mu_t^{k,i}$ and covariance $\Sigma_t^{k,i}$, one for each feature in the map. Here k is the particle index, i is the feature index, and t is the time step. The full SLAM

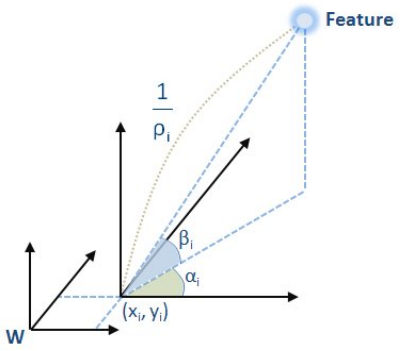


Fig. 6. Feature position parameterization

posterior is computed recursively in the following steps: robot pose prediction, measurement update, importance weighting, and resampling. Details of the algorithm are given in [18].

4.2 Parameterization for Feature Position Estimation

Since we use a single camera, it is not possible to obtain depth information from a single measurement of a natural feature. In the case of an artificial feature, it is possible to obtain depth from a single observation because we know the physical size of the feature. However, in the SLAM framework, we use the multiple measurements based scheme for position estimation of all features. The 3D feature position can be estimated by multiple measurements of the feature. To represent a feature location, we employ the inverse depth parameterization method [19]. In this method, the measurement function has low linearization errors even at low parallax, and the estimation uncertainty is accurately modeled with a multivariate Gaussian.

Five parameters are used to define a feature position as in Fig. 6. Here x_i, y_i are the robot position where the feature was first observed and α_i, β_i are azimuth and elevation, respectively, and ρ_i is the inverse depth.

4.3 Measurement Function

The measurement function is obtained in the following manner. The 3D position of a feature with respect to the world coordinate is calculated from the feature parameters, as given in Eq. (3).

$$\mathbf{x}_F^W = \begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix} = \begin{bmatrix} x_i + (1/\rho_i)\cos\beta_i \cdot \cos\alpha_i \\ y_i + (1/\rho_i)\cos\beta_i \cdot \sin\alpha_i \\ (1/\rho_i)\sin\beta_i \end{bmatrix} \tag{3}$$

Let the robot position $\mathbf{x}_R^W = \begin{bmatrix} x_r & y_r & 0 \end{bmatrix}^T$ and the robot orientation $\mathbf{R}_R^W$ with respect to the world coordinate. Then a ray $\mathbf{h}^R$ that describes the observation of the feature from the robot pose is obtained as in Eq. (4).

$$\begin{aligned} \mathbf{h}^R &= \begin{bmatrix} h_x^R & h_y^R & h_z^R \end{bmatrix}^T \\ &= \mathbf{R}_R^W (\mathbf{x}_F^W - \mathbf{x}_R^W) \end{aligned} \tag{4}$$

The ray is projected in the image plane according to the pinhole model and produces a point with pixel coordinates value x, y for the feature.

$$h_{x,n}^R = \frac{h_x^R}{h_z^R}, \quad h_{y,n}^R = \frac{h_y^R}{h_z^R} \tag{5}$$

$$h = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_x + f \cdot s_x \cdot h_{x,n}^R \\ c_y + f \cdot s_y \cdot h_{y,n}^R \end{bmatrix} \tag{6}$$

Here f is the focal length, s_x, s_y are the pixel size of the x, y axis, and c_x, c_y are the principal points.

4.4 Feature Position Update

The feature position described by the mean and the covariance of a feature state is updated by the standard EKF formula.

$$\mathbf{K}_g = \Sigma_{t-1}\mathbf{J}_\mathbf{h}^T(\mathbf{J}_\mathbf{h}\Sigma_{t-1}\mathbf{J}_\mathbf{h}^T + Q_t)^{-1} \tag{7}$$

$$\mu_t = \mu_{t-1} + \mathbf{K}_g(z_t - \hat{z}_t) \tag{8}$$

$$\Sigma_t = (\mathbf{I} - \mathbf{K}_g\mathbf{J}_\mathbf{h})\Sigma_{t-1} \tag{9}$$

Here $\mathbf{J}_\mathbf{h}$ is the Jacobian of the measurement function $\mathbf{h}$ with respect to the feature state, $\mathbf{K}_g$ is the Kalman gain, z_t is the real measurement, i.e. x, y coordinates value of the feature in the image, $\hat{z}_t$ is the expected measurement, Q_t is the measurement noise, μ_t is the feature state consisting of five parameters, and Σ_t is the covariance at time t.

4.5 Data Association

A mixture scheme made up of the measurement maximum likelihood method and the feature description based method is

employed for data association of all features. Fig. 7 shows the structure for this scheme. At each step, the features in the map are projected on the images, forming the expected features. Inside the projected uncertainty region, the current extracted features are compared with the expected features based on the position in an image space. By the measurement maximum likelihood method, for each feature measurement z_t, its correspondence $\hat{i}$ is determined by maximizing the likelihood of the measurement of z_t.

$$\hat{i} = \arg\max_i \left|2\pi L_t^k\right|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(z_t - \hat{z}_t^i)^T L_t^{k-1}(z_t - \hat{z}_t^i)\} \tag{10}$$

Here $L_t^k = Q_t + \mathbf{J_h} \sum_{t-1}^{k,i} \mathbf{J_h}^T$.

This data association is verified by the following feature description based method. In the case of a corner point in a natural scene, the matching is verified by normalized cross-correlation measure. If the matching is validated, the image patch, that is, the appearance of the corner point, is updated to the most recent one. When the appearance is updated, its image patch is saved in the map after alignment along the world coordinate. On the other hand, the KLT tracker [20] is also used for data association for a corner point in two consecutive frames. In the case of a blob in a natural scene, the matching is verified by the comparison of the areas. Matching with big difference in areas is invalidated. In the case of an artificial feature, the data association and the validation are strongly performed by analyzing the code. The code analysis is performed as a voting scheme for robustness to the noise or wrong instantaneous recognition of the code. The code is determined when the most dominant voting count of the code becomes large enough compared with the second dominant one.

4.6 Fusion with Prior Information

In order to build a highly accurate map in a large scale environment, we can utilize artificial features with prior known position. The position is a two dimensional x, y coordinate on a plane parallel to the ground, not relying on the height of the ceiling. The fusion of building a map with the prior information is done through weighting particles by observation of the artificial feature. When the artificial feature has small depth uncertainty by multiple measurements, each particle is weighted by the difference between the prior position and the estimated position. This scheme makes it easy to solve closing a large loop and helps us to get an accurate map in a large scale environment.

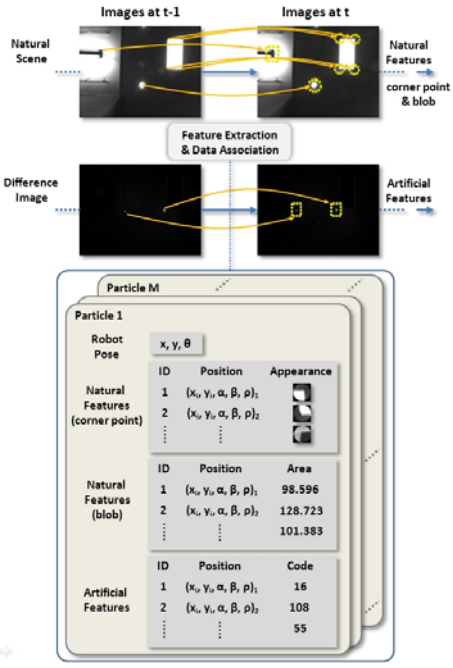


Fig. 7. Data association

5. Localization

5.1 Localization Framework

Our localization scheme also uses a particle filter based method [7], called Monte Carlo localization. The map that the maximum weighted particle possesses in the map building module is inputted into the localization module. In the localization module, each particle represents a hypothesis of the robot pose. The data association is performed similarly with the map building module, except that the localization module does not have feature initialization and position update. Every step each particle is scored by the difference between the real and the expected measurement of the features, and then re-sampled by the score.

The challenging problem in the localization is the global localization including initial robot pose estimation and the kidnapped problem. The common approach to these problems is to scatter particles uniformly in the whole area and to make the particles converge to the most probable area by obtained measurements while the robot moves around. However, in our case, in order to solve such problems, whenever the robot observes an artificial feature, the robot pose is re-estimated from the observation of the artificial feature. This relocation scheme prevents the divergence of localization and shows fast convergence speed in situations with large pose uncertainty. Our localization scheme is summarized in Fig. 8.

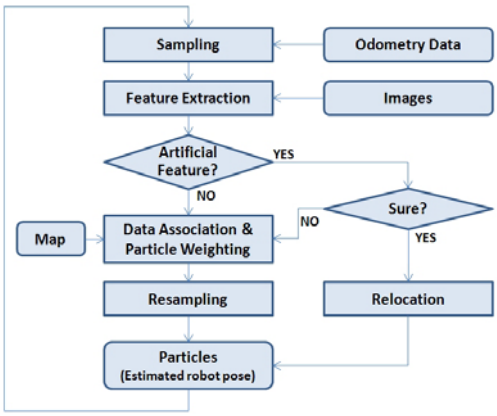


Fig. 8. The flowchart for our localization framework

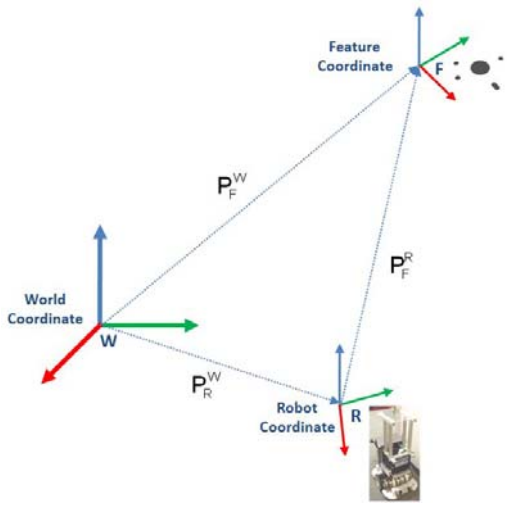


Fig. 9. Robot pose estimation from an observation of an artificial feature

5.2 Robot Pose Estimation from an Observation of an Artificial Feature

Since the position and orientation of an artificial feature are registered in the map, a robot pose can be directly estimated from an observation of the feature, as described in Fig. 9. Let P_F^W and θ_F^W be the registered position and the orientation of the feature with respect to the world coordinate. Since we know the real metric length L_{real} between the feature body and the direction indicator of the artificial feature, we can get P_F^R and θ_F^R from measurement of the feature, as in Eq. (12).

$$\begin{aligned} \mathbf{X}_{n,body} &= K^{-1}\mathbf{x}_{body} \\ \mathbf{X}_{n,indicator} &= K^{-1}\mathbf{x}_{indicator} \\ L_{n,\,feature} &= \left\| \mathbf{X}_{n,body} - \mathbf{X}_{n,indicator} \right\| \\ L_n^{real} &= \frac{L_{real}}{L_{n,\,feature}} \end{aligned} \tag{11}$$

Here $\mathbf{x}_{body}$ and $\mathbf{x}_{indicator}$ are the mean coordinate values of the feature body and the direction indicator area in the image, respectively, and K^{-1} is the inverse of the camera intrinsic matrix.

$$\begin{aligned} X_F^R &= L_n^{real} \cdot x_{n,body} \\ Y_F^R &= L_n^{real} \cdot y_{n,body} \\ Z_F^R &= L_n^{real} \\ P_F^R &= \begin{bmatrix} X_F^R & Y_F^R & Z_F^R \end{bmatrix}^T \\ \theta_F^R &= \text{atan2}(y_{n,indicator} - y_{n,body},\, x_{n,indicator} - x_{n,body}) \end{aligned} \tag{12}$$

Here $x_{n,body}$ and $y_{n,body}$ are the coordinate values of the feature body, $x_{n,indicator}$ and $y_{n,indicator}$ are the coordinate values of the direction indicator in the normalized image space.

Then, the robot position P_R^W and its orientation θ_R^W are obtained through Eq. (13).

$$\begin{aligned} \theta_R^W &= \theta_F^W - \theta_F^R \\ R_R^W &= \begin{bmatrix} \cos\theta_R^W & -\sin\theta_R^W & 0 \\ \sin\theta_R^W & \cos\theta_R^W & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ P_R^W &= P_F^W - R_R^W P_F^R \end{aligned} \tag{13}$$

Table. 1 shows pose estimation performance of our system from an observation of an artificial feature. The result shows that an observation of an artificial feature can be effectively used for the global localization including initial robot pose estimation and the kidnapped situation.

Table.1. Pose estimation performance from an observation of an artificial feature

Displacement (mm)	Position Error (mm)	Orientation Error (°)
0	26	0.79
100	29	0.61
200	16	0.60
300	15	0.57
400	16	0.26
500	11	0.32
600	7	0.37
700	9	0.39
800	15	0.18
900	17	0.33
1000	14	0.70
1100	10	0.47
1200	11	0.58
1300	32	1.18
1400	22	0.73
1500	16	0.62
Mean	**16.6**	**0.54**
Standard Deviation	**7.2**	**0.24**

Table 2. Performance of localization in a test environment

Checking Spot	Position Error (mm)	Orientation Error (°)
Spot 1	19	0.60
Spot 2	9	0.89
Spot 3	9	0.72
Spot 4	7	2.37
Spot 5	7	0.38
Spot 6	16	0.27
Spot 7	5	1.18
Spot 8	10	0.62
Spot 9	11	0.25
Spot 10	23	0.45
Spot 11	21	1.08
Spot 12	22	1.55
Spot 13	8	0.22
Spot 14	16	0.63
Spot 15	14	1.38
Spot 16	10	0.59
Spot 17	15	0.91
Mean	**13**	**0.83**
Standard Deviation	**5.6**	**0.56**

6. Experimental Results

6.1 Robot System

Our robot system is a mobile robot with a differential drive mechanism. Our robot is shown in Fig. 1. The robot controller drives the motors according to user request and provides odometry data from motor encoders. The hybrid feature sensing device is installed in the upper part of the robot to look at ceiling, and the images from the sensing device are transferred to a notebook PC through a IEEE1394A cable.

6.2 Environment for an Experiment

The proposed method was tested on the 4th and the A floor of Newell Simon Hall, Carnegie Mellon University. The size of the 4th floor is about 50.6m x 69.9m, and the height of the ceiling is approximately 2.92m. The 23 artificial features with each unique code were installed on the ceiling at the joint or intersection of the corridors. Each 2D position of the artificial features was known through the CAD model of the building. Meanwhile, the size of the A floor is 24m x 49m, and the height of the ceiling is about 2.3m. The total 11 artificial features were installed. During the experiment on the A floor, we measured the quantitative accuracy of localization.

6.3 Results & Discussion

In order to make a map, we manually drove the robot at an average speed of 0.2m/s along the corridor to visit the whole corridor area. The total length of trajectory is approximately 405m on the 4th floor and 175m on the A floor. The map building module was operated at about 7Hz with 100 particles. As natural features, corner points were usually extracted from corners of rectangular fluorescent lamps on the ceiling, as we extracted only corners with high cornerness. The circular patch with a radius of 11 pixels around a corner point is saved as the appearance. During artificial features extraction, it sometimes showed wrong instantaneous code decoding results. These errors are mainly caused by external lighting condition change and low resolution image region. The strong light from outside caused the sudden change of lighting condition and it deteriorates the difference images for extracting artificial features. As we use the wide angles lens, the un-distortion of the image by camera calibration parameters makes outer image region to have low resolution. When an artificial feature is located at the low resolution region, the feature shows the blurred appearance in the image, which makes it hard to recognize a code.

However the results finally converged to the correct results by our code voting scheme. As a result, all the artificial features were successfully detected, and each code was correctly decoded.

Since the robot moves parallel to the ceiling, almost all features from the ceiling images had significant parallax during the robot motion. Only incorrectly matched or unrepeatably detected features retain large depth uncertainty. In order to remove the effect from such features with large depth uncertainty, only features with small depth uncertainty are maintained in the map and used in particle weighting. Fig. 10 shows the map building result on the 4th floor. The result shows the 3D feature map. Each ellipsoid in the map indicates the uncertainty of the feature position. As shown in Fig. 11, the odometry based path is largely distorted due to accumulated errors of the motion sensors caused by the slippage on the carpet and uneven surface. On the other hand, the proposed method shows much better results compared with the odometry based method.

After building the map, we tested localization with the map. The same parameters, including the number of particles and the image patch size, were used. Fig. 12 shows the localization result on the 4th floor. The blue path indicates the particle distribution during robot movement. We drove the robot along the corridor. Initially, the robot was around point A. The robot did not have any knowledge of its initial pose. When observing an artificial feature installed at point A, the robot correctly recognized its pose from the observation, and the particles were moved to that place. While the robot moved to around point B, it successfully localized itself using natural features. In addition, relocation was done by two artificial features located at the lower middle part of the building, decreasing the robot pose uncertainty. When the robot reached point B, we kidnapped the robot and put it around point C. Upon seeing the artificial feature attached at point C, the robot realized that it was moved to around point C. Then, it moved to point D with successful localization.

Besides the localization test in the challenging situations described above, we also measured the error in localization. In order to measure the error, 17 spots on the A floor were set as error checking points, as in Fig. 13. The spot 1 to 9 were located at the artificial features, marked as red points in Fig. 13. On the other hand, the remaining spots were located between the artificial features, marked as blue points in Fig. 13. The ground truth position was manually measured, and was compared with the position and orientation from our localization system. We had experimented five times for each checking point in order to get the result. Table. 2 shows the localization performance of our system in the test environment. As indicated in the result, our localization system has shown a great performance.

7. Conclusion

We have presented a localization system using ceiling images in a large scale indoor environment. For a system with low cost and complexity, we proposed a single camera based system that utilizes ceiling images acquired from a camera installed to point upwards. For reliable operation in a large scale environment, we proposed a method using hybrid features, which include natural landmarks in a natural scene and artificial landmarks observable in an infrared ray domain. In the experiment with challenging situations in a real environment, our method has shown a great performance in terms of its robustness and accuracy. Our method was already successfully applied to the robot navigating in a real large scale environment [21]. Our system can be also easily utilized with a variety of service robot applications in a large indoor environment.

Reference

[1] R. Sim, P. Elinas, M. Griffin, and J. J. Little, "Vision-based SLAM using the Rao-Blackwellised Particle Filter", IJCAI Workshop on Reasoning with Uncertainty in Robotics, 2005.

[2] A. J. Davison, "Real-Time Simultaneous Localisation and Mapping with a Single Camera", International Conference on Computer Vision(ICCV)}, pp. 1403-1410, Oct. 2003.

[3] P. Jensfelt, D. Kragic, J. Folkesson and M. Björkman, "A Framework for Vision based Bearing Only 3D SLAM", IEEE International Conference on Robotics and Automation, pp. 1944-1950, May 2006.

[4] W. Y. Jeong, K. M. Lee, "CV-SLAM: A new Ceiling Vision-based SLAM technique", IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), pp. 3195-3200, Aug. 2005.

[5] W. Y. Jeong, K. M. Lee, "Visual SLAM with Line and Corner Features", IROS, pp. 2570-2575, Oct. 2006.

[6] Seo-Yeon Hwang, Jae-Bok Song, "Monocular Vision-Based SLAM in Indoor Environment Using Corner, Lamp, and Door Features from Upward-Looking Camera", IEEE Transactions on Industrial Electronics, 2011.

[7] S. Thrun, D. Fox, W. Burgard, F. Dellaert, "Robust Monte Carlo localization for mobile robots", Artificial Intelligence, Vol. 128, pp. 99-141, 2001.

[8] Y. Nakazato, M. Kanbara and N. Yokoya, "A Localization System Using Invisible Retro-reflective Markers", IAPR Conference on Machine Vision Applications, pp. 140-143, May 2005.

[9] S. Lee and J. B. Song, "Mobile Robot Localization using Infrared Light Reflecting Landmarks", International Conference on Control, Automation and Systems, pp. 674-677, Oct 2007.

[10] StarGazer™, http://www.hagisonic.com/

[11] A. Haro, M. Flickner and I. Essa, "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 163-168, June 2000.

[12] D.H. Yoo and M.J. Chung, "A Novel Non-intrusive Eye Gaze Estimation System using Cross-ratio under Large Head Motion", Computer Vision and Image Understanding, Vol. 98, No. 1, pp. 25-51, 2005.

[13] H. Kato and M. Billinghurst, "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System", in Proceedings of the 2nd IEEE/ACM International Workshop on Augmented Reality, pp. 85-94, Oct 1999.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.

[15] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359, 2008.

[16] J. Shi, C. Tomasi, "Good Features to Track", IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 593-600, June 1994.

[17] M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem", in Proceedings of the AAAI National Conference on Artificial Intelligence, pp. 593-598, 2002.

[18] S. Thrun, W. Burgard, D. Fox, Probabilistic Robotics, MIT Press, 2005.

[19] J. Civera, A. J. Davison and J. M. Martinez Montiel, "Inverse Depth Parametrization for Monocular SLAM", IEEE Transactions on Robotics, Vol. 24, No. 5, pp. 932-945, October 2008.

[20] C. Tomasi and T. Kanade, "Detection and tracking of point features", Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-91-132, April 1991.

[21] J. Kang, S. W. Bang, M. J. Chung, C. G. Atkeson, Y. Hong, J. Suh and J. Lee, "Ceiling Vision Based Autonomous Navigation of a Mobile Robot using Hybrid Visual Features in a Large Indoor Environment", The 7th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI 2010), pp. 653, Busan, Korea, November 24-27, 2010.

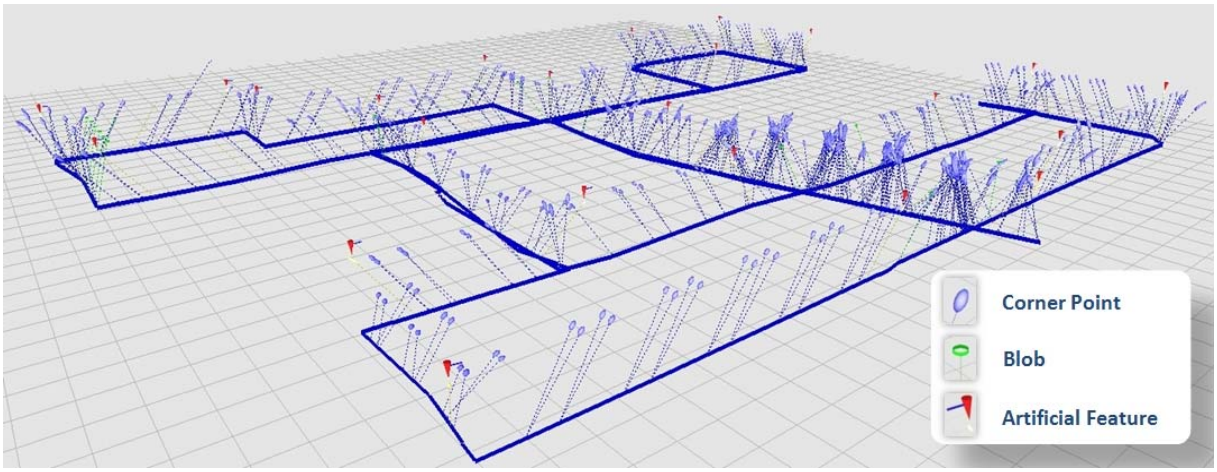


Fig. 10. The 3D feature map built by the proposed method

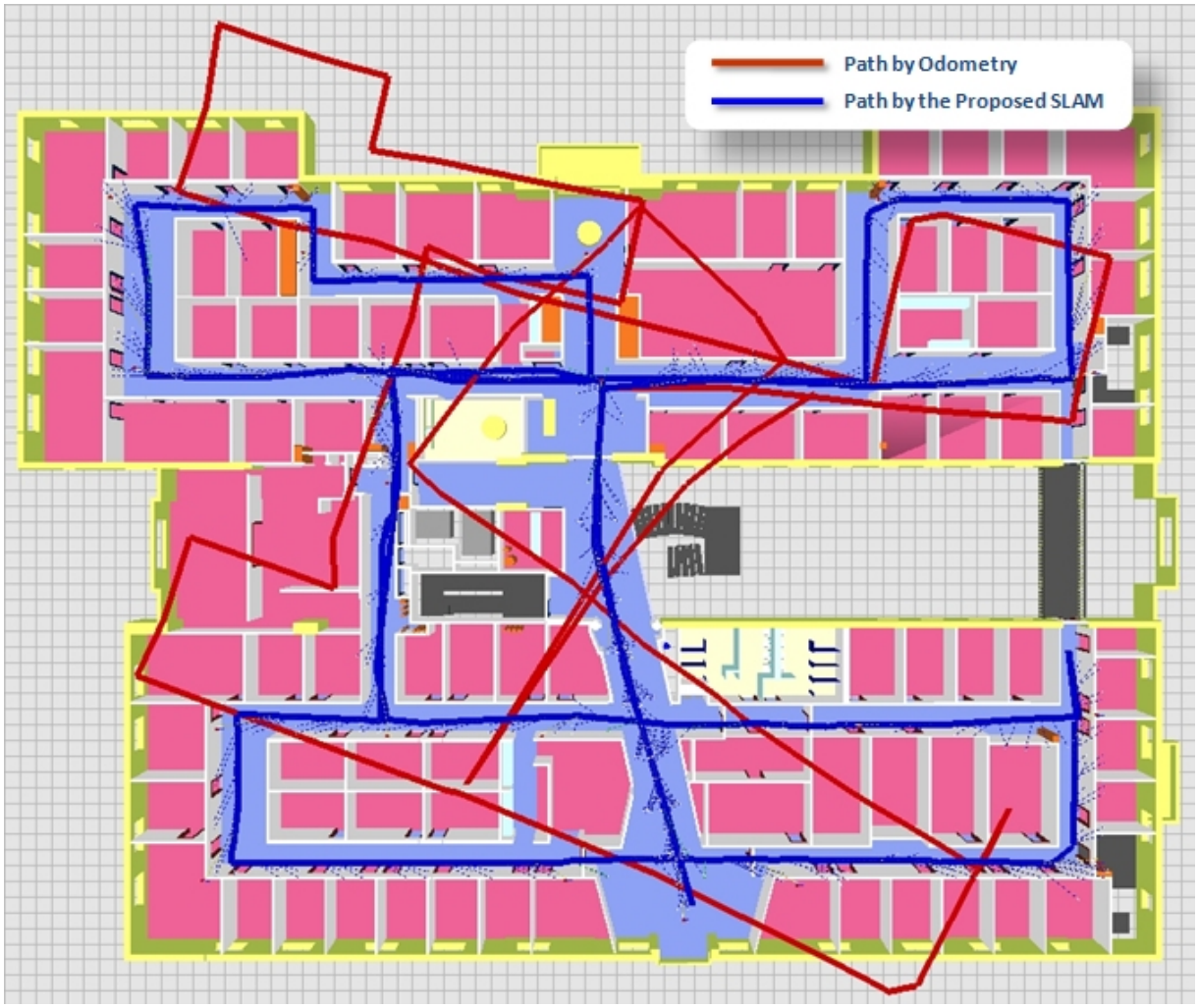


Fig. 11. The map building result on the 4th floor, Newell Simon Hall, Carnegie Mellon University

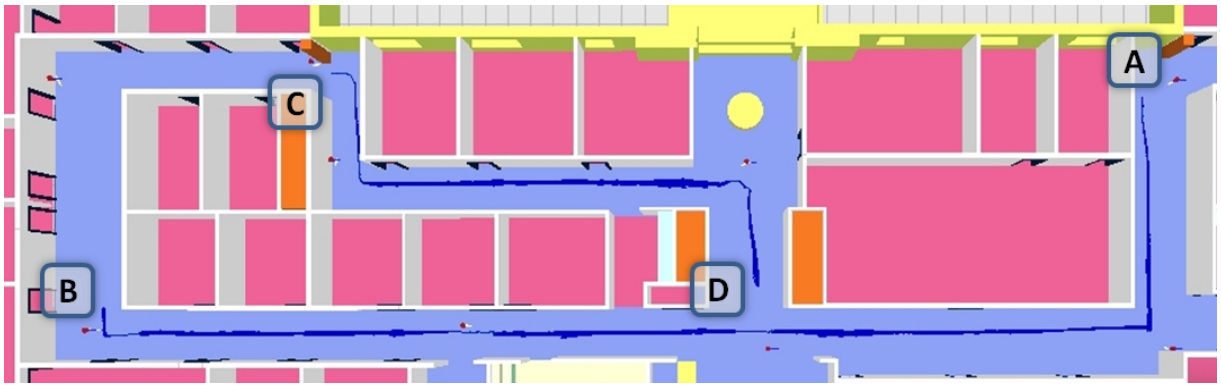


Fig. 12. A localization test in the north part of the 4th floor, Newell Simon Hall, Carnegie Mellon University

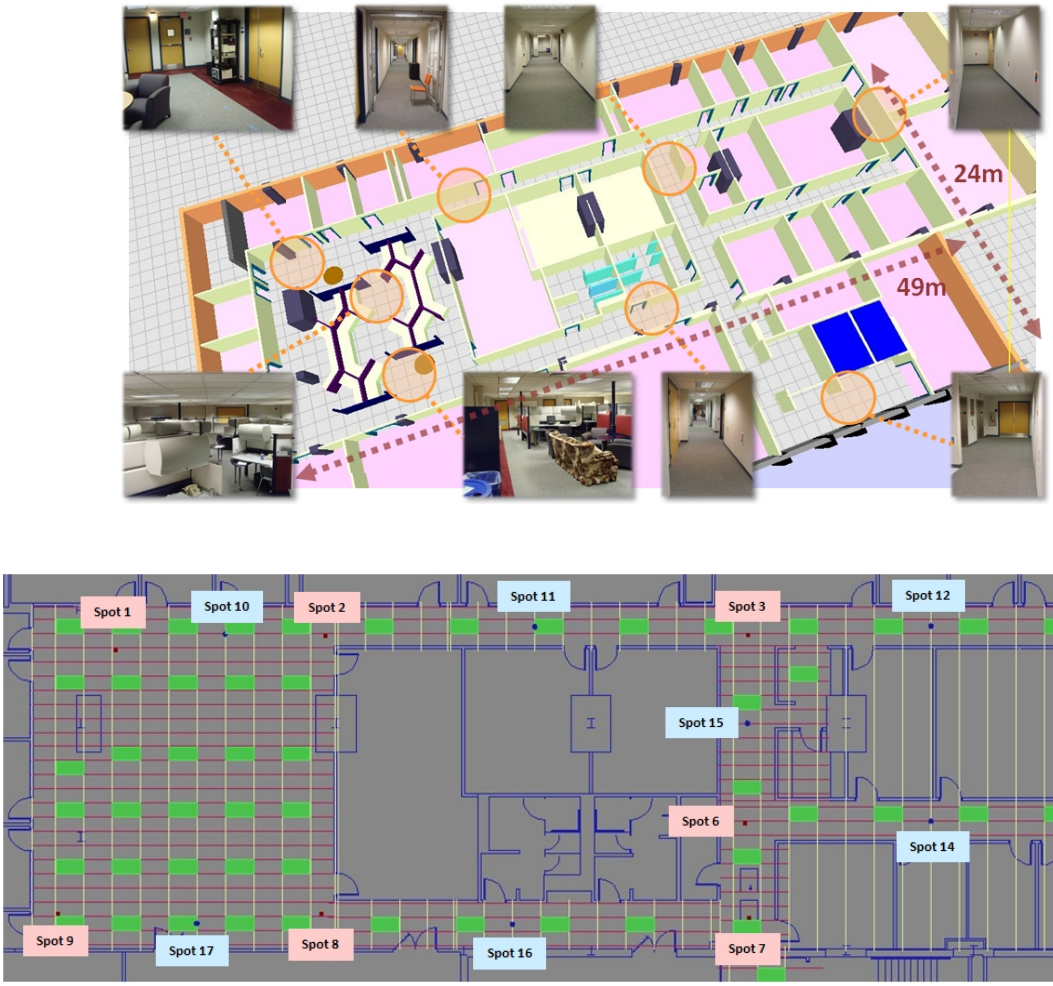
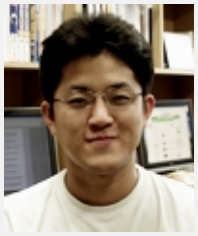


Fig. 13. The A floor (up) and the localization error checking spots (bottom)



Jungwon Kang

2004 Electrical Engineering, Korea University (B.S.)
2006 Electrical Engineering, KAIST (M.S.)
2006 ~ Present Ph.D. Candidate, Electrical Engineering, KAIST

Research Area : SLAM, Navigation, Computer vision

Seok Won Bang

1988 Electrical Engineering, Seoul National University (B.S.)
1991 Electrical and Electronics Engineering, KAIST (M.S.)
1996 Electrical and Electronics Engineering, KAIST (Ph.D.)
1995 ~ 1999 Samsung Motors Research Center
1999 ~ 2008 Samsung Advanced Institute of Technology
2008 ~ Present Robotics Institute, Carnegie Mellon University

Research Area : Robot navigation, Service robot, Consumer robotics, Artificial intelligence, 3D perception, Human robot interaction, Sensor fusion, Pattern recognition

Christopher G. Atkeson

1981 Biochemistry, Harvard University (A.B.)
1981 Applied Mathematics, Harvard University (S.M.)
1986 Brain and Cognitive Science, MIT (Ph.D.)
1986 ~ 1993 Associate Professor, MIT
1994 ~ 2000 Associate Professor, Georgia Institute of Technology
2000 ~ Present Professor, Carnegie Mellon University

Research Area : Humanoid robots, Machine learning, Human-computer interaction

Youngjin Hong

1996 Computer Science, Korea University (M.S.)
1995 ~ 1999 R&D Department, PSINet Korea
1999 ~ 2006 HCI Lab, Samsung Advanced Institute of Technology
2006 ~ 2007 Tsujii Lab, University of Tokyo
2007 ~ Present Applied Technology Division, Pohang Institute of Intelligent Robotics

Research Area : Underwater cleaning robot, Medical robot, Intelligent sensor

Jinho Suh

1992 Department of Mathematics, Hanyang University (B.S.)
1998 Department of Mechanical Engineering, Pukyong National University (M.S.)
2002 Department of Control Engineering, Tokyo Institute of Technology (Ph.D.)
2006 Post-Doctor, National Research Laboratory, Dong-A University
2006 ~ Present Director in Applied Technology Division, Pohang Institute of Intelligent Robotics

Research Area : Medical service robot, Underwater robot, Steel robotics

Jungwoo Lee

2005 Department of Electronics, Communications, Computer Science Engineering, Hanyang University (B.S.)
2005 ~ 2007 Korea Institute of Science and Technology
2007 ~ Present Pohang Institute of Intelligent Robotics

Research Area : Medical robot, Middleware, System Integration

Myung Jin Chung

1973 Electrical Engineering, Seoul National University (B.S.)
1977 Electrical Engineering, University of Michigan (M.S.)
1983 Control Engineering, University of Michigan (Ph.D.)
1976 Researcher, Agency for Defense Development
1981 ~ 1983 Research assistant, University of Michigan
1983 ~ Present Professor, Department of Electrical Engineering, KAIST
Research Area : Service robot, Human-robot interaction