

공간지능화에서 다중카메라를 이용한 이동로봇의 인간추적행위

Human-Tracking Behavior of Mobile Robot Using Multi-Camera System in a Networked ISpace

진 태 석¹, 하시모토 히데키²

TaeSeok Jin¹, Hideki Hashimoto²

Abstract The paper proposes a human-following behavior of mobile robot and an intelligent space (ISpace) is used in order to achieve these goals. An ISpace is a 3-D environment in which many sensors and intelligent devices are distributed. Mobile robots exist in this space as physical agents providing humans with services. A mobile robot is controlled to track a walking human using distributed intelligent sensors as stably and precisely as possible. The moving objects is assumed to be a point-object and projected onto an image plane to form a geometrical constraint equation that provides position data of the object based on the kinematics of the intelligent space. Uncertainties in the position estimation caused by the point-object assumption are compensated using the Kalman filter. To generate the shortest time trajectory to track the walking human, the linear and angular velocities are estimated and utilized. The computer simulation and experimental results of estimating and tracking of the walking human with the mobile robot are presented.

Keywords : Mobile robot, Distributed sensors, Intelligent space, Estimation, Kalman filter

1. Introduction

In recent years, numerous present and future tasks of technical systems require increasingly the perception and understanding of the real world outside the machine's metal box. Like the human's eyes, camera images are the preferred input to provide the required information, and therefore computer vision is the appropriate tool to approach these tasks. One of the most fundamental abilities in scene understanding is the perception and analysis of people in their natural environment, a research area often titled the "Looking at People" -domain in computer vision. Many applications rely on this ability, like autonomous service robots, video surveillance, person recognition, man-machine interfaces, motion capturing, video retrieval and content-based low-bandwidth video compression [1].

"People Tracking" is the image analysis element that all of these applications have in common. It denotes the task to detect humans in an image sequence and to keep track of their movements through the scene. Often this provides merely the initial information for following analysis, e.g. of the body posture, the position of people in the three-dimensional scene, the behavior, or the identity [2].

There is no general solution at hand that could process the images of any camera in any situation and deliver the positions and actions of the detected persons, because real situations are far too varied and complex to be interpreted without any prior knowledge. Furthermore, people have a wide variety of shape and appearance. A human observer solves the image analysis task by three-dimensional understanding of the situation in its entirety, using a vast background knowledge about every single element in the scene, physical laws and human behavior. No present computer model is capable to represent and use that amount of world knowledge, so the task has to be specialized to allow for more abstract descriptions of

¹ 동서대학교 메카트로닉스공학과 전임강사
(Email: jints@dongseo.ac.kr)

² 동경대학 생산기술연구소 교수 (Email: hashimoto@iis.u-tokyo.ac.jp)

the desired observations, e.g. “Humans are the only objects that move in the scene” or “People in the scene are walking upright and therefore have a typical shape”[3].

This robot is aimed at guiding a wheelchair in a hospital or a station and has the ability to estimate the position and velocity of humans and to avoid obstacles. However, these studies only addressed the problem of how to follow humans, not how to detect the presence of humans. They developed their studies on the premise that human detection is possible. Some technology which includes the recognition of humans, a position estimation technique for a mobile robot and humans [1], and a control strategy for tracking humans who are walking in a stable way, is required in order to realize robot human-tracking behavior. In order to recognize humans, most human-tracking robots are mounted with many sensors, such as charge-coupled device (CCD) cameras, ultrasonic sensors, etc [5]. These sensors detect the relative position from the mobile robot to the target human. The mobile robot in [6] recognizes a human’s skin color using a CCD camera, and traces the target human by combining pan-tilt control of a CCD camera. In addition to the vision sensor, a voice recognition sensor and LED sensors are mounted in the mobile robot [7], and is able to follow humans in an outdoor environment. Most of the proposed human-tracking robots burden the target human with special equipment. It is very difficult for a mobile robot to continue tracking a human in the shortest path and time while avoiding other obstacles, without missing the target while walking at a natural speed, since the stand-alone robot has limitations in terms of recognition performance.

In this research, an intelligent environment is used in order to solve these problems. A mobile robot cooperates with multiple intelligent sensors, which are distributed in the environment. The distributed sensors recognize the walking human and the mobile robot, and give control commands to the robot in order to follow walking human in the shortest time path. The mobile robot receives the necessary support for human-tracking control from the environmental sensors.

2. Configuration of iSpace

Experiments were performed in the laboratory room, which is about 7 m in both width and depth, is used for the iSpace. The iSpace has a mobile robot as a human-tracking agent, six DINDs which can obtain the situation in the environment, and a projector and a screen which

present suitable information to the human. Each module is connected through the network communication. Three DINDs are used in order to recognize the mobile robot and to generate the control commands. The other three DINDs are used to recognize the position of the human. DINDs are placed as shown in Fig. 1 [2].

The placement of the three DINDs for human recognition is optimized to expand the viewable area of the cameras [2] so that the head and hands of the human can be recognized over a wide area. On the other hand, the placement of DINDs for the mobile robot has to be decided by trial and error. It is desirable that the DINDs for the mobile robot recognize the whole of the area covered by three DINDs for human recognition in order to achieve the human-tracking system and reliable mobile robot control. Thus, three DINDs are placed so that the area for human recognition is completely covered. Human walking information is extracted by background subtraction and by detecting the skin color of a face and hands on captured images. In the coordinate system of the iSpace, the axis is parallel to the screen, the axis is perpendicular to it, and the axis follows the right-hand system. Each DIND measures 3-D positions, which are based on this coordinate system using kinematic relationship of consecutive image frames.

A mobile robot is connected to the DIND network via wireless LAN, as shown in Fig. 2, and shares the resources of the DIND’s. For recognizing the position of the robot, one color panels are installed around the mobile robot. The pattern of the color panel is recognized by the DIND and it estimates the posture and position of the robot by kinematics of robot projected onto an image plane. Since the height of the mobile robot is already known, the position of a mobile robot is reconstructed from one camera image.

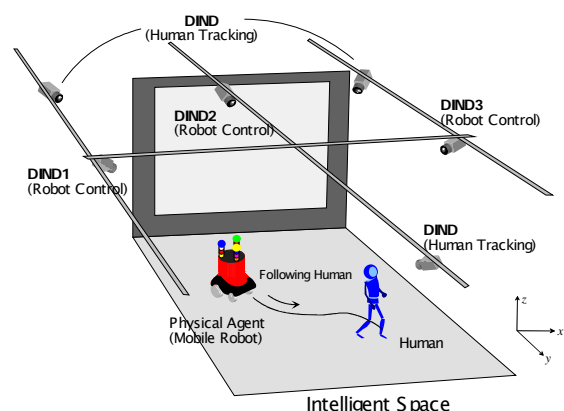


Fig. 1. Experimental environment.

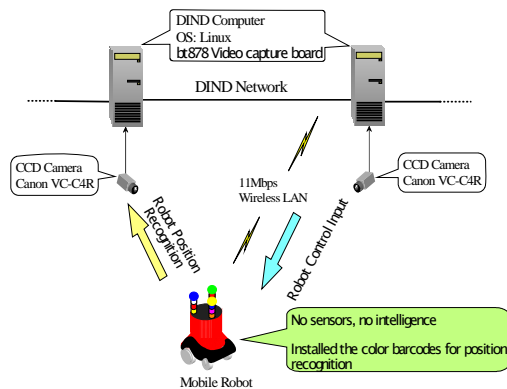


Fig. 2. Mobile robot and network system in ISpace.

3. Mobile Robot Control for Trajectory Generation

3.1 Tracking control

In order to follow a human, tracking control is performed. Many studies have been performed in the field of the tracking control. However, the trajectory which a mobile robot tracks is limited in most cases to continuous and smooth ones. A human tracking robot should be able to track actual human walking trajectories, including abrupt changes in velocity and direction. Therefore, stable tracking may not be achieved when conventional tracking control is used. A special control method for the mobile robot to follow humans is required.

In the control of a nonholonomic mobile robot, Brockett's theorem proved that a smooth state feedback law for an asymptotically stable to one point of the state space does not exist [5]. However, it is necessary to construct a closed-loop control system in which the error between the reference point and the state vector of a mobile robot should become zero for the tracking control of a mobile robot. In recent years, various closed-loop control systems which can overcome the feedback stabilization impossibility of Brockett's theorem, have been proposed. Reference [5] proposed piecewise-continuous controllers which neglect the "smoothness" and [10] proposed tracking control in which a mobile robot follows the trajectory planned as a function of time. These techniques are very effective in tracking control.

3.2 Kinematics of the camera system and coordinates

The camera system in ISpace has the ability of panning and tilting, as shown in Fig. 3. The position and posture of the camera are defined with respect to the base frame. According to the Denavit-Hartenberg convention, the homogeneous matrix can be obtained after establishing

the coordinate system and representing parameters and an attitude vector of the homogeneous matrix represents Roll(θ_R), Pitch(θ_P) and Yaw(θ_Y) angles by tilting and panning angles of the camera as [10].

To measure the distance from a camera to objects using the camera images, at least two image frames that are captured for the same object at different locations, are necessary. Usually a stereo-camera system has been used to obtain the distance information [6]. However there exist uncertainties in feature point matching and it takes too much time to be implemented in real-time. This approach requires only a frame to measure the distance to the object from the CCD camera. Since the approach becomes possible by assuming that a point-object is located on the floor, there also exist uncertainties in the position estimation. To minimize the uncertainty in the position estimation and to estimate the velocities of the moving object together, a state estimator is designed based on the Kalman filter. The image coordinates for the point object, (j, k) , is transformed to the image center coordinates which is orientation invariant in terms of the Roll angle, θ_R , and the size of the image frame, P_x and P_y , (j', k') :

$$\begin{bmatrix} j' \\ k' \end{bmatrix} = \begin{bmatrix} \cos(\theta_R) & -\sin(\theta_R) \\ \sin(\theta_R) & \cos(\theta_R) \end{bmatrix} \begin{bmatrix} j - \frac{P_x}{2} \\ k - \frac{P_y}{2} \end{bmatrix} \quad (1)$$

where P_x and P_y represent x and y directional size of the image frame in pixels, respectively.

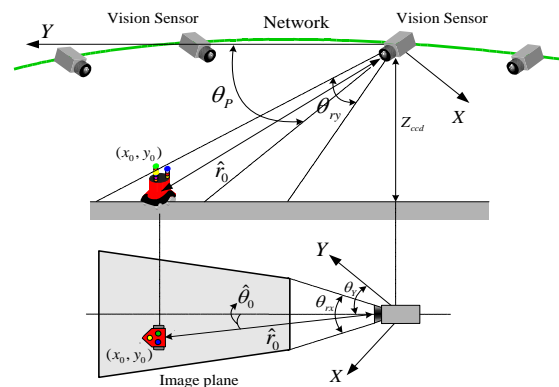


Fig. 3. Estimation of position information, \hat{r}_0 (upper figure), $\hat{\theta}_0$ (lower one).

To estimate the real location, (x_0, y_0) , $\hat{\theta}_0$ and \hat{r}_0 are estimated using the linear relationship between the real object range within the view angle and the image frame. That is, for a given set of $(\hat{\theta}_0, \hat{r}_0)$, there is one-to-one correspondence between the real object point and the image point. When a point image is captured at (j', k') on the image center frame, the real object position, $\hat{\theta}_0$ and \hat{r}_0 can be estimated as follows, as it is illustrated in Fig. 3:

$$\hat{r}_0 = \frac{z_{ccd}}{\cos(\frac{\pi}{2} - \theta_p + \frac{k'}{P_y} \theta_{ry})} \quad (2)$$

$$\hat{\theta}_0 = \frac{j'}{P_x} \theta_{rx} \quad (3)$$

where θ_{rx} and θ_{ry} represent the x and y directional view angles of the CCD camera, respectively. When (j, k) is a image coordinate for the point object, (j', k') is a image coordinate transformed to the image center coordinate. The position of the object with respect to the robot coordinates, (x, y) can be estimated using the $\hat{\theta}_0$ and \hat{r}_0 [7] as follows:

$$\hat{x}_0 = \cos(\theta_y) + \hat{r}_0 \cdot \cos(\theta_y + \hat{\theta}_0) \quad (4)$$

$$\hat{y}_0 = \sin(\theta_y) + \hat{r}_0 \cdot \sin(\theta_y + \hat{\theta}_0) \quad (5)$$

where θ_y represents the angle between the mobile robot and the camera of DINDs.

4. Localization and Tracking on the Ground Plane

In the preceding sections, proposed methods have been presented to track and describe people as flat image regions in the 2D image plane. These systems provide the information, where the tracked persons are located in the camera image. However, many applications need to know their location and trajectory in the real 3D scene instead, e.g., to evaluate their behaviour in sensitive areas or to merge the views from multiple camera. Additionally, it has been shown that knowledge about the depth positions

in the scene improves the segmentation of overlapping persons.

A mathematical model of the imaging process is used to transform 3D world coordinates into the image plane and vice versa. To this end, the prior knowledge of extrinsic (camera position, height, tilt angle) and intrinsic camera parameters (focal length, opening angle, potential lens distortions) is necessary. The equations derived in the following use so-called homogeneous coordinates to denote points in the world or image space:

$$\mathbf{x} = \begin{pmatrix} wx \\ wy \\ wz \\ w \end{pmatrix} \quad (6)$$

The homogeneous coordinate representation extends the original coordinates (x, y, z) by a scaling factor w , resulting in an infinite number of possible descriptions of each 3D point. The concept of homogeneous coordinates is basically a mathematical trick to represent broken-linear transformations by linear matrix operations. Furthermore, it enables the computer to perform calculations with points lying in infinity ($w=0$). A coordinate transformation with homogenous coordinates had the general form $\tilde{\mathbf{x}} = \mathbf{M}\mathbf{x}$, with \mathbf{M} being the transformation matrix:

$$\mathbf{M} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & x_{\Delta} \\ r_{21} & r_{22} & r_{23} & y_{\Delta} \\ r_{31} & r_{32} & r_{33} & z_{\Delta} \\ \frac{1}{d_x} & \frac{1}{d_y} & \frac{1}{d_z} & \frac{1}{s} \end{pmatrix} \quad (7)$$

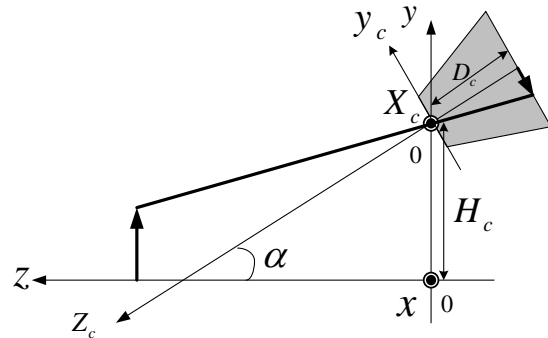


Fig. 4. Elevated and tilted pinhole camera model.

The coefficients r_{ij} denote the rotation of the coordinate system, $(x\Delta, y\Delta, z\Delta)$ the translation, $(1/d_x, 1/d_y, 1/d_z)$ the respective distortion and s the scaling.

Most camera can be approximated by the pinhole model. In case of pincushion or other image distortions caused by the optical system, additional normalization is necessary. The projection of 3D scene coordinates into the image plane of a pinhole camera located in the origin is given by the transformation matrix \mathbf{M}_H :

$$\mathbf{M}_H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/D_c & 0 \end{pmatrix} \quad (8)$$

D_c denotes the focal length of the camera model. Since the image coordinates are measured in pixel units while other measures are used for the scene positions (e.g. cm), measurement conversion is included in the coordinate transformation by expressing D_c in pixel units the width- or height resolution r_x or r_y of the camera image together with the respective opening angles θ_x or θ_y :

$$D_c = \frac{r_x}{2 \tan(\frac{\theta_x}{2})} = \frac{r_y}{2 \tan(\frac{\theta_y}{2})} \quad (9)$$

In a typical set-up, the camera is mounted at a certain height $y = H_C$ above the ground and tilted by an angle α . The according transformation matrices, \mathbf{M}_R for the rotation around the x-axis and \mathbf{M}_T for the translation in the already rotated coordinate system, are given by:

$$\mathbf{M}_R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{M}_T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -H_C \cos \alpha \\ 0 & 0 & 1 & H_C \sin \alpha \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (10)$$

The total transformation matrix \mathbf{M} results from the concatenation of the three transformations:

$$\mathbf{M} = \mathbf{M}_H \mathbf{M}_T \mathbf{M}_R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & -H_C \cos \alpha \\ 0 & -\sin \alpha & \cos \alpha & H_C \sin \alpha \\ 0 & \sin \alpha / D_c & -\cos \alpha / D_c & -H_C \sin \alpha / D_c \end{pmatrix} \quad (11)$$

the camera coordinates (x_C, y_C) can now be calculated from the world coordinates (x, y, z) as follows:

$$\begin{pmatrix} w_C x_C \\ w_C y_C \\ w_C z_C \\ w_C \end{pmatrix} = \mathbf{M} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = M \begin{pmatrix} x \\ y \cos \alpha - H_C \cos \alpha \\ -y \sin \alpha + z \cos \alpha + H_C \sin \alpha \\ \frac{1}{D_c} (y \sin \alpha - z \cos \alpha - H_C \sin \alpha) \end{pmatrix} \quad (12)$$

By dividing the homogenous coordinates by w_C , the final equations for the image coordinates are derived:

$$x_C = -D_c \frac{x}{z \cos \alpha + (H_C - y) \sin \alpha} \quad (13)$$

$$y_C = D_c \frac{(H_C - y) \cos \alpha - z \sin \alpha}{z \cos \alpha + (H_C - y) \sin \alpha} \quad (14)$$

The value of z_c is a constant and denotes the image plane, $z_c = -D_c$. The inverse transformation of the image coordinates into the 3D scene requires the prior knowledge of the value in one dimension due to the lower dimensionality of the 2D image. This has to be the height y above the ground, since the floor position (x, y) is unknown. In the analysis of the camera image, the head and feet coordinates of a person can be detected as the extrema of the silhouette, or, in a more robust way, with the help of body models. Therefore the height is either equal to zero or to the body height H_P of the person. The body height can be calculated, if they-positions of the head $y_{c,H}$ and of the feet $y_{c,F}$ are detected simultaneously:

$$H_P = H_C D_c \frac{y_{c,F} - y_{c,H}}{(y_{c,F} \cos \alpha + D_c \cos \alpha)(D_c \cos \alpha - y_{c,H} \sin \alpha)} \quad (15)$$

Besides improved tracking stability especially during occlusion, tracking on the ground plane also enables the inclusion of additional real world knowledge into the

system, like the human walking speed limits, minimum distance between two persons or valid ground positions defined by a floor map.

5. Motion Planning for Human Tracking

To follow a walking human, the mobile robot needs to be controlled by considering the relation between the position of the mobile robot and the position of the walking human. Fig. 5 shows the motion planning process of a mobile robot for tracking a walking human.

Each DIND computer system estimates the position of the moving object within m sampling time and selects the shortest distance from its current position to the walking human, assuming the its location is known a priori. The localization scheme of the mobile robot using the information on the walking human, which improves the accuracy in capturing, is developed in [8]. The target point of the mobile robot at k-th sampling time is denoted as $\hat{x}_R(k+m)$, which is one of the estimated points of the mobile robot after m sampling time.

$$\hat{x}_R(k+m_{opt}) = \min_{m=1 \sim M} \left\| \hat{x}_O(k+m) - \hat{x}_R(k+m) \right\| \quad (16)$$

where $\hat{x}_R(k+m)$ is the position of the mobile robot after m sampling time, and the mobile robot moves along the shortest path towards the target point $\hat{x}_O(k+m)$.

The position of the moving objects in the cartesian coordinate system is acquired using the relation between image frames. The linear and angular velocities of the moving objects are estimated by the state estimator, Kalman filter. After estimating the trajectory of the moving objects, the optimal trajectory and motion

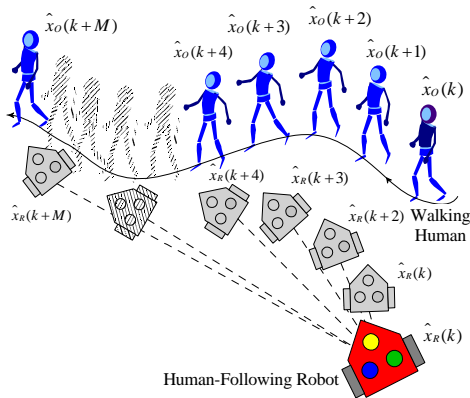


Fig. 5. Estimation of the trajectory for Human-tracking.

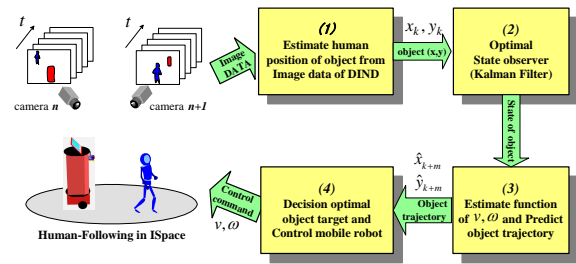


Fig. 6. Mobile robot control for estimating and tracking.

planning of the mobile robot are decided in order to follow the walking human in the shortest time. The tracking figure shows the overall structure of mobile robot control for human tracking.

6. Experiments

Experiment was performed to show the estimating and tracking a waling human. Fig. 7 shows the experimental results for estimating and tracking a walking human. Mobile robot is attached with 70x20[cm] red-colored panels. Human walks random velocities in the range of 45-50[cm/sec]. First, mobile robot detects the moving human using cameras in ISpace. When a walking human is detected within view, mobile robot tracks it tracking the proposed method. And Fig. 7 illustrates that the mobile robot tracking a human in the shortest time path as temporary position in experiments. The minimum path was estimated using the trajectories of the mobile robot and the human, while the mobile was tracking the walking human.

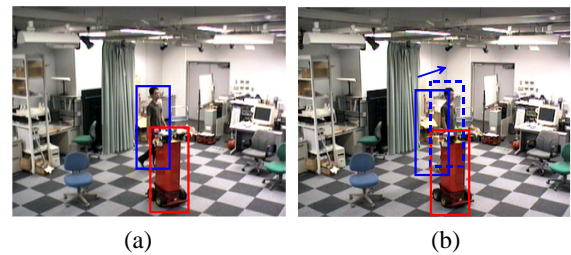


Fig. 7. Experimental results for Human-tracking.

7. Conclusion

This paper proposes a method of human-tracking for mobile robot to track and follow a walking human in ISpace environment. First, a control algorithm using a position estimation of moving objects, walking human and mobile robot, based on the kinematic relationship of consecutive image frames was proposed for a mobile

robot in order to follow a walking human whose position is estimated incompletely. The proposed model is able to absorb the gap between the motion of the human and the mobile robot. It was shown that movement estimation of the objects using a Kalman filter based DINDs for tracking and motion planning of a mobile robot to follow the walking human, based on its estimated trajectory, within the shortest time.

Acknowledgments

“This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-331-D00152).”

참고문헌

- [1] A. Adam, E. Rivlin, and I. Shimshoni, Computing the Sensory Uncertainty Field of a Vision-based Localization Sensor, Proc of the 2000 IEEE Int Conf on Robotics & Automation, 2000, pp. 2993-2999.
- [2] T. Akiyama, J.-H. Lee, and H. Hashimoto, Evaluation of CCD camera arrangement for positioning system in intelligent space, Proc of the Seventh Int Symp Artificial Life and Robotics (AROB'02), 2002, pp. 310-315.
- [3] G. Appenzeller, J.-H. Lee, and H. Hashimoto, Building topological maps by looking at people: An example of cooperation between intelligent space and robots, Proc of the IEEE/RSJ Int Conf Intelligent Robots and Systems (IROS'97), 1997, pp. 1326-1333.
- [4] R. F. Berg, Estimation and Prediction for Maneuvering Target Trajectories, IEEE Trans on Automatic Control 28 (3) (1983) 294-304.
- [5] C. C. de Wit and O. J. Sordalen, Exponential stabilization of mobile robots with nonholonomic constraints, IEEE Trans on Automatic Control 37 (1992) 1791-1797.
- [6] E. Grosso and M. Tistarelli, Active/Dynamic stereo vision, IEEE Trans on Pattern Analysis and Machine Intelligence 7 (1995) 868-879.
- [7] T. Inamura, T. Shibata, Y. Matsumoto, MasayukiInaba, and H. Inoue, Finding and following a human based on on-line visual feature determination through discourse, Proc of the Int Conf Intelligent Robots and Syst (IROS'98), 1998, pp. 348-353.
- [8] J. Jang, C. Sun, and E. Mizutani, “Neuro-Fuzzy and Soft Computing,” Prentice-Hall, 1997.
- [9] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Trans of the ASME, J of Basic Engineering 82D, pp. 35-45. 1960
- [10] Y. Kanayama, Y. Kimura, F. Miyazaki, and T. Noguchi, A stable tracking control method for a nonholonomic mobile robot, Proc of the IEEE/RSJ Int Conf Intelligent Robots and Syst, 1991, pp. 1236-1241.



진태석

1998 진주산업대학교 전자공학(공학사)
2000 부산대학교원 전자공학과(공학석사)
2003 동대학원(공학박사)

2004~2006 동경대학 생산기술연구소 연구원
2006~현재 동서대학교 메카트로닉스공학과 교수
관심분야: 센서융합, 지능로봇, 움직임추정



하시모토 히데키

1981 동경대학 전기공학과(공학사)
1984 동대학원(공학석사)
1987 동대학원(공학박사)
1989~1990 MIT 방문연구원

1990~현재 동경대학 생산기술연구소 교수
관심분야: 지능제어, 로보틱스, 공간지능화