Current Status of Intelligent Space

Hideki Hashimoto, Takeshi Sasaki, Laszlo Attila Jeni
(*Institute of Industrial Science, University of Tokyo, Tokyo* 153-8505, *Japan*)

Abstract – **Latest advances in network sensor technology and state of the art of mobile robotics and artificial intelligence research can be applied to develop autonomous and distributed monitoring systems. Intelligent Space (iSpace) is an environmental system, which is able to support human in informative and physical ways. iSpace observing the space with distributed sensors, extracts useful information from the obtained data and provides various services to users. This means that essential functions of iSpace are "observation", "recognition" and "actuation." In this paper, we focus on the observation function of iSpace. And we describe observation systems to get information of both human and mobile agents in the space to show new results.**

Key words – *current status*; *intelligent space*; *network sensor technology*

Manuscript Number: 1674-8042(2010)01-0086-07

dio: 10.3969/j.issn.1674-8042.2010.01.18

1 Introduction

In recent years, the demand for services such as human assistance and health care is increasing in countries where the society is aging. Before giving an actual service to people, we need a large number of information about them. To solve this issue, many works try to turn the daily environment into an intelligent one. Intelligent Space (iSpace) which has been studied in Hashimoto Laboratory at the University of Tokyo[1] is one of these works.

Fig. 1 shows the concept of iSpace, which is a space with multiple distributed and networked sensors and actuators. In iSpace, not only sensor devices but also sensor processing intelligence is distributed in the space because it is necessary to reduce the network load in the large-scale network, which can be realized by processing the raw data in each sensor node before collecting information. The sensor nodes distributed in the space are called Distributed Intelligent Network Device(DINDs). A DIND consists of three basic components: sensors, processors and communication devices. The processors deal with the sensed data and extract useful information about objects (type of object, position, etc.), users (identification, posture, activity, etc.) and the environment (geometrical shape, temperature, emergency, etc.). The network of DINDs can realize the observation and understanding of the events in the whole space.

Based on the extracted and fused information, actuators such as displayers or projectors embedded in the space provide informative services to users. In iSpace, mobile robots are also used as actuators to provide physical service to the users, and for them we use the name mobile agents. A mobile agent can utilize the intelligence of iSpace. By using distributed sensors and computers, the mobile agent can operate without restrictions for the capability of onboard sensors and computers as shown in Ref. [2]. Moreover, it can understand the request from people and offer appropriate service to them.

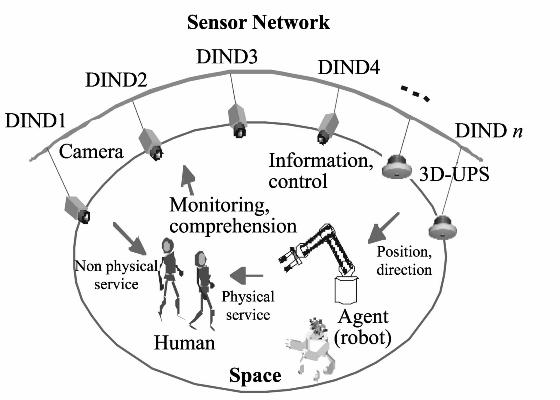

Fig. 1 Structure of intelligent space(iSpace)

Fig. 2 shows the present configuration of the iSpace in the Hashimoto Lab. The present configuration involves various sensors including Charge-Coupled Device (CCD) cameras, an ultrasound Zone Positioning System (ZPS), Laser Range Iinders (LRF). Moreover, the iSpace has mobile robots, a large size screen and speakers for presenting physical service and information to the users of the space. All the modules are connected through the local area network. Also, for achieving appropriate conditions for the operation of cameras, the lighting in the space can be easily adjusted.

In this paper, our current research topics on iSpace are explained. Especially, we focus on the observation function of iSpace. In section 2, 3 and 4, researches on human observation, position estimation using electric-field sensors, activity recognition using motion sensors, and facial expression recognition using cameras, respectively, are introduced. Section 5 shows observation of mobile robot for assisting the localization function of the robot. Section 6 introduces an application of an observation sys-

※ **Received:** 2010-01-06
Corresponding author: H. Hashimoto(hashimoto@iis.u-tokyo.ac.jp)

tem to surveying tasks in a construction field. Finally, a conclusion is given in section 7.

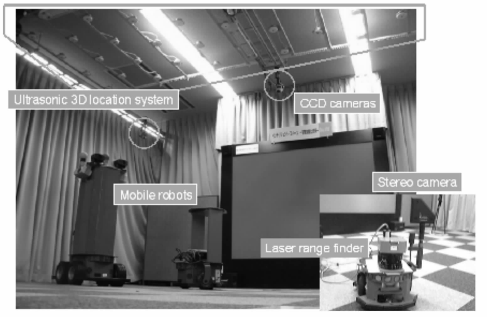

Fig. 2 Experiment environment

2 Human activity recognition based on 4W1H architecture

It is thought that the relation between a human and an object is described by observing the use history of the object[3]. The use history of the object is observed by focusing on the object's movement what is caused when it is used by the human[4]. The name, size, color, and shape etc. of the object are information given beforehand. On the other hand, there is information what occurs only after a person uses the object, such as the use history or the movement history.

Such information is vast. Therefore, considering the cost, it is not realistic to describe the use history information of a wide arrangement of objects that exist in the space. Hence, it is necessary that the object's information is written automatically without human interaction when the object is used by a person[5]. We try to describe human-object relations based on the following use history of the object (4W1H).

1) Where: the position of the object;
2) Who: the user of the object;
3) What: ID of the object;
4) When: the time of the object used;
5) How: the way of the object used.

The 4W1H context is used as a final application of the architecture that uses both Compressive Sensing (CS) and Self Organizing Maps (SOM) specifically for the "How" segment of the algorithm.

2.1 Hardware description

To perform the experiments we used an MTx sensor from the company Xsens, which is a small and accurate 3DOF inertial Orientation Tracker. It provides drift-free 3D orientation as well as cinematic data: 3D acceleration, 3D rate of turn (rate gyro) and 3D earth-magnetic field[6]. The system contains nine sensors which can be interlinked with each other in order to obtain a more complex set of data out of one specific object, as well as to provide a good architecture for setting referenced cinematic systems[6]. The data is retrieved using the Matlab toolbox that comes with the product, which allows us to acquire in real time all the needed data from the sensors.

The computer used was one with an Intel Core 2 Duo processor running Windows XP Professional Edition and Matlab 7.0. As well, we are making use of the L1 Magic toolbox created by Emmanuel Candes at Calthec to develop and make use of the Compressive Sensing Algorithms in a Matlab Environment.

We will be using as well position sensors and RFID tags for the further implementation of the 4W1H architecture, yet for this article we have not included experimental data regarding these sensors.

2.2 Algorithm description

The designed algorithm consists in 2 main blocks (Fig. 3), the sensing or sampling segment in which we make use of CS techniques to retrieve the data as fast as possible[7], and the classifying segment in which we apply the Self Organizing Map techniques[8].

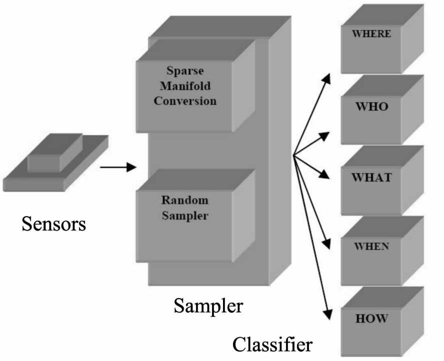

Fig. 3 Algorithm description

The first part obtains the data from the set of sensors and decide which mani-fold is the best option for the current working signal. And we apply the random sampling technique described when explaining the CS theoretical background. Afterwards we retrieve our signal in the main server and reconstruct it with L1 norm convex optimization. After the signal processing segment of the algorithm, a pre-classification done, we separate the data according to its nature: location, use, identification and generate five matrices; each of those matrices will have specific data for specific users and object interactions.

In this section we emphasize the processing of the "How" matrix in which we apply a SOM to classify the different inputs the system may present. And afterwards we will in future work apply an optimization algorithm in which we will be matching the resulting matrices with pre-recorded ones in order to identify the user and recognize or predict user actions in the space.

2.3 Results

When designing the sensing block of the algorithm we decided to apply simple Fourier transformation and use an orthogonal Fourier space for sampling. We decided to do this in order to comply with the sparsity of the signals, and being that the test signals where mainly circular-like movements of the hands a Fourier bases on seems correct to apply. And we expected to give good results.

In the experiment we tested the accelerometer sensor and apply the sensing algorithm to a single axis, expecting to get similar results from all of the other axes. We used an $N = 512$ for the signal elemental size. In Fig. 4 we can appreciate the original signal in red, which has two main natural frequencies and a shorter one at the end. The line in blue is the signal once it has being sensed with a signal compression of 20% that being only 100 of the signal elements were sampled. Visual inspection complies in that the sampled signal has the same harmonics as the original signal. Afterwards we will test the performance of this sampled signal in the SOM.

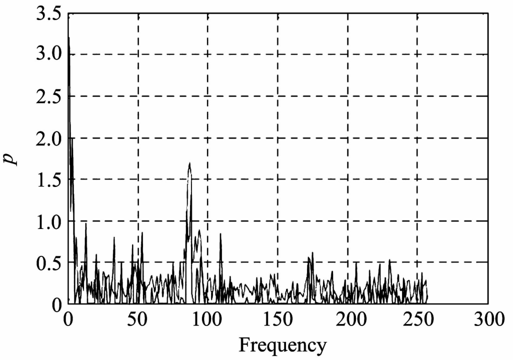

Fig. 4 Accelerometer X-ais data

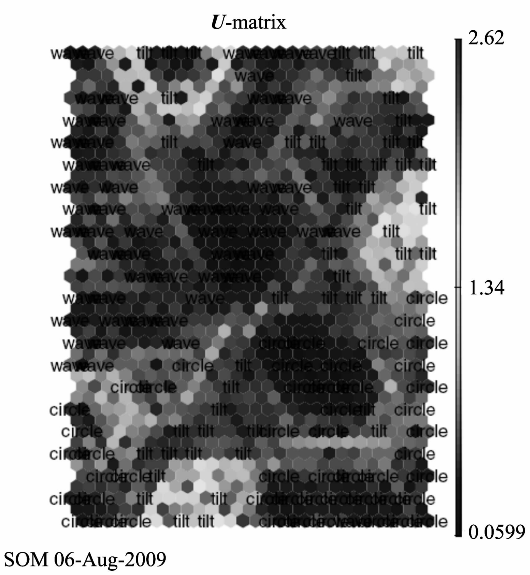

Fig. 5 20×20 map with 1% detection error

Now we will present some results of the classification output when sensing processed with the CS algorithm are applied to the overall architecture. In Fig. 5 we have the ***U***-Matrix (Graphical representation of a SOM where red elements represent farness and blue elements represent closeness) of a 20×20 map that showed 1% of error when recognizing previously trained patterns. It has been shown that this training error derives from the local minimum problem the SOM presents when being trained, since being a heuristic algorithm it is difficult to obtain a 100% recognition rate or a coherent result after training, yet it is very stable at the end of it.

As expected, the maps resulting from this work resembles those obtained in previous papers, showing that a CS technique effectively reduces the processing time and performs a very good classification.

3 Recognizing facial expressions based on machine learning technique

During the evolution of the human race several communication methods have been developed along with communication channels. These can be categorized into two main groups: verbal and non-verbal communication channels. Verbal communication can be attained easily, or transformed into another environment, so it came to existence early in the human-machine relationship. There are more than one way to communicate between humans (meta-communication) like expressions, gestures and postures. Nowadays the need for personal relationship (non-verbal communication) between humans and artificial tools is growing.

Different messages require different communication channels. Human beings use verbal, vocal and non-verbal signals to describe their emotional state. Facial expressions are a form of non-verbal communication, which is an outward reflection of a person's emotional condition. Recognizing these expressions help us to estimate the emotional state of a person.

3.1 Facial expressions

Facial expressions result from one or more motions or positions of the muscles of a face. These movements convey the emotional state of the individual to observers. Facial expressions are a form of nonverbal communication.

They are not only a primary means of conveying social information among humans, but also occur in other mammals as well as some other animal species. Facial expressions and their significance in the perceiver can, to some extent vary between cultures.

To describe these expressions Ekman et al. proposed an anatomically oriented coding scheme, the Facial Action Coding System[9]. This system is based on the definition of Action Units (AUs), what cause facial movements, of a face. Each action unit may correspond to several muscles that together generate a certain facial action.

As some muscles give rise to! more than one action unit, correspondence between action units and muscle units is only approximate. 46 AUs were considered responsible for expression control and 12 for gazing direction and orientation.

3.2 System configuration

The proposed system contains two main parts: a face tracking unit to extract the human faces and a learning system to learn and recognize the facial expressions (Fig. 6).

For the face tracking and extraction, the Dragonfly-

2 CCD Camera from Point Gray is used. In the research we used frontal images of the subjects, however their distance from the camera varies, therefore the size of the facial area does so too. Furthermore, sometimes some rigid head motion occurs during the tracking. This presents as rotation in the images, as well as some offsets, that need a solution.

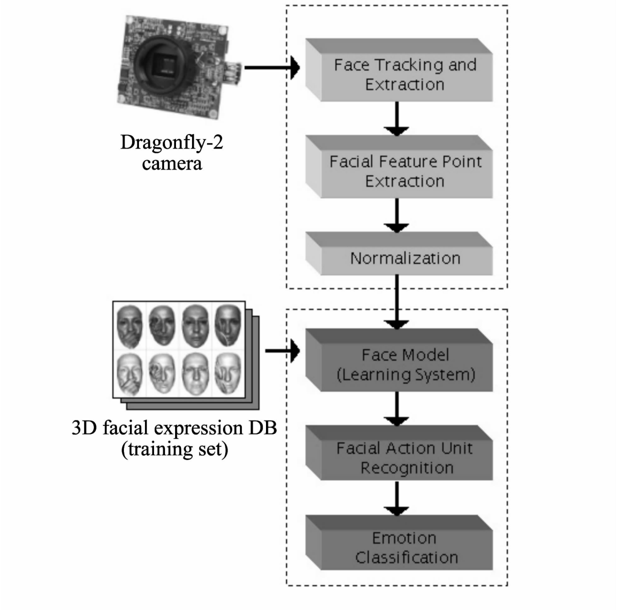

Fig.6 Procedure of facial expression recognition

The system, tracking the human face with high-speed cameras, extracts the pose-normalized image of the face and feeds it to a Bayesian Learning system[10], which provides a dynamic model of the human face and classifies the input image into AU codes.

4 Intelligent localization assistance for mobile robots

Research on mobile robot navigation focused on performing all subtasks of navigation by the robot itself. Meanwhile, there have been considerable researches done on tracking moving targets[11-14]. Often, this tracked information was utilized from the perspective of the tracker and was rarely used as an aid for the targets. Target tracking information can be successfully used as assistance for autonomous mobile robot navigation as active feedback information and is proposed here as "intelligent assistance." Intelligent assistance provides a means to aid mobile robots by reducing their computational overhead in navigation and enabling them to focus on their real application. Moreover, intelligent assistance can speed up robot navigation to achieve satisfactory service efficiencies. Most importantly, intelligent assistance will minimize any uncertainties in perceiving the environment by mobile robots.

4.1 System requirements and configuration

The main objective of this research is to introduce a novel scheme "intelligent assistance" as an aid for autonomous mobile robot navigation. An Intelligent Assistant (IA) with minimal set of sensing devices is expected to be developed with the following requirements.

1) assist mobile robots with localization by reducing their self-localization uncertainty (Fig.7);

2) reduce the computational overhead and occupy mobile robots solely on their applications;

3) provide global map information: mobile robots need not know the environment map nor build the map;

4) make mobile robots easily navigate in an unknown environment in the presence of an IA;

5) provide assistance for several mobile robots simultaneously;

6) both indoor and outdoor assistance;

7) expand assistance space by interconnected mobile intelligent assistants.

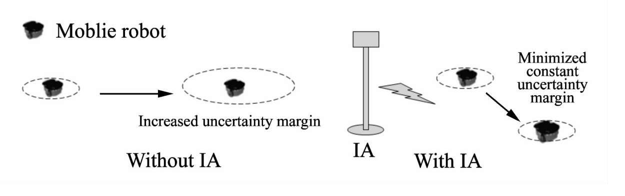

Fig.7 Reducing self-localization uncertainty using IA

The schematic diagram of the proposed system is shown in Fig.8. Background subtraction method is used to detect moving objects in both camera and laser range finder sensor data. The proposed sensor unit comprises an IEEE 1394 Point Grey Dragonfly-2 camera and a Hokuyo UTM-30LX laser range finder.

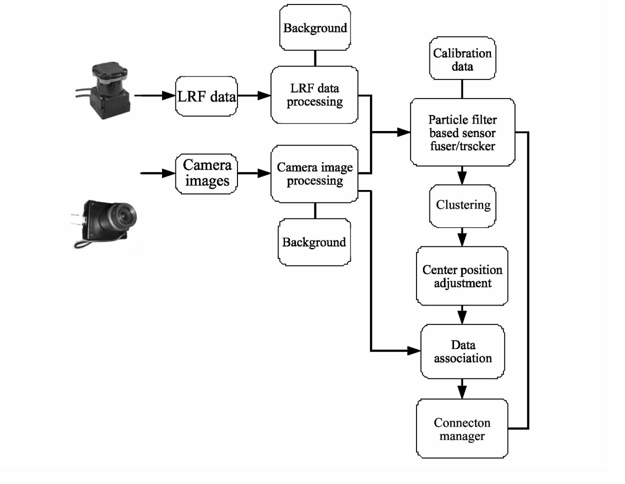

Fig.8 Schematic diagram of the proposed system

The two sensors are installed together in such a way that the optical axis of the camera is parallel to that of the LRF and are calibrated precisely beforehand. This implementation enables the coordinate transformation between laser scanning plane and camera coordinate system as simple as possible. Along with the sensor unit calibration information, the particle filter based sensor fuser tracks all those initiated targets. The feedback from the connection manager lets the tracker know which target to track (initiated targets) and which targets to be dropped. Output from the particle filter tracker is clustered and associated with respective to targets (mobile robots) and will be available to the connection manager, which keeps information about mobile robots. When assistance is needed, the mobile robots are connected to the connection manager, and requests localization and path planning information.

4.2 State-space representation of target dynamics and observation system

Effective extraction of useful information about the target's state from observations is the key to successful target tracking. To achieve this, two models, namely the system model and measurement model, are required.

Target State: The target state is represented with its X, Y coordinates as well its velocities in X and Y directions (Fig. 9).

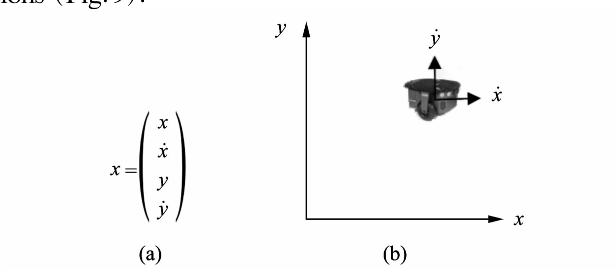

Fig. 9 (a) Target state; (b) Coordinate representation

By choosing the state like above, in addition to its position in the XY-coordinates, its velocity and thereby its heading direction can be estimated. This is because, as we have seen, most robots are non-holonomic, using differential-drive systems or Ackerman steered systems. For such robots, the non-holonomic constraints limit the robot's velocity in each configuration (x, y, θ), and as a result its heading angle can be computed using

$$\theta_k = \tan^{-1}\left(\frac{\dot{y}_k}{\dot{x}_k}\right). \quad (1)$$

System Model: We adhere a constant velocity (cv) white acceleration model described as follows (the equations only portrays with respect to X axis and equally applies to Y axis):

$$x_{k+1} = x_k + T\dot{x}_k + w_k,$$
$$x_{k+1} = x_k + w_k, \quad (2)$$

where x_k is the x coordinate at time k, T is sampling interval, w_k is process noise. As a result, we have the matrix equation:

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}_{\mathrm{cv}}\boldsymbol{X}_k + w_k,$$

for

$$\boldsymbol{F}_{\mathrm{cv}} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3)$$

Measurement Model: The likelihood is calculated for each particle. As camera observations and LRF observations are independent, the resulting likelihood can be given by

$$p(\boldsymbol{U}_k, \boldsymbol{Z}_k \mid \boldsymbol{X}_k) = p(\boldsymbol{U}_k \mid \boldsymbol{X}_k)p(\boldsymbol{Z}_k \mid \boldsymbol{X}_k), \quad (4)$$

where $\boldsymbol{U}_k$ is the processed image, $\boldsymbol{Z}_k$ is LRF data. $p(\boldsymbol{Z}_k \mid \boldsymbol{X}_k)$ is computed using

$$p(\boldsymbol{Z}_k \mid \boldsymbol{X}_k) = \exp(-d^2/2\sigma^2)/\sqrt{2\pi\sigma}, \quad (5)$$

for σ is position error for LRF and d is minimum Euclidian distance between $x_k^{(n)}$ (n^{th} particle) and z_i (i^{th} ray). Ref. [12] describes a method to evaluate $p(\boldsymbol{U}_k \mid \boldsymbol{X}_k)$ in which this probability is given by a simple function that returns a constant value $S(0 \leqslant S \leqslant 1)$ depending on the angle the target is presented in the image plane. Instead of such simple functions, statistical functions such as Gaussian kernels can also be utilized.

4.3 Preliminary results

Preliminary experiments were carried out tracking a single mobile robot-Pioneer 2DX. The Sampling Importance Resampling (SIR) particle filter is chosen for the state estimation for the sake of simplicity and effectiveness. Since camera data and LRF data are independent, only the LRF data is integrated in the measurement model.

The Pioneer 2DX is set to wander in the space, and the robot is tracked using the LRF measurements.

To validate the results, the pose of the robot is simultaneously tracked using the ZPS in the laboratory, via two ultrasonic tags attached to the robot.

Fig. 10 shows the results for position estimation. And Fig. 11 compares the heading angles estimated by the particle filter and ZPS. When the robot makes rotations with no translation velocity, the heading angle is not calculated, which explains the discontinuities in the graph.

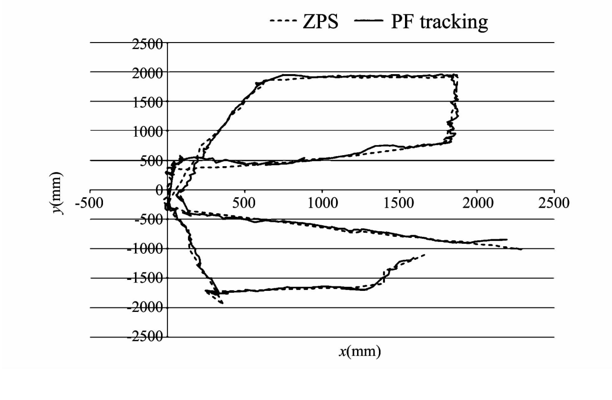

Fig. 10 Position estimation using ZPS system and PF tracker

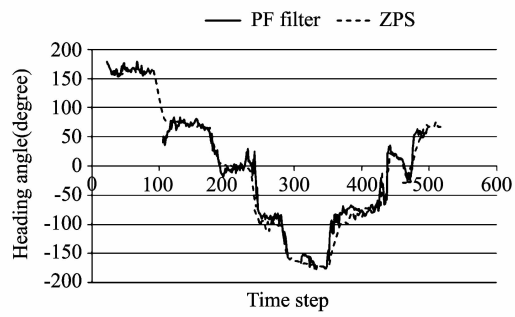

Fig. 11 Heading angle comparison

5 Surveying system in a construction field

In most construction fields, one often needs to put a mark on a certain position, and in most cases, these surveying tasks are performed by a Total Station. Although this device can survey with high accuracy, it has several disadvantages:

1) Impossible to track multiple objects at once;
2) Impossible to track in real-time;
3) Too expensive to purchase or even to rent.

To overcome such problems, we propose a position measurement system using LRFs.

5.1 System configuration

Fig. 12 shows an overview of the proposed system. In the proposed system, a worker moves in the construction field, carrying a reference bar, and it is detected by multiple LRFs set at a higher position than human height.

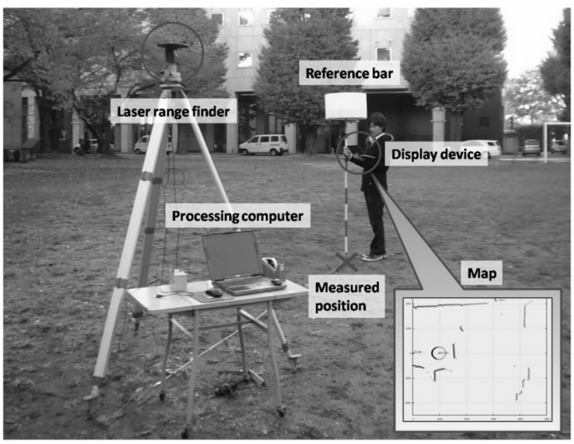

Fig. 12 Proposed position measurement system using LRFs

This real-time position measurement system is low-cost according to the price of the LRF, and can measure multiple positions at once when multiple reference bars are used, which leads to working hour efficiency. Also by carrying a mobile display device (e.g. a PDA or laptop), and displaying the positions of the reference bars on a map, the workers can easily gain information on where they are and where objective points are in the construction field through a wireless network.

However, since data from the LRFs are nothing more than the contour of the reference bar, we need to estimate its center position based on its contour. For this reason, we adopted cylindrical shaped reference bars, which will make the contour of the reference bar a circular arc irrespective of the direction from which it is scanned. We utilized the Least Square Method (LSM) and the Maximum Likelihood Estimation (MLE) to fit a circle equation to observed data points from an LRF[15]. Another problem using LRFs is that the number of sample points decreases if the distance becomes large, because the LRF emits a laser ray radially. Therefore, we set a moving head pan unit, SPU-01 from Sustainable Robotics, beneath the LRF as shown in Fig. 13(a). The unit can turn by the angular resolution of 0.015 degrees. Since the angular resolution of the LRF can be improved about 17 times by using the pan unit. By using 17 different scanned results, we can observe a dense contour as shown in Fig. 13(b).

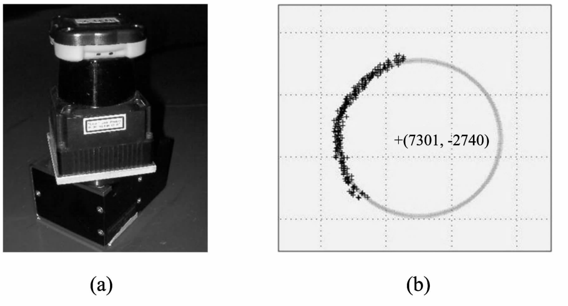

Fig. 13 (a) A combination of the LRF and the pan unit; (b) Obtained data by using the pan unit

5.2 Experiment

The experiment was to evaluate the proposed method. In this experiment, the target was placed on the points $(r,0)$, (r = 6 m, 8 m, 10 m, ⋯, 30 m) in the LRF coordinate system and the estimated error as compared among a conventional Constant Distance Method (CDM) proposed in Ref. [16] and the proposed LSM and MLE.

The estimated error of three methods without/with the pan unit is shown in Fig. 14 and 15, respectively. In the figures, an error bar shows the standard deviation of the estimated error at each distance. From Fig. 14 and 15, it can be said that the estimated error of the LSM and the MLE significantly decreases by improving the angular resolution using the pan unit.

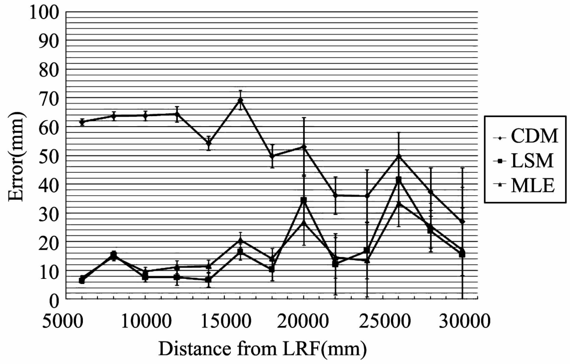

Fig. 14 Estimated errors of the three methods without pan unit

Especially, the estimated error of the MLE is stable through all distances and is less than 12 mm. Moreover, the variance of the estimate error at each distance could be decreased compared with Fig. 14, which is necessary for surveys in construction field to determine the measured position precisely.

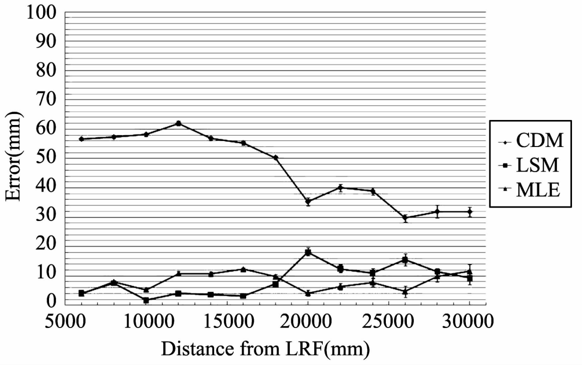

Fig. 15 Estimated errors of the three methods with pan unit

6 Conclusion

This paper described the ongoing researches on the observation function of Intelligent Space (iSpace) which has ubiquitous sensory intelligence. The ultimate goal of iSpace project is to accomplish an environment that comprehends human's intentions and satisfies them. Even though such a complete system cannot be achieved immediately, it is certain that a useful system can be achieved with current technology by proper system integration.

References

[1] H. H. Lee, H. Hashimoto, 2002. Intelligent space-concept and contents. *Advanced Robotics*, 16(3): 265-280.

[2] J. H. Lee, H. Hashimoto, 2003. Controlling mobile robots in distributed intelligent sensor network. *IEEE Trans. on Industrial Electronics*, 50(5): 890-902.

[3] M. Niitsuma, H. Hashimoto, 2009. Observation of human activities based on spatial memory in intelligent space. *Journal of Robotics and Mechatronics*, 21(4): 515-523.

[4] K. Kawaji, K. Yokoi, M. Niitsuma, H. Hashimoto, 2008. Observation System of Human-Object Relations in Intelligent Space. 6th IEEE Int. Conf. on Industrial Informatics, p. 1475-1480.

[5] M. Niitsuma, K. Yokoi, H. Hashimoto, 2009. Describing Human-Object Interaction in Intelligent Space. 2nd Int. Conf. on Human System Interaction, p. 395-399.

[6] Xsens. MTx. http://www.xsens.com/

[7] L. Palafox, H. Hashimoto, 2010. A Compressive Sensing Approach to the 4W1H Architecture. IEEE-ICIT 2010 Int. Conf. on Industrial Technology, p. 1579-1583.

[8] L. Palafox, H. Hashimoto, 2009. A Movement Profile Detection System Using Self Organized Maps in the Intelligent Space. IEEE Workshop on Advanced Robotics and its Social Impacts.

[9] P. Ekman, E. L. Rosenberg, 2005. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (2nd ed.). Oxford University Press, New York.

[10] L. A. Jeni, Gy. Florea, A. Lorincz, 2008. InfoMax bayesian learning of the furuta pendulum. *Acta Cybernetica*, 18: 637-649.

[11] S. Särkkä, A. Vehtari, J. Lampinen, 2007. Rao-blackwellized particle filter for multiple target tracking. *Information Fusion Journal*, 8(1): 2-15.

[12] R. Kurazume, H. Yamada, K. Murakami, Y. Iwashita, T. Hasegawa, 2008. Target Tracking Using SIR and MCMC Particle Filters by Multiple Cameras and Laser Range Finders. 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, p. 3838-3844.

[13] K. Nakamura, H. Zhao, R. Shibasaki, K. Sakamoto, T. Ohga, N. Suzukawa, 2006. Tracking pedestrians using multiple single-row laser range scanners and its reliability evaluation. *Systems and Computers in Japan*, 37(7): 1-11.

[14] X. Shao, H. Zhao, K. Nakamura, K. Katabira, R. Shibasaki, Y. Nakagawa, 2007. Detection and Tracking of Multiple Pedestrians by Using Laser Range Scanners. 2007 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, p. 2174-2179.

[15] H. Tamura, T. Sasaki, H. Hashimoto, F. Inoue, 2010. Circle Fitting Based Position Measurement System Using Laser Range Finder in Construction Fields. 2010 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems.

[16] D. Brscic, H. Hashimoto, 2008. Comparison of Robot Localization Methods Using Distributed and Onboard Laser Range Finders. IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics, p. 746-751.