Memory Recollection and Retrieval Based on Monitoring Human and Object in the iSpace

Kazufumi Saito, Akira Yoshimura, and Joo-Ho Lee

***Abstract*—In this paper, we propose a memory recollection and retrieval system using spatial log data in daily living which allows users to see behavior of the past, unconscious behavior and object-oriented behavior in the Intelligent Space. The proposed system consists of multiple camera devices and it utilizes various kinds of data; human identification, human motion, object instance, weather information, illuminance, heuristic information, etc. We constructed the system, focused on objects, in the Intelligent Space. To generate Metadata, based on collected log data to achieve the system of Recollection and Retrieval, several algorithms were used. For example, we used various image-processing algorithms of human detection, face recognition, object recognition, etc. for monitoring human and object. This paper presents the prototype of Recollection and Retrieval System, and updates original dataset, useful for various purposes in the Intelligent Space. Moreover, we improved the system performance of detecting object position compared to previous results.**

***Index Terms*—Intelligent space, environment monitoring, system integration, summarization.**

I. Introduction

In recent years, field of intelligent environment, which includes widely the Intelligent Space, smart home, sensor network, smart grid, etc, attracts people's attention. Many companies are interested in this field and they are investing lots of money and manpower. Equipments in such space can access to the network wherever and whenever. People may live more conveniently and more freely in this space. The most of projects in this field are related deeply with configuring the space to enrich the lives.

In recent years, as a result of various types of data digitized by the development of information technology, there is a huge amount of data diffusing in everywhere. The services such as twitter, facebook, pinterest, etc. are spreading, and the users are increasing too. The activity recorded to these services is similar to writing in one's diary, and a user is able to see the record like flipping an album of the photographs. On the other hand, recording information of daily living tends to create huge data. The user cannot see these records easily since the data is too big and it requires lots of time to see it. Moreover, there is a fundamental problem that recording user's memory is user-driven behavior. In fact, a user is able to record only intended information by the user himself. Thus, the user cannot recollect an unconscious action. Therefore, there are attempts to configure the space to record life-log. The purpose of these attempts is based on realizing smart, convenient life in the real environment. The space automatically records information of the persons in the space without persons' paying attention to the recording. For example, there are many researches which are health related concept of monitoring the pattern of person's living in long term observation. Ref. [1] presented that the method of unsupervised discovering activities using various sensors; motion sensor, thermometer, usage of phone, etc. There is a research which creates personal database of various information in daily living [2]. This database includes log of PC usage, phone call, watching TV, listening to music, screensaver occurred, light intensity, etc. observed by camera. In [3], in order to recognize the usage of an object, the first person who interacted with the object was recorded with a camera and it was analyzed. They utilized the appearance of user's hand and object in first person view to recognize "all about the objects being interacted with". From these researches, it can be concluded that one of the most important things is to monitor people. On the other hand, only monitoring people may limit the information for support people since the information which is focused only on the people is not enough. Especially, the unconscious events may occur around the people and there is possibility, even though it is small, that such unconscious events may influence the people.

Therefore, we propose Memory Recollection and Retrieval based on Monitoring Human and Object in the Intelligent Space that is focused on not only the person but also things. This paper is organized as follows. In Section II, we describe the Intelligent Space constructed in out laboratory, and our proposed system. Our approach of Metadata extraction is presented in Section III. We describe experiments and analysis in Section IV. Finally, conclusion and future works are described in Section V.

II. Intelligent Space

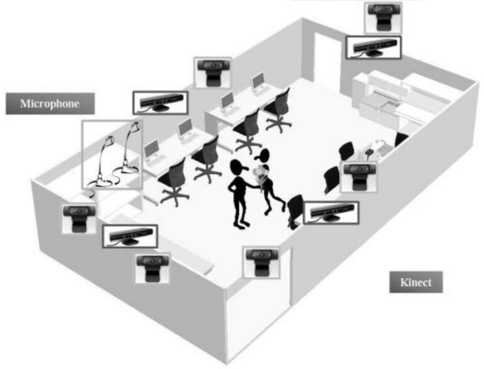


Fig. 1. System environment called "iSpace".

The Intelligent Space [4] is a room or area with sensors and

Manuscript received November 15, 2012; revised January 17, 2013.

The authors are with the Science and Engineering Department, University of Ritsumeikan, Japan (e-mail: cc005072@ed.ritsumei.ac.jp, is045089@ed.ritsumei.ac.jp, leejooho@is.ritsumei.ac.jp).

DOI: 10.7763/JACN.2013.V1.20

physical/virtual agents inside it. It supports humans in informative and physical ways. The Intelligent Space (hereafter, we call it iSpace) is able to perceive what happen to people by distributed sensors connected to the network and it provides information for the human and robots which are utilized to provide physical services to the people in the space. The software and hardware architecture of the iSpace must comply with a number of properties such as modularity, scalability, integration, realizability, low cost, and easy configuration. In order to organize the iSpace, sensors including cameras and microphones which can be obtained easily are mounted to the space. To satisfy the required properties of the iSpace, Distributed Intelligent Network Device (DIND) is required. DIND is composed of sensors, a processing unit and a network device. This DIND can recognize objects and events in the iSpace and it can share information by mutual communication through the network. Accordingly, it can be suitable for providing various services to people through agent robots. Moreover, it contributes for high accuracy recognition of the space by using the information obtained from multiple DIND through network (See Fig. 1).

Lots of the iSpace related researches have been conducted [4]-[6]. For example, there is research of storing a spatial history in the iSpace [5]. Lee proposed a system to store analytic information of what happened in a space and to extract a scene (e.g. human conversation) by people monitoring. In addition, this provides a digest movie of target scene automatically for memory recollection. Every research would create a new service to realize a high performance of the robot and the person interaction in the iSpace. On the other hand, because all these researches support human activity in the iSpace, they treat the information related to the person only. Consequently, we could provide only a service focused on human in the space, although, there are various changes which affect the person in daily living.

We propose the system to collect various kinds of information in the space. In particular, we focus on the object information. With respect to the advantages of the information obtained from the objects, if it is known that an object belongs to which object category in the space, the iSpace can collect information of what types of objects in it. In addition, if the object placement in the space is known, the robot is able to avoid it and provide services for human. The object information can be used for many researches in the iSpace. Moreover, the iSpace can store behavior conducted by human and robot. This indicates that the space can understand what happened between a human and an object.

Therefore, we propose Object Oriented Memory Recollection and Retrieval based on Spatial Log in the iSpace that is focused more on objects than the person.

III. The Proposed System

The proposed system's process is shown in Fig. 2; Collecting data, Metadata generation and Summary generation using integrated Metadata. We constructed the DIND with the sensors which are Kinect and C920 cam. First,

We collected basic data about color, pixel difference, two types of local features of HOG [7] and SURF [8] by recording video in the space. Second, we generated Metadata of human detection, human identification, human head orientation and body direction, and trajectory of movement based on the person. Moreover, as we mentioned previously in Section 2, we generated new information related to the objects. In particular, the system recognizes some robots in our laboratory. In addition, we tried to recognize the general objects, for example, chairs, mug cups, computers, industrial tools, drink bottles, cup noodles, etc as well as specific ones. We could recognize them in space using local features by computer vision algorithm. Moreover, we generate Metadata of heuristic and environment which consisted of the human relation, the specific object owner information, the layout of the iSpace. Preliminarily, we prepared many type of Metadata are constructed of various information in the iSpace, they are shown in Fig. 3. Finally, we summarize those Metadata to realize recollection and retrieval in long term observation. The following describes methods of generating some Metadata and summarize method, and original dataset.

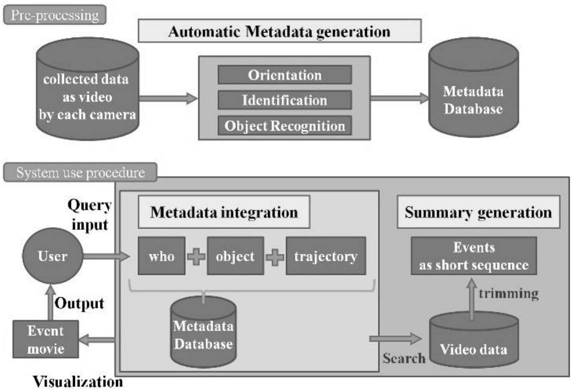


Fig. 2. Scheme of recollection and retrieval system

A. Metadata: Person Head Orientation and Body Direction

We collected the person head orientation as Metadata using proposed method by [6]. In this research, Tuan used HOG descriptor and then constructed the codebook which consisted of all extended templates based on HOG. As a result, the estimation of the head orientation has become possible robustly with only one camera. Moreover, we compute body orientation using Kinect sensors as well as head orientation. This is simple method which recognizes human trajectory in the short term for estimating head orientation. These methods have two advantages. One is that head orientation means human saliency to the objects (including human), and the other is that body direction means human trajectory. This information is used for a versatile storing event in the iSpace.

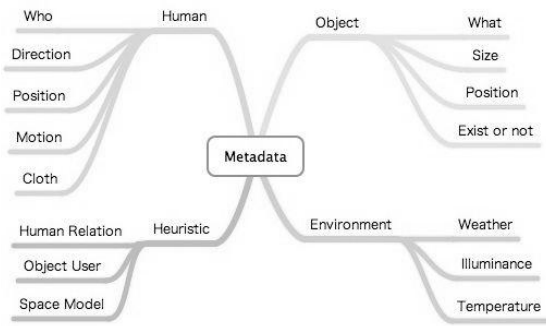


Fig. 3. Metadata component

B. Metadata: User Identification

We adopt a simple and robust method for user identification problem which utilizes the result of the matching process mentioned in [6]. In this method, we used 200 images per a person for training. The number of templates of person is 5.

C. Metadata: Object Recognition

Object information is indispensable in order to generate events in daily living. This is because human memory is related to various daily behaviors, for example, “Use an object for experiment in the iSpace”, “Take out an object”, “ROBOT activities (e.g. Our Lab. has a robot Ubiquitous Display called UD.)”, etc. The following is object recognition algorithm in our environment.

There are 2 approaches in our system. 1) General object recognition and 2) Specific object recognition. The former is used for object class categorization. For example, chairs, desks, computer, drink bottles, cup noodles, etc. The latter is used for Specific object recognition about UD, CEEE and MoMo which are the robots developed by our laboratory. These robots exist for various researches in the iSpace. Therefore, if the system can recognize them perfectly, we are able to make a reliable Metadata for existing object. Moreover, object and human identification Metadata are integrated to create new Metadata as event representation, for example, “When, Where, Who used this object” These Metadata are related to human activities very much.

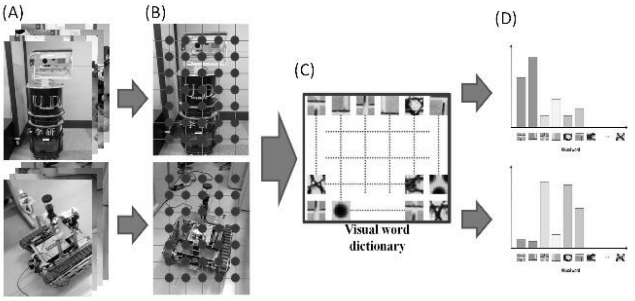


Fig. 4. Overview of the Bag-of-Features method. (A) collecting training images of learning for object recognition. (B) extracting sampling points and calculating descriptors for all training images. (C) clustering descriptors by all training images, and then deciding center of cluster as visualwords . (D) with each object category generating histograms of BoF based on visualwords.

Our system composed by recognition algorithm mainly used Bag-of-Features method shown in Fig. 4 in two approaches, but the method ignored spatial information from images because Bag-of-Features vector is amount of statistics. It is obvious that spatial information is very important to recognize objects from daily living data obtained by camera sensors in the iSpace. Therefore, we tried to extend this method by using Spatial Pyramid [9] shown in Fig. 5. This object recognition method consists of following two steps.

1) *Generation of BoF vector*

2) *Classification of object category*

Ref. [10] showed the detail of this algorithm and the result of accuracy of object recognition. However, the accuracy of object positioning by the algorithm of [10] was not satisfactory. Therefore, we extend the algorithm for detecting object position in the iSpace correctly shown in Fig.6. We introduced the evaluation value as likelihood. The value is computed by (1).

$$V(r_i) = Pos.(r_i) + Color(r_i) + Dis.(r_{i,f}, r_{i,f-1}) \qquad (1)$$

where *r* means a image in bounding box in a frame. *i* is the number of bounding box in a frame. *f* is the number of frame. *V(r)* is evaluation value. *Pos.* means the distance between object position and a bounding box in novel image. This object position is calculated by the center of same continuous pixels. *Color* is color similarity of each object using HSV histograms. *Dis.* means the distance between frames in time sequence. If the distance between two object regions is the shortest in previous frame and novel frame, this means the object is more likely same object in two regions. Therefore, *V(r)* becomes high value by increasing these parameters.

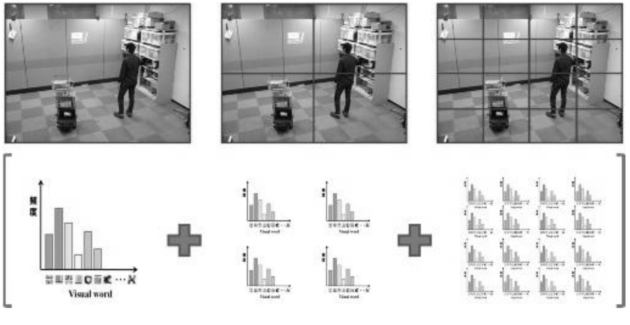


Fig. 5. Spatial Pyramid Representation

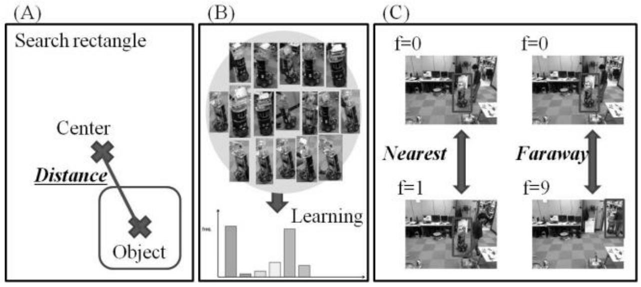


Fig. 6. The evaluation value computation algorithms for detecting object’s position. (A) the distance between object and rectangle’s center (B) the similarity of object color (C) the distance between frames in time sequence images

D. Metadata: Others

The other Metadata are described below. These Metadata are collected by user himself or by accessing WEB site of weather news automatically.

1) *Heuristic*

- Human relation
- Object owner
- Space model

2) *Environment*

- Weather
- Illuminance
- Temperature

Human relation of Metadata means a team member of research project and status such as seniors or juniors in the laboratory. The object owner means member of using the ROBOT for research in the iSpace (UD, CEEE, MoMo). Space model is layout of the iSpace which has the location of each fixed object; for example, desks, racks, working table, sensors such as cameras or microphone. Weather is collected by WEB site, and illuminance is observed by cameras, and temperature is measured by thermometer.

We collected these Metadata since these basic Metadata are useful for generating new Metadata which are augmented by integrating two Metadata of object name by recognition method and object owner of heuristic data. This means that the system can retrieve the event like "the object was taken away by its owner". If the object was taken away by not owner, it means that a big trouble occurred in the space. The iSpace can understand this problem by integrated Metadata. Therefore, we have to collect various types of Metadata for recollection and retrieval.

E. Dataset of Daily Living in the iSpace

We constructed original dataset (Fig. 7) of daily living to recognize various objects, humans and events in the iSpace. First, we recorded a forward-and-backward motion as images during a month in the space. The images were stored when background changed within 10sec. Second, we annotated human, names of objects and behavior of "human is taking away objects" or "using objects", "human didn't touch objects" or "human is bringing objects", "experimenting" on each image. Our dataset size was 7.4GB consisted of about 200,000 images and text data. It was increased about 50% more than previous dataset of [10]. In the future, we will construct dataset which consists of long term recording data. We aimed to generate Metadata of relation between human and objects by our dataset automatically.

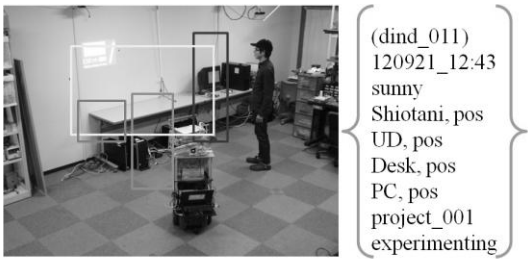
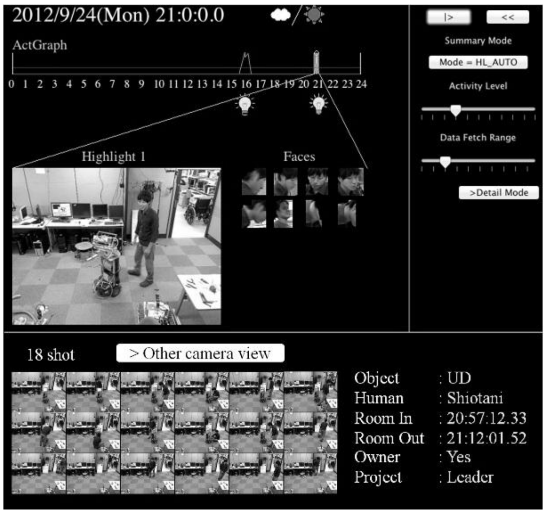


Fig. 7. A part of our dataset.



Object : UD
Human : Shiotani
Room In : 20:57:12.33
Room Out : 21:12:01.52
Owner : Yes
Project : Leader

Fig. 8. "What happened to the object in a day?" on our interface.

F. Summary of Metadata and Visualization

The procedure of creating summary in long term record is described. The summary is used for user to understand what happened in the space. We generated the summary of the space; for example, an amount of activity of human movement in a week, the conversation, the usage of the object by human, human activity that the human took away an object, etc. These summaries were generated by integrated Metadata which were based on human identification, object recognition, heuristic data, etc. The movie in time order is generated automatically by the system. An example is shown in Fig. 8.

IV. Experiments and Analysis

A. Procedure

In this experiment, we evaluate efficiency of our proposed system whether it can detect events (scenes) by integrating various Metadata which was generated beforehand. The experimental procedures are as follows. First, we collect the data as video for 24 hours by 10 cameras which are located in the space. Second, we generate Metadata by our proposed method described in Section 3 from the video of 24 hours. Finally, we set the events as detection tasks, and then the system detects those events by integrating various Metadata.

1) Who took away the object

2) What happened to the object in a day (Summary)

To validate performance of the system, we put the correct label on the events in the collected data beforehand, and then we check whether the system could detect the events or not.

B. Experiments and Evaluations

1) Initializations

The proposed method was implemented in C++ programming language, and all experiments in this section were performed under PCs with Intel Core i7-860 2.80GHz CPU and 4GB of RAM. We generated various kinds of Metadata by using each method as mentioned above. In particular, we adopted a dense regular grid sampling instead of interest points. We also used 64 dimensions SURF as local descriptors because we found that the difference of performance was little compared to SIFT and SURF. We used 500 visual words in BoF method and 3 levels area segmentation in Spatial Pyramid method because of little difference of performance comparison of 1000 visual words in previous our research. Moreover, we interpolated the result of an object recognition and detection of its location using the viewpoint of multiple camera devices when the objects are influenced by occlusion. In addition, we adopted the evaluation value as likelihood (Fig.6) for accurate object positioning. It improved accuracy more than simple additional approaches in previous research. We generated Metadata of the objects more correctly that is described in Section 3.C.

2) Result and analysis

Fig. 8 shows the results of event detection about "Who took away the UD outside" The results consisted of some sequence images constructed by using Metadata about object existence and extinction in the space, and the human existence near the object for a while. The results mean that the space situation is more likely to be "human used the object". Moreover, the results consisted of the situation of "the human is object owner" and "the object was distinct from human". Therefore, if the object was disappeared when the human go outside, this means the system detected event that was "The human more likely to have taken away the object". In previous research, the system had the performance about detecting movement of UD by human was 66.7%, but the new prototype system could get same performance. However, we found that our new method had better ability to detect object position in each frame than previous one. Therefore, we consider this problem is not related to the algorithm of detecting object position because we constructed our environment which has multi views by locating multiple cameras to avoid occlusion. From the above analysis, we found that the movement speed was related with the system performance. This means that the system cannot detect an object completely because human moved quickly with object. Therefore, our system could not record a tiny variation of object because we recorded images at intervals of 10 seconds during 24 hours. We need to improve the system in recording images and in extracting Metadata for more flexible representation of past scene summary.

The result is also shown in Fig. 8 of "how the object was used by humans in a day". The system visualized the summary by various kinds of information of not only the object owner but also weather and the amount of human activities in the space at same time. From the result of summary of human activity with object in a day, user can understand what happened to the object quickly. Moreover, if user is the object owner, he can recollect the past in the iSpace. But if the human saw the same event, what they perceived might different individually. However, the detected events can be adapted to various kinds of recognition and discovering. Followings are some daily examples to be applied; 1) The system informs the laboratory member to pull things together in the space if many objects are scattered all around the place. 2) The system generates summary of one's attitude change during a period such as a year or more. We consider it that these events are useful for recollection and retrieval by individuals.

From the above analysis, we conclude that the system had efficiency of object oriented memory recollection and retrieval by generated events both human and objects. However, there are many important things to be improved of the system. The way of visualization of the events needs to include more variations. Moreover, currently the number of specific objects which can be recognized is not satisfactory. We have to increase specific objects and improve classifier performance for every object. In addition, we need to detect the general object because we experimented to retrieve the events extracted by using only specific object. There are many types of general objects in common environment (e.g. scissors, industrial tool, etc.). Actually, we used pre-annotated dataset shown in Section 3.E to generate the classifier of each objects category. However, recognition rate of small general object is not good enough to use and we need to improve algorithm or to adopt a new method.

V. Conclusions and Future Works

We proposed the prototype of memory recollection and retrieval system based on monitoring human and object in daily living. We performed experiments on the system to detect some events like "Who took away the object" and "What happened to the object in a day". The result of experiments revealed that the proposed system is more flexible than the other past systems, since our new method is focused on the objects as new Metadata including human identification, estimation of head orientation, etc. Moreover, our method of detecting object position was improved in accuracy compared with previous system's approach. In the future, the system should be improved to cope with more flexible recollection and retrieval in long term data. The retrieval with user's original query without pre-defined constraint is also an important future work.

References

[1] P. Rashidi, D. J. Cook, L. B. Holder, and M. S. Edgecombe, "Discovering Activities to Recognize and Track in a Smart Environment," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 4, April 2011.

[2] J. Gemmell, G. Bell, and R. Lueder, "MyLifeBits: A personal database for everything," *Communications of the ACM (CACM)*, vol. 49, no.1, pp. 88-95, 2006.

[3] H. Pirsiavash and D. Ramanan, "Detecting activities of daily living in first-person camera views," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[4] J.-H. Lee and H. Hashimoto, "Intelligent Space – concept and contents," *Advanced Robotics*, vol. 16, no. 3, pp. 265-280, 2002.

[5] S.-O. Lee, R. Sakurai, T. Nishizawa, J.-H. Lee, and G.-T. Park, "A spatial history storing system," *IECON'08*, HF-013803, 2008.

[6] D. T. Tran, Y. Koizumi, R. Sakurai, and J.-H. Lee, "Robust methods for head orientation estimation and user identification based on hog and codebook," *SI International*, pp. 224-229, 2011.

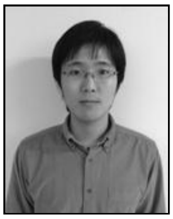
[7] N. Dalal and B. Trigss, "Histograms of oriented gradients for human detection," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886-893, 2005.

[8] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speed up rpbust features," *ECCV, Part1, LNCS3951*, pp. 404-417, 2006.

[9] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[10] K. Saito, A. Yoshimura, and J.-H. Lee, "Object oriented memory recollection and retrieval based on spatial log in the iSpace," *SI International*, pp. 271-276, 2012.

K. Saito was born in Kobe, Japan in 1987. He received the B.S degree in Computer Science from Ritsumeikan University, Shiga, Japan in 2011. He is studying computer vision mainly.



A. Yoshimura was born in Osaka, Japan in 1989. He received the B.S. degree in Computer Science from Ritsumeikan University, Shiga, Japan in 2012. He is studying computer vision mainly.

J.-H. Lee was born in Seoul, Korea in 1970. He received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, Korea in 1993 and 1995 respectively and the Ph.D. degree in electrical engineering from the University of Tokyo, Tokyo, Japan in 1999.

He was a Post-doctoral Researcher and a Japanese Society for the Promotion of Science Postdoctoral Researcher at the Institute of Industrial Science, University of Tokyo in 1999 and 2000 respectively, and a Research Associate at Tokyo University of Science in 2003. He Joined Ritsumeikan University in 2004 and he is currently a Professor in the College of Information Science and Engineering. He was Visiting Researcher in ATR and Visiting Scholar in the Robotics Institute of CMU, in 2006 and 2008, respectively. His research interests include intelligent space, intelligent interaction, service robots, and machine vision.

Professor Lee is a member of IEEE, IEEJ, SICE and RSJ.