We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Human System Interaction through Distributed Devices in Intelligent Space

Takeshi Sasaki\(^1\), Yoshihisa Toshima\(^2\), Mihoko Niitsuma\(^3\) and Hideki Hashimoto\(^1\)

\(^1\)Institute of Industrial Science, The University of Tokyo, \(^2\)Daikin Industries, Ltd., \(^3\)Chuo University

Japan

1. Introduction

Intelligent Space (iSpace) is a space with ubiquitous sensors and actuators (Lee & Hashimoto, 2002). iSpace observes the space using the distributed sensors, extracts useful information from the obtained data and provides various services to the users through the actuators. iSpace can be considered as an “invisible” robot that is united with the environment since it can carry out the fundamental functions of robots - observation, recognition and actuation functions.

This type of spaces is also referred to as smart environment, smart space, intelligent environment, etc. and recently there is a growing number of research work (Cook & Das, 2004). Some smart environments are designed for supporting the users in informative ways. For example, a meeting support system (Johanson et al., 2002) and a healthcare system (Nishida et al., 2000) using distributed sensors were developed. Other smart environments are used for support of mobile robots to provide physical services. Delivery robots with ubiquitous sensory intelligence were developed in an office room (Mizoguchi et al., 1999) and a hospital (Sgorbissa & Zaccaria, 2004). The functions of mobile robot navigation including path planning (Kurabayashi et al., 2002) and localization (Han et al., 2007), (Hwang & Shih, 2009) of mobile robots were assisted by using information from distributed devices. Fig. 1 shows the configuration of our iSpace which is able to support human in both informative and physical ways.

iSpace has to recognize requests from users to provide the desired services and it is desirable that the user can request the services through natural interfaces. Therefore, a suitable human-iSpace interface is needed. Gesture recognition has been studied extensively (Mitra & Acharya, 2007) and human motions are often utilized as an interface in smart environments. A wearable interface device, named Gesture Pendant, was developed to control home information appliances (Mynatt et al, 2004). This device can recognize hand gestures using infrared illumination and a CCD camera. Gesture pads are also used as input devices (Youngblood et al, 2005). Speech recognition is considered as another promising approach for realizing an intuitive human-iSpace interface. The smart environment research
project described in (Scanlon, 2004) utilizes distributed microphones to recognize spoken commands.

On the other hand, interaction can also be started by the space. If iSpace finds that a user is in trouble based on observation, for example, a mobile robot in the space would go to help the user. To realize this, human activity and behaviour recognition methods in smart environments are studied actively (Mori et al, 2007), (Oliver et al, 2004). It is also important to develop actuators including display systems, audio systems and mobile robots in order to provide services based on the observed situations.

Here both types of human-iSpace interaction mentioned above are described in the following sections. Section 2 and 3 introduce our human-iSpace interfaces - a spatial memory and a whistle interface. The spatial memory uses three-dimensional positions whereas the whistle interface utilizes frequency of sounds to activate services. Section 4 presents an information display system using a pan-tilt projector. Sections 2, 3 and 4 give also experimental results to demonstrate the developed system. Finally, a conclusion is given in section 5.

---

**2. Spatial memory**

Fig. 2 shows a schematic concept of the spatial memory. The spatial memory regards computerized information, such as digital files and commands, as externalized knowledge and enables human to store computerized information into the real world by assigning a 3-D position as the memory address. By storing computerized information into the real world, users can manipulate the information, as if they manipulated physical objects. For example, as shown in Fig. 2, conference proceedings can be organized in front of file cabinets or special memories might be stored into the second drawer from the top.
2.1 Definitions of terms

1) SKT (Spatial-Knowledge-Tag): We introduce a virtual tag, which associates computerized information with a spatial location. We will call them SKTs. SKT has three important parameters, namely: 1) stored computerized information; 2) 3-D position as a memory address; and 3) size of an accessible region. The details of an accessible region will be explained later. Stored computerized information is called spatial memory data.

Environmental information, such as arrangements of equipment and objects, is adopted as tags, which represent the whereabouts of externalized knowledge. Equipment placed in a working environment has visually distinguishable functions, respectively, and humans are able to recognize them easily by using their own cognitive abilities. Therefore, when real objects represent tags, they will play a role of a trigger to recall stored data and will be effective to memorize the whereabouts. In addition, the location of objects can be utilized to arrange externalized knowledge for easy recall.

There have been several approaches to associate computerized information with real objects, e.g., using Radio Frequency Identification (Kawamura et al., 2007), (Kim et al., 2005) or using 2-D barcode (Rekimoto et al. 1998). The approaches are useful to recognize the objects in a physical sense. However, there is a need to directly attach a hardware tag to each object in advance, and the user can only arrange computerized information for predetermined objects. Optional properties regarding accessibility, security, and mobility are not easily changed because they depend on hardware specifications, such as antennas or cameras.

There are two key differences between 3-D position-based method and hardware-tag-based method, such as 2-D barcodes. First, in the case of using 3-D position, we do not need to directly attach the hardware tag to each object, and therefore we can store information freely, if the position and the motion of human can be measured. Second, stored data manipulations, such as changing optional properties, are easier than using the hardware tag.

Another important aspect of our spatial memory is that the SKT can also take a human-centric approach to store the computerized information. More specifically, let us just consider the case where the coordinate frame is defined based on the human body. We can realize...
transformation from a human-centric coordinate frame to the real world, since the human can also take an option to carry the SKTs with his motion so that the relative position of SKTs will not be changed, for example, “my left side 2 m away to get access to file A,” “my right side 10 m away to call my friend B.” However, of course, if the user attaches the computerized information to a real object in the initial stage, the approach on the spatial memory needs to recognize an object with a human intuitive assist to register the object to the memory.

2) Human Indicator: The spatial memory whose address is represented by its 3-D position requires a new memory access method. In order to achieve intuitive and instantaneous access method that anyone can apply, the spatial memory adopts an indication of a human body as a memory indicator. Therefore, when using the spatial memory, a user can retrieve and store data by directly indicating the positions using his own body, for example, user’s hand or user’s body. The position based on the user’s body used for operating the spatial memory is called a “human indicator.” However, it is impossible for a human to indicate the exact position of the spatial memory address every time. To easily and robustly access an SKT by using the human indicator, it is necessary to define an accessible region for each SKT. The accessible region is also needed for the arrangement of several SKTs by distinguishing their locations from others. The size of accessible region is determined based on the accuracy of the human indicator and its action type. For example, when using a hand for the human indicator, the user can indicate more accurately than using the body position. Therefore, a small size of accessible region can be achieved when using a hand, whereas a large one will be needed for a body human indicator. We notice here that a guideline to determine the size of a human indicator using a hand has been obtained. In our previous work, we have investigated the accuracy of the human indicator (Niitsuma et al., 2004) using the user’s hand. The accuracy is defined by the indication error, which is the distance between a spatial memory address of SKT and the human indicator. Investigations of two cases of human activities were carried out, namely: 1) the case of performing only the indication task and 2) the case of performing the indication task during another task. The results show the different margin of indication error between two cases. The accuracy of case 2 is worse than case 1 because the error margin of case 2 is larger than case 1. In order to achieve both smooth access and arrangement of several SKTs, accessible region is defined as follows. The accessible region is the sphere whose origin is located at a spatial memory address of SKT, whereas the radius is determined according to human activities: the radius in the case of just the indication task was found to be about 20 cm, and the radius in the case of indicating while performing another task was found to be about 40 cm.

3) Spatial Memory Address: As explained above, spatial memory addresses define spatial locations of computerized information in the spatial memory. Addressing method of the spatial memory system adopts a human-indicator-centered method, i.e., a position indicated by a human indicator is used for a spatial memory address. Consequently, the action for storing data into spatial memory can be carried out intuitively by pointing a spatial location as well as the action for accessing SKT. The implemented spatial memory has a 3-D coordinate system whose origin is at an arbitrary point in a space.

2.2 Usability evaluation of the spatial memory
Memorizing both the contents of SKTs and their whereabouts are required to utilize the spatial memory. If the users learn SKT positions and contents, they can get access to aimed
SKTs quickly without errors. Namely, it can be assumed that access time will be limited only by the physical access time necessary to utilize a human indicator and indicate the position of an aimed SKT. Therefore, the accessibility of the spatial memory and the effectiveness of memorizing were investigated from the viewpoint of the time needed to access SKTs. Human subjects memorized some SKTs, which had been arranged in advance, then accessed them. The task started from the situation where each subject did not have any information about the whereabouts and ended when the subject could access all SKTs. We measured the task completion time of each subject by changing intervals of tasks; more specifically 1 h later, 3 h later, and up to 20 days later, in order to check how the subject would memorize the spatial arrangement of SKTs. We then analyzed the time variation of the task completion time. Here, the accessible region was determined as the sphere with radius of 20 cm. The experiment was carried out by six subjects (21–26 years old, science or engineering students). All subjects have used the spatial memory for about 30 min before the experiment, and they know the usage.

The details of the specified task are described as follows.

1. Initialization phase
   a) An experimenter stores seven SKTs whose spatial memory data are given by images.

2. Learning phase
   a) A subject accesses stored SKTs and memorize the contents and the whereabouts. In this step, the “access indicator” which shows colors based on the distance between the human indicator and the spatial memory address of the nearest SKT on a computer display is used.
   b) Learning is ended by the subject’s decision.

3. Test phase
   a) Experimenter specifies the content of the SKT.
   b) The subject accesses the specified SKT.
   c) Phases 3-a and 3-b are repeated until all SKTs were accessed by the subject.

The task completion time from phase 2-a to phase 3-c was measured for each subject. Fig. 3 (a) shows the completion time of each subject (Subjects 1–6). The horizontal axis represents logarithmic time [h] that had passed from the experiment start time, and the vertical axis represents the completion time [s]. All subjects learned the contents and the whereabouts of seven SKTs at the first performance, which resulted in a long completion time. The completion time of the first performance of Subject 6 is the shortest because he stored all SKTs before the experiment. All subjects completed accessing all SKTs at the performance after about four weeks from the first performance as short as the second performance. Although Subject 3 required learning at the third performance, the learning time was 50% less than the first performance. After the third performance, he completed the tasks in a time as short as the other subjects did.

These results show the easiness of accessing the stored SKTs by memorizing the spatial locations because almost all subjects did not require learning of SKTs after the first performance. The completion time at the last performance of all subjects became 18–24 s, which shows that the accessibility was maintained or even improved over time.

Fig. 3 (b) shows the completion time of each subject depending on the intervals between the performances in order to investigate the effectiveness of memorizing the stored SKTs. The horizontal axis represents the logarithmic interval time [h] of task executions, and the vertical axis represents the completion time [s] of each task performance. The figure shows the completion times from the second to the last performance. In the experiment, the interval
between performances was increased according to the number of performances, although the interval time is not exactly the same among subjects. Thus, the last performances of all subjects are performed with an interval time of about 500 h (about 20 days).

The completion times of three subjects fluctuated until the first half of the experiment, where the interval time was less than 20 h. The variations of the subjects’ completion time, however, decreases when the number of the execution times increase, and the completion times become shorter. Other subjects carried out the task in an almost fixed time through all performances. The time 18–24 of the last performances is close to the physically needed time to access SKTs. In addition, all subjects successfully completed getting access to all SKTs in the performance even after about 20 days from the first performance. As shown in Fig. 3(b), the performance after 20 days is as short as the performance of 2-h duration.

The results show that the subjects were able to recall the stored SKTs without forgetting them, and accessibility of the spatial memory has been maintained or even improved even if the interval time between usages increased. Therefore, the spatial memory approach in which the access method uses the human body and the storing method tags a real environment is effective for minimizing the forgetting of stored computerized information even if time has passed since it was stored.

![Graphs showing task completion times](a) ![Graphs showing task completion times focused on the task interval](b)

Fig. 3. Result of the usability evaluation experiment (a) time variation of the task completion time of each subject, (b) task completion time focused on the task interval

### 2.3 Service execution using spatial memory

By storing services into a space using the spatial memory, we can execute various services in iSpace. Fig. 4 shows an example of a service execution using the spatial memory. In the example, the spatial memory is used for sending commands to a mobile robot. A “call robot” service was stored behind a user and the user called a mobile robot by indicating the position (Fig. 4(a)-(c)). We also developed an interface to create and delete SKTs. This interface contains a speech recognition unit and SKTs can be managed using voice commands. In the example, another “call robot” service was stored in the user-specified position by using the interface (Fig. 4(d)-(f)).
Fig. 4. Service execution using the spatial memory (a) request for "call robot" service by indicating the specified position, (b)(c) execution of "call robot" service, (d) service management by using a voice command, (e) storing the "call robot" service, (f) execution of "call robot" service
3. Sound interfaces

Sound interfaces provide another method for activating services in iSpace. Although iSpace has speech recognition units as shown in the previous section, we introduce a simple but robust sound interface using a human whistling in this section. Here we consider the frequency of sounds as a trigger to call a service, i.e. a service is provided when the system detects a sound which has the corresponding frequency. The advantages of using a whistle as an interface are that humans do not have to carry any special devices and the range of the sound can be expanded through exercises to activate different types of services depending on the pitch. In addition, it carries a long way and can be easily detected by using distributed microphones. Fig. 5 shows an example of sound waveforms and their frequency spectrum for various sound sources obtained by Fourier analysis. As shown in Fig. 5 (d), the sound of a whistle is considered as a pure tone and easily recognized by considering the percentage of the power of the main frequency component among the total power of the sound.

![Sound waveforms and frequency spectrum](image)

Fig. 5. Result of frequency analysis for various sound sources (a) human voice, (b) melodica, (c) metallophone, (d) human whistle

Fig. 6 shows an example of human-robot interaction through the sound interface. In this example, each sound denotes different commands and a mobile robot is controlled based on the commands. Here we tested various sound sources and played a sound of a different pitch for each source. As shown in the figure, the system detected the sound and the mobile robot generated the corresponding motion successfully. We note that as also shown in Fig. 6, since the sound of the melodica contains rather large harmonic components, the system sometimes failed to detect the sounds. On the other hand, a whistle is robustly recognized even in the presence of environmental noise. By associating the frequency of sounds with services, this interface can be used to activate various services by the users.
4. Information display using a pan-tilt projector

Based on observation of users in the space, iSpace can actively provide information which is expected to be useful for the users. Here we consider an interactive information display system using visual information. The information display system uses a projector with a pan-tilt unit, which is able to project an image toward any position according to human movement in the space. By utilizing the interactive information, many applications can be developed, for example, the display of signs or marks in public spaces, or various information services in daily life.

However, main issues in active projection are distortion of the projection image and occlusion of the image. These issues are addressed in the following subsections.

4.1 Compensation of projection image

When projection direction is not orthogonal to the projection surface, projection distortion occurs. Moreover the size of the projected image depends on the distance to the projection surface. Therefore with change of the projection point it is not possible to provide a uniform
image to a user. The projector provides uniform projection toward any position by compensation of the projection image by using a geometric model and inverse perspective conversion. The resize ratio $\gamma$ for compensation of the image size is given as follows.

$$\gamma(d) = \frac{W}{t(d)}$$  

(1)

where $d$ denotes distance between the projector and the projection surface, $W$ is the desired image size and $t(d)$ is a image size on the projection surface. Distortion is also caused by the angle between the optical axis of the projector and the projection surface. The geometrical definition is shown in Fig. 7. As shown in this figure, the pan-tilt projector projects an image toward $O_p$. The plane $Q$ is the projection surface and the plane $R$ is orthogonal to the projection direction. The points $r_1$ to $r_4$ denote the corners of the non-distorted image whereas the points $q_1$ to $q_4$ are the corresponding points on the distorted image. A relation between a point $p_Q$ on plane $Q$ and a point $p_R$ on plane $R$ is obtained based on perspective conversion:

$$\begin{bmatrix} p_Q \\ 1 \end{bmatrix} = H_{QR} \begin{bmatrix} p_R \\ 1 \end{bmatrix}$$  

(2)

This conversion matrix $H_{QR}$ is a $3 \times 3$ matrix and the degree of freedom is 8. Therefore, if four or more sets of corresponding points of $p_Q$ and $p_R$ are given, we can identify $H_{QR}$ and represent image distortion. The corresponding points can be found by the intersection of the plane $Q$ with the line through $r_i$ from the projection origin (lens). The inverse matrix of $H_{QR}$ represents compensation of image distortion and we can get pre-compensated output image.

![Fig. 7. Geometrical definition of image distortion](image)

4.2 Occlusion avoidance

Projection occlusion occurs when human enters into the area where the human obstructs the projection. This problem sometimes happens in active projection due to human movement or change in projection environment. Hence, by creating an occlusion area and a human
(obstruction) model and judging whether they overlap with each other, occlusion can be detected and avoided.
We modelled the shape of projection light and human as cone and cylinder, respectively. We judge the overlap between these two models to detect occlusion. Moreover, not only humans but also other objects including chairs and tables could cause the occlusion problem. Our occlusion avoidance algorithm can be used by considering the object shape model.
The avoidance method needs to modify the projection position so that the user can easily view the image. Fig. 8 shows the determination of the modified position. In the situation that the projection position is on the left side of the human model, the projection direction is moved to the left to avoid occlusion since it requires less angular variation compared to the rightward movement. On the contrary, when the projection position is on the right side of the human model, it moves to the right for the same reason. If the calculated correction angle is greater than the limit correction angle $\theta_{\text{max}}$, the projection position is moved away from the human.

4.3 Visitor guidance application
We developed a visitor guidance application as an example of interactive informative services. Fig. 9 shows the procedure of the visitor guidance. When iSpace detects a visitor using distributed sensors, the projector displays a guidance panel in front of the visitor. In this case, two messages “call robot” and “view map” are shown (Fig. 9 (a)(b)). If the visitor stands on the “call robot,” the projector provides a message “calling robot” and a mobile robot comes toward a user (Fig. 9 (c)). On the other hand, if the visitor stands on “view map,” the projector displays the map of the space in front of the visitor (Fig. 9 (d)). In addition, the projector indicates the direction of the place that is selected by the visitor (Fig. 9 (e)(f)).
Fig. 9. Guidance application using a pan-tilt projector (a) detection of a visitor, (b) display of a guidance panel (c) “call robot” service, (d) “view map” service, (e) selection of a place where the visitor wants to go (in the “view map” service), (f) indication of the direction of the place selected by the visitor (in the “view map” service)

5. Conclusion

Intelligent Space (iSpace) is an environmental system, which has multiple distributed and networked sensors and actuators. Since a variety of sensors including cameras,
microphones, laser range finders and pressure sensors are taken into account as sensor devices of iSpace, the users can interact with the space in various ways. The spatial memory was presented as an interface between users and iSpace. We adopt indication actions of users as operation methods in order to achieve an intuitive and instantaneous way that anyone can apply. A position of a part of user’s body which is used for operating the spatial memory is called a human indicator. When a user specifies digital information and indicates a position in the space, the system associates the three-dimensional position with the information and manages the information as Spatial-Knowledge-Tag (SKT). Therefore, users can store and arrange computerized information such as digital files, robot commands, voice messages etc. into the real world. They can also retrieve the stored information in the same way as on storing action, i.e. indicating action.

Sound interfaces are also implemented in iSpace. The whistle interface which uses frequency of a human whistling as a trigger to call a service was introduced. Since a sound of a whistle is considered as a pure tone, the sound is easily detected by iSpace. As a result, this interface works well even in the presence of environmental noise.

An information display system was also developed to realize interactive informative services. The system consists of a projector and a pan-tilt enabled stand and is able to project images toward any position. In addition, this system can provide easily viewable images by compensating the image distortion and avoiding occlusions.

6. References


Rapid advances in the field of robotics have made it possible to use robots not just in industrial automation but also in entertainment, rehabilitation, and home service. Since robots will likely affect many aspects of human existence, fundamental questions of human-robot interaction must be formulated and, if at all possible, resolved. Some of these questions are addressed in this collection of papers by leading HRI researchers.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:


InTech Europe
University Campus STEP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 886 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
Chinese Shanghai 65号, Shanghai International Building, 405 Unit
Phone: +86-21-62489820
Fax: +86-21-62489821