IJBSCHS (2010-15-2-15)
[Original Article]
Biomedical Soft Computing and Human Sciences, Vol.15, No.2, pp.99-105

(Accepted on 2009.1.10)

A Human-machine Interface for Manipulating a Virtual 3-D Object Using Natural Hand Gestures

Keisuke MORI*, Joo Kooi TAN** and Seiji ISHIKAWA**

** Department of Information Science, Faculty of Engineering Kyushu Kyoritsu University*
*** Faculty of Engineering, Kyushu Institute of Technology*

Abstract: *In this paper, we propose a new human-machine interface for manipulating 3-D contents in a display. Motions of a hand or both hands of a user are taken images by a simple USB camera. The motions are recognized by a PC and transformed into corresponding instructions such as translation, rotation and scaling. In addition, unintentional movement and unconscious movement are eliminated in the method and smoother operation is realized. The result of the manipulation is displayed in the screen. Experiments were performed using Google Earth and 3-D CG contents, and the proposed system successfully manipulated its rotation and scaling by some hand motions of the user.*

Keywords *Human-machine interface, hand gesture, pattern recognition, computer graphics.*

1. Introduction

The support service of many kinds of jobs with a computer has come to be used in various situations nowadays. Moreover, it has become more and more popular to handle a three-dimensional (3-D) space in a computer. The coordinates operation on a computer is generally done by using a mouse. As other devices, there is a track ball that turns a ball, and a pen tablet, etc., by which the coordinates on the screen are absolutely specified. As for such devices, natural operations can be done according to usage. However, certain skills are required particularly for 3-D operations. So, they are not always a natural interface for some purposes.

To operate a computer more naturally as well as freely, various methods are proposed. They include a method of detecting the movement of the radiant of a laser pointer [1] and a method of detecting gesture of an arm [2,3]. The use of a laser pointer is basically limited to the movement of the spotted point. On the other hand, a gesture of an arm is limited to the registered movement for instructions. Those who operate the system cannot perform natural operation directly, as they must get accustomed to the gesture.

In this paper, we propose a new man-machine interface [4,5] for manipulating 3-D contents in a display in a natural way. Motions of a hand (or both hands) of a user are taken images by a simple USB camera. The motions are recognized and transformed into corresponding instructions by a PC. The instructions are then applied to a 3-D object in a display to produce some geometrical transformation. Through this procedure, a user can feel that he/she handles a 3-D virtual object in the display dynamically and naturally by hand.

1-8 Jiyugaoka, Yahatanishi-Ku,Kitakyushu-Shi, Fukuoka, 807-8585, Japan.
phone : 81-93-693-3251; Fax : 81-93-693-1553
kmori@kyukyo-u.ac.jp

2. Related study

Many articles are reported that use hand gestures of a user for some operations in the field such as virtual reality or computer vision. In artificial reality, glove-based gestural interfaces have been developed. Among such devices, DataGlove of VPL Research company [6] is famous. From a view point of natural interaction, such glove-shaped devices, have a limit in free action because of the connection cable or even the trouble to wear and take it off. Hence, a method employing image processing has been studied as a non-contact method. In one of those methods, there is a method which adds markers to a hand, but it is undesirable to add something to a hand. The present study intends to detect hand movement without adding anything to a hand. The system with a sensor as the screen surface [7,8] is suggested. When we detect a hand, preprocessing is necessary. All methods have both advantages and disadvantages, and it is necessary to choose a suitable method depending on usage. In this study, we are based on real application and we will introduce a suitable technique to realize natural interaction by hand gesture. Generally, the hand depth from a camera is detected with stereo cameras [9]. In this study, we propose a method to detect parameters to operate a 3-D object based on the image from a single viewpoint. In addition, we introduce a method to detect a hand or hands stably, even if the background changes during the hand area detection.

3. Method

The proposed technique is realized simply by a PC and a USB camera. Figure 1 shows the system setup where a USB camera is placed in front of the display of a PC. The camera detects hand gestures of a user who sits facing at the display. The camera setup is simple like the one used in web chat.

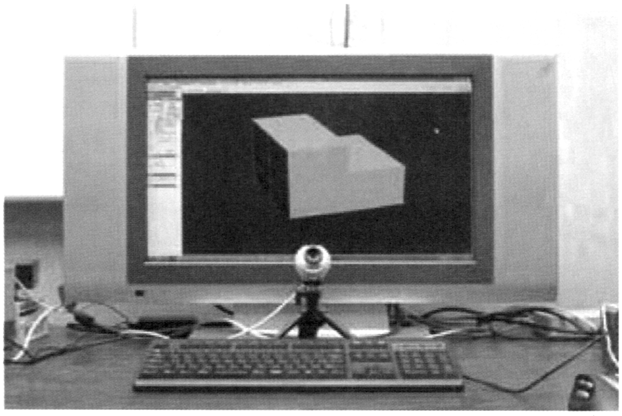


Figure 1: System setup: A USB camera is placed in front of a PC display.

The image range of the camera is adjusted to the area which includes the face and the both hands of the user. Even if a screen is not fixed, the hand can be detected stably. The parameter of the camera is dynamically adjusted even if the camera and the background move. It is assumed that the system always detects the hand area.

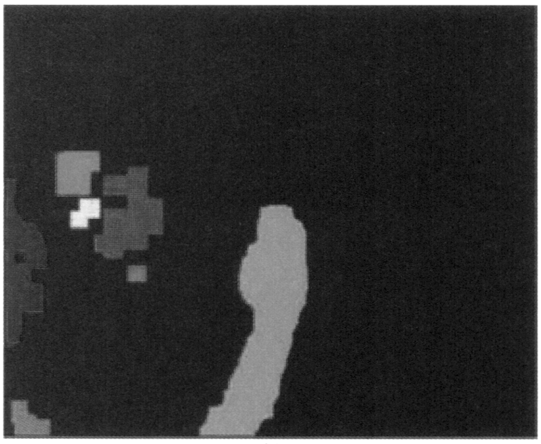
Figure 2(a) is an example of a camera image. The area of the hand is extracted from this image. To do the installation and the operation easily, the numerical value of a prior parameter is detected while taking an image. The background image is detected real-time and the area with movement is obtained by differencing the background image from the original image. The movement image is separated into small areas of 8×8 pixels. The vector representing the small area of a present input image is denoted by Fn. Then the background image denoted by BFn is given by the following formula;

$$BF_n = (1 - w)BF_{n-1} + wF_n \quad (1)$$

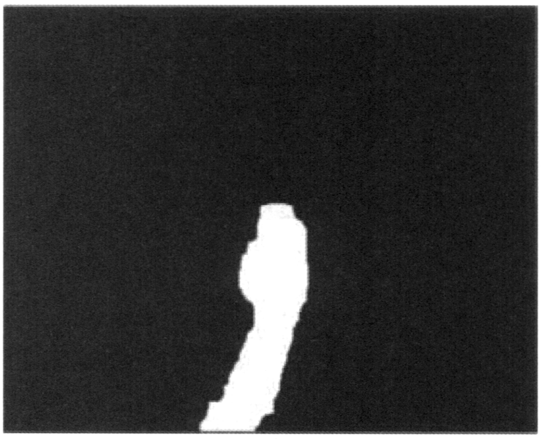
Here w is a weight. The area without movement is calculated by this equation. The weight is assumed to be 0 for the area with movement. The hand area is detected using this background together with the difference picture of a consecutive image.

The area of the hand is detected by using color information and a background subtraction algorithm. For expressing color information, the HSL (Hue, Saturation, Lightness) representation is employed. The parameter used to detect the hand is dynamically extracted from the input image. The palm can be derived under the following conditions.

(a)



(b)

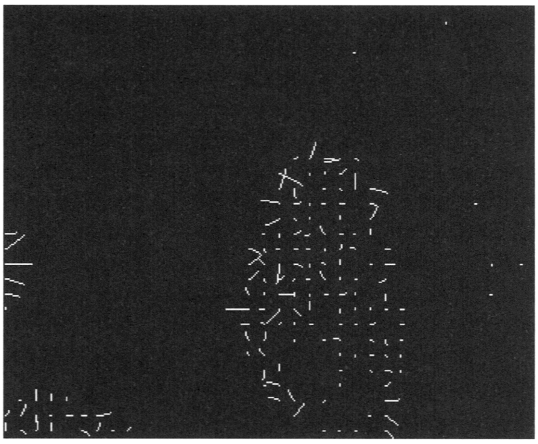


(c)

Figure 2: Area of a hand detected from movement area and color information: (a) Original image, (b) detected image, (c) hand area.

(a)

(b)



(c)

Figure 3: Movement area detected by optical flow processing: (a) Present image, (b) the latest image before the present, (c) optical flow image.

- The hand area occupies a large area in the image.
- The fingers point upwards.

Figure 2(b) shows the hand area detected from the movement area and the color information. The areas other than the hand area are also detected as shown in the figure. The palm can be detected under the above conditions. Finally, we obtain the hand area as shown in Figure 2(c).

The palm is obtained from the detected hand area, and is approximated as an ellipse. The center of gravity $G(x_{sn}, y_{sn})$ of the ellipse $E(a_{sn}, b_{sn})$ is then detected by moment calculation. Here a_{sn} is the length of the longer axis, whereas b_{sn} is the length of the shorter axis. The position $G(x_{sn}, y_{sn})$ is employed as an operation parameter. Area A_{sn} of the ellipse E is also calculated.

Optical flow is as well employed between successive image frames for smooth movement and for excluding unnecessary motion such as tremble. The optical flow is calculated using a gradient-based method. Figure 3(a) and (b) show original successive images. Figure 3(c) shows the movement vectors detected by the optical flow processing. The hand movements have been detected in the image.

Two operation methods are examined from the detected hand area. The first is a method that decides a standard position of the hand like a joystick of a computer (See Figure 4). It is employed for detecting the displacement of the hand from the standard position. Such operation looks like a usual joystick. Mouse click is realized by the operation that opens and closes the palm (See Figure 5). This shape change is detected by calculating the roundness (or the complexity) of the hand silhouette. The roundness R ($0 \leqq R \leqq 1$) of a region is defined by the ratio of the area A of the region to its squared peripheral length L, i.e.,

$$R=4\pi A/L^2 . \tag{2}$$

The movement toward the depth direction can be extracted by the change of the area of the hand. Obviously, the area is larger in a nearer view, whereas it is smaller in a farther view (See Figure 6). The allocation of the movement toward the depth orientation is normally different according to application. Ordinary 3-D software often allocates the movement to the rotation of a mouse wheel. The proposed system also allocates the movement toward the depth orientation in a screen to the change in the area of a hand. The detected hand area is approximated in the form of an ellipse. Since the ellipse is enlarged or reduced its magnitude according to the movement of the hand toward the depth orientation, we extract the movement by evaluating the magnitude change.

As well as one hand operation, the operation of both hands is taken into account, since two hands operation is sometimes easier as well as more natural than one hand operation. The distance between both hands is defined as a parameter of expansion and reduction (See Figure 7). The operation begins by opening and closing both hands, simulating mouse click (See Figure 8). The rotational operation with both hands is defined as well as real natural hands operation. The operation is moving one hand forward and at the same time moving the other hand backward.

We define a virtual camera equivalent to a viewpoint to operate a virtual 3-D object. The position of the virtual camera is expressed by polar coordinates expression. The polar coordinates vector of the virtual camera is $VC_n(r, \theta, \phi)$. Here r is the distance from the origin to a given point VC_n, θ is the angle the positive z-axis and the line connecting the origin and the point VC_n makes; and ϕ is the angle the positive x-axis and the above line makes on the xy-plane, i.e., the rotation of the line around the z-axis. In practice, in the case of one hand, the change of the area A_{sn} is simply employed in place of calculating the change of the distance r. In the same way, the change in the center of gravity position G is used in place of obtaining the change in the angles of viewpoint θ, ϕ.

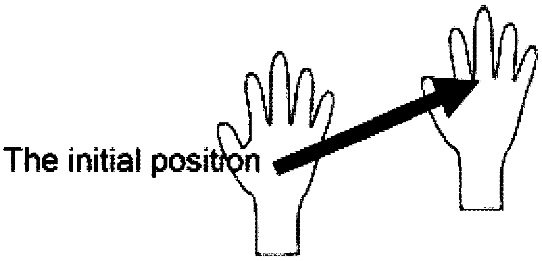


Figure 4: A method that decides a standard position of the hand like a joystick.

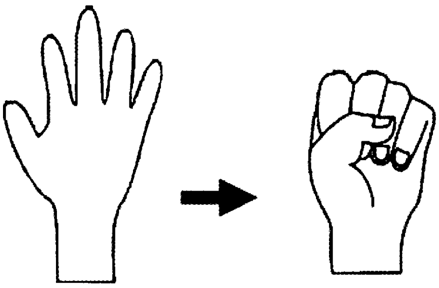


Figure 5: Mouse click is realized by the operation that opens and clothes the palm.

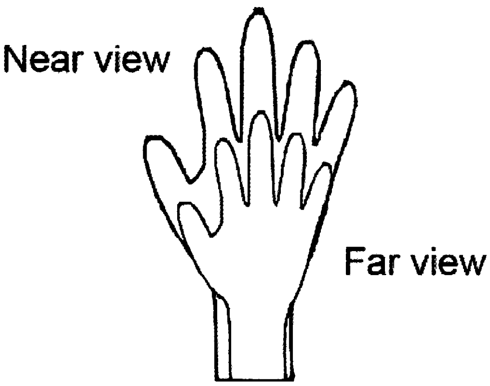


Figure 6: The movement toward the depth direction.

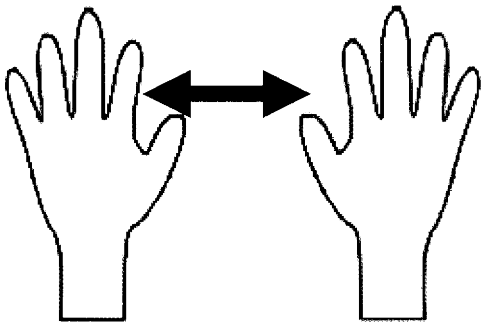


Figure 7: The distance between of both hands is detected for scaling.

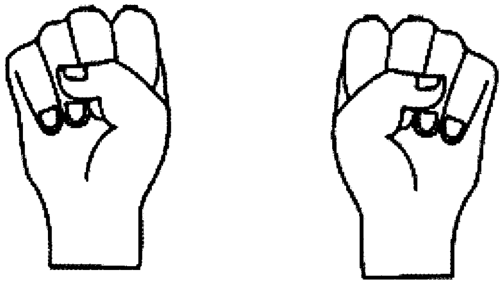


Figure 8: Mouse click is represented by the operation that opens and closes the both palms.

When a hand motion is detected from the optical flow between successive images, because of poor precision of the optical flow image, the hand operation is not performed smoothly. To achieve smooth movement, the vectors having large different values are excluded from the detected optical flow image and the average movement vector is calculated. Furthermore, a median value of the detected average movement vectors among consecutive N frames is used to maintain the continuousness of the hand operation. Smooth hand movement and therefore smooth hand operation is realized in this way.

4. Experiment

In order to show performance of the proposed technique, we performed an experiment in which Google Earth in a PC display was operated remotely by a user's left hand. The Google Earth retrieves various kinds of information including maps, satellite- and aero-photographs, etc., and allocates them on a virtual ball like a world globe [10]. This Google Earth is going to be rotated and scaled by a hand operation of a user. The allocation of the movement of the detected hand to the application software goes as follows. The upper, lower, right and left movements are allocated to the operation that turns the displayed globe in the same direction. The operations that move the hand backward and forward are allocated as those which bring the globe closer or further, respectively. In this way, the selected hand operations well imitate human actual movements.

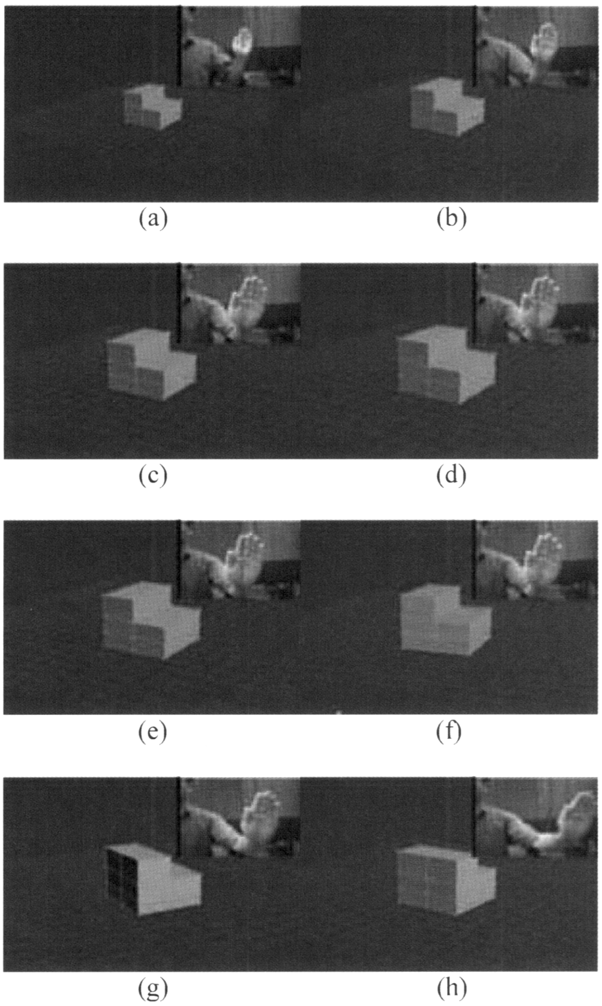


Figure 9 : The experiment using the software Metasequoia in a virtual 3-D environment. (a) to (d) shows enlargement, and (e) to (h) shows rotation.

The second experiment was done on an object created using a 3-D computer graphics software called Metasequoia [11] (See Figure 9). The objective of this experiment is to realize remote 3-D operation of an object in a display by a user's hand. In the movement of the detected hand, upper, lower, right and left movements correspond to usual mouse operation, and the movement of the back and forth is allocated to the movement of mouse's wheel. Thus, by using the proposed system, an object in a display is put right or left, moved upper or lower, and placed closer or distant by a user's hand operation performed in front of the display, as if we grasp the object directly and bring it to an arbitrary position.

The operation experiment by one hand is shown in Figure 9. The center of the figure is the operated 3-D object. The upper right of the figure is the original video image of the hand operation.

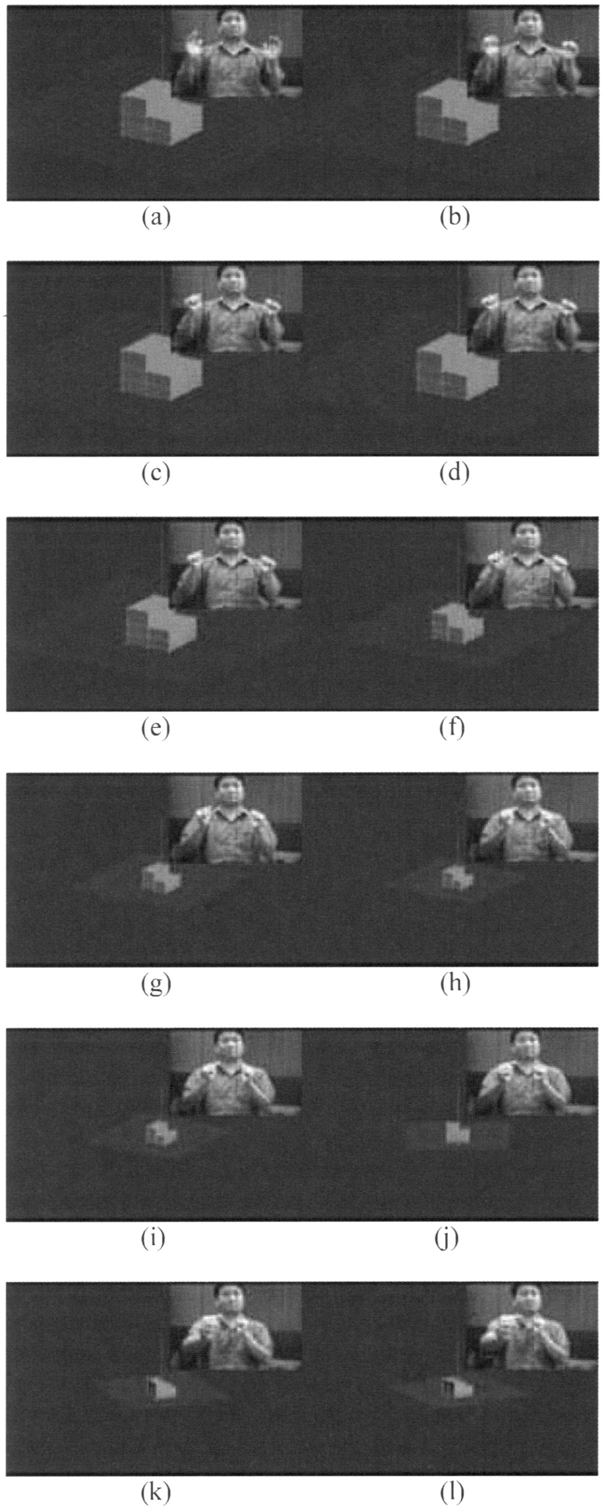


Figure 10 : The experiment using both hands. (a) to (d) shows start operation; (e) to (h) shows reduction operation; and (i) to (l) shows rotation operation.

The experiment first enlarged the object, and rotated the object afterwards. Figure (a) to (d) shows the operation of enlargement, whereas Figure (e) to (h) shows the operation of rotation. The 3-D object can be operated according to the movement of the hand as shown in the figure. It is possible to operate it more intuitively and easily than the operation that employs the mouse.

Finally, Figure 10 shows the operation experiment that uses both hands. In this experiment, the object is reduced, and it is rotated afterwards. Figure (a) to (d) shows start operation. Then Figure (e) to (h) shows the operation of reduction. Afterwards, Figure (i) to (l) shows the operation of rotation.

5. Discussion and Conclusions

A novel 3-D man-machine interface has been proposed employing a USB camera and a PC. One- and two-hand operations were defined for giving geometric transformation to a virtual 3-D object in a display. The result of the manipulation was displayed in the screen. Experiments were performed using Google Earth and a 3-D virtual object, and the proposed system successfully manipulated its rotation and scaling by some hand gestures of the user.

One can use a cursor, a mouse or a track-ball, for example, in order to specify some operations to the object interested in the virtual 3-D space in a display. It is then necessary to make correspondence between the actual operation and the operation of these devices. This interpretation is not necessary in the proposed system as similar and therefore natural hand motions are employed irrespective of in the real world or in the virtual world. This therefore realizes a human-friendly interface between computer and a human. In a practical sense, however, we can choose an appropriate device among a mouse, a track ball, the proposed system, etc., in order to realize the most convenient way of accessing an object in a 3-D virtual world that computer provides. Since the proposed system defines hand motions in place of a usual pointing device, it can be employed in any software without modifying it.

It is, however, still necessary to adjust the system so that the speed and the amount of the change of hand movement may fit a user's own sense. As a next step of the study, we need to take account of more natural hand operations. The hand motions employed in this particular paper only realizes almost parallel displacement of the hand. On the other hand, we often do a repeated hand motion when we call a person or turn the world globe right or left, for example. These movements will be taken into account in the next version of the system. It is also requested to realize a man-machine interface that allows the hand motions of not only adults but also kids or elderly people.

Referrences

[1] Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., and Long, A. C. (2002): "Interacting at a distance: measuring the performance of laser pointers and other devices", Proceedings of the SIGCHI conference on Human factors in computing systems, Input Devices , pp. 33 - 40

[2] Shimada, N., Kimura, K., and Shirai, Y. (2001): "Real-time 3-D hand posture estimation based on 2-D appearance retrieval using monocular camera", Proc. Int. Workshop on RATFG-RTS, pp. 23-30.

[3] O'Hagan, R. G., Zelinsky, R., and Rougeaux, S., (2002): " Visual gesture interfaces for virtual environments", Interacting with Computers, Elsevier Science B.V., Vol. 14, Issue 3, pp. 231-250.

[4] Mori, K., et al. (2006): "A study on the operation of a virtual 3-D space using hand motions", Proc. of the 25th SICE Kyushu Branch Annual Conference, pp.123-126. (in Japanese)

[5] Mori, K., et al. (2007): "A study on the operation of a virtual 3-D space using hand motions", Proc. of the 1st ISICE 2007, pp. 285-290.

[6] Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S. and Harvill Y.(1987): " A hand gesture interface device", ACM SIGCHI Bulletin, Vol. 17, Issue SI, pp. 189-192.

[7] Rekimoto, J. (2002):" SmartSkin: an infrastructure for freehand manipulation on interactive surfaces", Proceedings of the SIGCHI conference on Human factors in computing systems, Two-Handed Interaction, pp. 113-120.

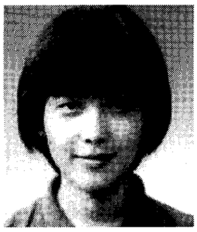
[8] Oka, K. Sato, Y., Koike, H.,(2002): " Real-time fingertip tracking and gesture recognition", IEEE Computer Graphics and Applications, Vol. 22, No. 6, pp. 64-71.

[9] Sato, S., Shibayama, E., Shin, T.,(1999): " A portable presentation system using stick-based interaction", Nippon Software Kagakukai Taikai Ronbunshu, ISSN:0913-5391, Vol.16, pp.405-408.(in Japanese)

[10]http://earth.google.com/

[11]http://www.metaseq.net/

Keisuke MORI
Keisuke MORI received B.E. degree in Electronics engineering and M.E. degree in Electrical Engineering from Kumamoto University. He is presently with Information Science, Faculty Engineering, in Kyushu Kyoritsu University as Associate Professor, and Inter-University Information Processing and Educational Research Institute in the same university as Vice manager. His current research interests include three-dimensional shape/motion recovery, human motion analysis, human sound production/synthesis, and educational computer network systems. He is a member of IEEE, The Society of Instrument and Control Engineers, The Institute of Electronics, Information and Communication Engineers, and The Acoustical Society of Japan.



Joo Kooi TAN
Joo Kooi TAN received B.E. and M.E. degrees in Computer Science and Ph.D degree in Control Engineering from Kyushu Institute of Technology. She is presently with Department of Mechanical and Control Engineering in the same university as Assistant Professor. Her current research interests include three-dimensional shape/motion recovery, human motion analysis, and human activities recognition. She received the SICE Kyushu Branch Young Author's Award in 1999, the AROB10th Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008. She is a member of IEEE, The Society of Instrument and Control Engineers, and The Information Processing Society of Japan.

Seiji ISHIKAWA
Seiji Ishikawa obtained B.E., M.E., and D.E. from The University of Tokyo, where he majored in Mathematical Engineering and Instrumentation Physics. He joined Kyushu Institute of Technology and he is currently with Department of Control & Mechanical Engineering, KIT, as Professor. Professor Ishikawa was a visiting research fellow at Sheffield University, U.K., from 1983 to 1984, and a visiting professor at Utrecht University, The Netherlands, in 1996. He was awarded BMFSA Best Paper Award in 2008. His research interests include three-dimensional shape/motion recovery, human motion analysis, and medical image analysis. He is a member of IEEE, The Society of Instrument and Control Engineers, The Institute of Electronics, Information and Communication Engineers, and The Institute of Image Electronics Engineers of Japan.