3D Measurement by Distributed Camera System for Constructing an Intelligent Room

Kota Irie, Masaki Wada, and Kazunori Umeda
Dept. Precision Mechanics, Faculty of Science and Engineering
Chuo Univ. / CREST, JST
Tokyo, Japan
Email: {kotairie,wada}@sensor.mech.chuo-u.ac.jp, umeda@mech.chuo-u.ac.jp

***Abstract*— We are constructing an intelligent room in which an operator can control home appliances such as a TV set with intuitive gestures. In this paper, we discuss three dimensional (3D) measurement of the intelligent room using a distributed camera system. The room carries CCD cameras with pan, tilt, and zoom functions. An operator makes hand waving and the room detects it. 3D position of the waving hand is measured using the cameras, and then pan, tilt and zooming-up of the cameras are carried out with the 3D information. A hand region is extracted using the color information, and the operator's gestures are recognized. 3D Finger pointing to select an appliance is measured using the cameras. Experiments verify the effectiveness of the proposed 3D measurement methods.**

***Keywords*— distributed camera system; stereo vision; intelligent room; gesture recognition; image processing**

I. Introduction

These days, computerization of our living environment is progressing, and the home appliances are becoming more intelligent and function-rich. On the other hand, the increase of their functions makes their operation complicated. For such appliances frequently used in everyday life, intuitive operation is desirable for users. Gestures, which we use frequently and intuitively in our everyday communication, can be one of such human-machine interfaces. Until now, many studies which recognize gestures from a sequence of images have been reported [1]. A shortage of many of the gesture recognition studies is that the place where gestures can be recognized is limited and thus they are not suitable for practical applications. There is a trend of studies to make a room or some space itself intelligent [2], [3], [4], [5], [6]. This approach is effective for realizing natural human-machine interface. Gesture recognition is one of the important technologies for such an intelligent room or space. We are also constructing an intelligent room in which an operator can control home appliances such as a TV set with intuitive gestures [7]. Characteristics of our intelligent room are that, it focuses on operation of home appliances, it allows users to operate without restriction to the place and without having any instrument, and it uses pan-tilt-zoom cameras only as sensor information.

In this paper, we discuss three dimensional (3D) measurement of the intelligent room using a distributed camera system, especially 3D measurement of hand waving and finger pointing.

II. Outline of Intelligent Room to Operate Home Appliances

The intelligent room in this paper has some intelligent functions that are intended for a general house or an office. In the room, we can operate home appliances such as a television set and a lighting by gestures. The operator (i.e., a person with an intention to operate an appliance) does not need a special attachment such as a glove or a microphone and can make operations in a natural state. The room specifies the operator autonomously even when two or more persons exist. The operator can make operations wherever in the room without any restriction. Gestures in the room are supposed to be by a hand or fingers.

The room carries CCD cameras with pan, tilt, and zoom functions, and detects an operator and recognizes his/her gestures autonomously with them. Fig.1 shows the conceptual figure of the intelligent room. First, the system performs "detection of waving hands" with two or more cameras to detect the operator and then acquires 3D position information. It carries out pan, tilt and zooming-up of the cameras with the 3D information and restricts the region to process. Then it extracts a hand region using the color information. The skin color of the operator is registered in this stage, which improves the robustness of extracting the hand region to the difference of individual skin colors and the change of lighting environment. Then gestures are recognized for the hand region that is extracted using the color information. The recognized result is presented on a PC monitor and by a speaker, and interaction by the operator is made possible. Based on the recognized operation, a control signal is transmitted to the target appliance by the infrared remote controller connected to PC. For example, turning on/off the TV set, inputting the channel and turning up/down the volume using gestures are possible.

III. Field of View of Pan-Tilt-Zoom Camera

The camera's field of view is an important parameter to use in the intelligent room. When watching the wide space to find an operator, it should be wide. At the same time, too wide field of view is not appropriate for detecting a waving hand. And when focusing on the detected operator to recognize his/her gestures, the field of view should be narrow

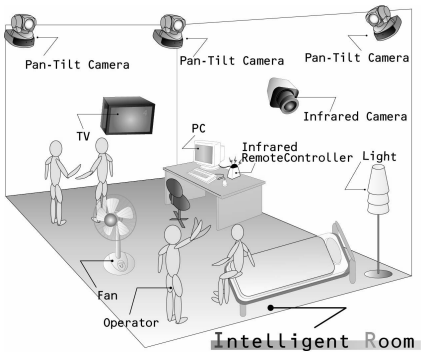


Fig. 1. The conceptual figure of our intelligent room

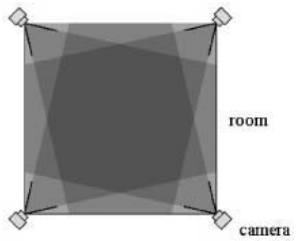


Fig. 2. Covering of a room by the field of view of four cameras

(a) widest field of view (b) narrowest field of view

Fig. 3. Range of field of view of a pan-tilt-zoom camera EVI-D100 at 5[m]

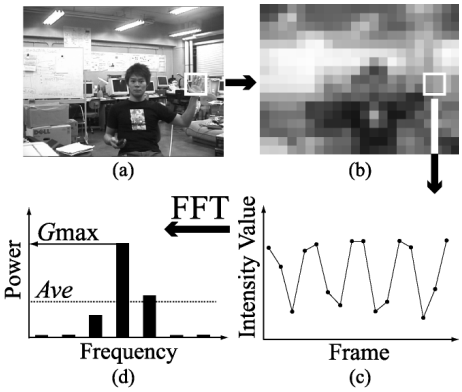


Fig. 4. Detection of hand waving using FFT

enough. These days several kinds of pan-tilt-zoom cameras are commercially available. Pan, tilt and zoom of these cameras can be controlled by a PC, and these cameras are suitable for usage in the intelligent room. For example, Sony EVI-D100, which we adopted in our intelligent room, has 10X optical zoom and its horizontal field of view varies from 65 to 6.6[deg]. Other pan-tilt-zoom cameras have similar zoom specifications. Suppose four cameras are set at the corners of a square room and the field of view is 65[deg]. Then the room is covered by the cameras' field of views as shown in Fig.2. It can be seen that the room is mostly (95.8[%]) covered by two or more cameras, which means that 3D measurement is possible almost everywhere in the room.

And as for detecting a waving hand, suppose the horizontal number of pixels of the low-resolution image is 25, the width of hand waving is 0.3[m], and the observed hand waving should be equal to or larger than 1[pixel] of the low-resolution image. Then the distance to the waving hand can be up to 5.9[m], which is long enough in an ordinary room. And when the field of view is set to the narrowest 6.6[deg], the corresponding size at the distance 5.9[m] becomes as small as 0.58[m], which means that a hand or other regions can be observed large enough.

Fig.3 shows an example. The two images are captured with the widest and narrowest field of view at 5[m] respectively. We can see this kind of camera is suitable for both watching the whole scene and recognizing each person's gesture.

IV. 3D Measurement of Waving Hands

The 3D position of waving hands is measured by detecting it with two or more cameras.

A. Outline of Detecting Waving Hands

We use hand waving to detect an operator in the intelligent room. For human-human interface to communicate one's intention to other person, hand waving is often used and thus it is thought to be appropriate to use in the intelligent room, too. The outline of the method to detect waving hands [8] is as follows (see Fig.4). The images are converted to low-resolution ones, and FFT is applied to each pixel of the low-resolution images. Pixels with high power at the frequencies of hand waving are detected as the pixels of hand waving. The method is robust to lighting condition and individual difference of skin color, because it uses intensities only.

B. Calibration of the Cameras

The intrinsic and extrinsic parameters of the cameras are obtained by the following calibration method that uses a known pattern [9].

(1) Obtain the projection matrix between the 3D points and their projected points on a 2D image.

(2) Obtain the intrinsic and extrinsic parameters from the projection matrix. For the point on the image plane $\tilde{\mathbf{m}} = [u, v, 1]^T$ and the point in 3D space $\tilde{\mathbf{M}} = [X, Y, Z, 1]^T$, the projection equation for the perspective projection is

$$s\tilde{\mathbf{m}} = \mathbf{P}\tilde{\mathbf{M}} = \mathbf{A}[\mathbf{R}, \mathbf{t}]\tilde{\mathbf{M}} \tag{1}$$

where

$$\mathbf{A} = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u_0 \\ 0 & \alpha_v / \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \mathbf{r}_1{}^T \\ \mathbf{r}_2{}^T \\ \mathbf{r}_3{}^T \end{bmatrix}, \mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{2}$$

When $\mathbf{P}$ is obtained, the matrix $\mathbf{A}$ that consists of intrinsic parameters, and the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ that are extrinsic parameters, are calculated.

C. Stereo Vision

3D coordinates of the matched point are obtained from the projection matrices of the two cameras. When a 3D point is observed by camera i at $\mathbf{m}_i = [u_i, v_i]^T$! $, $$s_i \tilde{\mathbf{m}}_i = \mathbf{P}_i \tilde{\mathbf{M}} \tag{3}$$ is satisfied. Then for 2 or more cameras, $$\mathbf{BM} = \mathbf{b} \tag{4}$$ is introduced. The 3D coordinates $\mathbf{M}$ of the point is given by least squares method as $$\mathbf{M} = \mathbf{B}^{+}\mathbf{b}. \tag{5}$$

V. 3D Measurement of Finger Pointing

We often point at an object with a finger to select it. Therefore, we use finger pointing in the intelligent room to select an appliance. When it is observed by two cameras, 3D finger pointing can be obtained. Then the object that is closest to the pointing is selected and is operated.

The 3D pointing is obtained as follows. When a finger, hand or arm region is extracted in each image by using skin color extraction, the 2D pointing can be obtained as the principal axis of the extracted region as shown in Fig.5. Here we use the extracted skin color region without distinguishing whether it is a finger, hand or arm. Furthermore, when a face region is also extracted, utilizing the vector from the face region to the hand region as the pointing is possible as shown in Fig.5(b). The principal axis and the origin of the camera makes a plane in 3D space. When such a plane is obtained by two cameras, the 3D pointing can be obtained as the intersection of the two planes as shown in Fig.6. The normal vector of the plane in Fig.5 becomes $$\mathbf{N} = [f \tan\theta, -f, b]^T \tag{6}$$ where f is the focal length of the camera, θ represents the direction in 2D, and b is the v-intercept. By multiplying $\mathbf{R^T}$, the transpose of the rotation matrix $\mathbf{R}$ in (2), the normal vector in the world coordinate system is obtained. Then the 3D pointing direction is obtained as the outer product of the two normal vectors.

Yamamoto et al. [10] realizes selection of appliances by arm pointing using multiple stereo cameras. Our method does not use stereo cameras but multiple monocular cameras to realize the 3D measurement of finger pointing.

VI. Experiments

The experiments were performed with the constructed intelligent room illustrated in Fig.1. Fig.7 shows the overview of the experiments. MVTec Halcon is used for image processing and other calculations and controls are performed by a DELL PC (Pentium 4 2.2GHz). Three SONY EVI-D100 CCD cameras were set at the corners of a $6.9 \times 7.8[m^2]$ room, at $2.3[m]$ height. The three images from the three cameras are composed by a Panasonic WJ-MS488 picture division unit and inputted into the PC with a Leutron PicPort Color image capture board (640×480[pixel]). A Sugiyama Electron Crossam 2+USB infrared remote-controller is used that can be controlled by the PC to operate home appliances.

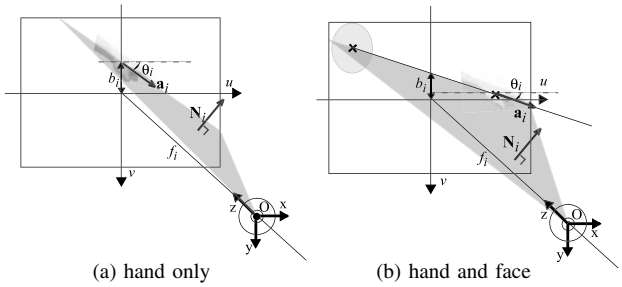


Fig. 5. Measurement of 2D pointing

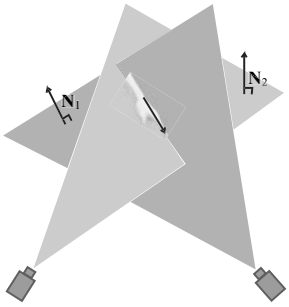


Fig. 6. Measurement of 3D pointing

A. Measurement of 3D Position of Hand Waving

This experiment was carried out with the two cameras of the intelligent room. The zoom was set to the widest.

Fig.8 shows the results of error analysis. The hand waving at every 0.5[m] positions for x and y directions was detected. Hand waving was carried out by one subject 5 times for each position with changing the height. The error bars are the standard deviations. The results show that the errors are about 0.1[m], which is small enough for obtaining the 3D

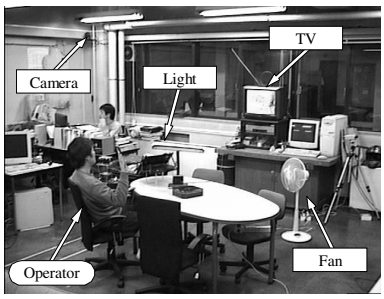


Fig. 7. Overview of experiments for operating home appliances

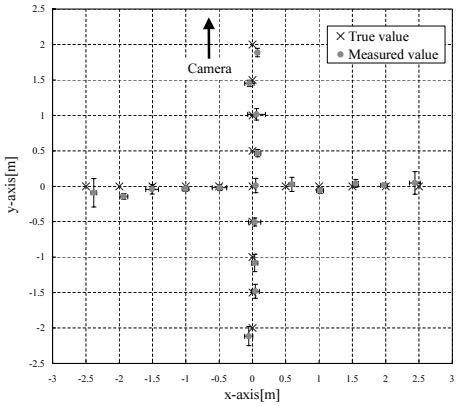


Fig. 8. Measurement of position of hand waving

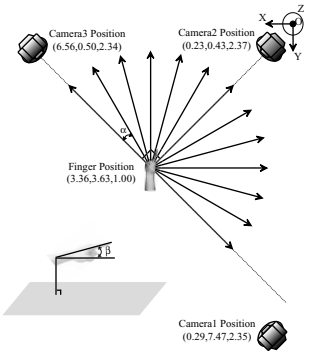


Fig. 9. Experimental condition of measuring pointing direction

position of the operator in the room.

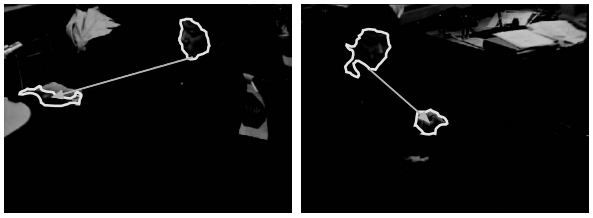
B. Measurement of Finger Pointing

The experiments were performed as shown in Fig.9. The angle between the camera 3 and the pointing direction α was changed every 15[deg] from 0 to 180[deg]. The angle of elevation of the pointing direction β was fixed to 0[deg], i.e. parallel to the floor. Fig.10 shows an example of detected pointing in the images. As can be seen, the vector from the face region to the hand region was used as the pointing. Fig.11 show the results of the accuracy of detected pointing direction. The best result in the results by three pairs of cameras was selected.

It is shown that the pointing direction was roughly acquired. The reason β has some bias is utilizing the vector from the face region to the hand region as the pointing. However, this was more robust than using the principal axis of the extracted hand region.

VII. Conclusion

We have discussed three dimensional (3D) measurement of the intelligent room using cameras with pan, tilt and zoom functions. Concretely, 3D measurement of position of operator's waving hand and finger pointing to select an appliance were proposed. Additionally, we showed that the



(a) Camera 3 (b) Camera 2

Fig. 10. Example of detected pointing using face and hand regions

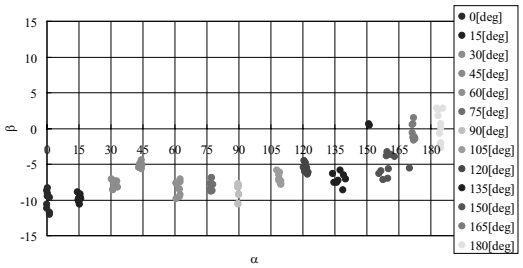


Fig. 11. Experimental results of measuring pointing direction

pan-tilt-zoom camera is appropriate to use in the intelligent room. Experiments verified the effectiveness of the proposed 3D measurement methods.

Future works include more experimental evaluation, and improvement of the intelligent room.

Acknowledgment

We thank Mr. Naohiro Wakamura, who was a master's student at Chuo University, for his support to this study.

References

[1] V. I. Pavlovic, R. Sharma, T. S. Huang: "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", Trans. PAMI, vol.19, no.7, pp.677-695, 1997.

[2] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.1, pp.107-118, 2000.

[3] M. Coen, "The Future Of Human-Computer Interaction or How I learned to stop worrying and love my Intelligent Room", IEEE Intelligent Systems, vol.14, no.2, pp.8-19, 1999.

[4] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer, "EasyLiving: Technologies for Intelligent Environments," Proc. Int. Symposium on Handheld and Ubiquitous Computing, pp.12-27, 2000.

[5] T. MORI and T. SATO, "Robotic Room: Its concept and Realization," Robotics and Autonomous Systems, vol.28, no.2!$pp.141-144!$1999.

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

[7] K. Irie, N. Wakamura, and K. Umeda, "Construction of an Intelligent Room Based on Gesture Recognition -Operation of Electric Appliances with Hand Gestures-," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.193-198, 2004.

[8] K. Irie and K. Umeda, "Detection of Waving Hands from Images Using Time Series of Intensity Values," Proc. 3rd China-Japan Symposium on Mechatronics, pp.79-83, 2002.

[9] O. Faugeras, Three-dimensional computer vision: a geometric viewpoint, MIT Press, 1993.

[10] Y. Yamamoto, Y. Yoda, and I. Sakaue, "Arm-Pointing Gesture Interface Using Surrounded Stereo Cameras System," Proc. Int. Conf. on Pattern Recognition (ICPR 2004), vol.4, pp.965-970, 2004.