Improvement of Key Functions of the Intelligent Room

Takeshi Nagayasu, Hidetsugu Asano, Masahito Takahashi, Kenji Terabayashi, and Kazunori Umeda

Abstract— **In this paper, key functions of the intelligent room that we formerly proposed have been improved, and a new intelligent room has been constructed. An intelligent room is defined as one in which the operation of appliances is conducted by gesture without any other equipment or restrictions on location. Three key functions, i.e., hand detection, skin color registration, and recognition of the number of fingers, have been improved. To detect hand motion, the resolution and sensitivity were improved along with the insignificant finger motion. The detection of color registration was improved with simultaneous processing along with the detection of finger motion. And for "recognition of number of fingers", a new method that uses active contour models (Snakes) is introduced to identify the number of fingers used in gesturing, which increases the robustness of the process.**

Index Terms— **Intelligent Room, Gesture Recognition, Image Processing**

I. Introduction

Currently, living environments are becoming more electronically sophisticated and networked. On the other hand, the sophistication in function complicates the operations of the devices. Many apparatus around us are operated using a button or a remote control. However, disadvantages of a remote control are that they must be accessible and a button must be pushed by the operator at specific location. It frequently takes much time to locate a remote control. The operation of home appliances is enhanced by simplicity and the absence of restrictions on the location of the operator. Intuitive gestures that do not require additional equipment are a possible solution to these challenges. Until now, many studies on gesture recognition using images have been published [1][2]. Some commercial products using gestures have also been released [3]. Moreover, intelligent rooms that take advantage of gestures have been studied previously [4][5]. Irie et al. constructed an intelligent room in which a person can operate home appliances without any additional equipment or restrictions on location using gesture recognition techniques starting from the detection of hand waving [6][7].

T. Nagayasu and M. Takahashi are with the Precision Engineering, Chuo Univ. / CREST, JST, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan (e-mail: {nagayasu, takaha}@sensor.mech.chuo-u.ac.jp).

H. Asano is with the R&D Division, Pioneer, 1-1 Shin-ogura, Saiwai-ku, Kawasaki-shi, Kanagawa 212-0031, Japan (e-mail: hidetsugu_asano@post.pioneer.co.jp).

K. Terabayashi and K. Umeda are with the Department of Precision Mechanics, Chuo Univ. / CREST, JST (e-mail: {terabayashi, umeda}@mech.chuo-u.ac.jp).

Improvements in the key functions for the operation of devices in an intelligent room [6][7], i.e., detection of hand motion, skin color registration, and recognition of the number of fingers, are reported in this paper.

II. Outline of Intelligent Room

For the purpose of this paper, an intelligent room is defined as one with cameras equipped to recognize human gestures. The room is intended to use as an office or living room.

In this study, appliances, such as television sets, are operated by gestures. Functions, such as the detection of finger motion, skin color registration, and gesture recognition, are used in an intelligent room in this study. A finger or hand motion can be readily identified as that of the operator even when multiple individuals are in the room. Moreover, the function is robust to change of the color according to the difference in the color of individuals, and change of lighting environment by registering the skin color at the beginning of each operation. Furthermore, since the appliances are operated by gesture, no physical contact is necessary, and intuitive operation is possible. Gesture recognition includes the identification of the number of fingers and hand motion. The system requires the use of a pan-tilt-zoom camera, a personal computer (PC), and an infrared remote control. The camera and PC are connected on the network, and the remote control is connected to the network. Therefore, an operator would be able to use the system by preparing the camera and remote control.

III. Finger waving detection for Selection of Operator

For selection of the operator, finger waving detection is performed within an image. The Fourier transform is applied to each pixel of a low-resolution image. If a periodic motion is detected at a pixel, the pixel is voted. When the number of votes reaches a threshold value, the pixel is specified as a candidate pixel of finger waving. Two or more cameras perform the above processing, and the candidate pixel that satisfies the epipolar constraint is specified as a finger waving pixel [7]. Moreover, the 3-dimensional position is measured simultaneously.

Formerly, whole hand motion was required for detection; however, now, a slight motion of a finger can be detected even when the finger is 5m or more away from a camera. Moreover, in the new system, the accuracy of three-dimensional position measurement is also improved because the spatial resolution

of detection increased. Furthermore, position adjustment after zooming that was necessary in the former system can be omitted in the new system, which reduces the processing time of the system.

A. Fourier transform to each pixel

When performing finger waving, the finger and the background are seen alternately. Therefore, the intensity of corresponding region changes periodically. The method utilizes this effect. First, the color images are converted to low-resolution gray images to reduce the noise and the calculation cost.

Then, the Fourier transform is applied to the time series of the intensity values of each pixel of a low-resolution image. A hamming window is used to reduce the noise, such as influence of minute variations.

B. Finger waving detection

Finger waving detection is achieved by detection of specific periodic motion to each pixel based on the result of the Fourier transform. These processes are performed by two or more cameras, and the pixel is specified as a finger waving pixel if the pixel satisfies the epipolar constraint.

An example of finger motion detection is shown in Fig.1. The distance from the camera is approximately 6m. (a) is an original image, in which a person is moving his finger in the position inside of the white circle. (b) is a low-resolution image, and the pixel where finger motion is detected is displayed as a bright pixel within the white circle.

(a) Input image (b) Low-resolution image

Fig.1 Finger waving detection

C. Three-dimensional measurement with two cameras

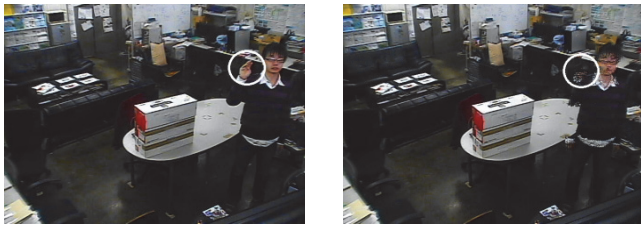
Triangulation is applied when two or more cameras are used to detect finger motion. The information of position is used to calculate the distance from the camera and to adjust the zooming [8].

IV. Skin color registration

Skin color registration is performed using the pixel of finger waving region [9]. In a previous study, skin color registration required several seconds to focus on the operator's palm. However, the modifications introduced here permit a smooth operation with the simultaneous capture of finger waving and skin color registration.

A. Background image registration

The pixels that are changeless for several frames are registered as a background image and sequentially updated to separate the region of finger waving and other regions. The background image is updated successively. Because change occurring in a short time period is assessed, still objects, such as furniture, walls, and the body of a relatively motionless operator, are registered as background. On the other hand, the region of finger waving is not registered as background because it is moving at high speed. This process is illustrated in Fig.2. (a) is an input image, and (b) is the obtained background image. The region of finger waving is shown with a white circle. Background objects, as well as the operator's body and arms, are registered as background, but the region of finger waving is not registered.



(a) Input image (b) Registered background image

Fig.2 Background image registration

B. Background subtraction

Image values averaged for several frames within the region of finger waving, and difference values between background and captured image are calculated. The difference value is large when the finger is in the frame. So we can select frames including finger by difference value. In selected frames, pixels of finger region are selected by difference values of background and captured pixels. And color data of selected pixels are registered as skin color.

C. Skin color registration

V (Value) tends to be influenced by lighting conditions. Therefore, we use H (Hue) and S (Saturation) values, which are more robust to lighting conditions, for skin color registration and extraction. Mahalanobis distance of the registration data and input data is calculated, and the pixel is extracted as a skin color when the distance is under a threshold.

V. Gesture recognition

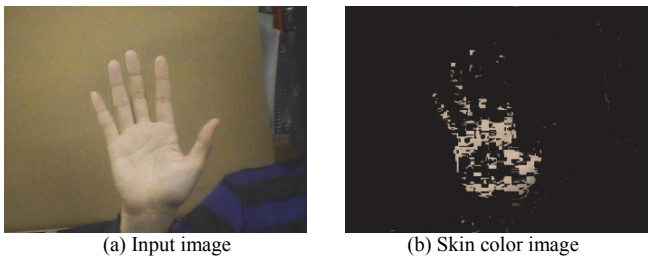
A. Finger number recognition

A hand region is extracted from skin color data, and the number of fingers is estimated based on the area of hand region. In the previous study [6][7], only finger region is extracted based on morphology process, and number of fingers are estimated from counting regions. Therefore, the hand region had to be extracted correctly. However, correct extraction of hand region other than outline is difficult in a real environment. The new method estimates number of fingers from area of hand region by using Snakes. Therefore, robust extraction is possible in a real environment.

1) Extraction of a skin color region

A hand region is extracted based on the registered skin color information. As shown in Fig.3, when color of background and hand are similar, only hand region is

extracted if we set the threshold small. The extraction is incorrect but the outline of hand region can be extracted.

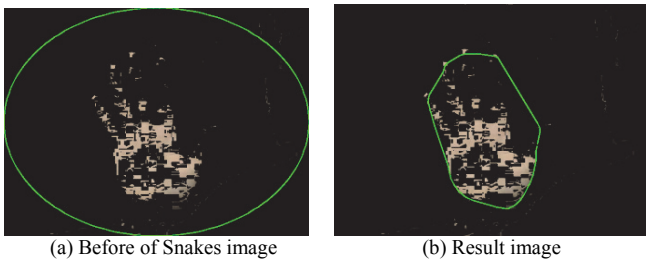


(a) Input image (b) Skin color image

Fig.3 Extracted skin color region

2) Measurement of the hand region by Snakes

The area of hand region changes according to the number of fingers. Therefore, we can estimate the number of fingers from the area of hand region. We calculate the area enclosed by Snakes. Snakes is a method for obtaining the closed curve near the shape of the object as a energy minimization problem. By giving an initial closed curve and a parameter suitably, the given initial closed curve is deformed sequentially, and we can obtain the outline near an arbitrary form. The condition in which the initial outline is given is shown in Fig.4(a). Fig.4(b) is the outline curve obtained by Snakes



(a) Before of Snakes image (b) Result image

Fig.4 Recognition of number of fingers by Snakes

3) Estimation of the number of finger

The area of hand region is given as model data of each number of fingers. The area of input data is compared with model data to estimate number of fingers. Average $\mu_i$and standard deviation $\sigma_i$of the hand area of each finger number are obtained by model data. We estimate number of fingers from area A about input data based on following two methods.

・Mahalanobis distance

We obtain the Mahalanobis distance D_i about each number of fingers from the following equation.

$$D_i = \frac{|A-\mu_i|}{\sigma_i} \quad (i = 1,\cdots,5) \tag{8}$$

Number of fingers with the smallest distance is selected as the recognition result.

・Accumulated likelihood

We obtain the likelihood l_i to each number of fingers from the following equation.

$$l_i = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(A-\mu_i)^2}{2\sigma_i^2}\right) \quad (i = 1,\cdots,5) \tag{9}$$

This process is repeated, and the likelihood is accumulated for each finger number. When the accumulated likelihood reaches a threshold, then the corresponding finger number is selected as the recognition result.

B. Recognition of hand gesture

DP (Dynamic Programming) matching is used for recognition of a hand gesture. A hand region is extracted in each frame using the skin color extraction. The centroid of the extracted hand region is compared with the previous frame, and the direction of the movement of the hand centroid is obtained; eight directions of upper and lower, right and left, and slanting. DP matching is performed using the obtained directions as features. The similarity to each registered model is measured, and the model with the largest similarity is recognized as the input operation. An arbitrary model can be used if it can be registered. The system makes use of four instructions: Up, Down, Right, and Left.

VI. EXPERIMENT

A. Experimental system

The system uses OpenCV for image processing. A PC (Core2Quad Q9400 2.66GHz, DDR2 6GB) is used to implement each process and control the intelligent room. Cameras with a pan-tilt-zoom function AXIS 233D are used to observe the gestures. They are operated via the network by the PC. To operate home appliances, an infrared remote control KURO-RS is used. The remote control has a learning function and can be controlled by the PC.

B. Finger waving recognition

Experiments to detect finger waving were conducted using the method in Section III. The recognition rate and time for recognition were evaluated for different distances for five subjects. When finger waving is detected by two cameras at corresponding positions, we assume that finger waving is recognized. Twenty experiments were conducted at each distance. Table 1 shows the time for recognition for each distance. Ave. and S.D. represent average and standard deviation, respectively. The time is about 2 seconds regardless of the distance and subject. The recognition rate was 100% for every distance. These results demonstrate the robustness of the method.

Table 1 Time for recognition of finger waving (s)

	4m	5m	6m
Ave.	2.04	2.50	2.32
S.D.	0.12	0.28	0.18

C. Finger waving recognition

Experiments of skin color registration were conducted using the method explained in Section IV. One hundred experiments were performed, and the processing times for background registration, skin color registration, and skin color extraction were evaluated. The results are shown in Table 2. Both background registration and skin color registration are fast and do not impede real-time operation. In contrast, the extraction of skin color region required more time. However, for the pre-processing of gesture recognition, the speed is sufficient.

Table 2 Processing time for skin color registration (ms)

	Background registration	Skin color registration	Extraction of skin color region
Ave	45.6	40.1	310
S.D.	4.5	6.7	5.5

D. *Recognition of the number of fingers*

Experiments to estimate the number of fingers were conducted with five subjects using the method described in Section V-A. First, skin color registration explained in Section IV was performed with each subject. Then, the recognition rate and processing time were evaluated with one through five fingers. One hundred experiments were conducted for each of one through five fingers for each subject. The two methods described in Section V-A and the previous method using morphological processing were compared.

For training, the hand region was measured 20 times for fingers one through five from subjects. The average and standard deviation of the area were calculated and used for estimating the number of fingers. The distance to the subject was set to 6m from camera in each case at obtaining the training data and at estimating the number of fingers. The results are shown in Fig.5 and Table 3. The recognition rate is shown in Fig.5, and the processing time, in Table 3.

For each finger one through five, the new method provides better results. Especially, the method by accumulation of likelihood gives high recognition rate. On the contrary, the results from the previous method are not good. This is because of the low quality of the extracted skin color region (see Fig.3(b)). The new method is much more robust than the previous one. For two fingers, the result of the previous method is good. However, this is not a result of correct recognition because the regions other than finger are counted. This means that the reliability of the previous method is inferior.

More processing time is required with the new method. However, the amount of time does not affect the system much. The reason that the standard deviation of the method using the accumulation of likelihood is large is that the data required for recognition changes according to the input. In this experiment, the number was 3 or less.

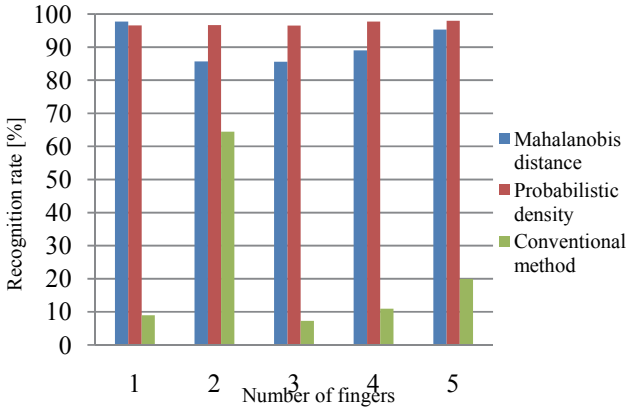


Fig.5 Recognition rate of the number of fingers

Table 3 Processing time (s)

	Mahalanobis distance	Accumulation of likelihood	Previous method
Ave	0.30	0.41	0.20
S.D.	0.02	0.23	0.02

E. *Recognition of hand gesture*

Experiments for hand-gesture recognition were conducted for five subjects using the method explained in Section V-B. Twenty experiments were conducted for each operation. The distance from the camera to the operator was the same as that for the skin color registration. First, skin color registration was performed for subjects. Four gestures were then performed, and the recognition rate and processing time were evaluated.

The average and standard deviation of the processing time were 2.5s and 0.1s, respectively. Table 4 shows the results of the recognition rate. They are high for each operation.

The recognition rates for the Left and Right operations are lower because the Right and Left operations tend to fluctuate more.

Table 4 Recognition rate of hand motion (%)

Operation	UP	Down	Left	Right
Ave	98	99	87	87
S.D.	4	2	7	5

VII. CONCLUSION

In this study, we improved key functions used for operation in an intelligent room [6][7], i.e., hand motion detection, skin tone registration, and recognition of the number of fingers. We then constructed an intelligent room with enhanced performance, better processing time, and more convenience for the operator. Future works include further improvement of the convenience and usability of the new intelligent room. We are now considering the introduction of lip reading and 3D gesture recognition using a range image sensor.

REFERENCES

[1] V. I. Pavlovic, R. Sharma, and 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] P. Hong, M. Turk, and T. S. Huang, “Gesture Modeling and Recognition Using Finite State Machines,” IEEE In. Conf. on Automatic Face and Gesture Recognition

[3] Microsoft, KINECT, http://www.xbox.com/en-US/

[4] T. Mori and T. Sato, "Robotic Room: Its Concept and Realization," Robotics and Autonomous Systems, vol.28, no.2, pp.141-144, 1999.

[5] J. H. Lee and H. Hashimoto, "Intelligent Space - Concept and Contents," Advanced Robotics, vol.16, no.4, pp.265-280, 2002.

[6] 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.

[7] K. Irie, M. Wada, and K. Umeda, “3D Measurement by Distributed Camera System for Constructing an Intelligent Room,” BoF paper, Fourth International Conference on Networked Sensing Systems (INSS2007), pp.118-121, Braunschweig, Germany, June 2007.

[8] Olivier Faugeras, "Three-Dimensional Computer Vision," MIT Press, 1993.

[9] K. Irie, M. Takahashi, K. Terabayashi, H. Ogishima, and K. Umeda, "Skin Color Registration Using Recognition of Waving Hands,” J. Robotics and Mechatronics, Vol.22, No.3, pp.262-272, 2010.

[10] M. Kass, A. Witkin, and D. Terzopouls, "Snakes: Active Contour Models," International Journal of Computer Vision, pp.312-331, 1988.