SICE Annual Conference 2011
September 13-18, 2011, Waseda University, Tokyo, Japan

Improvement of an Intelligent Room that Detects Hand Waving Motion for Operation of Home Appliances

Takeshi Nagayasu[1], Hidetsugu Asano[2], Kenji Terabayashi[3], and Kazunori Umeda[3]

[1]Course of Precision Engineering, Chuo Univ., Tokyo, Japan
(E-mail: nagayasu@sensor.mech.chuo-u.ac.jp)
[2] R&D Division, Pioneer, Tokyo, Japan
(E-mail: hidetsugu_asano@post.pioneer.co.jp)
[3]Department of Precision Mechanics, Chuo Univ. / CREST, JST, Tokyo, Japan
(E-mail: {terabayashi, umeda}@mech.chuo-u.ac.jp)

Abstract: In this paper, the intelligent room that was formerly proposed by our group is improved by improving the key functions of the room. An intelligent room is a room in which home appliances are operated using gestures, without any additional equipment and any restriction for positions. Three key functions: detection of hand waving, skin color registration, and recognition of number of fingers, are improved. For "detection of hand waving", resolution and sensitivity are improved, and consequently, detection of small finger waving motion is enabled. For "skin color registration", simultaneous processing with detection of finger waving is achieved. And for "recognition of number of fingers", a new method that uses Snakes is introduced and robustness of the process is greatly improved. Additionally, a system to control a TV set with only hand waving is constructed.

Keywords: Intelligent Room, Gesture Recognition, Image Processing

1. INTRODUCTION

In recent years, living environments are becoming more intelligent and networked. On the other hand, the increase of functions complicates the operations of the apparatus. 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 reported [1][2]. Some commercial products using gestures have also been released [3]. Moreover, intelligent rooms that take advantage of gestures have been studied [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].

In this study, we improve the three key functions [8] for the operation of appliances in an intelligent room [6][7], i.e., detection of hand waving, skin color registration, and recognition of the number of fingers. Additionally, we construct a system to control a television (TV) set with hand waving by applying the concept of “Spatial Memory” [9].

2. OUTLINE OF INTELLIGENT ROOM

The intelligent room in this study is a room in which cameras are equipped to recognize gestures of a person. The room is intended to use as an office or living room.

In this study, appliances, such as a TV set, are operated by gestures. Functions, such as the detection of finger waving, skin color registration, and gesture recognition, are used in an intelligent room in this study. A finger or hand waving 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 when a hand waving is observed. 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 has pan-tilt-zoom cameras, 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.

3. "FINGER WAVING DETECTION" FOR SPECIFICATION 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 three-dimensional (3D) position is measured simultaneously.

In the previous system, 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 3D 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.

3.1 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 image 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 influence of minute variations.

3.2 Finger waving detection

Finger waving detection is achieved by applying the following procedure to each pixel based on the result of the Fourier transform.

1) Search for peak intensity of amplitude within 3-6Hz.
2) Obtain difference between the peak intensity and the intensity of amplitude at 2Hz.
3) Accumulate the value in time.
4) When the accumulated value exceeds a threshold value, the pixel is detected as a candidate of finger waving pixel.
5) The above 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) Detected finger waving

Fig.1 Detection of finger waving.

3.3 3D measurement with two cameras

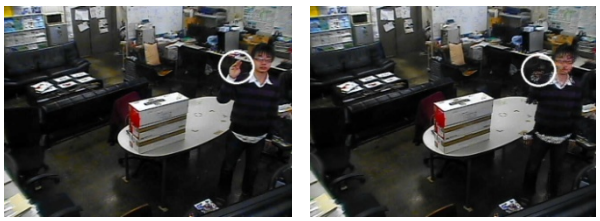
Triangulation is applied when two or more cameras are used to detect finger waving. The information of position is used to calculate the distance from the camera and to adjust the zooming [10].

4. SKIN COLOR REGISTRATION

Skin color registration is performed using the pixel of finger waving region [11]. In the previous system, 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.

4.1 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.

4.2 Background subtraction

Fingers and skin color are identified by analyzing the input and background images. First, differences in background and captured images are computed for several frames within the region of finger waving.

The sum of the absolute values of normalized hue, saturation, and intensity value is used to evaluate the difference. Suppose $I_H(0)$, $I_S(0)$ and $I_V(0)$ as the normalized hue, saturation and intensity value of the background image respectively, and $I_H(t)$, $I_S(t)$ and $I_V(t)$ as these values of the t-th captured image respectively. Then the evaluation function is given by

$$S_d(t) = |I_H(t) - I_H(0)| + |I_S(t) - I_S(0)| + |I_V(t) - I_V(0)| \quad (t = 1,2,\dots,n). \tag{1}$$

And when S_d of a pixel satisfies

$$S_d(t) \geq \max(S_d)/2\,, \tag{2}$$

the pixel is then registered as skin color data.

The process above is applied in two stages. It is first applied to the low-resolution image and then to the original image. In the second stage, the background image is smoothed using a Gaussian filter so that the background subtraction is not too sensitive to the difference of the background and captured images.

4.3 Skin color registration

The obtained skin color candidates are not perfect, because they contain noises and background pixels whose color is similar to skin color. To remove the outliers, the H, S, and V values are limited in fixed ranges for the median of H, S, and V. Furthermore, values in either end of the histogram with a fixed rate are eliminated, and then the remaining pixels are registered as skin color pixels.

V (Intensity 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. Then the feature vector is given as

$$\mathbf{X}=[\mathrm{H}, \mathrm{S}]. \tag{3}$$

The mean vector **A** and the covariance matrix **V** of **X** are given as follows.

$$\mathbf{A} = \frac{1}{\mathrm{N}}\sum_{\mathrm{t}=1}^{\mathrm{N}} \mathbf{X}_{\mathrm{t}} \tag{4}$$

$$\mathbf{V} = \begin{bmatrix} \sigma_{\mathrm{H}}^2 & \sigma_{\mathrm{HS}} \\ \sigma_{\mathrm{HS}} & \sigma_{\mathrm{S}}^2 \end{bmatrix} \tag{5}$$

where N is the number of components of the pixel values to register. When the skin color region is extracted from a captured image, the Mahalanobis distance d_m for the skin color data is calculated from the following equation:

$$\mathrm{dm}^2 = (\mathbf{X}-\mathbf{A})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{X}-\mathbf{A}) \tag{6}$$

When the Mahalanobis distance satisfies

$$\mathrm{dm}^2 \leq \mathrm{k} \tag{7}$$

the pixel is extracted as a skin color pixel. k is a threshold value.

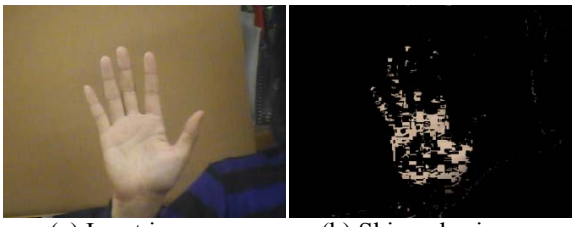
5. GESTURE RECOGNITION

5.1 Finger number recognition

A hand region is surrounded using Snakes [12] to extract skin color data, and the number of fingers is obtained using the surrounding region. In the previous system, finger regions were extracted using a morphology process (opening), and, therefore, the hand region had to be extracted correctly. However, in a real environment, correct extraction of hand region is difficult: the method may extract not only hand region but also background that has similar color of skin or extract a hand region insufficiently. Because Snakes is used in the new system, the finger number recognition is possible if only the outline of the hand region is extracted.

1) Extraction of a skin color region

A hand region is extracted using the registered skin color information. When the background color is similar to the color of a hand, as shown in Fig.3 (a), it is difficult to extract the hand region only. As the proposed method that uses Snakes works if only an outline is obtained, we set the threshold k small so that background regions are not included (see Fig.3(b)), which enables correct recognition.

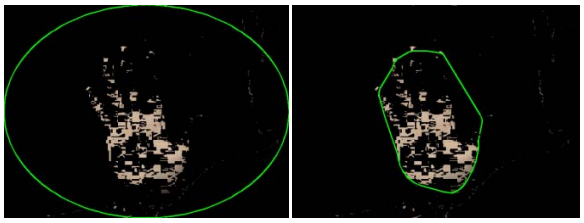


(a) Input image (b) Skin color image

Fig.3 Extracted skin color region.

2) Measurement of the hand region by Snakes

We surround a hand region 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 for an arbitrary shape. The initial outline is given as shown in Fig.4 (a). Fig.4 (b) is the outline curve obtained by Snakes. Then the area surrounded by the outline is measured.



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

Fig.4 Recognition of number of fingers by Snakes.

3) Estimation of the number of finger

The number of fingers is estimated using the area of the hand region. We introduce two methods.

・Recognition using the Mahalanobis distance

The method obtains average μ_i and standard deviation σ_i of the area for each finger number using training data. When an area of hand region A is measured, we obtain the Mahalanobis distance D_i to each finger number using the following equation.

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

The finger number with the smallest distance is selected as the recognition result.

・Recognition by accumulation of likelihood

The method also obtains the average μ_i and standard deviation σ_i. When an area of hand region A is measured, we obtain the likelihood l_i to each finger number using the following equation.

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

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.

5.2 Recognition of a 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, as illustrated in Fig.5.

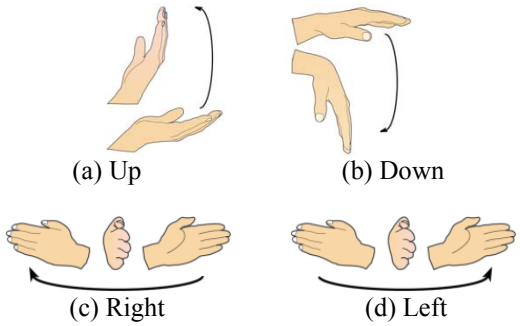


Fig.5 Gesture models.

6. EXPERIMENTS TO EVALUATE KEY FUNCTIONS

We evaluated the improved key functions by experiments [8].

6.1 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.

6.2 Finger waving recognition

Experiments to detect finger waving were conducted using the method in Section 3. The recognition rate and time for recognition were evaluated for different distances for five subjects. We assume that finger waving is recognized when it is detected by two cameras. 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

6.3 Skin color registration

Experiments of skin color registration were conducted using the method in Section 4. 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	310.0
S.D.	4.5	6.7	5.5

6.4 Recognition of the number of fingers

Experiments to estimate the number of fingers were conducted with five subjects using the method in Section 5.1. First, skin color registration in Section 4 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 number of fingers for each subject. The two methods in Section 5.1 and the previous method using morphology process were compared.

For training, the hand region was measured 20 times for each number of fingers from each subject. The average and standard deviation of the area were calculated and used for estimating the number of fingers. The training data are shown in Fig.6. The distance to the subject from the camera was set to 6m both at obtaining the training data and at estimating the number of fingers. The results are shown in Fig.7 and Table 3. The recognition rate for each subject is shown in Fig.7, and the processing time, in Table 3.

For each number of fingers, 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. Even in this case, extraction of the fingers failed (the method counted the number of regions other than fingers), which 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 number of data required for recognition changes according to the input. In this experiment, the number was 3 or less.

To estimate the number of fingers, prior skin color extraction is necessary. The total time in the experiment was less than 2 seconds, which is feasible for real use.

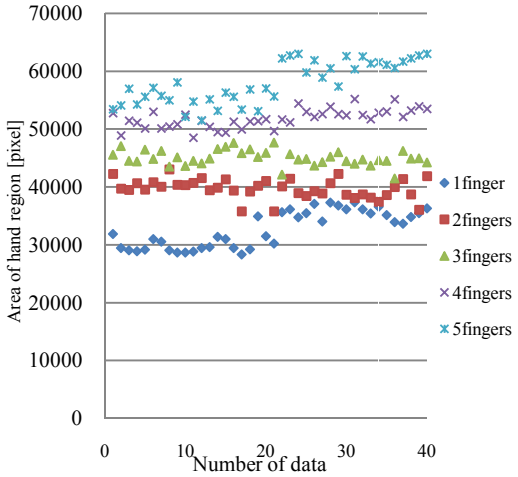


Fig.6 Model data of the number of fingers.

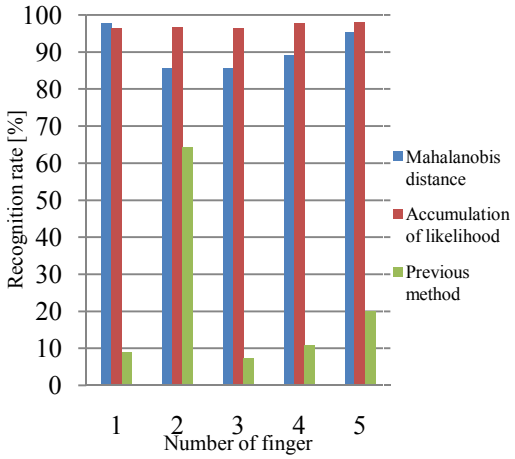


Fig.7 Recognition rate of the number of fingers.

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

6.5 Recognition of a hand gesture

Experiments for hand-gesture recognition were conducted for two subjects using the method in Section 5.2. 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 each subject. 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 these operations tend to fluctuate more.

Table 4 Recognition rate of hand gesture (%).

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

7. APPLICATION OF HAND WAVING RECOGNITION

The method of hand/finger waving detection can detect small motion robustly, and can calculate its 3D position with high accuracy using two cameras. We take advantage of the characteristics of the method and constructed a system to control a home appliance with only hand waving. We introduced the concept of "Spatial Memory" proposed by Niitsuma et al. [9] and allocated a cuboid space with 0.3-0.4m sides to each operation. When the hand waving is observed in a cuboid space, the corresponding operation is selected. As a guide for an operator, we put sheets (A4 size) with printed operation.

We used a TV set as an example of a home appliance and constructed a system. Fig.8 shows the floor map of the experiments. Fig.9 and 10 show the overview of the experiments, and Table 5 shows the kind of operations and assigned space to each operation. The spaces for TV on/off and channel inputs are allocated just above each sheet and 0.3m above for volume control commands. Fig.9 shows the scene in which the operator selects a channel by waving hand above the corresponding sheet. Fig.10 shows the scene in which he selects a volume control command.

Table 6 shows the experimental results of recognition rate and processing time for each operation for 50 times. It is shown that recognition rates are high except for Ch. 3. The reason why the result for Ch. 3 is not good is that vibration of arm was detected in an incorrect space. The processing time is about 1 second for each operation. This time is short enough and does not give a stress to the operator. In total, the results confirmed the usefulness of this system.

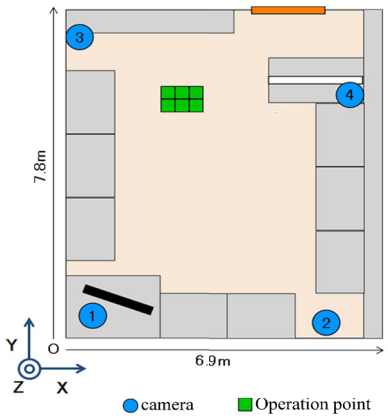
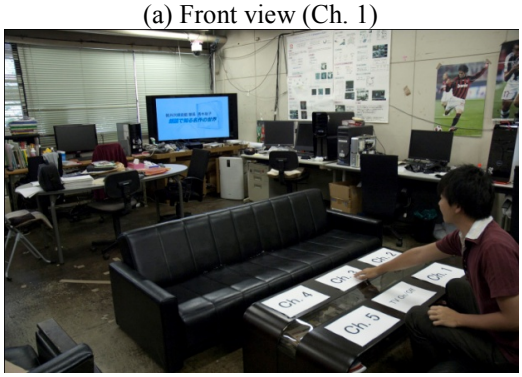


Fig.8 Floor map of the room for experiments.

(a) Front view (Ch. 1)



(b) Side view (Ch. 3)

Fig.9 Channel input using hand waving.


(a) Up


(b) Down

Fig.10 Volume control using hand waving.

Table 5 Kind of appliance operations and its positions.

	X	Y	Z
TV on/off	2.4~2.8	5.3~5.6	0.3~0.6
Ch. 1	2.8~3.2	5.3~5.6	0.3~0.6
Ch. 2	2.8~3.2	5.0~5.3	0.3~0.6
Ch. 3	2.4~2.8	5.0~5.3	0.3~0.6
Ch. 4	2.0~2.4	5.0~5.3	0.3~0.6
Ch. 5	2.0~2.4	5.3~5.6	0.3~0.6
Vol. up	2.0~2.6	5.0~5.6	0.6~1.0
Vol. down	2.6~3.1	5.0~5.6	0.6~1.0

Table 6 Recognition rates of experimentation.

	Recognition rate [%]	Processing time [s]
TV on/off	100	1.19
Ch. 1	98	1.13
Ch. 2	100	1.21
Ch. 3	72	1.08
Ch. 4	100	1.10
Ch. 5	98	1.05
Vol. up	100	0.92
Vol. down	100	1.00

8. CONCLUSION

In this study, we improved key functions used for operation in the intelligent room that was formerly proposed by our group [6][7], i.e., detection of hand waving, skin color registration, and recognition of the number of fingers. Additionally, we constructed a system to control a TV set as an application of detection of hand waving. With the system, power on/off, change of channels and volume control of a TV set is achieved with only hand waving.

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,” *Proc. Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 410-415, 2000.

[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,” *Fourth International Conference on Networked Sensing Systems (INSS2007)*, pp.118-121, Braunschweig, Germany, 2007.

[8] T. Nagayasu, H. Asano, M. Takahashi, K. Terabayashi, and K. Umeda: “Improvement of Key Functions of the Intelligent Room,” *Proc. of Eighth International Conference on Network Sensing Systems (INSS2011)*, 2-2, 2011.

[9] M. Niitsuma, H. Hashimoto, H. Hashimoto, "Spatial Memory as an Aid System for Human Activity in Intelligent Space," *IEEE Trans. on Industrial Electronics*, Vol.54, No.2, pp.1122-1131, 2007.

[10] O. Faugeras, "Three-Dimensional Computer Vision," MIT Press, 1993.

[11] 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.

[12] M. Kass, A. Witkin, and D. Terzopouls, "Snakes: Active Contour Models," *International Journal of Computer Vision*, Vol.1, No.4, pp.312-331, 1988.