Paper:

Skin Color Registration Using Recognition of Waving Hands

Kota Irie[*,*], Masahito Takahashi[*], Kenji Terabayashi[**,***], Hidetoshi Ogishima[**], and Kazunori Umeda[**,***]**

[*]Graduate School of Science and Engineering, Chuo University
E-mail: kotairie@sensor.mech.chuo-u.ac.jp
[**]Faculty of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
[***]CREST Program, Japan Science and Technology Agency (JST)
[Received September 7, 2009; accepted February 1, 2010]

This paper proposes skin color registration using the recognition of waving hands. In order to recognize hand gestures from images, skin colors are useful information. The proposed method can register skin colors simply and quickly because it uses just a few waves of the hand. The method consists of 2 steps. First, the regions of the waving hands are extracted from low-resolution images without using color information. Second, the color values of the extracted regions are classified into background colors and hand colors depending on time series of color images. The color information classified as hand colors is registered as skin colors. The proposed method is robust against lighting conditions and individual differences in skin color, because the skin color is registered as an adapted skin color in each case. Several experiments are conducted to demonstrate the effectiveness of the proposed method.

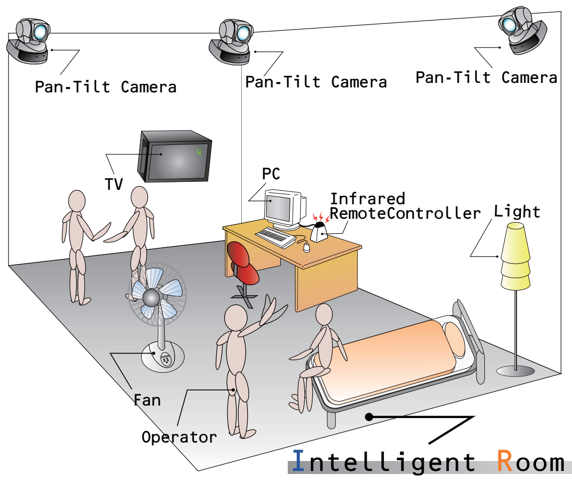


Fig. 1. The conceptual figure of our intelligent room [18].

Keywords: image processing, gesture recognition, color extraction, intelligent room, hand waving

1. Introduction

To ensure smooth communication between humans and machines, including robots, there are an increasing number of researches underway on Human-Machine Interfaces (HMI) using gesture recognition [1–6]. In gesture recognition using images, such features as motions or poses of human heads or hands need to be quantified from images. The processes involved in gesture recognition are first to identify subjects of motions or to extract the regions of such motions. Various methods have been proposed to extract regions of human heads or hands from information on "motions," "shapes," "colors," etc. [7]. The method using "motions," though suited to tracking human motions in the scene, has difficulty distinguishing different moving objects in the scene or in accurately positioning such objects. The method using "shapes," though capable of accurately positioning objects based on outlines or internal structures, is susceptible to effects of changes in views from different viewpoints, distances to objects, or occlusions, and it generally requires complex calculations. The method using "colors" is robust against changes in views and is capable of relatively high-speed operation. A large number of reports are available on methods of extracting skin-color regions from images [8–16]. If there are many similar colors in the scenes, however, it is difficult to cut heads or hands out of them and combined use of clue information is required. Other methods already proposed include a method of using sensors to convert into images thermal information from far-infrared cameras [17]. The extraction of human heads or hands by this method may be possible by narrowing down regions into the temperature range corresponding to that of human body temperatures. This may allow relatively stable recognition in the environments with less noise from similar temperatures in the ambiance. It is still difficult to put a system using such expensive sensors into general practice, however.

We have constructed an intelligent room as a system in which HMI is carried out by electric appliances using gesture recognition [18] (see **Fig. 1**). In this system, we have architected gesture-recognition applications using skin color information, so that stable performance of recognition can be assured without the use of expensive sensors or high-performance computers. In using color information, we have to beware that in addition to the above-mentioned concerns, colors will appear to vary depending on the brightness of photographing environments, types of light sources, and cameras. Moreover, since skin colors tend to depend on individual differences or races, the system should be robust against such issues. As a solution to those issues, we have made it a rule to carry out skin color registration operations with operators each time the system is operated. In architecting a system effective only in the room as in **Fig. 1**, one potential method may be to install a system to register skin color information on users, perhaps at the entrance of the room, so that users should not be admitted before completion of the registration of skin color information. Inside the room, however, skin colors vary owing to lighting conditions that may not be uniform or to relative positional relations between light sources and users. In short, the system should also be robust against variations in the ambience or users' positions inside the room.

In the skin color registration operation [19] implemented in previously constructed intelligent rooms, the extraction of hand regions has required the palm of the hand to rest stationary towards camera for a given time. Moreover, the skin color registration operation has relied on the extension of regions of the hand by means of the zoom function of camera; shifts in the positions of palms have extracted colors from the wrong regions. To ensure smooth operation of the system, it should require the users to make as few troublesome motions as possible.

In this paper, we propose a method of registering skin colors using hand-waving motions. In the intelligent room, the recognition of waving hands [20] is used to identify the positions of operators upon a signal to commence operations of the system. Combined use of waving-hand recognition and skin color registration is aimed at reducing troublesome burdens on users and at achieving smoother operations of the system.

The proposed method, though primarily intended for applications to intelligent rooms, is also applicable to such systems that register arbitrary colors at arbitrary places, such as in gesture games; hence, it is a general-purpose method.

2. Skin Color Registration Method

Skin color registration, as shown in **Fig. 2**, takes the following flow of processes. First, register in real-time background images and recognize waving hands. After recognizing waving hands, capture differences in subtraction between several frames of images in the waving-hand region and registered background images. After excluding outliers in the set of subtraction pixel values, obtain average vectors and covariance matrices for Hues (H) and Saturations (S) to register as final skin color data.

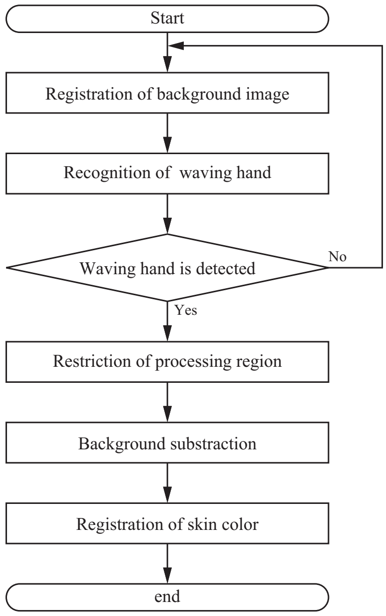
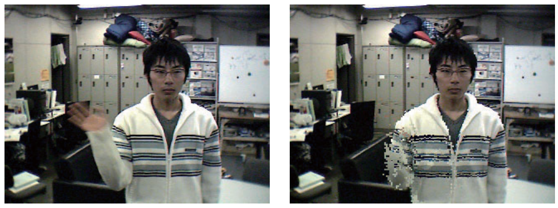


Fig. 2. Flowchart of skin color registration.



(a) Hand waving image (b) Background image

Fig. 3. Registration of background image.

2.1. Registration of Background Images

The registration of background images is intended to extract the pixel values of waving hands. Pixel values that do not vary for several frames are registered as background images and updated on a real-time basis. To allow quick determination of the existence or non-existence of variations, what is registered as background images are generally static objects such as furniture or walls, as well as the operator's body, which is temporarily made to stand still during hand-waving. On the other hand, since hand-waving motions are relatively quite rapid, it may be possible that backgrounds are not updated within the waving-hand region, as shown in **Fig. 3**, nor that hands be registered as background images.

2.2. Recognition of Waving Hands

Hand-waving motions that provide a signal to commence operations of the system in the intelligent room are to be detected in accordance with the detection principles given in the Reference [20]. The outline of the detection method is as follows (see **Fig. 4**):

2.2.1. Low Resolutions of Images

When hand-waving recognition is carried out using gray images (**Fig. 4(a)**), some differences arise in inten-

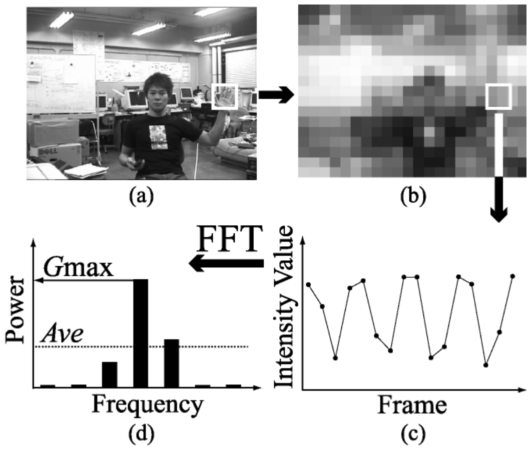


Fig. 4. Detection of hand waving using FFT.

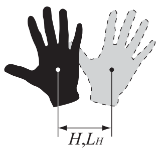
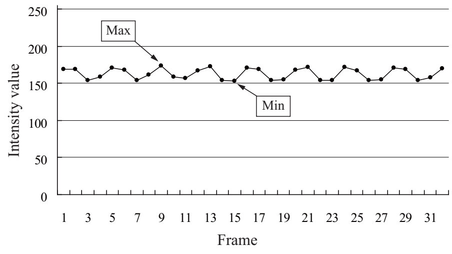
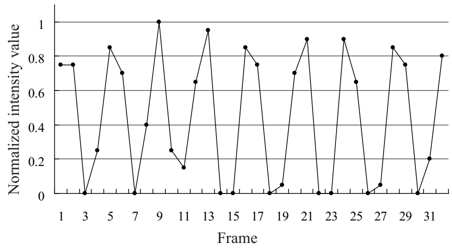


Fig. 5. Width of waving hands.



(a) Intensity value before normalization



(b) Intensity value after normalization

Fig. 6. Normalization of intensity variation.

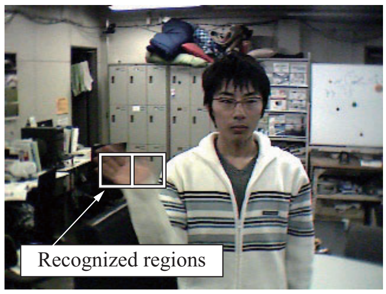


Fig. 7. Regions recognized as hand waving.

sity values between the hand and background in resolutions of the hand region. In preparation for capturing such time-series variations in intensities, low resolutions of images are conducted as shown in **Fig. 4(b)** in expectation of such effects as reductions in calculation volumes, improved robustness against noises, and smooth patterns of variations in intensities.

With camera angles denoted by θ [rad], horizontal resolutions of captured images by a [px], measuring distances by L [m], hand-waving widths by L_H [m], and angles between the hand-waving motion plane and the images plane by α [rad], shifts in the center of gravity H [px] of the hand in captured images are expressed by Eq. (1). **Fig. 5** shows a schematic diagram of waving hands.

$$H = \frac{aL_H \cos\alpha}{2\tan\frac{\theta}{2}} \cdot \frac{1}{L} \quad \text{. . . . . (1)}$$

For variations in intensities between the hand region and background to emerge in low-resolution images, hand-waving images may not be less than 1 pixel in width in such low-resolution images. With sampling resolutions denoted by P_{lim}, they need to satisfy $H > P_{lim}$.

2.2.2. Recognition of Waving Hands

Next, time-series variations in intensity values for pixels of low-resolution images are captured, and differences between the maximum and minimum values are normalized to reference values, as shown in **Fig. 6**. Variations in normalized intensity values undergo a Fast Fourier Transform (FFT) (**Fig. 4(c)**). In the hand-waving regions, intensity values of low-resolution images oscillate in a form close to sinusoidal waves. It is therefore fairly easy to recognize waving hands by analyzing oscillation components of waving hands from power specters obtained through FFT (**Fig. 4(d)**). Then, from the power specters of obtained frequencies are calculated the maximum value $G_{\max}$, average value Ave, skewness Sk, and kurtosis Ku, to conduct discriminant analyses [21, 22] by the Mahalanobis distance with multivariate features. Now, with feature vectors $\boldsymbol{X}_{wh}$ comprised of these four variables, average vectors $\boldsymbol{A}_{wh}$ and covariance matrices $\boldsymbol{V}_{wh}$ for the hand-waving class ω_{true} and other classes ω_{false}. With the Mahalanobis distance for ω_{true} being denoted by Dm_{true} and that for ω_{false} by Dm_{false}, they are expressed by Eq. (2), and hand-waving is recognized if Eq. (3) is satisfied.

$$Dm^2_{(true,\,false)} = (\boldsymbol{X}_{wh} - \boldsymbol{A}_{wh})^T \boldsymbol{V}_{wh}{}^{-1} (\boldsymbol{X}_{wh} - \boldsymbol{A}_{wh}) \quad . \ (2)$$

$$Dm^2_{true} \leq Dm^2_{false} \quad \text{. . . . . (3)}$$

2.3. Background Subtraction

Within the hand-waving regions shown in **Fig. 7**, subtraction between background images and captured images in the past n-frames is obtained. **Fig. 8** shows the flow of

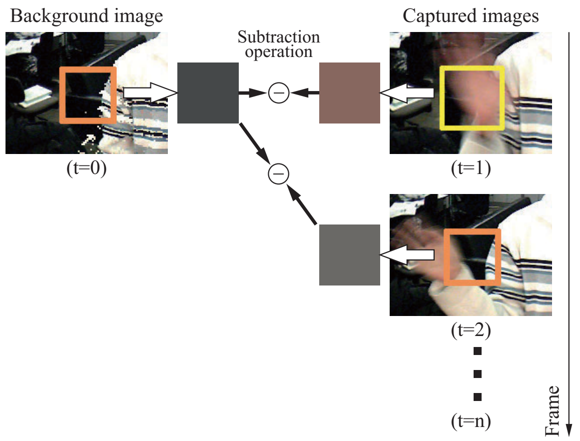


Fig. 8. Background subtraction in the region of the waving hand.

background subtraction operations. After averaging pixel values in the hand-waving regions (squared in the figure), comparison is made between the background image ($t = 0$ in the figure) and captured images in time frames $t = 1 - n$. Greater subtraction with the background is obtained when the hand is within the regions of waving-hand recognition (e.g., $t = 1$ in the figure), while subtraction with the background is smaller when the hand is not within the regions of waving-hand recognition or only a small part of the hand is within the recognition regions (e.g., $t = 2$ in the figure). In other words, extracting only frames containing hand-regions to the degree possible will allow us to narrow down candidate skin colors.

Evaluation functions in subtraction calculations are defined by Eq. (4) using normalized Hues (H), Saturations (S), and intensities (V) as parameters.

$$S_d(t) = |I_H(t) - I_H(0)| + |I_S(t) - I_S(0)| + |I_V(t) - I_V(0)| \quad \ldots\ldots\ldots\ldots (4)$$

where $I_H(t)$, $I_S(t)$, and $I_V(t)$ denote color components, saturation components, and intensity components, respectively, in captured images in the t_{th} frame; they are normalized so that their domains are within the range of 0-255.

Based on the evaluation functions as defined by Eq. (4), frames that satisfy Eq. (5) are determined to be candidate skin colors.

$$S_d(t) \geq \max(S_d)/2 \qquad (t = 1, 2, \ldots, n) \quad \ldots \quad (5)$$

Eq. (5) has been so established through experiments as to define conditions on which to ensure extraction of only frames containing hand regions. Frames that satisfy Eq. (5) then go through a much stricter processes of color extraction. Candidate skin colors are registered on the basis of pixel by pixel determination of such extracted colors against the original resolutions (320×240 px) (see **Fig. 9**). In this operation, background images are smoothed with Gaussian filters to reduce any small shifts between background images and captured images that may be caused by swaying. Kernel size $m \times m$ is defined by Eq. (6) using the widths H_w [px] of the regions

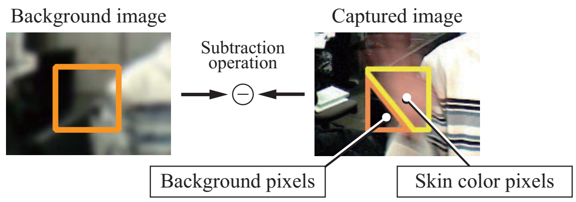


Fig. 9. Background subtraction at the pixel level.

where waving hands are recognized; standard deviations σ at Gaussian filters are defined by Eq. (7).

$$m = H_w/2 \quad \ldots\ldots\ldots\ldots (6)$$

$$\sigma = (m/2 - 1) \times 0.3 + 0.8 \quad \ldots\ldots\ldots\ldots (7)$$

Pixel values excluded through the above-mentioned subtraction operations (triangle regions in **Fig. 9**) are registered as background data.

2.4. Registration of Skin Colors

To improve the accuracy of color information to register, pixel values of erroneously registered clothes or outliers due to noise are excluded. Hues (H), Saturations (S) and intensities (V) are restricted to within certain limits of respective medians as reference values; certain ratios of pixel values from the both ends of respective histograms are excluded as outliers. Skin color pixel values are thus determined and registered. Out of the registered pixel values, intensities (V) vary greatly depending on ambient lighting conditions. To ensure a robust skin color registration, Hues (H) and Saturations (S) that are not susceptible to lighting conditions are adopted as feature vectors $\boldsymbol{X}$. Average vectors $\boldsymbol{A}$ and covariant matrices $\boldsymbol{V}$ are obtained from Eqs. (8) and (9) to calculate Mahalanobis distances and register final skin color data.

$$\boldsymbol{A} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{X}_t \quad \ldots\ldots\ldots\ldots (8)$$

$$\boldsymbol{V} = \begin{bmatrix} \sigma_H^2 & \sigma_{HS} \\ \sigma_{HS} & \sigma_S^2 \end{bmatrix} \quad \ldots\ldots\ldots\ldots (9)$$

where N denotes the number of elements in the set of pixel values.

In extracting skin colors from captured images, Mahalanobis distances for skin color data and background data are obtained from Eq. (10) and pixel values that satisfy Eqs. (11) and (12) are extracted as skin colors.

$$dm^2 = (\boldsymbol{X} - \boldsymbol{A})^T \boldsymbol{V}^{-1} (\boldsymbol{X} - \boldsymbol{A}) \quad \ldots\ldots (10)$$

$$\frac{dm_{true}^2}{dm_{false}^2} \leq c \quad \ldots\ldots\ldots\ldots (11)$$

$$dm_{true}^2 \leq k \quad \ldots\ldots\ldots\ldots (12)$$

where c denotes a threshold to ensure adequate proximity to skin colors as compared with distances to background data, and k denotes a threshold to ensure adequate proximity to skin colors. Extraction operations are conducted within a restricted range eight times the longitudinal and horizontal dimensions of the waving-hand recognition region on the premise that operators gesture from the same position that they wave. Furthermore, extracted results undergo morphology operations and labeling operations for the denoising of images and correction of skin color regions.

3. Verification Experiments on Registration of Skin Colors

3.1. Experimental System

The experimental system consists of a PC (DELL XPS 420, Intel Core2 Quad Q6600), capturing software (Microsoft DirectShow), image-processing software (Intel Open CV), and a USB camera (ELECOM UCAM-E130, 30 fps).

3.2. Experimental Conditions

Parameters for various operations have been established as follows:

For the background registration described in Section 2.1, variations in five frames (about 0.17 sec) have been used. Minimum frequencies in hand-waving in Section 2.2 have been experimentally set to 1.9 Hz, with individual differences in hand-waving speeds taken into consideration so as not to erroneously recognize any motions other than hand-waving. The number of frames sampled for FFT has been set to 64. Resolutions of images used for the recognition of waving hands have been determined by Eq. (1). The angle of the camera has been set to $\theta = 0.93$ rad and horizontal resolutions to $a = 320$ px. In these experiments, the upper limit of the measuring distance has been set to $L = 8$ m, hand-waving width to $L_H = 0.2$ m, and angles between the hand-waving motion plane and the image plane to $\alpha = 0$ rad. These parameters have led us to determine image resolutions to 40×30 px. The number of frames used for background subtraction operations, as described in Section 2.3, has been set to 64. For restrictions by medians, we have adopted ± 9 for Hues (H) and ± 50 for Saturations (S) and intensities (V) each, and 5% at each end of the histogram has been excluded. Discrimination thresholds for skin colors, as described in Section 2.4, have been set to $c = 4$ and $k = 10$.

Registrations of skin colors have been determined to be successful only if hand areas S_{true}, as measured in true images, and average hand areas $\overline{S}_{ext}$, as extracted from registered data, have satisfied Eq. (13), where $\overline{S}_{ext}$ is given by Eq. (14).

$$\frac{|\overline{S}_{ext} - S_{true}|}{S_{true}} \leq 0.4 \quad \quad (13)$$

$$\overline{S}_{ext} = \frac{1}{100} \sum_{k=1}^{100} S_{ext}(k) \quad \quad (14)$$

The threshold of 0.4 is a value experimentally set on approximate criteria that enable the extraction results of hand regions to present the number of fingers from a short distance and the direction of movement of the hand from a long distance.

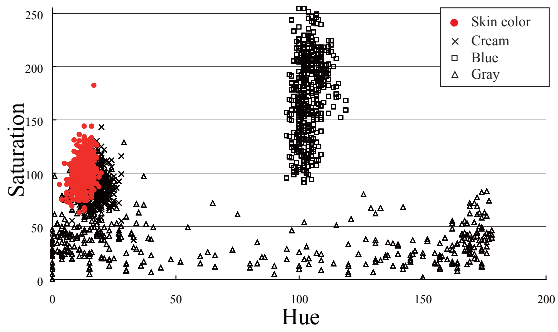


Fig. 10. Distributions of skin color and ambient colors.

3.2.1. Validation of Registered Values

To validate registered values of skin colors, we have conducted a comparison between values registered by our proposed system and manually registered values of hand regions. The comparison was conducted for the distance of 3 m in indoor environments containing plural colors.

3.2.2. Distances

To evaluate the performance of the proposed method against distances, we have verified the success rates of skin color registrations in 20 times of hand-waving movements at three distances from the camera: 1.5 m, 3 m, and 5 m.

3.2.3. Background Environments

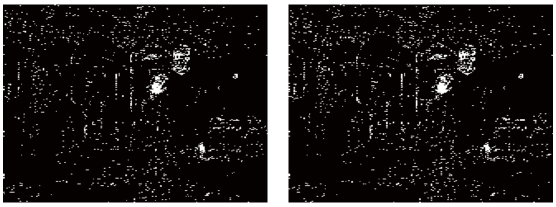
To evaluate the performance of the proposed method against background colors, we have conducted experiments using different colored clothes and backgrounds. We have verified the success rates of skin color registrations through 20 times of hand-waving movements at a measuring distance of 1.5 m and with gray, blue, and cream-colored backgrounds. **Fig. 10** shows distributions of manually-captured ambient colors.

3.2.4. Individual Differences

To evaluate differences in skin colors due to individual differences, we have attempted to register the skin colors of five subjects. In the experiments, we have set the measuring distance to 1.5 m and have only used fluorescent lamps for lighting. The five subjects have waved their hands at the same position and angle, so there should be no difference in lighting conditions among the subjects.

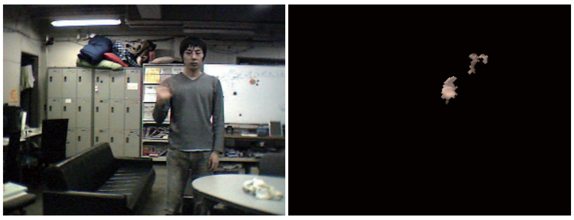
3.2.5. Lighting Conditions

To evaluate differences in skin colors with different lightings or lighting intensities, we have attempted, using fluorescent lamps and incandescent lamps, to register skin colors for one subject at the measuring distance of 1.5 m. We have used the following four lighting conditions:

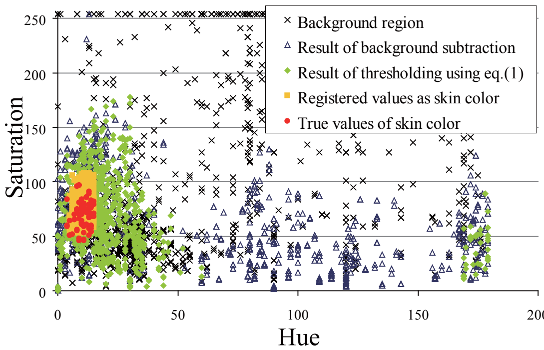


(a) Proposed method (b) Manual operation

Fig. 11. Examples of regions extracted as skin colors.



(a) Captured image and extracted regions as skin color



(b) Distribution of colors in captured image

Fig. 12. Example of a successful trial.

(a) All fluorescent lamps in the room were lit.

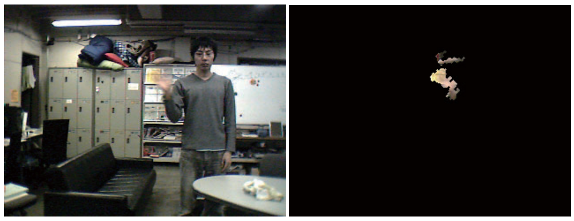
(b) About two-thirds of the lamps were then turned off.

(c) Condition (b) and one 40 w incandescent lamp was turned on and set at a distance of 1.0 m from the subject.

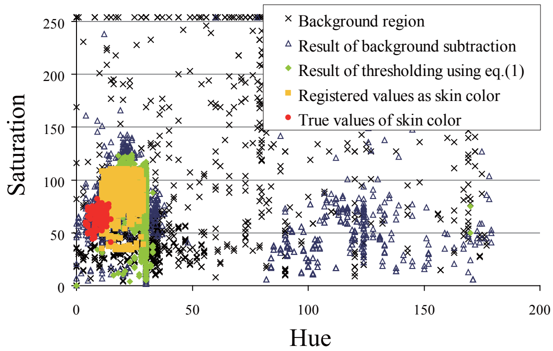
(d) One 40 w incandescent lamp was turned on and set at a distance of 1.0 m from the subject.

3.2.6. Background with Mixed Moving Objects

To verify the practicality of the proposed method, we have conducted experiments with mixed moving objects behind the subject. The experiments were conducted on one subject at the measuring distance of 1.5 m with 1-4 persons walking freely behind the subject.



(a) Captured image and extracted regions as skin color



(b) Distribution of colors in captured image

Fig. 13. Example of a failure.

3.3. Experimental Results

3.3.1. Evaluation of Values Registered as Skin Colors

We have validated the skin colors registered by the proposed method. **Fig. 11** compares the extraction results of skin colors between the proposed method (**Fig. 11(a)**) and manual operation (**Fig. 11(b)**). Both figures, representing binarized images prior to denoising, show almost equal extraction results. To quantitatively evaluate the registration results of skin colors, we have computed distances in the H-S space between the skin color values registered by the proposed method (class of registered values as skin colors) and those manually registered (class of true values of skin colors) as group average distances. From the said computation, we have obtained a group average distance of -4.41 for the examples shown in **Fig. 11**. Moreover, the absolute value is smaller than the standard deviation of the class of true values. We have therefore determined that the distance between the two classes is short. Such results seem to suggest that registered skin colors fairly represent actual hand colors.

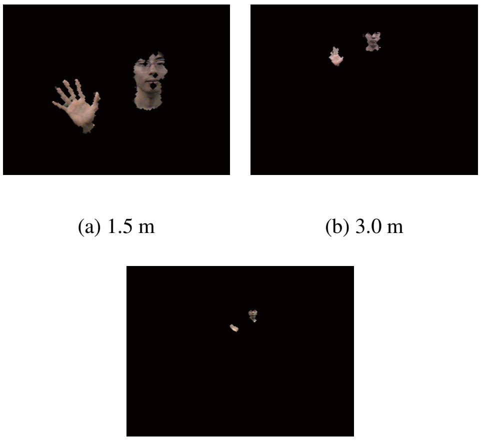
Figure 12 is a graph on which specific data on pixel values in **Fig. 11** are plotted in the H-S space; the plotted data are related to background regions, results of background subtraction, results of threshold operation, skin color values registered using the proposed method, and manually registered skin color values. While the background subtraction operation has produced broadly distributed results, the threshold operation has made the results converge into nearly the same range as the manually registered values. On the other hand, **Fig. 13** shows a failure, as the fingertips of the waving hand are selected for

Table 1. Success and failure rates at various distances.

Distance	Success	Failure	
		Position	Skin color extraction
1.5 m	17/20	3/20	0/20
3 m	15/20	3/20	2/20
5 m	13/20	4/20	3/20

Table 2. Success and failure rates with various ambient colors.

Color	Success	Failure	
		Position	Skin color extraction
Gray	17/20	3/20	0/20
Blue	18/20	2/20	0/20
Cream	15/20	3/20	2/20



(a) 1.5 m (b) 3.0 m

(c) 5.0 m

Fig. 14. Examples of regions extracted as skin colors.

capture and are not properly registered as skin colors. The failure in registration of skin colors may be attributed to the fact that the speed of the waving hand is fast enough to cause blurring in the vicinity of the fingertips, mixing hand colors with background colors. This has resulted in values representing colors different from actual skin colors. To cope with fast waves of the hand, one possible measure could be to raise the frame rate of the camera. The group average distance between the class of registered values and the class of true values in the case of failure in **Fig. 13** is -25.3, much bigger than that in the case of success.

3.3.2. Performance Evaluations from Various Distances

The performance results of the proposed system from various distances are given in **Table 1**, where cases of failure are divided into cases where recognized hand regions are shifted from actual waving-hand positions and cases where proper skin colors are not extracted; most of the failures are due to shifts in regions of waving-hand recognition irrespective of distances. In some cases, wrists, arms, or hand shades of color that oscillate together with waving hands are recognized, and in other cases, only very small parts of images for the waving-hand regions are recognized. Lack of recognition of the palm of the hand may have contributed to the failure in proper extraction of skin colors in the subsequent background subtraction operations. In some cases, longer distances, which may not affect the proper recognition of the regions, have caused a failure in registration of skin colors; some registrations of apparent skin colors do not present adequate separation from the background and others can only identify parts of the skin.

All of these failures seem attributable to relatively deteriorated resolutions, reduced amounts of information, insufficient capture of pixel values of skin colors, and resultant registration of deviated values. Those are the reasons why accuracy decreases with distance.

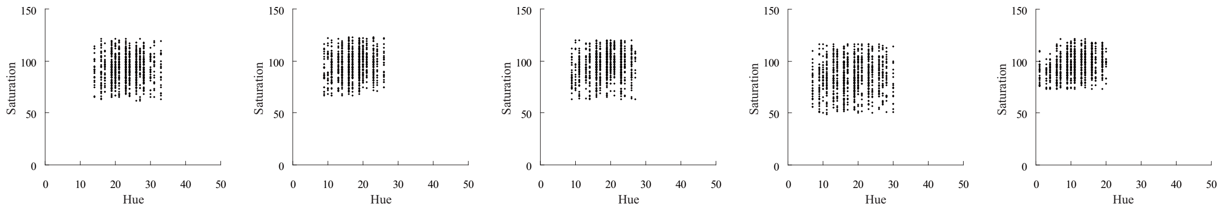
In the case of proper registration of skin colors, as shown in **Fig. 14**, the extraction results are such as to be able to identify fine shapes of the fingers of the hand at the measuring distance of 1.5 m. If the measuring distance is 3 m or more, however, though fine shapes cannot be extracted owing to inadequate resolutions of the images, we may duly expect to be able to identify general motions, such as the direction of movement of the hand regions. The hands in the images vary in size with the measuring distances as follows: 70×78 px in **Fig. 14(a)**, 25×29 px in **Fig. 14(b)**, and 13×14 px in **Fig. 14(c)**.

3.3.3. Performance Evaluations with Various Ambient Colors

Table 2 shows the results of performance evaluations on ambient colors. Most of the failures are not attributable to differences in ambient colors but to shifts in recognition regions for the waving hand. With gray and blue background colors, proper recognition of the regions has led to successful extractions of skin colors in all cases. However, when the background is a cream color similar in color distributions to the waving hand, there seem to be considerable effects on deviations in registration of data. In the two failures in which the hand region and background are mixed in the extracted regions, proper extractions of skin colors have been found to be possible with adjustments of c and k as described in the experimental conditions. Dynamic settings of thresholds depending on registered values may ensure improvements.

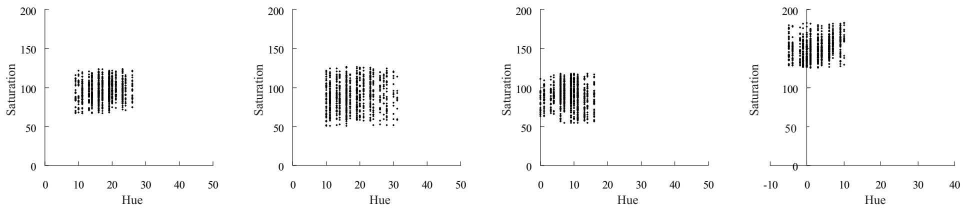
3.3.4. Verification of Individual Differences

Figure 15 shows distributions of registered data on skin colors by Subjects A-E. **Table 3** shows component values in average vectors $\boldsymbol{A}$ and covariance matrices $\boldsymbol{V}$ for hues and saturations of Subjects A-E. While Subjects B and C present similar distributions, there are differences in average vectors of hues or saturations among other subjects, presenting different skin colors. We can confirm with these results that there are individual differences in skin



(a) Subject A (b) Subject B (c) Subject C (d) Subject D (e) Subject E

Fig. 15. Distributions to express the effects of the individual differences.



(a) Bright fluor lamp (b) Dark fluor lamp (c) Dark fluor lamp and incandescent lamp (d) Incandescent lamp

Fig. 16. Distribution to express the effects of lighting conditions.

Table 3. Skin color data registered with each subject.

	$\boldsymbol{A}$		$\boldsymbol{V}$		
	$\overline{H}$	$\overline{S}$	σ_H^2	σ_S^2	σ_{HS}
Subject A	23.0	92.6	17.4	157.4	1.8
Subject B	17.6	96.6	17.6	152.3	8.3
Subject C	18.3	95.8	18.5	157.5	10.9
Subject D	17.7	85.0	32.0	249.6	8.6
Subject E	11.4	96.2	21.1	126.5	17.1

Table 4. Skin color data registered in each lighting condition.

	$\boldsymbol{A}$		$\boldsymbol{V}$		
	$\overline{H}$	$\overline{S}$	σ_H^2	σ_S^2	σ_{HS}
Condition (A)	17.6	96.6	17.6	152.3	8.3
Condition (B)	17.8	90.0	30.0	333.4	5.2
Condition (C)	8.3	90.2	14.5	208.8	−2.4
Condition (D)	2.4	148.2	14.7	169.4	16.0

colors. In other words, the proposed system, capable of registering skin colors specific to operators, has proved a flexible system without any restrictions on operators.

3.3.5. Verification Under Various Lighting Conditions

Figure 16 shows distributions of skin color data registered under the four lighting conditions given in Section 3.2.5. (**Fig. 16(d)** shows values of hues after reducing 256 from values of 128 or more for the convenience of graphic display.) **Table 4** shows component values of hues and saturations for average vectors $\boldsymbol{A}$ and covariance matrices $\boldsymbol{V}$ under the different lighting conditions. Comparison between conditions (a) and (b) reveals greater distributions of hues and saturations as well as slightly smaller average values of saturations under condition (b). We find from a comparison between fluorescent lamps and incandescent lamps that there is a big difference in the center of gravity of distributions. These results tell us that differences in brightness or color temperatures are bound to affect skin colors. In other words, the proposed system, capable of registering skin colors according to lighting conditions, allows robust registrations of skin colors even under varying levels of brightness or different lightings (color temperatures).

3.3.6. Evaluations with Mixed Moving Objects in the Ambience

We have conducted experiments to evaluate performance of the proposed system when there are plural moving objects within the imaging range. **Fig. 17** shows the results of extracted skin colors in the ambience with mixed moving objects. **Fig. 17(a)** shows the captured image at $t = 30$ frames of commencement of the program, and subsequent images captured at every 30 frames are displayed in time series up to $t = 210$ frames (**Fig. 17(g)**). **Fig. 17(h)** shows the results of the hand region extracted from the registered information on skin colors. We can see that hand regions are accurately extracted in the situations where one to four persons are moving freely behind the subject. These results confirm the practicality of the proposed system.

(a) $t = 30$ (b) $t = 60$ (c) $t = 90$ (d) $t = 120$

(e) $t = 150$ (f) $t = 180$ (g) $t = 210$ (h) $t = 240$

Fig. 17. Experiment in which one or more people are moving.

3.3.7. Operation Time

Prior to using the proposed system, the registration of skin colors used to take about 0.2-0.3 sec. Moreover, controlling the camera for pans, tilts, and zooms has required an additional 1-5 sec. On the other hand, the proposed system, which requires 0.05-0.10 sec for registrations of skin colors and requires no controlling of the camera, makes it possible for operators to focus on registrations of skin colors without regard to operation time. In addition, the proposed system, capable of recognizing hand-waving motions in about 2 sec on average, is expected to provide smooth operation inside intelligent rooms.

4. Conclusion

We have proposed a method of registering skin colors using the recognition of waving hands and have confirmed through plural experiments the effectiveness of the proposed method for applications to intelligent rooms. The proposed method has also eliminated the trouble involved in registration movements for skin colors, which used to be an issue to be solved, as described in Section 1 of this paper. However, it still remains an issue that longer distances between operator and camera result in lower recognition accuracies. Therefore, future issues to be further addressed need to include the correction of recognized regions for waving hands, dynamic settings of thresholds for the extraction of skin colors, and the adoption of higher-resolution and lower-noise cameras.

References:

[1] V. Pavlovic, R. Sharma, and T. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Vol.19, No.7, pp. 677-695, 1997.

[2] J. Sherrah and S. Gong, "VIGOUR: A System for Tracking and Recognition of Multiple People and their Activities," Proc. of the Int. Conf. on Pattern Recognition, pp. 179-183, 2000.

[3] P. Hong, M. Turk, and T. S. Huang, "Gesture Modeling and Recognition Using Finite State Machines," IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 691-694, 2000.

[4] H. Wu, T. Shioyama, and H. Kobayashi, "Spotting Recognition of Head Gestures from Color Image Series," Proc. of the Int. Conf. on Pattern Recognition, pp. 83-85, 1998.

[5] A. Nishikawa, A. Ohnishi, M. Nishimura, A. Hirano, K. Koara, and F. Miyazaki, "Systematic selection of local correlation parameters for opticalflow-based gesture recognition," Proc. of the 8th IEEE Int. Workshop on Robot and Human Interaction (ROMAN'99), pp. 183-188, 1999.

[6] T. Yamagishi, Y. Nakanishi, and K. Umeda, "Gesture Recognition for Realizing Intelligent Room," Mecatronics'01 5th Franco-Japanese Congress & 3rd European-Asian Congress Besancon, pp. 293-298, 2001.

[7] S. Akamatsu, "Computer Recognition of Human Face – A Survey –," The Trans. of the Institute of Electronics, Information and Communication Engineers D-II, J80-DII(8), pp. 2031-2046, 1997.

[8] K. Asanuma, M. Onishi, A. Kojima, and K. Fukunaga, "Extracting Regions of Human Face and Hands considering Information of Color and Region Tracking," The Trans. of the Institute of Electrical Engineers of Japan C, A publication of Electronics, Information and System Society, Vol.119-C, No.11, pp. 1351-1358, 1999. (in Japanese)

[9] S. Kawato and J. Ohya, "Automatic skin-color distribution extraction for face detection and tracking," Proc. of 5th Int. Conf. on Signal Processing (WCCC-ICSP 2000), Vol.2, pp. 1415-1418, 2000.

[10] G. Gomez and E. Moralez, "Automatic feature construction and a simple rule induction algorithm for skin detection," In Proc. IMCL Workshop on Machine Learning in Computer Vision, pp. 31-38, 2002.

[11] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.27, No.1, pp. 148-154, 2005.

[12] P. Kakumanua, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," Pattern Recognition, Vol.40, No.3, pp. 1106-1122, 2007.

[13] M. R. Girgis, T. M. Mahmoud, and T. Abd-El-Hafeez, "An Approach to Image Extraction and Accurate Skin Detection from Web Pages," Int. J. of Computer Science and Engineering, pp. 88-96, 2007.

[14] J. H. Shin, J. S. Lee, S. K. Kil, D. F. Shen, J. G. Ryu, E. H. Lee, H. K. Min, and S. H. Hong, "Hand Region Extraction and Gesture Recognition using entropy analysis," IJCSNS Int. J. of Computer Science and Network Security, Vol.6, No.2A, pp. 216-222, 2006.

[15] Y. Shi, R. Taib, and S. Lichman, "GestureCam: A Smart Camera for Gesture Recognition and Gesture-Controlled Web Navigation," 9th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV), pp. 1-6, 2006.

[16] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A Hidden Markov Model-Based Continuous Gesture Recognition System for Hand Motion Trajectory," 19th Int. Conf. on Pattern Recognition (ICPR), pp. 1-4, 2008.

[17] Y. Sato, Y. Kobayashi, and H. Koike, "Fast tracking of hands and fingertips in infrared images for augmented desk interface," Proc. 2000 IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp. 462-467, 2000.

[18] K. Irie, M. Wada, and K. Umeda, "Construction of an Intelligent Room using Distributed Camera System," Stereo Vision edited by Asim Bhatti, InTech Education and Publishing, pp. 121-130, 2008. ISBN 978-953-7619-22-0

[19] 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.

[20] K. Irie and K. Umeda, "Detection of Waving Hands from Images Using Time Series of Intensity Values," The 3rd China-Japan Symposium on Mechatronics (CJSM), 2002.

[21] K. Fukunaga, "Introduction to Statistical Pattern Recognition (Second Edition)," Academic Press, 1990. ISBN 978-0122698514

[22] G. Taguchi and R. Jugulum, "The Mahalanobis Taguchi Strategy: A Pattern Technology System," John Wiley & Sons Inc., 2002. ISBN 978-0471023333

Name:
Kota Irie

Affiliation:
Hitachi Automotive Systems, Ltd.

Address:
2520 Takaba, Hitachinaka-shi, Ibakaki-ken 312-8503, Japan

Brief Biographical History:

2001 Received B. Eng. degree in precision mechanics from Chuo University

2003 Received M. Eng. degree in precision engineering from Chuo University

2003-2009 Joined Hitachi, Ltd.

2009- Transferred to Hitachi Automotive Systems, Ltd.

2010 Received Ph.D. degree in precision engineering from Chuo University

Main Works:

• Kota Irie, Masaki Wada, and Kazunori Umeda, "Construction of an Intelligent Room using Distributed Camera System," Stereo Vision edited by Asim Bhatti, InTech Education and Publishing, pp. 121-130, November 2008.
ISBN 978-953-7619-22-0

• 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, Sep. 2004.

Membership in Academic Societies:

• The Robotics Society of Japan (RSJ)

Name:
Masahito Takahashi

Affiliation:
Course of Precision Engineering, Graduate School of Science and Engineering, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:

2009 Received B. Eng. in precision mechanics from Chuo University

Membership in Academic Societies:

• The Robotics Society of Japan (RSJ)

• The Japan Society of Mechanical Engineers (JSME)

Name:
Kenji Terabayashi

Affiliation:
Assistant Professor, Department of Precision Mechanics, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
2004 Received M. Eng. degree in systems and information engineering from Hokkaido University
2008 Received Ph.D. degree in precision engineering from the University of Tokyo
2008- Assistant Professor at Chuo University

Main Works:
• K. Terabayashi, N. Miyata, and J. Ota, "Grasp Strategy when Experiencing Hands of Various Sizes," eMinds: Int. J. on Human-Computer Interaction, Vol.1, No.4, pp. 55-74, 2008.
• K. Terabayashi, H. Mitsumoto, T. Morita, Y. Aragaki, N. Shimomura, and K. Umeda, "Measurement of Three Dimensional Environment with a Fish-eye Camera Based on Structure From Motion – Error Analysis," J. of Robotics and Mechatronics, Vol.21, No.6, pp. 680-688, 2009.

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)
• The Japan Society for Precision Engineering (JSPE)
• The Japan Society of Mechanical Engineers (JSME)
• The Virtual Reality Society of Japan (VRSJ)
• The Institute of Image Electronics Engineers of Japan (IIEEJ)
• The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Kazunori Umeda

Affiliation:
Professor, Department of Precision Machanics, Chuo University

Address:
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Brief Biographical History:
1994 Received Ph.D. degree in precision machinery engineering from the University of Tokyo
1994- Lecturer at Chuo University
2003-2004 Visiting Worker at National Research Council of Canada
2006- Professor at Chuo University

Main Works:
• N. Hikosaka, K. Watanabe, and K. Umeda, "Development of Obstacle Recognition System of Humanoids Using Relative Disparity Maps from Small Range Image Sensors," J. Robotics and Mechatronics, Vol.19, No.3, pp. 290-297, June 2007.
• M. Tateishi, H. Ishiyama, and K. Umeda, "A 200Hz Small Range Image Sensor Using a Multi-Spot Laser Projector," Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 3022-3027, May 2008.

Membership in Academic Societies:
• The Robotics Society of Japan (RSJ)
• The Japan Society for Precision Engineering (JSPE)
• The Japan Society of Mechanical Engineers (JSME)
• The Horological Institute of Japan (HIJ)
• The Institute of Electronics, Information and Communication Engineers (IEICE)
• Information Processing Society of Japan (IPSJ)
• The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Hidetoshi Ogishima

Affiliation:
Tera International Inc.

Address:
Ikebukuro East Bldg., 3-4-3 Higashi Ikebukuro, Toshima-ku, Tokyo 170-0013, Japan

Brief Biographical History:
2009 Received B. Eng. in precision mechanics from Chuo University