Adaptive Skin Color Prediction Using Multi Skin Color Models

Keiichi Horio and Tomohiro Katayama
Graduate School of Life Science and Systems Engineering,
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196 Japan
email:horio@brain.kyutech.ac.jp, katayama-tomohiro@edu.brain.kyutech.ac.jp

***Abstract*—In this paper, a new skin color detection method, in which a skin color model for a given image is adequately selected from a set of models to realize adaptive detection, is proposed. In the proposed method, multiple skin color models are tuned by a learning based on a concept of self-organizing adaptive controller. The skin color models for various lighting conditions can be obtained from small number of images. The effectiveness of the proposed method is verified by simulation results.**

I. Introduction

Recently, human-machine interfaces using some parts of bodies such as face recognition and gesture recognition are paid attention. In these studies, detection of pixels with skin color is widely used for region extraction of objects, because skin color does not greatly vary from individual to individual. Color of hands and faces, however, changes depending on lighting condition, and the simplest method with fixed threshold can not achieve stable detection of pixels with skin color. To cope with this problem, many extended methods, in which the thresholds were adaptively adjusted, were reported. Cho et al. proposed an adaptive thresholding technique in HSV color space[2]. Furthermore, the methods that skin color was estimated based on skin color models constructed with Gaussian models were proposed. Yang and Waibel used a gaussian mixture model (GMM) for skin modeling in *rg* space[3]. The mean and covariance of the model as the illumination changes dynamically are approximated using a liner combination. But, in these methods, a lot of images taken in various kinds of conditions are needed to realize the stable detection.

In this paper, a new skin color detection method which can achieve stable detection adjusted using small number of imaged is discussed. The proposed method is based on multiple skin color models constructed with radical basis function (RBF) network which can perform stable learning and the appropriate skin color model for inputted image is selected. We assume that the skin color models should continuously change in accordance with change of lighting condition. To construct multiple skin color models, a concept of a self-organizing adaptive controllers (SOAC)[4], which is an extension of modular network self-organizing map (mnSOM), is employed. By using SOAC, various skin models can be constructed in self-organization from learning images of some lighting conditions, and stable detection of skin color in various conditions can be expected.

II. Related Techniques

In this section, SOAC and color space used in this study is briefly introduced.

A. Self-Organizing Adaptive Controller

SOAC is one of extensions of mnSOM. The mnSOM which is a general model of self-organizing map (SOM) consists of arrayed functional modules. Each functional module can be represented with various models such as multi-layered perceptron. A learning of the modules are achieved in a self-organizing manner, *i.e.* iteration of evaluation, competition, cooperation and adaptation. In the SOAC, the functional module includes a pair of a predictor and a controller. The pair of the predictor and the controller are specialized for the special case, and adaptive control can be achieved by adequately selecting the module. Here the prediction is used as a switcher of controllers to select the appropriate controller for the present system under control. One of the characteristics of SOAC is a function of interpolation. Such mnSOM can do that. That is, SOAC learned using some learning data can discover the system corresponding to the hidden data in self-organization. Thus the framework and the characteristics of the SOAC coincide with requirement of the skin color detection with multiple skin color models.

Tani et. al. proposed recurrent neural network with parametric biases (RNNPB)[5]. RNNPB can generate the model according to parametric bias (PB) by adding the recurrent neural network. Function of RNNPB is similar to that of an mnSOM with MLP-modules. RNNPB processes PB value and control value by single neural network, thus generalization ability declines.

B. YCrCb Color Space

YCrCb color space is one of frequently used color space for skin color detection, because it is robust for change of lighting condition[6]. Let components of RGB color be R, G and B, respectively. Actually the components of YCrCb can be obtained by:

$$\begin{aligned} Y &= 0.29891R + 0.58661G + 0.11448B, \\ Cr &= 0.50000R - 0.41869G - 0.08131B, \\ Cb &= -0.16874R - 0.33126G + 0.50000B. \end{aligned} \tag{1}$$

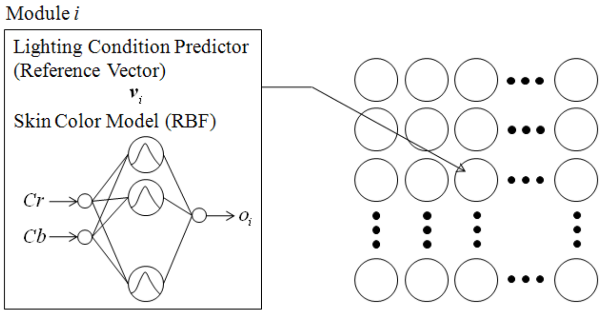

Fig. 1. Outline of the proposed ASCD.

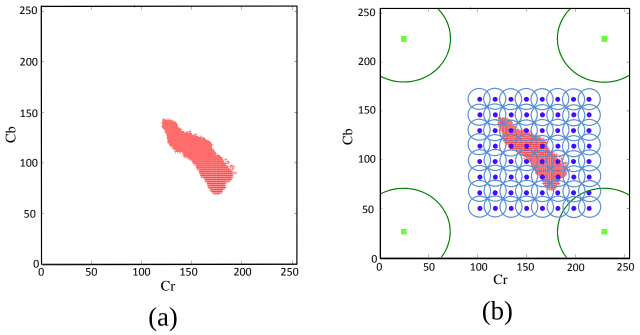

Fig. 2. (a) Distribution of skin color in Cr-Cb space of all learning images. (b) Center points and widths of basis of RBF. Blue and green points display the centers of basis of RBF and blue and green circles display the widths of those.

Y means brightness of the pixel, and Cr and Cb are color differences. Cr and Cb are usually employed to detect skin color. In this study, Cr and Cb is used in accordance with the past studies.

III. Adaptive Skin Color Detection with Multiple Skin Color Models

In this section, the proposed skin color detection method called as adaptive skin color detection with multiple skin color models (ASCD) is introduced.

A. Overview

The outline of the proposed ASCD is shown in Fig. 1. A basic component of the proposed method is a module which consists of a pair of a lighting condition predictor and a skin color model. The predictor is just a reference vector $\boldsymbol{v}_i \in R^2$, and the skin color model is constructed with RBF.

A target image is given to ASCD, and a feature vector $\boldsymbol{y} \in R^2$, called as key vector, is calculated by averaging S and V components in HSV color space for all pixel of the image. The key vector is compared to the reference vector of all the modules, and the best matching module c is decided based on minimum distance between $\boldsymbol{y}$ and $\boldsymbol{v}_i$. The skin color model of the best matching module is selected as the skin color model for the given image. In the skin color model, inputs are Cr and Cb of each pixel, and an output is whether the pixel is skin color or not, *e.g.* O_c is 1 or 0 for skin color or not, respectively.

B. Learning of ASCD

To tune the parameters of the ASCD such as reference vectors and weights in RBF, learning is needed. As mentioned above, the learning is achieved in self-organizing manner. The processes of evaluation, competition, cooperation and adaptation are iteratively done. Before the learning N images, of which all the pixels are known as whether skin color or not, are prepared. The reference vectors of predictors are randomly initialized, and all weights in all RBFs are also initialized with random value of [0:1]. Center points and widths of basis of RBFs are fixed for that the bases cover the area in Cr-Cb space where pixels with skin color exist as shown in Fig. 2.

1) Evaluation Process: Image n are picked up from the image set, and all the pixels of the image are transformed to HSV color spaces. The mean of S and V components for all the pixels are calculated to get the key vector $\boldsymbol{y}^{(n)} = [\bar{S}, \bar{V}]^T$, and the key vector is inputted to all the module. The distance between the key vector and the reference vectors are calculated by

$$E_i^{(n)} = \parallel \boldsymbol{y}^{(n)} - \boldsymbol{v}_i \parallel . \qquad (2)$$

The above processes for all the images are achieved.

2) Competition Process: The best matching module $c^{(n)}$ for image n is defined by

$$c^{(n)} = \arg\min_i E_i^{(n)}. \qquad (3)$$

The above processes for all the images are achieved.

3) Cooperation Process: Amount of learning for each module should be determined based on position of the best matching modules and arrangement of the modules. The amount of learning for module i for image n is given by

$$\Psi_i^{(n)} = \frac{\phi(d_{i,c^{(n)}};t)}{\sum_{i=1}^{M} \phi(d_{i,c^{(n)}};t)}, \qquad (4)$$

where, $d_{i,c^{(n)}}$ means distance between module i and the best matching module $c^{(n)}$ on the module array, and $\phi(d(i,c^{(n)}))$ is called as neighborhood function at learning step t given by

$$\phi(d_{i,c^{(n)}};t) = \exp\left(-\frac{d_{i,c^{(n)}}}{2\sigma^2(t)}\right), \qquad (5)$$

where, $\sigma(t)$ represents a width of the neighborhood function, and monotonically decrease with learning progresses.

4) Adaptation Process: Based on the amount of learning for each module, the reference vector of the predictor and weights of the RBF are updated. The weights j of the RBF of the module i is updated by

$$w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}, \qquad (6)$$

TABLE I
SUBSTANCE OF IMAGES USED IN THE SIMULATION.

Case	Time	Place	Remarks	Learning	Testing
1	Daytime	Indoor	No light	10	5
2	Nighttime	Indoor	Flour lamp	10	5
3	Nighttime	Indoor	Spotlight	10	5
4	Daytime	Outdoor1	Fine Whether	10	10
5	Nighttime	Outdoor1	Street lamp	10	6
6	Daytime	Outdoor2	Fine Whether	0	9
7	Nighttime	Outdoor2	Street lamp	0	5

where

$$\Delta w_{i,j} = -\eta \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{n=1}^{N} \Psi_i^{(n)}(O_i^{n,p,q} - T^{n,p,q})h_{i,j}(\boldsymbol{x}^{n,p,q}), \tag{7}$$

where, P and Q are horizontal and vertical sizes of image, respectively. $O_i^{n,p,q}$ and $T^{n,p,q}$ are output of the RBF of the module i and required output for pixel (p,q) of the image n, respectively. And $h_{i,j}(\boldsymbol{x}^{n,p,q})$ is the output of the basis function j of the module j for pixel (p,q) of image n, and η is a learning rate. This update equation means that the weights are updated based on gradient descent methods, and amount of update is proportional to $\Psi_i^{(n)}$.

The reference vector of the predictor of the module i is updated by the same equation of the batch learning of the SOM by

$$\boldsymbol{v}_i = \sum_{n=1}^{N} \Psi_i^{(n)} \boldsymbol{y}^{(n)} \tag{8}$$

C. Skin Color Detection by ASCD

When a target image is given to the ASCD, all pixels of the image are transformed to HSV color space, and means of S and V components for all pixels are calculated to obtain the key vector $\boldsymbol{y}^* = [\bar{S}^*, \bar{V}^*]^T$. The key vector is compared to the predictors of all the modules, and the module c^* which has the closest reference vector $\boldsymbol{v}_{c^*}$ to $\boldsymbol{y}$ is set as the best matching module. The skin color model of the best matching module c^* is used for predction of skin color of the image. A pixel (p,q) of the target image is transformed to YCrCb color space, and the components of Cr and Cb, represented as $\boldsymbol{x}^* = [Cr^*_{p,q}, Cb^*_{p,q}]^T$, are inputted to the RBF of the module c^*. The output of the RBF $O_{c^*}^{*,p,q}$ is obtained. If $O_{c^*}^{*,p,q}$ is larger than the threshold Th, the pixel (p,q) is estimated as skin color. On the other hand, the pixel (p,q) is judged as background, if $O_{c^*}^{*,p,q}$ is smaller than the threshold Th.

IV. SIMULATION RESULTS

To verify an effectiveness of the proposed method, simulation of skin color prediction in various illumination conditions was achieved.

A. Data Set

Table I shows a substance of images used in this simulation. Conditions that photographs are taken are roughly classified into 7 categories as shown in Table I. In the table, learning and testing mean numbers of images used for learning of skin color model and testing. Each image includes hand which should be detected, and Fig. 3 shows some examples of images.

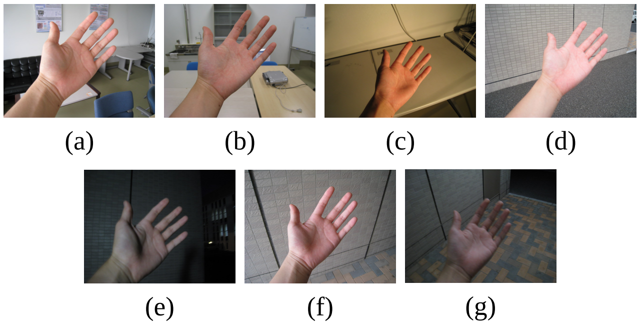
(a) (b) (c) (d) (e) (f) (g)

Fig. 3. Some examples of images taken in various conditions. (a) to (g) correspond to case 1 to 7, respectively.

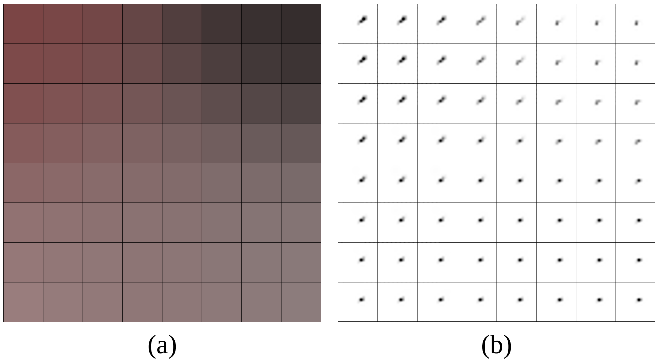
(a) (b)

Fig. 4. (a) Reference vector of the predictors, the modules are colored based on S and V components. (b) outputs of the RBFs of modules. In each module, horizontal and vertical axes are Cr and Cb components, respectively, and black and white correspond to 1 and 0, respectively.

B. Learning Results of ASCD

The learning of ASCD was achieved using 50 images as shown in Table I. In the learning, number of modules were 8 × 8. Fig. 4 (a) shows the reference vectors of the predictors, and each module is colored based on the S and V components of the predictor. Fig. 4 (b) shows the output of the RBF of each module. In each module, horizontal and vertical axes mean the Cr and Cb components, respectively, and black and white correspond to 1 and 0, respectively. It is shown the input/output characteristics of the RBFs, *i.e.* skin color models, continuously vary on the map. Fig. 5 (a) shows the arrangement of the best matching modules for images used for the learning. It is shown that the similar images make the modules located near each other the best matching modules.

C. Skin Color Detection

To confirm the effectiveness the proposed ASCD, skin color detection simulation using 45 images that were not used for learning of ASCD was achieved. To evaluate the results, some variables are defined as shown in Table II. Four criteria are used to evaluate the performance of the skin color detection, and Table III shows the results, and these values are average

TABLE III
SKIN COLOR DETECTION RESULTS.

		Case 1		Case 2		Case 3		Case 4		Case 5		Case 6		Case 7	
Th	Criterion	ASCD	Fixed	ASCD	Fixed	ASCD	Fixed	ASCD	Fixed	ASCD	Fixed	ASCD	Fixed	ASCD	Fixed
0.1	Ss/Rs	0.99	0.95	0.81	0.69	0.27	0.23	0.90	0.92	0.22	0.98	0.97	0.93	0.21	0.70
	Ss/Ns	0.95	0.97	0.99	1.00	1.00	0.70	0.98	0.99	1.00	0.40	0.96	0.98	1.00	0.78
	Sb/Rs	0.01	0.05	0.19	0.31	0.73	0.77	0.10	0.08	0.78	0.02	0.03	0.07	0.79	0.30
	Sb/Nb	0.05	0.03	0.01	0.00	0.00	0.30	0.02	0.01	0.00	0.60	0.04	0.02	0.00	0.22
0.3	Ss/Rs	1.00	0.95	0.84	0.69	0.41	0.23	0.99	0.92	0.23	0.98	1.00	0.93	0.24	0.70
	Ss/Ns	0.91	0.97	0.98	1.00	0.91	0.70	0.96	0.99	1.00	0.40	0.88	0.98	0.98	0.78
	Sb/Rs	0.00	0.05	0.16	0.31	0.59	0.77	0.01	0.08	0.77	0.02	0.00	0.07	0.76	0.30
	Sb/Nb	0.09	0.03	0.02	0.00	0.09	0.30	0.04	0.01	0.00	0.60	0.12	0.02	0.02	0.22
0.3	Ss/Rs	1.00	0.95	0.93	0.69	0.73	0.23	1.00	0.92	0.28	0.98	1.00	0.93	0.30	0.70
	Ss/Ns	0.86	0.97	0.97	1.00	0.76	0.70	0.93	0.99	0.99	0.40	0.81	0.98	0.90	0.78
	Sb/Rs	0.00	0.05	0.07	0.31	0.27	0.77	0.00	0.08	0.72	0.02	0.00	0.07	0.70	0.30
	Sb/Nb	0.14	0.03	0.03	0.00	0.24	0.30	0.07	0.01	0.01	0.60	0.19	0.02	0.10	0.22

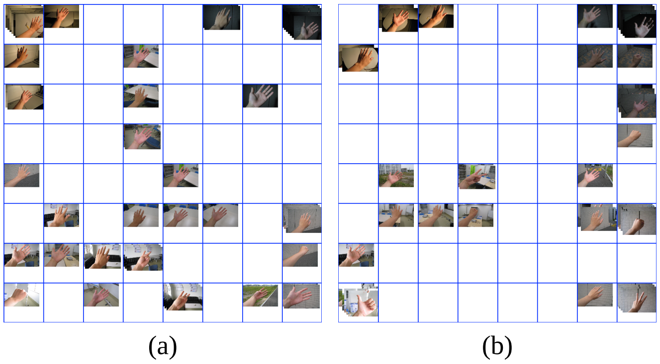

Fig. 5. Arrangement of the best matching modules for images used for learning (a) and testing (b).

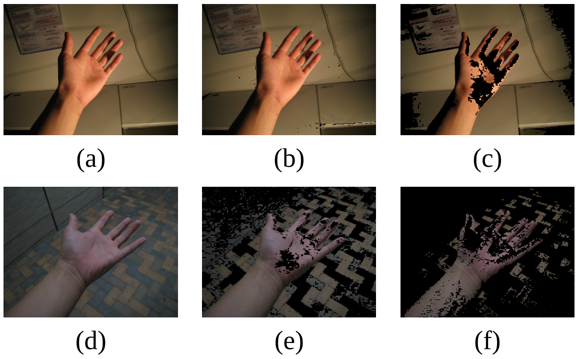

Fig. 6. Some examples of skin color detection. (a) Inputted image of case 3. (b) Result by the ASCD. (c) Result by the fixed threshold. (d) Inputted image of case 5. (e) Result by the ASCD. (f) Result by the fixed threshold.

TABLE II
VARIABLES USED FOR EVALUATION OF THE PERFORMANCE.

Ns	# of pixels of skin color (actual)
Nb	# of pixels of non skin color (actual)
Rs	# of pixels estimated as skin color
Rb	# of pixels estimated as non skin color
Ss	# of pixels correctly estimated as skin color
Bs	# of pixels correctly estimaged as non skin color
Sb	# of pixels of skin color estimated as non skin color
Bb	# of pixels of non skin color estimated as skin color

for images included in each class. In this Table, 'Fixed' means the skin color detection method with fixed thresholds ([133:173] for Cr and [77:127] for Cb components)[6], and Th is the threshold for output of the skin color model. First two criteria Ss/Rs and Ss/Ns indicate correct detection rates. On the other hand, last two criteria Sb/Rs and Sb/Nb represent the wrong detection rates. From Table III, it is shown that skin color detections for cases 1, 2, 3, 4 and 6 are superior to those of the fixed threshold. Especially, the wrong detection rates are improved. The fixed threshold is widely set to detect the skin color in various kinds of lighting conditions. The first criterion Ss/Rs becomes small, because non skin color in the background is picked up as the skin color. Fig. 6 (a) to (c) show examples of skin color detection of an image of case 3. Spotlight used in this simulation is yellowish, and it is difficult to detect the skin color with high accuracy using fixed threshold. On the other hand, the proposed method can select the skin color model adequately, then the accurate detection can be achieved. However, for images of cases 5 and 7, the ASCD is inferior to the fixed threshold. Fig. 6 (d) to (f) show examples of skin color detection of an image of case 7. The reason is thought as follows. The distribution of the pixels in CrCb space since the images of cases 5 and 7 are very dark. In such case, the boundary between skin color area and non skin color area might be sensitive. The arrangement of the bases of RBF should be reconsidered.

V. CONCLUSIONS

In this paper, we proposed the ASCD which selects the adequate skin color model for the inputted image and detects skin color with high accuracy. The skin color models can be obtained by the learning which is based on the concept of the SOAC.

The averages of S and V components for all pixels to select the skin color model, and the results are strongly influenced by the background. The construction of the predictor should be considered in the future work. RBF is employed as the skin color model. The simulation results indicate a limitation of RBF, and it is also future work that a new skin color model is considered.

References

[1] P. Kakumanu, S. Makrogiannis, N. Bourbakis. ”A survey of skin-color modeling and detection methods”, *Pattern Recognition*, Vol.40, pp.1106-1122, 2007.

[2] K.M. Cho, J.H. Jang, K.S. Hong, ”Adaptive skin-color filter”, *pattern recognition* Vol.34 (5), pp.1067-1073, 2001.

[3] J.Yang, A. Waibel, ”A real-Time Face Tracker”, *Proceedings of the third IEEE Workshop on Applications of Computer Vision (WACV96)*, pp.142-147, 1996

[4] T. Minatohara, T. Furukawa. ”Self-Organizing Adaptive Controllers: Application to the Inverted Pendulum”, *Proc. of 5th Workshop on Self-Organizing Map (WSOM’05)*, 2005.

[5] J. Tani, M. Ito, Y. Sugita, ”Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB”, *Neural Networks, Volume 17, pp. 1273-1289,* ,2004.

[6] D.Chai, K.N. Ngan, ”Face segmentation using skin-color map in videophone applocations”, *IEEE Trans. Circuits Syst. Video Technol*, Vol.9 (4), pp. 551-564, 1999.