# FACE DETECTION USING SKIN-COLOR MODEL AND SUPPORT VECTOR MACHINE

*Yoko Seki, Ichiro Yuyama, Hiroshi Hasegawa and Yu Watanabe*

Department of Information Systems Science
Graduate School of Engineering, Utsunomiya University
7-1-2 Yoto, Utsunomiya, Tochigi, Japan
E-mail: (seki@durian., yuyama@, hasegawa@, yu@)is.utsunomiya-u.ac.jp

## ABSTRACT

In this paper, we propose a face detection technique for still pictures which sequentially uses a skin-color model and a support vector machine (SVM). SVM is a learning algorithm for solving the classification problem. Some studies on face detection have reported superior results of SVM over neural networks. The SVM method searches for a face in a picture while changing the size of the window. The detection accuracy and the processing time of SVM vary largely depending on the complexity of the background of the picture or the size of the face. Therefore, we apply a face candidate area detection method using a skin-color model as a preprocessing technique. We compared the method using SVM alone with that of the proposed method in respect to face detection accuracy and processing time. As a result, the proposed method showed improved processing time while maintaining a high recognition rate.

**Keywords:** face detection, skin-color model, support vector machine

## 1. INTRODUCTION

Face detection technology was initially used for consumer electronics, security equipment, and robot interfaces. Several face detection techniques have been proposed [1] and are broadly classified into feature-based approaches and image-based approaches. Most of the feature-based approaches detect a face by using the facial features of the eyes, nose and mouth in the face candidate area. The face candidate area is calculated by using the skin-color model [2][3] or stereo systems [4][5]. It is easy for the facial features used to be affected by a beard, the hairstyle, or the orientation the face, and as such consistent face detection is difficult. On the other hand, in most image-based approaches, a face is detected by image scanning with a scanning window which has the step size, scale-up rate and the number of scale-up iterations set beforehand. This is known as the window scanning technique. However, since the scanning window scans an image several times, the processing time increases.

Support Vector Machine(SVM) [6][7][8] is a learning algorithm for solving the classification problem, and one of the image-based approaches applied to face detection. SVM decides a boundary which divides the face class and non-face class by learning face images and non-face images, and distinguishes which class an input image belongs to using the window scanning technique. Some studies on face detection have reported superior results of SVM over neural networks [9]. To detect faces correctly in an image with various backgrounds, SVM learns from many non-face images. However, it is difficult to learn all of the possible backgrounds. Therefore, when there are not enough images to learn from, the number of false detections increases. In the window scanning technique, the step size, scale-up rate and the number of scale-up iterations must be set beforehand. However, because SVM cannot detect a face which is larger or smaller than the scanning window, the step size, scale-up rate and the number of scale-up iterations must be set to the size of the face in an image. In automatic face detection, to detect all of the faces in an image, a small step size, small scale-up rate and large number of scale-up iterations are set. The processing time increases because the number of scale-up iterations is large. In order to improve these problems, we propose a technique which uses a skin-color model as preprocessing.

## 2. PROPOSED METHOD

### 2.1 Detection of the Face Candidate Area

Figure 1 shows the flow of the proposed face detection method. The area where pixels are evaluated as skin-color by a skin-color model is detected as a face candidate area. This process uses only color information. Therefore, the processing time is short. Several kinds of skin-color models evaluate pixels by setting an appropriate threshold in the color space, using Gaussian distribution or calculating a histogram using sample images. To set an appropriate threshold in the color space, upper and lower thresholds are set, and a distribution of skin-color is considered to be a rectangle area in the color space. When using the Gaussian distribution method, a distribution of skin-color is considered to be an oval area in the color space, but in reality human skin-color is not distributed in the shape of an rectangle or oval. When using the method of calculating a histogram, skin-color distribution is not limited, so we adopt the histogram model in this work. The skin-color model which we propose has a common distribution histogram and a total distribution histogram. The common distribution histogram expresses the distribution of the skin-color pixel values which some sample face images show in common.

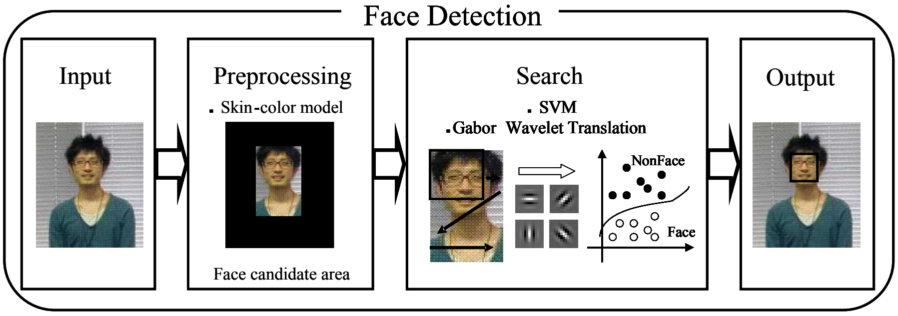



Fig.1: The face detection flow

The total distribution histogram expresses the distribution of all of the skin-color pixel values which some sample images show. To detect the face candidate area, the corresponding pixels to common distribution are detected from an image. The detected pixels are set as the skin-color area. The scanning window searches the image from the upper left to the lower right. In the process, pixel values of eight neighboring pixels of the identified skin-color area are compared with the pixel values in the total distribution. When those correspond, the pixel is integrated as the skin-color area. The rectangle that comprises the detection skin-color area is considered to be the face candidate area.

## 2.2 Face Detection Using SVM

SVM is applied to the detected face candidate area. The size of the scanning window is automatically set according to the size of the face candidate area. Consequently, we achieve consistent face detection and a shorter processing time. SVM basically is a linear identification machine, and decides an identification super-plane that maximizes the margin. The margin is the minimum distance from the super-plane and the learning data.

When input space $\chi \in \boldsymbol{R}^n$ and a learning data set $S = ((\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_r, y_r))$ are given, the identification function of linear SVM is expressed by Eq. (1).

$$f(\boldsymbol{x}) = \sum_{i=1}^{r} \alpha_i y_i \left(\boldsymbol{x}^T \boldsymbol{x}_i\right) + b \quad \text{------------ (1)}$$

Here, $\alpha_i$ is the solution of the quadratic programming problem, and $b$ is the bias. These are decided by learning. Many $\alpha_i$ will become 0. The learning data with $\alpha_i$ other than 0 is called a support vector.

When two classes can be separated linearly, linear SVM will show a high recognition rate. However, there are generally few cases that can be separated linearly. Thus, nonlinear SVM, which uses map $\phi(\boldsymbol{x})$, was devised. Nonlinear SVM maps the learning data to the higher dimension feature space, and identifies in the space. A known kernel function $k(\boldsymbol{x}, \boldsymbol{x}_i)$ is used to map the learning data, which reduces the frequency of the calculations of the inner product. The function meets the conditions of the Mercer kernel. The identification function of nonlinear SVM is expressed by Eq. (2).

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i=1}^{r} \alpha_i y_i \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}_i) + b \\ &= \sum_{i=1}^{r} \alpha_i y_i k(\boldsymbol{x}, \boldsymbol{x}_i) + b \end{aligned} \quad \text{------------ (2)}$$

We use the Gabor wavelet transform image as an input of SVM. The image is obtained by the convolution of the Gabor kernel and an image of the face candidate area, and shows the character of the simple cells in the primary visual cortex. The cells detect the direction and the width of the outline on the retina. The feature obtained is effective in recognizing faces and 3-dimensional objects. In the following experiment, we treat the feature vector of 3600 dimensions. These vectors were obtained by using the Gabor wavelet transform in 4 directions for an area of 30x30 pixel. Moreover, we use the RBF kernel as a function for the map.

## 3. EXPERIMENT

In order to show the effectiveness of the proposed method, we compared the method using SVM alone with that of the proposed method with respect to face detection accuracy and processing time.

## 3.1 Development of Skin-Color Model

In order to develop the skin-color model, we calculate the histogram from some sample images. First, we take some photographs of the faces of 60 Japanese men and women in their twenties. Optical reflection was suppressed by using a black cloth as background and as a cover for the body. Next, the skin-color area of the images is detected. The parts without skin-color such as the eyes and mouth are painted out black, and remaining area is assumed to be the skin-color area. The common distribution histogram and the total distribution histogram are calculated from pixel values of the skin-color area. The shape of the distribution in the histogram is different according to which color space is used. We experimented to determine which color space is more effective for the proposed method. We calculated skin-color models by using RGB color space, XYZ color space, L*a*b* color space, YCbCr color space, YIQ color space, HSV color space, and Modified HSV color space. By comparing the results of the face area detection by using all of the models, the RG model and XZ model were selected for use.

## 3.2 Images for the Experiment

We used 1500 30x30 pixel face images of and 2000 30x30 pixel non-face images for the learning of SVM. The face images contain the facial features of the eyes, nose, and

Table 1: Group of non-face images

| Group | Plant (green) | Plant (not green) | Room | Town | Road | Waterside | Building | Night view | Animal | Food | Vehicle |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Contents | Tree, Lawn, etc. | Colored leaves , Flower, etc. | Furniture, Store's interior, etc. | Town | Road | Sea, River, etc. | Building House, etc. | Night view, Illumination | Dog, Cat, etc. | Vegetable, Fruit, etc. | Car, Ship, etc. |

mouth. In addition, for verification, 100 30x30 pixel face images of and 1100 350x230 pixel (230x350 pixel) non-face images were used. The images for learning are different from the images for verification. We classified the non-face images for verification into the 11 groups shown in Table 1.

There are two criteria for face detection. One is True Positive Rate (TPR) which identifies the face area as a face class. TPR is obtained from the number of images identified as face class in the face images for verification. The other criterion is False Positive Rate (FPR) which identifies the non-face area as a face class. When SVM is used alone, the scanning window scans five times while resizing the image size to 1/1.2 in order to detect a face of a different size. Therefore, 10,584,800 non-faces are identified in 100 images. FPR is obtained by regularizing the number of the misidentified images by 10,584,800. In the proposed method, SVM is applied to the face candidate area. Therefore, the number of non-face identification decreases, but similar processing is done for the comparison. We used a PC with Pentium 4 3GHz CPU in our experiment.

## 3.3 Results

Figure 2 shows the results of TPR and processing time.

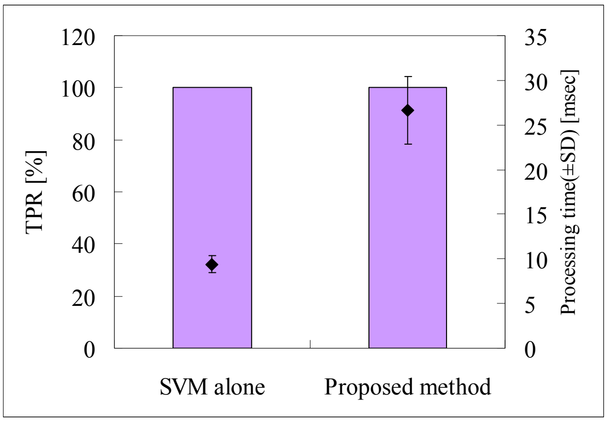



Fig. 2: TPR and processing time obtained from the number of images is identified as a face class in the face images for the verification. The bar graph indicates TPR, and the two diamonds indicate the processing time.

The value of TPR was 100% with both methods. Even though some images of faces wearing glasses, smiling faces, and serious faces were used, we were able to obtain good results from the face detection. The processing time of the proposed method was 17 milliseconds slower than the processing time of SVM alone. Figure 3 shows the FPR of the two methods.

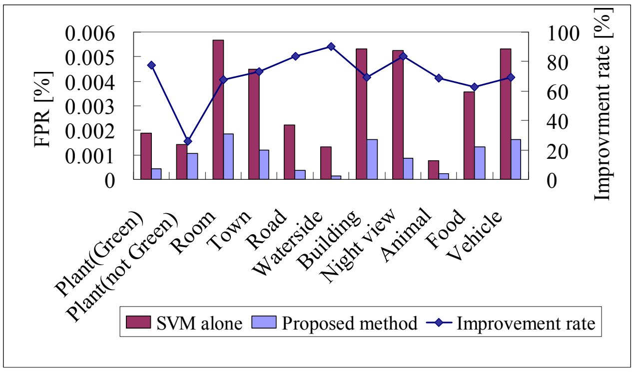



Fig. 3: FPR obtained by regularizing the number of misidentified images by 10,584,800. The bar graph indicates TPR, and the line graph indicates the improvement rate.

In every group, the FPR of the method using SVM alone was larger than that of the proposed method, and the improvement rate was about 26 to 90%. With the method using SVM alone, a difference was apparent in the FPR. A large FPR means that many areas are detected as having features similar to that of a human face, whereas a small FPR means there are few areas detected as having features similar to the human face. On the other hand, in the proposed method, all the FPR were less than 0.002%. For example, most of the images in the Waterside group were colored blue, having a small face candidate area. In the case that the image had a small face candidate area and little area with features similar to the human face, the FPR was especially low with the proposed method. Figures 4 and 5 show the processing time per image.

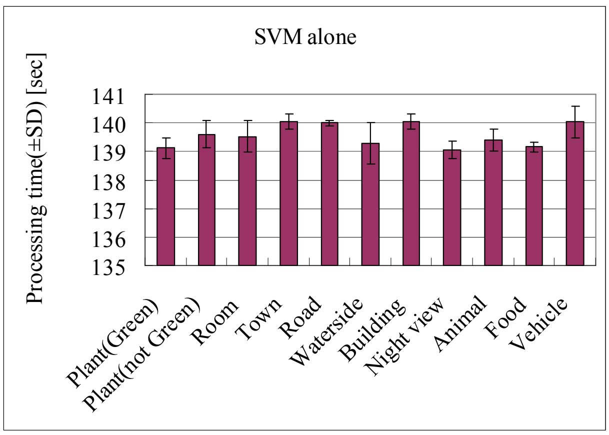



Fig. 4: Processing time per image with the method using SVM alone.

When the method using SVM alone was used, the processing time was 138-141 seconds. Because the window scanning technique is used in SVM, the step size, scale-up rate and the number of scale-up iterations are set beforehand, so the difference in processing time was small.

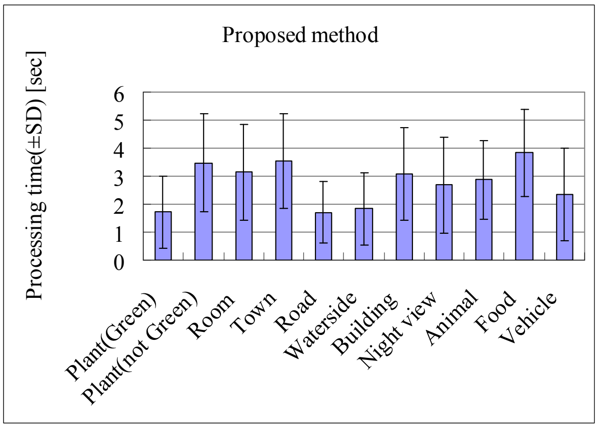
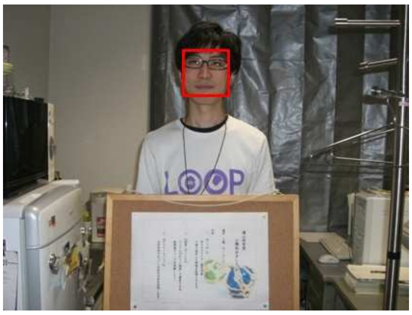
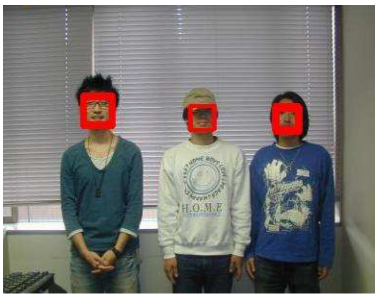



Fig. 5: Processing time per image in the proposed method.

On the other hand, when the proposed method was used, the processing time was 0.5 to 6 seconds. Because neither the step size, scale-up rate nor the number of scale-up iterations are set beforehand, and the window scanning technique was applied only to the face candidate area. When the size of the face candidate area is different, the processing time will also be different. When comparing the two processing times, the deletion of an area unnecessary to the preprocessing led to a shortening of the total processing time. However, it takes a maximum processing time of 6 seconds and if the proposed technique is to be applied to a -time system and a dynamic scene, further improvement of the processing time is required. Figure 6 shows the face detection results from the images.




(a) Face detection for 1 person



(b) Face detection for three people

Fig. 6: Examples of face detection

Both images show good results. In the 420x320 pixel image (a) s, it took the processing time of about 3 seconds. In the image of (b) of 390x300 pixels, it took the processing time of about 4 seconds.

## 4. CONCLUSION AND FUTURE WORK

We proposed a face detection method to improve detection accuracy and processing time. We compared the method using SVM alone with that of the proposed method in respect to face detection accuracy and processing time. As a result, the proposed method showed an improvement in FPR from about 26 to 90% without decreasing the value of TPR. Although the processing time of the method using SVM alone is 138 to 141 seconds, the processing time of the proposed method is 0.5 to 6 seconds. Because the size of the scanning window was automatically set according to the size of the face candidate area in the images with various backgrounds, the area of input of the SVM was limited, and the FPR and processing time were improved. We have thus demonstrated that the proposed method is effective for face detection.

We will examine the relationship between the number of dimensions and the processing time of the amount of features used, and will work to improve the processing time in the future.

This research was supported in part by the Hoso Bunka Foundation, Inc. in Japan.

## 5. REFFERENCES

[1] Erik Hjelmas and Boon Kee Low, "Face Detection: A Survey," Computer Vision and Image Understanding, Vol. 83, No.3, pp. 236-274, 2001.

[2] Kawato Shinjiro and Tetsutani Nobuji, "Skin-color Model and Face Region Extraction," Technical Report of IEICE, PRMU2001-59, pp. 143-148, 2001.

[3] Kawato Shinjiro and Tetsutani Nobuji, "Two-step Approach for Real-time Eye Tracking," Technical Report of IEICE, PRMU2000-63, pp. 15-22, 2000.

[4] Sang-Hoon Kim, Nam-Kyu Kim, Sang Chul Ahn, and Hyoung-Gon Kim, "Object Oriented Face Detection Using Range and Color Information, " Proc. IEEE 3rd Int. Conf. on Automatic Face and Gesture Recognition, pp. 76-81, 1998.

[5] Darrell T, Gordon G. Harville M. and Woodfill J., "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," International Journal of Computer Vision, Vol. 37, No. 2, pp. 175-185, 2000.

[6] Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," Proc. 5th Annual ACM Workshop on Computational Learning Theory, pp. 144-152, 1992.

[7] Edgar Osuna, Robert Freund and Federico Girosi, "Training Support Vector Machines: an Application to Face Detection," Proc. IEEE. Conf. on Computer Vision and Pattern Recognition, pp. 130-136, 1997.

[8] Kazuhiro Hotta, "A Robust Face Detection Method under Partial Occlusion, " IEICE Trans. D-II, Vol. J88-D-II, No. 8, pp. 1624-1633, 2005.

[9] B. Schölkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio and V. Vapnik, "Comparing Support Vector Machines with Gaussian Kernels to Radial Basis Function Classifiers," IEEE Trans. Signal Processing, Vol. 45, No. 11, pp. 2758-2765, 1997.