Using HVS Color Space and Neural Network for Face Detection with Various Illuminations

CHIUNHSIUN LIN
National Taipei University
69 Sec. 2 Chian_Kwo N. Road, Taipei, Taiwan, 10433
TAIWAN

CHING-HUNG SU
The Department of Electronic Engineering,
Lunghwa University of Science and Technology,
No.300, Sec. 1, Wanshou Rd., Gueishan Shiang, Taoyuan County, Taiwan, 33306
TAIWAN

Abstract: - We introduce an efficient face detection system that can detect multiple faces in color images with various illuminations. The proposed system consists of two parts. The first part utilizes color and triangle-based segmentation to search the potential face regions. The second part performs face verification task by a multilayer neural network. The system can conquer distinct size, changed lighting condition, various pose, orientation, facial hair, and partial occlusion, presence of glasses, and diverse hair styles at the same time. Moreover, the system significantly speeds up the execution time of face detection with complex backgrounds. The experimental results reveal that the proposed method is better than traditional methods in terms of altered circumstance and efficiency.

Key-Words: - Face detection; Skin color segmentation; Triangle-based segmentation; Neural network

1 Introduction

Automatic face detection and recognition of human faces is one of the most intricate and important problems in computer vision. It can be used as the security mechanism to replace metal key, plastic card, and password or PIN number. However, most face detection and recognition systems require the number of the input face to be free of environment and the size to be approximately consistent. Moreover, they cannot tolerate varying pose, expression and lighting at the same time. These constraints greatly hinder the usefulness of the system. The key distinction between face localization and face detection centers on the number of faces in an image. Face localization is to locate the only one human face embedded in an image. Face detection is to detect one or more human faces embedded in an image. Abundant of researches have been conducted on human face detection. Some successful systems have been proposed in the literature, such as [1-12]. Rowley [11] presented a neural network-based face detection system by using a retinal connected neural network to check the small windows of an image, and judge whether each window contains a face or not. They adopted a small window (20 * 20) to slide over all portions of an image at various scales and used oval mask for ignoring background pixels. Their system arbitrates between multiple networks to improve the performance over a single network. However, the inefficient search is a time-consuming procedure. Lin [12] presented a triangle-based approach system that can locate multiple faces embedded in complicated backgrounds. However, when the condition is complicated, their system becomes slow and time-consuming.

Our designed system is shown as Fig. 1. The first part is to search the potential face regions. The second part is to perform the task of face verification for each potential face region. The rest of the paper is organized as follows. In section 2, segmentation of potential face regions based on HSV color space and triangle-based approach is described. In section 3, each of the normalized potential face regions is fed to the neural network to verify whether the potential face region really contains a face. Experimental results are demonstrated in section 4 to verify the validity of the proposed face detection system. Finally, the conclusions are given in section 5.

2 Searching for Potential Face Regions

The main purpose of this process is to find the regions in an input image that might potentially contain faces.

The process consists of four steps. First, read in a color image. Then, convert this input color image to a binary image using HSV color segmentation. Second, label all 4-connected components in the image to form blocks and find out the center of each block. Third, detect any 3 centers of 3 different blocks that form an isosceles triangle. Fourth, clip the blocks that satisfy the triangle criteria as the potential face regions.

Because the RGB color space is very sensitive to intensity difference, many color spaces have been proposed to get better color consistency or color segmentation. In our system, we utilize the HSV color space instead of the RGB color pace. The HSV color space [13] was chosen because: (1) In the HSV color space, the brightness component (V) is independent factor, so we can use this characteristic to conquer the problem of illumination variation. (2) The HSV color model over alternative models such as RGB or CMYK, because of its similarities to the way humans tend to perceive color. The HSV (Hue, Saturation, and Value) model, also called HSB (Hue, Saturation, and Brightness), defines a color space in terms of three constituent components: Hue — the color type (such as red, blue, or yellow): Ranges from 0 to 360 (but normalized to 0 to 100% in some applications). Saturation — the "vibrancy" of the color: Ranges from 0 to 100%, and occasionally is called the "purity". The lower the saturation of a color, the more "grayness" is present and the more faded the color will emerge. Value — the brightness of the color: Ranges from 0 to 100%. The HSV model was created in 1978 by Alvy Ray Smith. The HSV model can be present by a cone. In this representation, the hue (H) is depicted as a three-dimensional conical formation of the color wheel. The saturation (S) is represented by the distance from the center of a circular cross-section of the cone, and the value (V) is the distance from the pointed end of the cone. The HSV model is shown as Fig. 2.

The HSV color space is a nonlinear transformation of the RGB color space, and it can be transformed from RGB coordinates. Given a pixel of color defined by (R, G, B) where R, G, and B are between 0.0 and 1.0, with 0.0 being the least amount and 1.0 being the greatest amount of that pixel of color. An equivalent (H, S, V) color can be determined by a series of formulas. Let MAX equal the maximum of the (R, G, B) values, and MIN equal the minimum of those values. The formula can then be written as:

$$H = \begin{cases} \left(0 + \frac{G-B}{MAX-MIN}\right) \times 60, & \text{if } R = MAX \\ \left(2 + \frac{B-R}{MAX-MIN}\right) \times 60, & \text{if } G = MAX \\ \left(4 + \frac{R-G}{MAX-MIN}\right) \times 60, & \text{if } B = MAX \end{cases}$$

$$S = \frac{MAX - MIN}{MAX}$$

$$V = MAX$$

If S = 0 then the color lies along the central line of grays, so naturally it has no hue, and the angular coordinate is meaningless. If MAX = 0 (i.e. if V = 0), then S is undefined. If V = 0, then the color is pure black, so naturally it has neither hue, nor saturation. Thus the conical diagram collapses to a single point and both the angle and radius coordinates are meaningless at that point. Therefore, we took samples (by random) from huge pixels of human faces and obtained the human skin distribution in HSV color space. We collected the human skin colors with different illumination circumstances from 1000 faces and 200 light colors from the background of images. We collected 900 (30*30) pixels from each face and each light color of the background. Therefore, there are more than 1,080,000 pixels of the human skin colors were used to acquire the cluster of skin color. Our principle is completely different to [6] that tried to effectively separating skin color regions from seemingly similar, but different color regions. Our principle is that the output of human skin color segmentation must include the entire seemingly similar skin color (even some of them are not real human skin color).

When we read in a color image, we will go on human skin color segmentation task firstly. The process will classify pixels in the HSV color space. Examples illustrating that our skin color segmentation scheme is robust to different people/race and illumination conditions are shown as Fig. 3. In other words, the seemingly similar light colors should be considered as real skin color, and should be kept the original color in the human skin color segmentation process because the seemingly similar light colors will be eliminated during the binarization process (color image $\rightarrow$ gray level image $\rightarrow$ binary image with threshold = 100) later. Here, we use raster scanning (left-to-right and top-to-bottom) to get 4-connected components, label them, and then locate the center of each block. Two pixels p and q with values from V (the set of pixels) are 4-connected if q is in the set N_4 (p). In other words, if a pixel p at coordinates (x, y) has four horizontal and vertical neighbors whose coordinates are (x+1, y), (x-1, y), (x, y+1), (x, y-1). This set of pixels, called the 4-neighbors of p, is denoted by N_4 (p). The detail of raster scanning can be found in the textbook written by Gonzalez et al. [14].

From careful observation, we discover that two eyes and one mouth in the frontal view will form an isosceles triangle. This is the rationale on which the finding of potential face regions is based. We could search the isosceles triangles that are gotten from the criteria of "the combination of two eyes and one mouth". The detail of finding an isosceles triangle can be found in our previous work [12]. Since we think that the real facial region should cover the eyebrows, two eyes, mouth and some area below the mouth, the coordinates can be calculated as follows:

$$X1 = X4 = Xi - 1/4*D(i, k); \quad (1)$$
$$X2 = X3 = Xk+1/4*D(i, k); \quad (2)$$
$$Y1 = Y2 = Yi+1/4*D(i, k); \quad (3)$$
$$Y3 = Y4 = Yj -1/4*D(i, k); \quad (4)$$

Assume that (Xi, Yi), (Xj, Yj) and (Xk, Yk) are the three center points of blocks i, j, and k, that form an isosceles triangle. Then (X1, Y1), (X2, Y2), (X3, Y3), and (X4, Y4) are the four corner points of the face region as shown in Fig. 4. X1 and X4 locate at the same coordinate of (Xi - 1/4*D (i, k)); X2 and X3 locate at the same coordinate of (Xk + 1/4*D (i, k)); Y1 and Y2 locate at the same coordinate of (Yi + 1/4*D (i, k)); Y3 and Y4 locate at the same coordinate of (Yj - 1/4*D (i, k)); where D (i, k) is the Euclidean distance between the centers of block i (right eye) and block k (left eye).

3 Face Verification

The third part of the designed system is to perform the task of face verification. We propose an efficient neural networks function that is applied to decide whether a potential face region contains a face. There are two steps in this part. The first step is to normalize the size of all potential facial regions. The second step is to feed every normalized potential facial region into the neural networks function and perform the verification task. Normalization of a potential face region can reduce the effects of variation in the distance and location. Since all potential faces will be normalized to a standard size (i.e. 60 * 60 pixels) in this step, the potential face regions that we have selected in the previous section are allowed to have different sizes. Here, we resize the potential facial region using “bicubic” interpolation technique. The detail of “bicubic” interpolation technique can be found in the textbook written by Gonzalez et al. [14].

The neural networks function employed here work as the followings. Train the Network: 1. Present the training data (e.g.1200 faces and 1200 non_faces) to the network. 2. Network output compared to desired output (e.g. If it is a face the output should be 1 — is_face. If it is not a face the output should be 0 — non_face.). 3. Network weights are modified to reduce error. Use the Network: 1. Present new data (e.g. a potential facial region) to the network. 2. Network computes an output based on its training data (e.g. If it is a face the output should be 1 — is_face. If it is not a face the output should be 0 — non_face.).

Our neural network function is a multilayer feedforward neural network with three layers. The input layer contains 3600 nodes (e.g. a real human faces) and 3600 nodes (e.g. not a real human faces) which are obtained in our previous work [12]. The hidden layer has 15 nodes. The output unit gives a result of 0 (non_face) or 1 (is_face). The block diagram of our neural network function is shown as Fig. 5. An elementary neuron with R inputs is shown below. Each input is weighted with an appropriate w. The sum of the weighted inputs and the bias forms the input to the transfer function f. Every normalized potential facial region is fed into the neural network function and performs the verification task. Once a face region has been confirmed, the last step is to eliminate those regions that overlap with the chosen face region, and then output the result.

4 Experimental Results

In this section, a set of experimental results is demonstrated to verify the effectiveness and efficiency of the proposed system. There are 1200 test images (include 620 different persons). Some test images are taken from digital camera, some from digital video, some from scanner, and some from videotape. Moreover, we also use some parts of the "AR face database"[15] to verify our system. Therefore, there are totally 1400 faces that are used to verify the validity of our system. The sizes of the test images range from 80*80 to 768*576 pixels. In these test images, human faces were presented in various environments. Among them, only 39 faces cannot be found correctly. Experimental results demonstrate that an approximately 97% (1361/1400 = 97.21%) success rate is achieved and the relative false rate is below 3 % (39/1400 = 2.79%). The environments of the experimental set are described as follows. Fig. 6: Experimental results with different conditions: various pose, orientation, facial hair, variation in lighting conditions (AR face database), partial occlusion, presence of glasses, and diverse hair styles. The execution time required to locate the precise locations of the face in the test image set is dependent upon size and complexity of images. However, our new system speeds up a lot than our previous work [12] in the condition of complicated backgrounds. Fig. 7 shows experimental results depict ***without/with*** skin color segmentation in complex backgrounds. Fig. 7(a) shows the original input RGB color image; 7(a_Lin) depicts the result of binary image of 7(a) by using the scheme ***without*** skin color segmentation in Lin [12], and the number of blocks **is**

99; 7(a_Proposed) illustrates the result of the skin color segmentation; 7(b) displays the result of binary image of 7(a_Proposed) by using the scheme in the proposed system ***with*** skin color segmentation, and the number of blocks **is only 7;** 7(c) illustrates the result of triangle; 7(d) shows the potential facial region of 7(b); 7(e) illustrates the potential facial region of 7(a_Proposed); 7(f) depicts the potential facial region with normalized size of 7(d); 7(g) displays the result of human face detection of 7(a_Proposed); 7(h) shows the final result of human face detection of 7(a). Due to the decrease of the number of blocks (from 99 to 7), we can expect that the speed will be improved drastically in the complicated background case. Fig. 7(a) depicts a images with 70*161 pixels, and it needs **387.3750 seconds** in Lin [12] (the number of blocks **is 99)** to locate the correct face position by using a P4 CPU 3.0 GHz PC. In this proposed work to locate the correct face position need **only 0.0625 seconds**. Moreover, the proposed scheme is also significantly faster than Rowley [6] that adopted a small window (20*20) to slide over all portions of an image at various scales is a time-consuming procedure.

5 Conclusion

In this paper, a robust and effective face detection system is presented to extract face in various conditions of face images. The color and triangle-based segmentation process can reduce the background part of a cluttered image up to 98%. Since human skin color segmentation can replace the complex background with the simple (black) background, human skin color segmentation can reduce much executing time in the complicated background case. Moreover, it can manage distinct size, changed lighting condition (normal light variation excluding fluorescent light, incandescent light, or extraordinary light …), various pose, orientation, facial hair, and partial occlusion, presence of glasses, and diverse hair styles at the same time. The experimental results reveal that the proposed method is better than traditional methods (e.g. [11 and 12]) in terms of altered circumstance and efficiency. In the future, we plan to use this face detection system as a preprocessing for solving face recognition problem.

References:

[1] E. Hjelmas and B. K. Low, Face detection: A survey, Computer Vision and Image Understanding, Vol. 83, 2001, pp. 236–274.

[2] Ming-Hsuan Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, Issue 1, Jan. 2002, pp. 34 – 58.

[3] Y. Dai and Y. Nakano, Face-texture model based on SGLD and its application in face detection in a color scene, Pattern Recognition, Vol. 29, no. 6, 1996, pp. 1007-1017.

[4] Karin Sobottka, Ioannis Pitas, A novel method for automatic face segmentation, facial feature extraction and tracking, Signal Processing: Image Communication Vol.: 12, Issue: 3, 1998, pp. 263-281.

[5] E. Saber and A. M. Tekalp, Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions, Pattern Recognition Letters, Vol. 19, 1998, 669-680.

[6] Kyung-Min Cho, Jeong-Hun Jang, Ki-Sang Hong, "Adaptive skin-color filter," Pattern Recognition, Vol.: 34, 2001, pp. 1067-1073.

[7] H. M. El-Bakry, Face detection using fast neural networks and image decomposition, Neurocomputing, Vol. 48, 2002, pp. 1039-1046.

[8] J. Fang and G. Qiu, Human face detection using angular radial transform and support vector machines, Int. Conf. on Image Processing, Vol. 1, 2003 pp. I - 669-672.

[9] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Networks, Vol. 16, 2003, pp. 555-559.

[10] H. Fashandi and M. S. Moin, Face detection using CMAC neural network, in Artificial Intelligence and Soft Computing - Icaisc 2004, Vol. 3070, Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag Berlin, 2004, pp. 724-729.

[11] H. A. Rowley, S. Baluja, and T. Kanade, Neural network-based face detection, IEEE Transactions on PAMI, Vol. 20, Issue 1, 1998, pp. 23 – 38.

[12] Chiunhsiun Lin, Kuo-Chin Fan, Triangle-based Approach to the Detection of Human Face, Pattern Recognition, Vol. 34, No. 6, 2001, pp. 1271-1284.

[13] Information of HSV color space, http://encyclopedia.laborlawtalk.com/HSV_color_space

[14] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Addison-Wesley Publishing Company, 1992.

[15] A.M. Martinez and R. Benavente. The AR Face Database. CVC Technical Report #24, June 1998

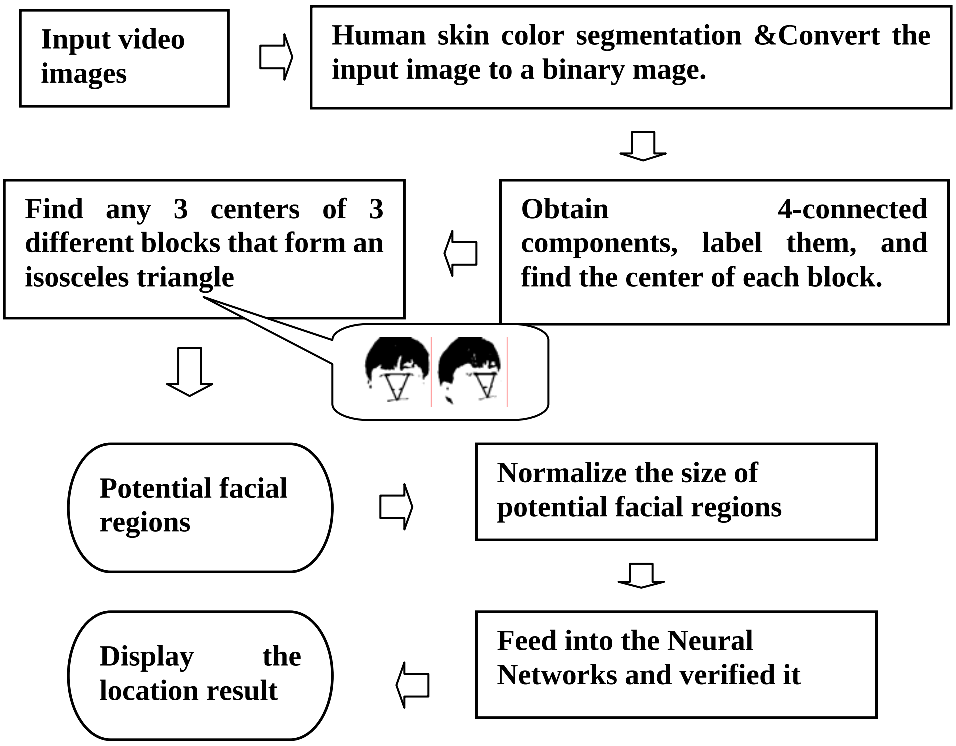

Fig. 1 Overview of our system

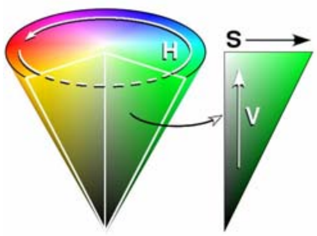

Fig. 2 The HSV model

Fig. 3 Examples illustrating that our skin color segmentation scheme is robust to different people/race.

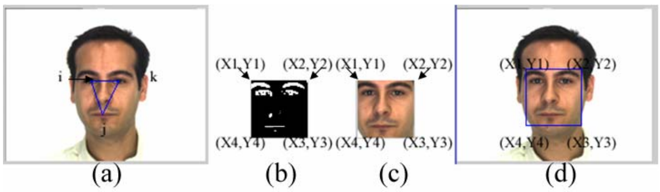

Fig. 4 (Xi, Yi), (Xj, Yj) and (Xk, Yk) are the 3 center points of blocks i, j, and k, respectively. The four corner points of the face region will be (X1, Y1), (X2, Y2), (X3, Y3), and (X4, Y4).

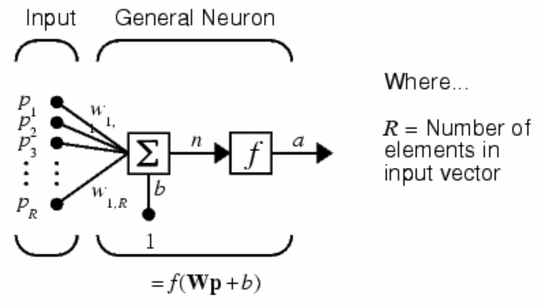

Fig. 5 Block diagram of our neural network function

Fig. 6 Experimental results with different conditions.

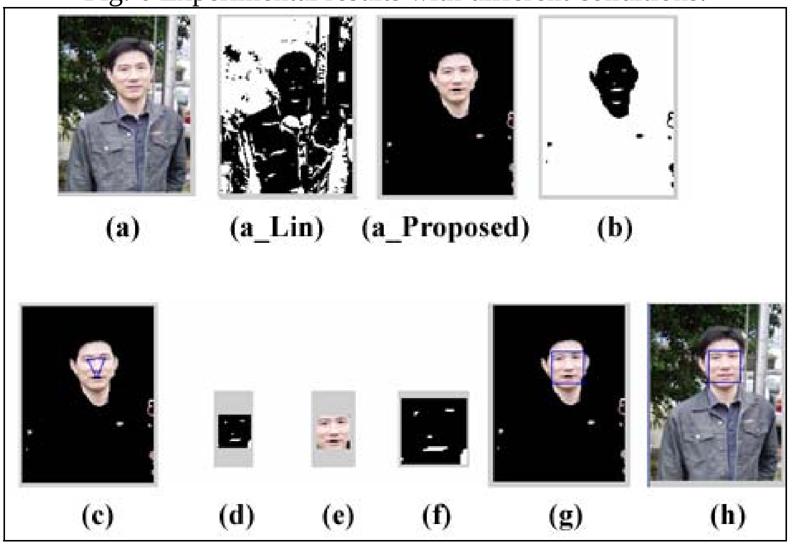

Fig. 7 Experimental results depict ***without/with*** skin color segmentation in complex backgrounds.