Face Location in Non-Uniform Illumination Conditions Using Dynamic Thresholds, Lighting Normalization and Weighting Mask Function

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: - In this paper, a robust and efficient human face location system that uses dynamic thresholds, lighting normalization and a weighting mask function to locate the human face embedded in photographs with non-uniform illumination conditions is presented. The designed system is composed of two principal parts: The first part is to use dynamic thresholds to obtain a satisfied binary image, and then to search the potential face regions based on the rules of "the combination of two eyes and one mouth". The second part of the proposed system is to perform the face verification task by a lighting normalization function and a weighting mask function. The experimental results reveal that the proposed method is better than previous methods in non-uniform illumination conditions.

Key-Words: - Face location; Non-uniform illumination conditions; Lighting normalization; Weighting mask function; Internet security

1 Introduction

Face localization is necessary for face recognition or identification system. Its definition is to locate only one human face in a frame/image. Some successful systems have been reported in the literatures [1~14]. However, none of them researched into the various illuminations problem. The proposed system is composed of two principal parts as shown in Fig. 1.

The first part is to use dynamic thresholds to obtain satisfied binary image, and then to search the potential face regions based on the rules of "the combination of two eyes and one mouth". The second part of the proposed system is to perform the face verification task. The proposed face location system can locate the human face in various illumination conditions automatically. Excellent location results were observed for images obtained from photographs, and the relative false location rate is very low. The rest of the paper is organized as follows. In section 2, searching for face candidates is illustrated. In section 3, each of the normalized potential face regions is fed to a lighting normalization function and a weighting mask function to verify whether the potential face region really contains a face or not. Experimental results are demonstrated in section 4 to verify the validity of the proposed face location system. Finally, conclusions are given in section 5.

2 Searching Face Candidates Using Dynamic Thresholds

The main purpose of this process is to search face candidates in an input image. The task of exploring face candidates can be divided into the following six steps. We assume the images with various illuminations and simple backgrounds. First, a color or a gray scale image is acquired by the system; subsequently it is converted to a satisfied binary image by using dynamic thresholds with try error tactic. The dynamic thresholds algorithm with try error tactic for acquiring a satisfied binary image that is used for finding out any three centers of three different blocks that satisfy an isosceles triangle is described as follows:

Step (0) Initiate the threshold = 80;
Step (1) Label all 4-connected components in the image to form blocks and count the number of the blocks. If the number of the blocks (N) is equal to 15 or larger than 15, then we will find out the center of each block;
Step (2) Else if N < 10, change the threshold n = 150. Label all 4-connected components in the image to form blocks and count the number of the blocks. If the number of the blocks is equal to 15 or larger than 15, then we will find out the center of each block;
Step (3) Else if N < 13, change the threshold n = 130. Label all 4-connected components in the image to form blocks and count the number of the blocks. If the number of the blocks is equal to 15 or larger than 15, then we will find out the center of each block;
Step (4) Else if N < 15, change the threshold n = 100. Label all 4-connected components in the image to form blocks, then we will find out all centers of all blocks.

In other words, we will change the threshold from 80, 150, 130 and 100 respectively.

Second, 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 R. C. et al. [15].

Third, find out any three centers of three different blocks that satisfy an isosceles triangle that contains two eyes and one mouth in the human face. This is the rationale on which the finding of potential face regions is based. We could search the potential face regions that are acquired from the criteria of "the combination of two eyes and one mouth (construct an isosceles triangle)". If the triangle i j k is an isosceles triangle as shown in Fig. 2(a), then it should possess the characteristic of "the distance of line i j = the distance of line j k". From observation, we discover that the Euclidean distance between two eyes (line i k) is about 90% to 110% of the Euclidean distance between the centers of the right/left eye and the mouth. Due to the imaging effect, imperfect binarization result and various poses of human faces, a 25% deviation is given to absore the tolerance. Here, “abs” means the absolute value, "D (i, j)" denotes the Euclidean distance between the centers of block i (right eye) and block j (mouth), "D (j, k)" denotes the Euclidean distance between the center of block k (left eye) and block j (mouth), "D (i, k)" represents the Euclidean distance between the centers of block i (right eye) and block k (left eye).The first matching rule can thereby be stated as (abs(D(i, j)-D(j, k)) < 0.25*max(D(i, j), D(j, k))), and the second matching rule is (abs(D(i, j)-D(i, k)) < 0.25*max(D(i, j), D(j, k))). Since the labeling process is operated from left to right then from top to bottom, we can get the third matching rule as “i < j < k”. For example, as shown in Fig. 2(a), if three points (i, j, and k) satisfy the matching rules, then we think that they construct an isosceles triangle. Assuming that the real facial region should cover the eyebrows, the eyes, the mouth and some area below the mouth, the coordinates can be determined as follows:

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

Assuming 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. 2(d). 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).

Fourth, normalizing a potential face region can decrease the effects of variation in the distance and location. Since all potential faces are normalized to a standard size (i.e. 60 × 60 pixels) in this step, the potential face regions selected in the previous section can have different sizes. Herein, the potential facial region is resized by the bicubic interpolation technique as described in the textbook written by Gonzalez R. C. et al. [15]. In this step, we have already removed the restriction of changeable sizes.

Fifth, the illuminations of the potential facial regions are normalized by utilizing a lighting normalization function. The illumination conditions of the detected facial regions are modified by regulating the average grayscale value to a constant (which in the proposed system is 135). A simple global thresholding function is then performed with a threshold value of T (in the proposed system, T = 100) to generate satisfactory binary images. The average grayscale value of the detected facial image is changed to a constant 135, because the average grayscale values of the detected facial images are about 110–160. Hence, the average gray value of 110 and 160 [(110+160)/2 =135] is selected to adjust the average grayscale value of a facial image, which in this case is 135. If the average gray value of a facial image is adjusted to a constant between 125 and 145 should be all right. The average grayscale value of different images seems to be impossible to regulate to a constant. Nevertheless, regulation is possible in the detected facial region, since the outline of the detected facial region that all covers the eyebrows, two eyes, and one mouth is outstandingly stable, as demonstrated in Fig. 3. Fig. 3(a) shows the original detected facial region. Fig. 3(b) illustrates the original detected facial region by a simple global thresholding process with a threshold value of T (T = 100), which produces unstable binary images (the block of mouth is almost invisible). Fig. 3(c) depicts the original detected facial region from adjusting the average grayscale value to a constant (constant = 135), and then by simple global thresholding with threshold T (T = 100) to obtain stable binary images. Figs. 3(d), 3(g), 3(j) and 3(m) display the original detected facial region with different intensity of illumination, respectively. Figs. 3(e), 3(h), 3(k) and 3(n) depict the original detected facial region with different intensity of illumination, and then by simple global thresholding with threshold T (T = 100) to acquire unstable binary images (the block of mouth is almost invisible). Figs. 3(f), 3(i), 3(l) and 3(o) show the original detected facial region with different intensity of illumination. The average grayscale value was first adjusted to a constant value of 135, and then converted to a stable binary image by a simple global thresholding process with a threshold value of T (T = 100, and the block of mouth is easy to find)). In this task, the problems of non-uniform illumination conditions can be resolved using the lighting normalization function.

Sixth, feed every normalized potential face region into a weighting mask function that is applied to decide the actual location of the face region. In this part, we have already removed the restriction of changeable sizes and diverse illumination conditions.

3 Face Verification

Each face candidate obtained from the previous process has already normalized to a standard size and a regular illumination. Subsequently, each of these normalized potential face regions is fed to the weighting mask function to verify whether the potential face region really contains a face. If the normalized potential facial region really contains a face, it should have high similarity to the mask that is formed by 100 binary training faces. The method for generating a mask is to read in 100 binary training masks that are cut manually from the facial regions of images. Then add the corresponding entries in the 1000 training masks to form an added mask. In other words, we use 100 binary faces to structure a template mask that is used for comparing with each normalized potential facial region. The idea is similar to the neural network concept, but it does not need so many training samples. Each normalized potential facial region is fed into the weighting mask function that is used to compute the similarity between the normalized potential facial region and the mask. The computed value can be utilized in deciding whether a potential region contains a face or not.

The algorithm for obtaining the weight of the potential facial region is stated as follows:

Input: the potential facial region and the mask of the training samples

Output: the weight of the potential facial region

Begin

For all pixels of the potential facial region and the mask of the training samples

Step 0) If the pixel of potential facial region is black and the pixel of the mask is also black, then weight = weight + 6.

Step 1) If the pixel of potential facial region is white and the pixel of the mask is also white, then weight = weight + 2.

Step 2) If the pixel of potential facial region is black and the pixel of the mask is white, then weight = weight - 4.

Step 3) If the pixel of potential facial region is white and the pixel of the mask is black, then weight = weight –2.

Step 4) Calculate the weight of the potential facial region for all pixels.

End

A set of experimental results demonstrates that the threshold values of the human face should be set between 4,000 and 5,500.

4 Experimental Results

In this section, a set of experimental results is demonstrated to verify the effectiveness and efficiency of the proposed system. We use the "AR face database"[16] to verify our system: This face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains 135 people's faces (75 men and 60 women). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. However, we don’t cover the part of the occlusions (sun glasses and scarf). In other words, we only use the parts of (1) neutral expression, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) all side lights on, but exclude the others. Moreover, there are two sessions per person (2 different days) in the AR face database. There are 1875 test images (include 135 different persons * 7 each person * 2 sessions per person = 1890, but there are 15 images are damaged during Internet transference from the original "AR face database". Therefore, there are 1875 test images for testing) that are employed to verify the validity of our system. Among them, only 45 faces cannot be found correctly. Experimental results demonstrate that an approximately 98% (1830/1875 = 97.6%) success rate is achieved and the relative false rate is below 3 % (45/1875 = 2.4%). The whole procedures is shown in Fig. 4, where Fig. 4(a) is the original image; Fig. 4(b) is the gray level image; Fig. 4(c) is the unsatisfied binary image with threshold = 80 (Initiate threshold is 80); Fig. 4(d) is the satisfied binary image with threshold = 150 by using dynamic thresholds function; Fig. 4(e) is the isosceles triangle formed by the three centers of three blocks; Fig. 4(f) is the best binary potential face region cropped from the binary image that covers the isosceles triangle; Fig. 4(g) is the best potential facial region cropped from the original image that covers the isosceles triangle; Fig. 4(h) is the best binary potential binary face region which is normalized to a constant gray level intensity and a standard size (60 * 60 pixels); Fig. 4(i) is the result of face location. Fig. 5 displays face images with altered illumination conditions and different problems. The environments of the experimental set are described as follows. Fig. 5(a) neutral expression; Fig. 5 (b) smile; Fig. 5 (c) anger; Fig. 5 (d) scream; Fig. 5 (e) left light on; Fig. 5 (f) right light on; Fig. 5 (g) all side lights on respectively. The sizes of the test images are 768*576 pixels. The execution time required to locate the precise locations of the face in the test image set is dependent upon complexity of images. For example, the test image as shown in Fig. 5 (d) need only 0.1720 second and Fig. 5 (g) need 0.6720 second to locate the correct face position by using a P4 CPU 3.0Ghz PC.

5 Conclusion

We proposed an effective face location system that uses dynamic thresholds and lighting normalization to locate the human face embedded in images with different illuminations. The experimental results confirm that the proposed method is better than previous methods in non-uniform illumination conditions. In the future, we plan to use this face location system as a preprocessing for solving face recognition problem.

References:

[1] V. Govindaraju, S. N. Srihari, D. B. Sher, "A computational model for face location", Proc. Computer Vision and Pattern Recognition, 1990, pp. 718-721.

[2] H. L. Choong, S. K. Jun, H. P. Kyu, "Automatic human face location in a complex background using motion and color information", Pattern Recognition Vol. 29, Issue: 11, 1996, pp. 1877-1889.

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

[4] P. Juell, R. Marsh, "A hierarchical neural network for human face location", Pattern Recognition, Vol. 29, no. 5, 1996, pp. 781-787.

[5] J. W. Kim, B. H. Kang, P. M. Kim, M. S. Cho, "Human face location in image sequences using genetic templates", IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, 1997, pp.2985 - 2988.

[6] L. S. Shen, K. Q. Wang, X. Xing, "Automatic human face location and tracing in a complex background", Chinese Journal of Electronics, Vol. 9, 2000, pp. 65-69.

[7] D. Maio, D. Maltoni, "Real-time face location on gray-scale static images", Pattern Recognition, Vol. 33, 2000, pp. 1525-1539.

[8] Y. J. Wang, B. Z. Yuan, "Robust face location and tracking using optical flow and genetic algorithms", Chinese Journal of Electronics, Vol. 10, 2001 pp. 450-454.

[9] Y. Wang, B. Yuan, "Fast method for face location and tracking by distributed behaviour-based agents", Vision, Image and Signal Processing, IEE Proceedings, Vol. 149, Issue 3, 2002, pp. 173 – 178.

[10] P. Sharma, R. Reilly, "Fast marching methods applied to face location in videophone applications using colour information",IEEE International Conference on Multimedia and Expo, Vol. 2, 2002, pp. 141 – 144.

[11] Tianxiang Yao, Hongdong Li, Guangyao Liu, Xiuqing Ye, Weikang Gu, Yiqing Jin, "A fast and robust face location and feature extraction system", International Conference on Image Processing, Vol. 1, 2002, pp. I-157 - I-160.

[12] J. S. Tang, S. Acton, "Locating human faces in a complex background including non-face skin colors", Journal of Electronic Imaging, Vol. 12,

2003, pp. 423-430.

[13] B. Raducanu, M. Grana, F. X. Albizuri, A. d'Anjou, "A probabilistic hit-and-miss transform for face localization", Pattern Analysis and Applications, Vol. 7, 2004, pp. 117-127.

[14] Wang Xianbao, Cao Wenming, Li Guojun, "A Method of Face Location in Complex Background", International Conference on Neural Networks and Brain, Vol. 3, 2005, pp. 1507 – 1510.

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

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

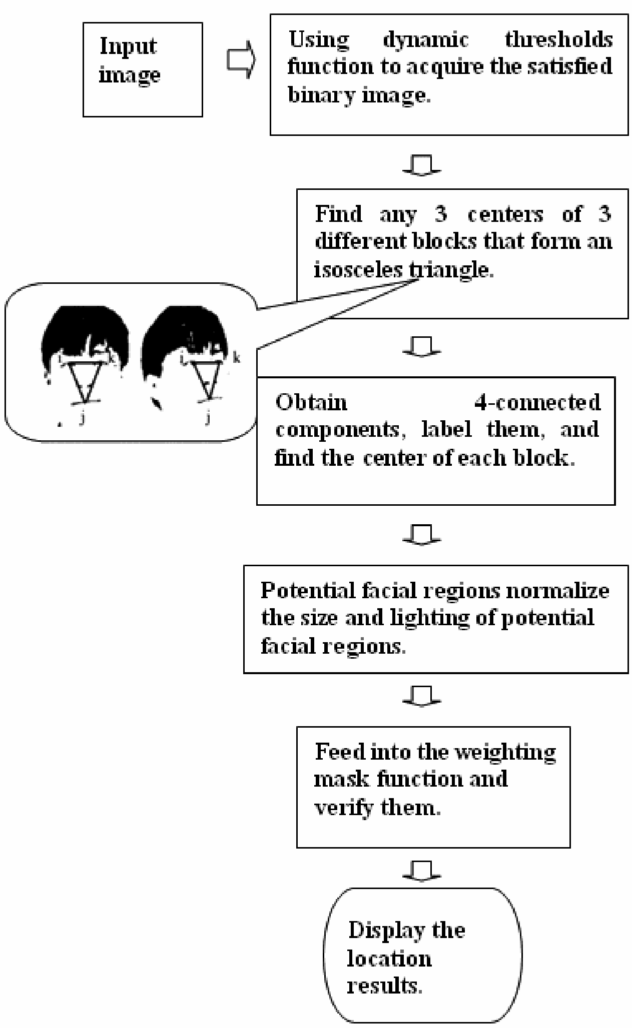

Fig. 1: Overview of our system

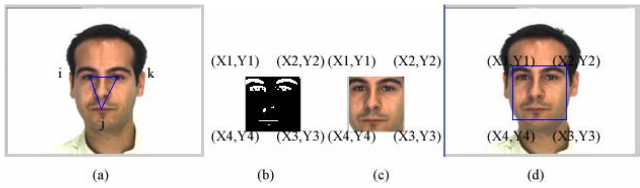

Fig. 2: (X_i, Y_i), (X_j, Y_j) and (X_k, Y_k) are the 3 center points of blocks i, j, and k, respectively. The four corner points of the face region will be (X1, Y1), (X2, Y2), (X3, Y3), and (X4, Y4).

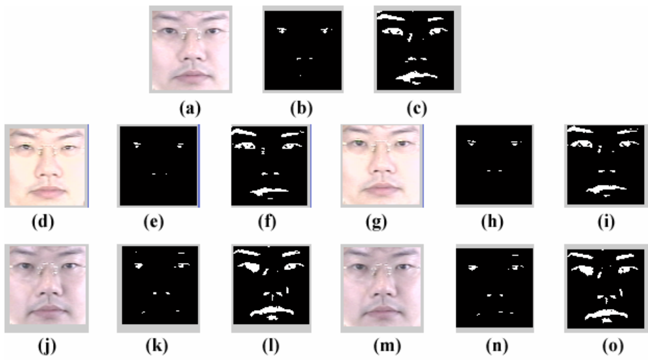

Fig. 3: The process of the original detected facial region by adjusting the average gray-level value to a constant (constant = 135), and then by simple global thresholding with threshold T (T = 100) to acquire stable binary images.

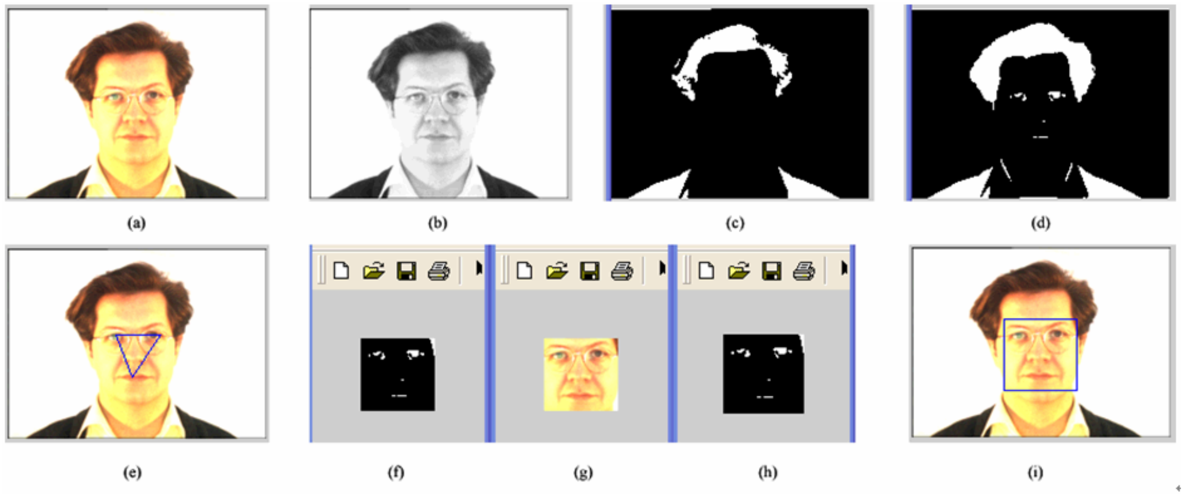

Fig. 4: Example illustrating the process of how to obtain the verified face from an image.

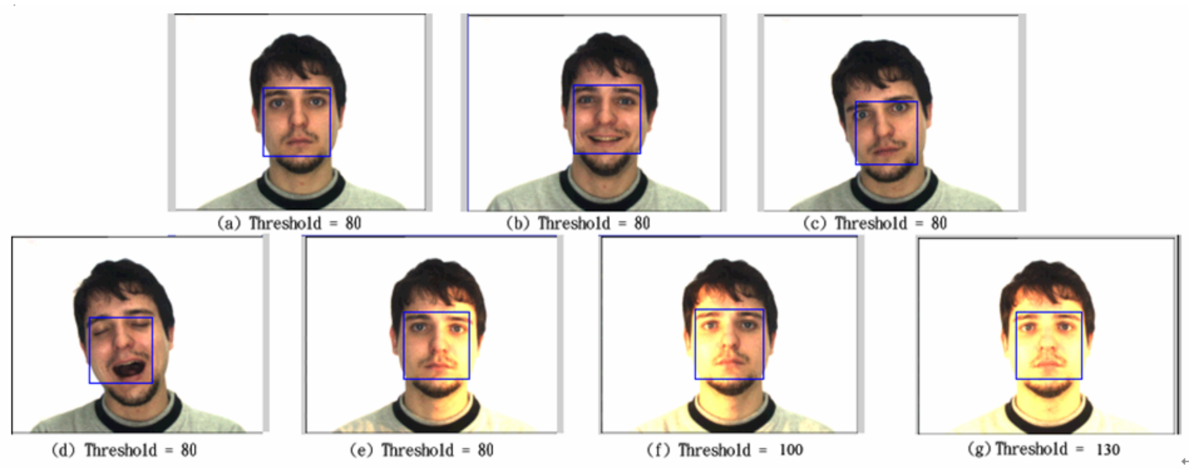

Fig. 5: Example illustrating the face images with altered lighting conditions and different problems by using dynamic thresholds. (a) neutral expression, (b) smile, (c) anger, (d) scream, (e) left light on, (f) right light on, (g) all side lights on respectively.