A ROBUST APPROACH FOR EYE LOCALIZATION UNDER VARIABLE ILLUMINATIONS

Shan Du and Rabab Ward

Department of Electrical and Computer Engineering
The University of British Columbia, Vancouver, BC, Canada
{shand, rababw}@ece.ubc.ca

ABSTRACT

Illumination variation is a main obstacle in facial feature detection. This paper presents a novel automated approach that localizes eyes in gray-scale face images and that is robust to illumination changes. The approach does not require prior knowledge about face orientation and illumination strength. Other advantages are that no initialization and training process are needed. Based on an edge map obtained via multi-resolution wavelet transform, this approach first segments an image into different inhomogeneously illuminated regions. The illumination of every region is then adjusted so that the features' details are more pronounced. To locate the different facial features, for every region, Gabor-based image is constructed from the re-lit image. The eyes sub-regions are then identified using the edge map of the re-lit image. This method has been applied successfully to the images of the Yale B face database that have different illuminations.

Index Terms: face recognition, eye detection, wavelets, biometrics.

1. INTRODUCTION

Facial feature localization is crucial to many applications such as face recognition, face expression analysis, and face tracking. Illumination variations form a main obstacle in facial feature detection. Unfortunately, there exists little work on automated detection under variable illumination conditions, especially for extremely bad illuminated images such as shown in Figure 1.

Figure 1. Images with extremely bad illumination conditions

Many feature localization methods have been proposed previously. Kass *et al.* [1] introduced Active Contour Models and an energy minimization approach for edge detection. These models can be used to detect face boundary and facial features. Wiskott *et al.* [2] used Gabor wavelets to generate a data structure, named the Elastic Bunch Graph, to locate facial features. Yuille *et al.* [3] first proposed the use of a deformable template for locating human eyes. This method designs an eye model and the eye position is then obtained through a recursive process. Active Shape Model (ASM) [4] and Active Appearance Model (AAM) [5], proposed by Cootes *et al.*, are two popular shape and appearance models for feature localization. Feature search based on the above models will however become unstable under significant illumination variations. All the above model-based algorithms need a good initialization that is close to the correct solution; otherwise, they are prone to getting stuck in local minima. Illumination effects can also make their performance worse.

In this paper, we propose a different automated eye detection approach that is robust even under severe illumination changes. First, the lighting of the image is regionally adjusted. The re-lighting of the regions makes the features details more pronounced. The regions are obtained by segmenting the image into different sections of different illumination strengths. The number of regions may be greater than 1 or equal to 1 where the lighting is uniform. This segmentation is based on the edge map E obtained from the original image I using multi-resolution wavelet transform. Then the illumination of the original image I is regionally adjusted using histogram equalization to generate the re-lit image I_r.

Based on the re-lit image I_r, another edge map E_r and a Gabor image G_r are both generated regionally. I_r, E_r, and G_r are used together to localize eyes. The Gabor image G_r is first used to locate windows within which facial features lie. This is because the convolution of Gabor wavelets with an image results in the salient facial features with high magnitudes, such as eyes, nose and mouth. The windows corresponding to eyes are identified using both the adjusted image I_r and the new edge map E_r. After the eyes' windows are detected, the centroids of the windows form the approximate positions of the eye balls. This method assumes that the face region is given, but the orientation of the face and the illumination of the image are unknown. Another advantage of our method is that no initialization is needed. Training data and a training process are not required either. The block diagram of the proposed method is shown in Figure 2.

2. THE PROPOSED APPROACH

Lighting conditions are not usually uniform when imaging a picture. The worst case is the side lighting effect. Normally, the contrast enhancement method is used to deal with this problem. However, the conventional histogram equalization method does not work well when extreme conditions are present, especially in the case of side lighting effects. In this paper, the lighting of the image is first adjusted on a region-by-region basis. The image I is partitioned into n regions with different illuminations based on an edge map E resulting from the multi-resolution wavelet transform. n can be equal to 1 when the image has even illumination and can be greater than 1 when the image has uneven illumination. The regional contrast-enhancement method we proposed in [6] is then applied on each region separately.

2.1. Regionally Illumination Adjusted Image I_r

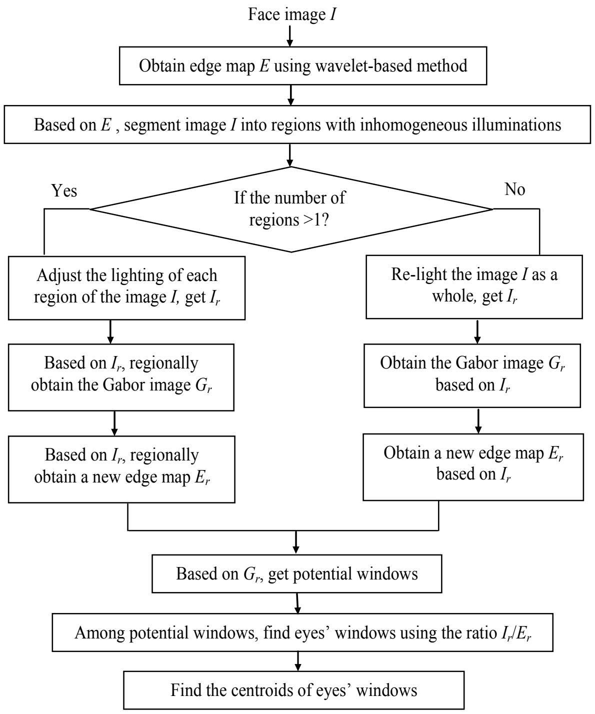

Figure 2. Block diagram of the proposed method

2.1.1. Edge Map E Obtained Using Wavelet Transform

Edges involve the high frequency components of an image. And often, edges occur at different resolutions; both strong edges and weak edges exist in the same image. It is appropriate to extract edges at different scales or resolutions.

Wavelet decomposition of an image provides a good solution to obtaining the edge map. Here we use the redundant wavelet transform. The decomposition procedure for a redundant wavelet transform is different from the normal one in that the scaling of the wavelet is not achieved by sub-sampling the image in each step, but rather by an up-sampling of the filters. The four wavelet sub-bands at scale j are of the same size as the original image, and all filters used at scale j are up-sampled by a factor of 2^j (padding $2^j - 1$ zeros) compared with those at scale zero.

The edges of an image are full of high frequency information that scatters into several scales or resolutions. In order to take advantage of the multi-resolution property of wavelet transforms, two corresponding sub-bands at adjacent resolutions are multiplied to enhance image edges and suppress noise. This is based on the fact that edge structures are present at each scale while noise decreases rapidly along the scales. For more details, please refer to [6]. Figure 3a and 3c show the original images and Figure 3b and 3d show the edge maps respectively.

2.1.2. Region Segmentation

Now we use our observation that for poor illumination, the edges have extremely weak intensity strengths. Actually lack of light or over lighting both weaken real edges. Therefore, we examine the edge map. If some edges in the edge map are weak, it means the illumination in this region may be either over-lit or under-lit. We differentiate these regions from the well-lit regions. Then we apply histogram equalization on each region separately.

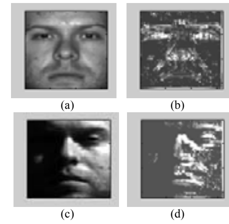

Figure 3. Edge maps

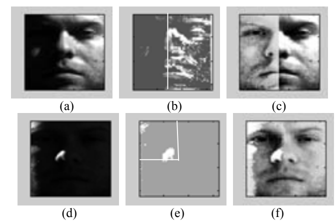

Figure 4. Regionally re-lit intensity image I_r (c), (f)

2.1.3. Regionally Re-lit Intensity Image I_r

The re-lit (illumination adjusted) image I_r is obtained by applying histogram equalization to each region separately. Figure 4a and 4d show the original images. Figure 4b and 4e show their edge maps. The white lines show the segmented regions. Figure 4c and 4f show the resulting regionally re-lit images, respectively.

2.2. Localizing the Eyes' Features

2.2.1. Gabor Wavelet Convolution and Regionally Obtained Gabor Image G_r

Gabor images are the results of convolving the face image with a set of Gabor wavelets. A Gabor image exhibits high magnitudes on the salient facial features, such as eyes, nose and mouth. We thus use Gabor images to locate windows within which facial features lie.

The 2-D Gabor wavelets are defined as follows:

$$\psi_j(\vec{x}) = \frac{\|\vec{k}_j\|^2}{\sigma^2} \exp\left(-\frac{\|\vec{k}_j\|^2 \|\vec{x}\|^2}{2\sigma^2}\right)\left[\exp\left(i\vec{k}_j\vec{x}\right) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (1)$$

where

$$\vec{k}_j = k_m e^{i\varphi_n}$$

$$k_m = \frac{0.5\pi}{\left(\sqrt{2}\right)^m} \quad \varphi_n = n\frac{\pi}{8}$$

In this paper, three different scales' Gabor wavelets (Figure 5) of horizontal orientation with $m \in \{0,1,2\}$ and $n = 0$ are used to derive the Gabor wavelet convolved images.

The Gabor wavelet representation $O_j(\vec{x})$ of an image is the convolution of the image $I(x, y)$ with a family of Gabor wavelets $\psi_j(\vec{x})$,

$$O_j(\vec{x}) = I(\vec{x}) * \psi_j(\vec{x}) \quad j = 1 \to 3 \quad (2)$$

where $(\vec{x}) = (x, y)$, and * denotes the convolution operator.

Figure 6a shows the original image and Figure 6b shows the three-scale Gabor convolved images. The three images are then multiplied and thresholded to highlight the high responses and suppress noise. Thus,

Figure 5. 3 horizontal Gabor wavelets

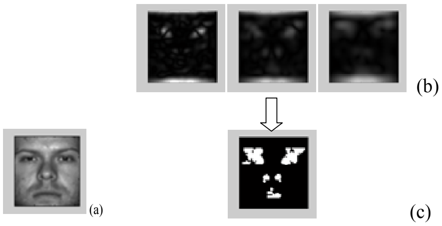

Figure 6. Gabor convolved image

$$G_r(\vec{x}) = O_1(\vec{x}) \times O_2(\vec{x}) \times O_3(\vec{x}) \tag{3}$$

Figure 6c is the Gabor image G_r obtained by multiplying the three-scale convolved images and then thresholding the product. Normally, we always get high magnitudes at images' upper and bottom outer edges. As they are not associated with any facial feature, they are removed.

For illustration purpose, in Figure 6, a well-lit image which has one illumination region (i.e., n=1) is used. G_r is obtained using the whole image. However, if I has more than one region (n>1), then Gabor image G_r is obtained regionally, i.e., by applying the above Gabor convolution to each region separately. Figure 7a and 7d show the original images. Figure 7b and 7e show the Gabor images obtained using the whole image. Figure 7c and 7f show the regionally obtained Gabor images G_r.

2.2.2. Localization of Facial Features' Windows

The Gabor image G_r is used to locate the windows within which different facial features lie. The regions of G_r with high magnitudes are clustered, based on their positions to form isolated windows. These windows are the potential windows that contain the facial features. In order to localize the separate windows containing the facial features, the Gabor image G_r is first partitioned into small blocks so as to filter out the scattered isolated high magnitude points and to speed up the following clustering process. The blocks with magnitudes higher than a threshold are marked as candidate blocks. The candidate blocks are clustered based on their coordinates. Eight nearest neighbours method is used to connect the high magnitudes' blocks. The connected blocks form the potential features' windows. We label these windows as $w_1, w_2, ..., w_n$. The clustering method is described in Figure 8. The potential windows are shown in Figure 9.

2.2.3. Regionally Obtained New Edge Map E_r

Similarly to G_r, the regionally obtained new edge map E_r is obtained by applying the above wavelet transform method to different regions of I_r separately. Figure 10a and 10d show the original images. Figure 10b and 10e show the edge maps obtained using the whole image. Figure 10c and 10f show the regionally obtained edge maps I_r.

2.2.4. Determining the "Eyes" Windows

To identify the windows in G_r that contain the eyes, we rely on I_r and E_r. Compared with the other features, the eyes have the lowest average gray-scale intensity since eyes normally have low intensity, but the highest average edge intensity since there are a lot of edge points near the eyes' positions. Thus, we compute the ratio of the

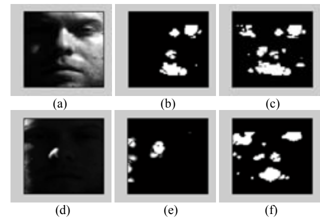

Figure 7. Regionally obtained Gabor image G_r (c), (f)

1. Partition the Gabor image G_r into blocks, e.g., 4×4.
2. For each block, compute the average magnitude: If the average is larger than a pre-defined threshold, merge it with the adjacent blocks with higher magnitudes using the eight nearest neighbours method.
3. The connected blocks form isolated regions. The isolated regions form the potential features' windows.

Figure 8. Clustering algorithm

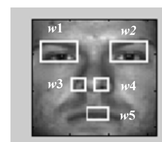

Figure 9. Potential features' windows

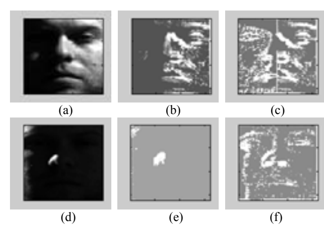

Figure 10. Regionally obtained new edge map E_r (c), (f)

average gray-scale intensity to edge intensity for each window i to obtain

$$\frac{\bar{I}_r^{\,i}}{\bar{E}_r^{\,i}}$$

where $\bar{I}_r^{\,i}$ is the average intensity value of window i, and $\bar{E}_r^{\,i}$ is the average edge intensity value of window i. The eyes' windows are chosen as the two windows with the first two lowest ratios.

The eye balls in the "eyes" windows have the lowest gray-scale intensity. We can use this information to roughly detect the positions of eye balls. The rough positions of the eye balls are located by locating the centroids of the eyes' windows. The eyes' windows and centroids (the eye balls) are shown in Figure 11.

2.2.5. Possible Windows Corresponding to Nose and Mouth

When the eyes' locations are detected, these locations can be used to detect the nose and mouth's positions. The eyes' locations can be used to detect the orientation of a face. The windows related to nose and mouth can be defined as being below the eyes in the direction of the vertical axis and between the two eyes in the horizontal axis direction. The nose is above the mouth. In our method, the nose is sometimes represented by two small windows

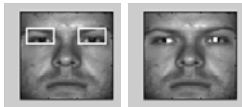

Figure 11. Eyes' windows and centroids (the eye balls)

corresponding to each nostril; sometimes only one nostril is located. Such information can be further used to detect precise locations of the nose and mouth. In this paper, we do not consider them.

3. EXPERIMENTS

In order to examine the effectiveness of the proposed method, we apply it to the images of different persons with different illuminations and different poses. The images used are from the Yale B face database [7]. Figure 12 shows the detection results on an image with the side lighting effect: (a) is the original image I; (b) is the original Gabor image G; (c) is the original edge map E; (d) is the regionally re-lit image I_r; (e) is the regionally obtained Gabor image G_r; (f) is the regionally obtained new edge map E_r; (g) shows the potential features' windows; (h) shows the eyes' positions. Figure 13 shows the detection results on an image that is almost all dark: (a) is the original image I; (b) is the original Gabor image G; (c) is the original edge map E; (d) is the regionally re-lit image I_r; (e) is the regionally obtained Gabor image G_r; (f) is the regionally obtained new edge map E_r; (g) shows the potential features' windows; (h) shows the eyes' positions. Figure 14 shows the detection results on images with different poses. Figure 15 shows the detection results on images of different persons.

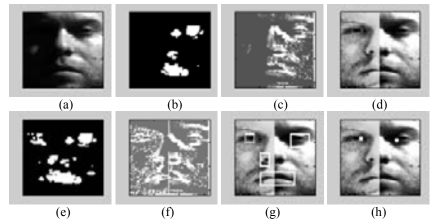

Figure 12. Detection on the image with side lighting effect

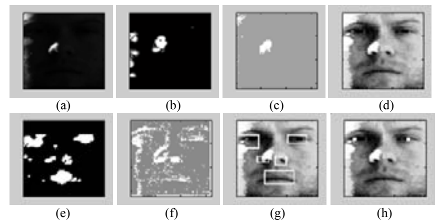

Figure 13. Detection on the image that is almost dark

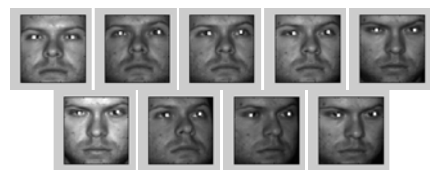

Figure 14. Detection on images in different poses

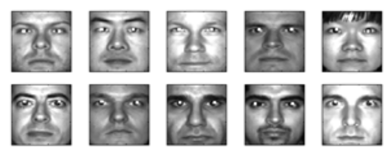

Figure 15. Detection on images of different persons

4. DISCUSSIONS AND FUTURE WORK

In this paper, we propose a simple and fast method that can quickly detect the windows that contain facial features, mainly the eyes. The method is shown to be robust to illumination changes and to a certain degree of pose orientations. Our method detects the approximate features' locations. More precise locations could be detected by adding a further step such as template matching. An eye model obtained by training a number of sample eyes can be used to obtain the precise eyes' positions through a recursive optimization process as in [3]. The original template matching in [3] is time-consuming and is only feasible if the initial position of the eye model is placed near the actual eye position. Thus, a combination of such a scheme with our method will guide the recursion so as to speed up the method and make it more robust. In the improved template matching method [8], corner detection is used to guide the recursive process. The corner detection algorithm however is based on the edge map, and unfortunately a good edge map is hard to obtain when the contrast of the image is relatively low. Our regional contrast enhancement method can offer a satisfactory solution to this problem.

This paper concentrated on finding a scheme that can detect the approximate positions of facial features and that is robust to illumination variation, thus we did not take precise localization into account. In our future work, the precise localization of eye positions needs to be considered, also the nose and mouth precise localization.

5. CONCLUSIONS

In this paper, we propose a simple yet efficient method to find the windows containing facial features and detect the approximate positions of the eyes. The regionally re-lit gray-scale image I_r, the regionally obtained Gabor image G_r, and the regionally obtained edge map E_r are used together to facilitate this task. The experimental results show that this method works well for face images of different people and under various conditions such as changes in lighting and pose.

6. REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," 1st International Conference on Computer Vision, pp. 259-268, 1987.

[2] L. Wiskott, J. M. Fellous, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," IEEE Trans. Patt. Anal. Mach. Intell., vol. 19, pp. 775-779, 1997.

[3] A. L. Yuille, D. S. Cohen, P. W. Hallinan, "Feature extraction from faces using deformable templates," International Conference on Computer Vision and Pattern Recognition, pp. 104-109, 1989.

[4] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic face identification system using flexible appearance models," Image Vis. Comput. vol. 13, pp. 393-401, 1995.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681-685, 2001.

[6] S. Du, and R. K. Ward, "Adaptive region-based image enhancement method for face recognition under varying illumination conditions," IEEE International Conference on Acoustics, Speech and Signal Processing, pp.II353-II356, 2006.

[7] A. Georghiades, P. Belhumeur and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 643-660, 2001.

[8] X. Xie, R. Sudhakar, and H. Zhang, "On improving eye feature extraction using deformable templates," Pattern Recognition, vol. 27, no. 6, pp. 791-799, 1994.