Eye-verifier Using Ternary Template for Reliable Eye Detection in Facial Color Images

Nguyen Van Huan, Nguyen Thi Hai Binh and Hakil Kim

Abstract—This paper introduces an eye-verifier for reliable detection of eyes in facial color images. At first, the eye region candidates are searched by the circular filter to the binary face image. The eye candidates are then fed into an eye-verifier. The eye-verifier uses a ternary template, generated from the eye area consisting of iris, sclera and skin. Then the template matching is made by ternary Hamming distance. Experimental results evaluating the proposed method on our own face database and FERET database show promising performance in terms of detection rate. The proposed algorithm showed a robust performance to the face poses and eye directions.

I. INTRODUCTION

Biometrics is undoubtedly contributing to the human ID management in the modern society. With several advantages compared to other biometrics, face recognition has been finding many applications since its birth about two decades ago. A robust and accurate detection of eyes is a crucial step in applications related to face recognition as the eyes are considered one of the most salient and distinguishable features on the human face[1, 2]. Once the two eyes are located reliably, the other facial features (e.g. mouth, nose, ears etc.) can be roughly estimated based on the position of eyes as they have reference values to each other [3].

Many researchers have dedicated their efforts to propose several approaches to the problem. The existing methods can coarsely be classified into 3 categories: template-based approaches, feature-based approaches and appearance-based approaches. As to the template-based approaches, an eye template is firstly generated using the knowledge of the eye shape, and then, the most likely eye position is located by matching the template to the extracted window in the image [4 - 6].

The appearance-based approaches aim at producing a classifier that is learned from a large amount of sample images in different subjects, under different orientations of head and eye itself, as well as different states of illumination using Neural networks or SVM [7 - 9].

The feature-based approaches make use of the distinctive features of eyes such as dark pupil, white sclera, the circular shape of iris and the ellipse-like shape of eyelids, to specify the eyes on human face [10, 11]. Besides, symmetrical characteristics of eye pair are also utilized in a number of eye detection methods [12, 13].

The previous methods try to search for the two eyes on the face image. Verification is commonly done by using the symmetric characteristics of the face or the relative position of the two eyes and mouth. In the latter situation, the detection of mouth is further required. However, detection of eyes encounters difficulty when the face is not captured in the frontal view or when one eye is visible in the face image or in cases the two eyes are not horizontally level.

In this paper, a template-feature based eye-verifier is proposed for detection of eyes in facial color images. The proposed method is highly appropriate for the new generation of high quality facial photos used for biometric passport, recommended by International Civil Aviation Organization Report [27]. The proposed method consists of two stages: eye region candidate detection and eye verification. For the eye region candidate detection, the face region is firstly detected by using skin color in the HSV and YCrCb color spaces. Thereafter, only the face region is cropped for further processes. The circular filter [14] is applied to the binary face image after binarization of the face region image. The eye region candidates are selected from the peaks in the filtered image.

The eye candidates are then fed into an eye-verifier. A ternary eye mask is generated for each eye candidate by overlapping three sub-regions: iris, sclera and the skin around eye and then compared with a ternary eye template which is generated on the basis of geometric features of an eye. Verification is performed by measuring the ternary Hamming distance between the template and the ternary eye mask to decide whether the eye candidate is truly eye or not. After the eye is verified, the detailed information about the eye such as pupil center and eye corners are all retrieved. The algorithm diagram can be figured out as in Fig. 1 below.

The remainder of this paper is organized as follows. Section II presents the detection of eye candidates by applying circular filters. The main section describing the eye-verifier is explained in Section III. Section IV illustrates the experimental results. Conclusion and future work are described in Section V.
II. EYE REGION CANDIDATE DETECTION

A. Face detection and binarization

Detection of face region is done by using skin color information that is known as one of the most important features for face localization. It is said that skin color after removing the brightness distributes in a small area of the chromatic color space [15]. Motivated by the work of Kim, et al. [16], instead of detecting skin color directly in RGB color model, Hue component of HSI color model and C_b, C_r components of YC_bCr are used. The decision rules originate from Sobottka, et al. [17] (for the hue image) and Hsu, et al. [18] (for the C_b, C_r images). Pixel \((i,j)\) is assigned as skin color pixel if

\[
\begin{align*}
0 & \leq \text{hue}(i,j) < 50^0, \\
137 & \leq C_r(i,j) < 177, \\
77 & \leq C_b(i,j) < 127, \\
190 & < C_b + 0.6 \times C_r < 215
\end{align*}
\]

Thereafter, small connected components in the map are removed by simple morphological operations to obtain the face mask. After the face region is detected, only the face region is cropped for further processes.

The face binarization is a minor modification of Khosravi and Safabakhsh [22]. By observing that the eyes are dark regions on the face, the morphological closing operation is applied to the grayscale face image. Then the difference between the closed image and the grayscale face image is computed. The size of the structuring element \(R\) is estimated by the same method. The difference image exhibits outstanding pixels for those in the low-intensity regions such as eyes, eyebrows, nostrils or beards. In order to generate the binary image, the threshold is selected by considering the ratio of the eye region in the human face. Thus, 5 percent of the highest intensities in the difference image are considered to belong to the low-intensity areas. By normalizing the difference image to \([0, 255]\), the binarization process can be presented as

\[
\begin{align*}
\Gamma_b(x, y) = \begin{cases} 
1 & \text{if } I_{\text{diff}}(x, y) > \text{th} \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

where \(cH\) is the cumulative histogram of the difference image \(I_{\text{diff}}\). Then, the morphological operations, dilation and erosion, are utilized to fill the reflections in the pupil regions (if present) and remove small connected components. In order to increase the discriminating power of the circular filter subsequently applied, the binary face image is re-assigned:

\[
\Gamma_b'(x, y) = \begin{cases} 
1 & \text{if } \Gamma_b(x, y) = 1 \\
-1 & \text{otherwise}
\end{cases}
\]

B. Eye candidate region detection

With the circular shape of eye in the binary face image, the eye-candidates are located by convolving the binary face image with a series of circular filters. Different filters at different radii (namely, 4 filters in our experiments) are assessed and the measure in Eq. (4) is used to find out the most fitting filter that corresponds to the maximum \(k\).

\[
k = \frac{\text{Max}(I_f)}{s_f}
\]

where \(\text{Max}(I_f)\) is the maximum value of the filtered image and \(s_f\) is the size of the filter. The radii of the candidate filters \(R_{cf}\) are selected in a range \(R - \delta < R_{cf} < R + \delta\). An example of the filtered image can be seen in Fig. 2 below as the eye-peaked image.

Four peaks corresponding to local maxima are extracted to feed into an eye-verifier module. If the two eyes were already found before the four peaks are verified, the process finishes. The size of the eye window to be extracted is \(4 \times R_{cf}\) (height) \(\times 8 \times R_{cf}\) (width) centered at the peak, to assure that the eye components are included in the window, where \(R_{cf}\) is the radius of the selected circular filter. The overall diagram for detection of eye region candidates is depicted as Fig. 2 below.

III. EYE-VERIFIER

A. Ternary eye-mask generator

Under low illumination, the separability between the iris, sclera and skin classes is degraded. Thus, the light compensation method proposed in [19] is used to enhance the separability, since for the eye window, the two assumptions of the method are satisfied. Due to small size of eye window, the window is up-sampled to three times using the pixel interpolation. While interpolation enlarges the window size, it also makes each class more homogeneous. Both effects are
beneficial for the eye-mask generation process. Then, a small area at the bottom of the eye window is referred to the skin sample area.

Fig. 3. Skin sample area with a height of 1/3 of pupil radius in the eye window

1) Iris mask detection: The iris mask in this paper includes pixels that are in the pupil or iris region of the eye. Comparing to the other parts of the eye window, those pixels have the lowest intensities, even for African descent. From the skin sample area, the averages of RGB components are computed, and the 3D distance map is computed using the Euclidean distance as

\[ I_{3D}(x, y) = \sqrt{\sum_{\kappa=\text{red, green, blue}} (I_{EW}^\kappa(x, y) - m^\kappa)^2} \]  

where \( I_{EW}^\kappa \) refers to each color component of the eye window image and \( m^\kappa \) indicates the intensity mean value of each color component of skin. It is obvious that the iris class is more separable from the skin class in color space. As a result, the high intensities in the 3D image concentrate in the iris region.

After normalizing \( I_{3D} \) to \([0, 1]\), the following rule is used to detect an iris mask \( M_{iris} \),

\[ I_{3D}^\text{Norm}(x, y) = \frac{I_{3D}(x, y) - \min_{p\in\text{iris}} (p(i, j))}{\max_{p\in\text{iris}} (p(i, j)) - \min_{p\in\text{iris}} (p(i, j))} \]

\[ M_{iris}(x, y) = \begin{cases} 1 & \text{if } I_{3D}^\text{Norm}(x, y) > 0.6 \\ 0 & \text{otherwise} \end{cases} \]

Eq. (8) explains that the pixels in the iris region should have a long 3D distance to the skin and low intensity; however, the eyelids are sometimes visible in the mask. Therefore, a morphological closing operation with a small size structuring element is utilized to remove them. In addition, the morphological operation can also fill the hole in the iris region that represents specular reflections. After all, the biggest connected region is kept to be considered as the iris mask.

For a fine estimation of the iris parameters (i.e., the center and radius), the iris mask is re-assigned as following in order to be adaptive to a semi-circular filter,

\[ M'_{iris}(x, y) = \begin{cases} 1 & \text{if } M_{iris}(x, y) = 1 \\ -1 & \text{otherwise} \end{cases} \]

Then, the center and radius of the iris region \((x_{iris}, y_{iris})\) and \(r_{iris}\) is determined by applying a semi-circular filter (as depicted in Fig. 5) to the iris mask. The reasoning for using a semi-circular filter instead of a circular filter is that the upper half of the iris is very often occluded.

Fig.5. Semi-circular filter

2) Sclera mask detection: In order to find sclera, the saturation image in the HSV color space is utilized. Regardless of the color of skin, the pixels belonging to sclera in the saturation image are always in low-intensity due to its white color. A morphological closing operation with a structuring element with the size of iris’s radius is applied to the saturation image, and then the resulted image is subtracted from the saturation image to generate the difference image. Compared to the saturation image, the sclera class in the difference image to the skin class is more separable. Then, the threshold for binarizing the difference image is defined by

\[ th_s = m_{sk} + 4\sigma_{sk} \]

where \( m_{sk} \) is the mean value of the pixels of difference image in the skin sample area and \( \sigma_{sk} \) is their standard deviation. Assuming the skin pixels in the difference image has a normal distribution; the \( 4\sigma_{sk} \) distance can guarantee occupancy of nearly all skin pixels. Then, the raw sclera mask is found by

\[ M_{sk}(x, y) = \begin{cases} 1 & \text{if } S_{\text{diff}}(x, y) > th_s \\ 0 & \text{otherwise} \end{cases} \]

where \( S_{\text{diff}} \) is the difference image. The raw sclera mask generated by Eq. (10) may include pixels from iris and specular reflections in the iris region.

Fig.6. Illustration of sclera mask detection: a) the eye window; b) the saturation image; c) the difference image; d) the raw sclera mask; e) the \( E_{eye} \) image; and f) the sclera mask
Therefore, an Eye image is constituted from the raw sclera mask and iris mask by $E_{eye} = M_{iris} \| M_{sc}$ where $\|$ indicates the OR operation. The $E_{eye}$ image is smoothed by morphological dilation and erosion operations, and then the largest connected region is kept. The sclera mask is then refined by $M_{sc} = E_{eye} - M_{iris}$.

3) Eye-mask generation: The ternary eye-mask is the overlapping of the three classes (i.e. iris, sclera and skin) with different amplitudes for each class: 0 for iris, 0.5 for skin and 1 for sclera. The ternary eye-mask having the same size as the masks is generated by

$$TEM(x,y) = \begin{cases} 
0 & \text{if } M_{iris}(x,y) = 1 \\
1 & \text{if } M_{sc}(x,y) = 1 \\
0.5 & \text{otherwise}
\end{cases}$$

From the binary Eye image, eye features (e.g. eye corners and eye height) can be determined. The leftmost and rightmost edge points in $E_{eye}$ are considered to be the eye corners. Based on the two corners, the $TEM$ image is rotated so that the two corner points are horizontally level and then, it is cropped to compact the eye to be ready for matching as shown in Fig. 7 below.

![Fig. 7. a) TEM image, and b) ternary eye-mask after cropping.](image)

**B. Ternary eye-template generation**

In modeling an eye template, an eye usually consists of a circle that represents the boundary of the iris to the surroundings, and parabolic curves to represent the upper and lower eyelids [4, 20, 21]. From the binary Eye image, parameters needed to represent an eye can be obtained. In this method, a ternary eye template having the same size as the eye-mask is generated that also consists of three components: iris, sclera and skin. The upper eyelid is modeled by a parabolic

$$y = h_U - \frac{4*h_U}{w^2} x^2;$$

and the lower eyelid by

$$y = \frac{4*h_L}{w^2} x^2 - h_L.$$  

The region inside the two boundaries is assigned with the value 1. The iris circular region is modeled by a circle of radius $r_{iris}$ centered at the same center position in the eye-mask by

$$(x - x_{iris})^2 + (y - y_{iris})^2 = r_{iris}^2.$$  

The iris region bounded by the two eyelids and the circle is assigned with the value 0. The remaining pixels are assigned with 0.5 to represent skin. The ternary eye-template is as depicted in Fig.8 below.

![Fig.8. Ternary eye-template](image)

**C. Hamming distance-based ternary eye matching**

Hamming distance measures the disagreements of symbols in corresponding positions between two strings of equal length. For verification of the eye candidate, the ternary Hamming distance is computed to assess the similarity between the ternary eye-mask and the ternary eye-template as

$$HD = \|TEM \oplus_T TET\|$$

where $TET$ is the ternary eye-template. The operation $\oplus_T$ indicates a ternary XOR-like that is defined as

$$p \oplus_T q = \begin{cases} 
0 & \text{if } p = q \\
1 & \text{otherwise}
\end{cases}$$

**IV. EXPERIMENTAL RESULTS**

The proposed algorithm is evaluated on a collection of 1,000 face images from the color FERET face database [24, 25] and 353 facial images collected by the authors. The collection from the FERET database includes frontal and half-profile images and involves European, Asian and African descents.

The experiment on this collection shows that the eye-verifier is sensitive to the quality of the face images, especially in the eye regions since the proposed method requires the distinction between the three classes of sclera, pupil and skin. The detection rate is 87.2%, and the main reason for failed cases is the failure in skin detection. In all the experiments, a Hamming threshold of 0.4 is used for verification.

![Fig.9 Examples from FERET database](image)
In order to illustrate the robustness of the proposed method to the variety of eye directions and face poses, a face database is created from 10 volunteers. For each volunteer, 35 to 40 images were captured to cover almost all face and eye directions such as face roll, yaw, pitch, expressions and eye look-left, look-right, look-up, look-down, and different illumination. Also the database includes the cases in which only one eye is visible due to wearing eye-patch or one eye is closed as depicted in Fig. 10. The summary results are reported in table 1 below. The result of the proposed method is not only the eye positions but also the accurate eye feature points which are valuable for evaluating eye conditions.

Table 1: Performance summary on our face pose database

<table>
<thead>
<tr>
<th>Image type</th>
<th>No. of images</th>
<th>Detection rate (%)</th>
<th>Error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head poses (roll, yaw, pitch)</td>
<td>108</td>
<td>87.04</td>
<td>1.85</td>
</tr>
<tr>
<td>Eye directions (look left, right, up, down)</td>
<td>80</td>
<td>97.5</td>
<td>0</td>
</tr>
<tr>
<td>Expressions (neutral, smile, surprise, sad, eye squint, angry, mouth open)</td>
<td>70</td>
<td>92.86</td>
<td>4.28</td>
</tr>
<tr>
<td>Wearing eyeglasses</td>
<td>95</td>
<td>85.26</td>
<td>3.15</td>
</tr>
</tbody>
</table>

Fig. 11 shows some examples of failed detection. The proposed method failed to verify the eyes in cases the face image is nearly profile, then the eyes are deformed and different from the template. Besides, the eye-verifier can not work for cases the eye is too narrowly opened to detect sclera and pupil.

V. CONCLUSION

The geometric symmetry information is widely used in eye detection methods for verification of eye pairs. However, this information is sensitive to the pose of face and requires the appearance of both two eyes. This paper proposes an eye verification method that verifies each eye separately based on the knowledge of the eye geometry and the eye features. The proposed method showed a robust performance to the variety of the face poses and eye directions. However, the method is still not robust enough to overcome the difficulty of bad quality images. Further effort will be dedicated to solve this problem and then to evaluate the method on more public databases to have a thorough comparison to other methods.

REFERENCES


