Eye Detection in the Middle-Wave Infrared Spectrum: Towards Recognition in the Dark

Thirimachos Bourlai[1] and Zain Jafri[2]

Lane Department of Computer Science and Electrical Engineering, West Virginia University
Evansdale Drive, Morgantown, WV 26506-6070, U.S.A.
[1] `Thirimachos.Bourlai@mail.wvu.edu`
[2] `zjafri@mix.wvu.edu`

Abstract In this paper, the problem of eye detection in the Middle-Wave Infrared (MWIR) spectrum is studied in order to demonstrate the importance of performing eye detection in the thermal band. While currently there are methods that are capable of performing automatic eye detection efficiently in the visible and active infrared (IR) spectrum (i.e., Near-IR and Short-Wave IR), eye detection in the thermal band is a very challenging problem. This is because in the thermal domain limited features can be extracted from the eye region, mainly eyelashes and eyebrows, while features such as human irises, pupils, and superficial blood vessels of the conjunctiva are not clear. Our proposed eye detection method operates in the MWIR band by combining a set of methodological steps such as face normalization, integral projections, and template-based matching. In this paper, a face problem of thermal-based eye detection using still frontal face images is being investigated.

I. INTRODUCTION

Thermal imaging cameras are devices capable of detecting radiation in the passive infrared range of the electromagnetic spectrum (see Fig. 1) and produce images of that radiation. They are particularly useful for law enforcement, enhancing the police effectiveness as powerful tools in the fight against crime. Practically, this technology can be employed on the street for night time operational scenarios (from drug activity surveillance to evidence retrieval) to make the job of law enforcement officers easier and safer. These cameras do not require external illumination, allowing the officers to identify potential threats in total darkness, from a distance, without revealing their location. They are also used by tactical teams to gain an additional surveillance advantage, especially in low-light or smoky conditions.

MWIR cameras, when compared to Long-Wave IR (LWIR) cameras, are considered more suitable to see at longer ranges [1], while LWIR imagery has been used for face recognition (FR) with promising results [2], [3]. MWIR imaging was also used for determining the identity of an individual (DARPA Human ID Project 2001-2005), but did not yield promising results (due to e.g., sensor limitations available at that time, environmental factors etc. [4]). Recently, MWIR sensors have significantly improved and they reveal different image characteristics of the facial skin when compared to LWIR sensors. The importance of using thermal sensors by law enforcement officers is significant, e.g., they are used not only for night time detection of human activity, but also for collecting mug shots in the thermal band[1]. Such mug shots can be used for identity verification by first, applying eye detection, and then, face normalization techniques.

978-1-4577-1019-3/11/$26.00 ©2011 IEEE.

and the horizontal integral projection of the right face side 2D region $A^R(i)$ of $I(x;y)$, denoted by $P^H_{A^R(i)}$, is a discrete and finite 1-D signal given by the following equation:

$$P^H_{A^R(i)} : \{x^R_{min}; \ldots; x^R_{max}\} \to \mathbb{R};$$

$$P^H_{A^R(i)}(x) := |A^R_x(i)|^{-1} \sum_{y \in A^R_x(i)} I(x;y)$$

where x_{min}, x_{max}, and $A_x(i)$ for each face side (A^L and A^R) are defined as follows:

$$x^L_{min} = \min(x) \text{ of } (x;y) \in A^L(i)$$

$$x^L_{max} = \max(x) \text{ of } (x;y) \in A^L(i)$$

$$x^R_{min} = \min(x) \text{ of } (x;y) \in A^R(i)$$

$$x^R_{max} = \max(x) \text{ of } (x;y) \in A^R(i)$$

$$A^L_x(i) = \{y = \forall y; (x;y) \in A^L(i)\}$$

$$A^R_x(i) = \{y = \forall y; (x;y) \in A^R(i)\}$$

The sets $\{x^L_{min}; \ldots; x^L_{max}\}$, and $\{x^R_{min}; \ldots; x^R_{max}\}$ are defined as the domains of the horizontal integral projections of the left and right face side of image $I(x;y)$ respectively.

We know that the local minimum and maximum values of $P^H_{A^L(i)}$, and $P^H_{A^R(i)}$ equate to facial features of either face side respectively. To perform these integral projections, first, horizontal log edge detection is applied to each face side of the input thermal face image, resulting in a binary image with all horizontal edges detected. Then, for each face side, the highest frequency in the radon transform corresponds to the area with the most horizontal lines. This area determines the location of each eyebrow (see Fig. 3(e)). Note that integral projections can be applied on any direction, thus they can be adopted to handle tilted faces. The main problem though is handling noise formed due to human hair.

(e) Eye Detection: The generated eye or ocular templates (see Fig. 4) are passed either (a) through each face side of an input face image, or, (b) when integral projections are employed, through the reduced search space that was determined by the location of the detected left and right eyebrows. The similarity score between the average eye or ocular templates and a searched region is calculated using the Pearson Product Moment correlation coefficient (r). This measure is defined in equation 2, where X and Y are the image and template pixel intensity values, respectively, N is the total number of pixels, σ_X and σ_Y are their respective standard deviations, and μ_x and μ_y are the expected values of x and y, respectively.

$$r = \frac{\sum_{i=1}^{N}(X_i - \mu_{X_i})(Y_i - \mu_{Y_i})}{N \sigma_X \sigma_Y} \quad (2)$$

We evaluate this measure exhaustively throughout the face until the coordinates of the two highest correlation coefficients are found [15], i.e., one for each eye (Fig. 3(h)).

(f) Geometric Normalization: A geometric normalization scheme was applied to images acquired after face detection. The normalization scheme compensates for slight perturbations in the frontal pose, and consists of eye detection and affine transformation. When using the visible dataset, automated eye detection was performed. The method used was based on a template-based matching algorithm where the coordinates of the eyes were automatically obtained [16]. Traditional eye detection techniques were used but did not work when employing the face image dataset acquired in the thermal band. Thus, the eyes centers were located by our automatic eye detection approach. When the eyes were not found, we used manual annotation to indicate the eye centers. In either the visible or the MWIR dataset, after the eye centers were found, the canonical faces were automatically constructed by applying an affine transformation. Eventually, all faces were warped to the same dimension of 150 × 130 pixels. Then, these canonical face images were used for the FR experiments.

(g) Face Recognition Methods: Both commercial and academic software was employed to perform the FR experiments: (i) Commercial software Identity Tools G8 provided by L1 Systems (www.l1id.com); and (ii) CSU Face Identification Evaluation System [17]: Principal Components Analysis (PCA) [18]–[20], a combined PCA and Linear Discriminant Analysis algorithm (PCA+LDA) [21], and the Bayesian Intra-personal/Extra-personal Classifier (BIC) using the Maximum Likelihood (ML) hypothesis [22].

IV. EXPERIMENTAL SCENARIOS

By using the visible and MWIR images in the assembled database, four experiments were performed: (i) eye detection in the visible dataset of the WVU-VT face database (baseline); (ii) eye detection in the MWIR dataset of the WVU-VT face database, (iii) face identification using visible face images, i.e., Visible vs. Visible (baseline) test; and (iv) face identification using MWIR face images, i.e., MWIR (indoor) vs. MWIR (indoor or outdoor) test (see Table I and Fig.6).

In the first experimental scenario, our eye detection method proposed in [16] was used to establish a baseline for comparison. Then, our new proposed MWIR eye detection method is evaluated, when using different eye templates, before and after applying the integral projections methodological steps (see Fig. 3 (e) and (f)). To validate the performance of our eye detection system, we used the relative error measure (D_{eye}) based on the distances between the expected (true eye positions acquired by manual annotation), and the estimated (using our method) eye positions [23]. First, for each eye, we compute the distance between the true (manually annotated) eye center $C_l; C_r \in \mathbb{R}^2$, and the estimated one $\tilde{C}_l; \tilde{C}_r \in \mathbb{R}^2$ (Fig. 5), i.e., d_l for the left and d_r for the right eye. Then, we determine the maximum distance between d_l and d_r, i.e., $\max(d_l; d_r)$, which is, finally, normalized by dividing it by the distance between the expected eye centers, denoted as $\|C_l - C_r\|$. The measure is shown in the following equation:

$$D_{eye} = \frac{\max(d_l; d_r)}{\|C_l - C_r\|} \quad (3)$$

Based on the fact that in an average human face the distance between the inner eye corners equals the width of a single eye,

Fig. 5. (a) The relations between true (C_l and C_r) and estimated eye positions ($\tilde{C}_l$ and $\tilde{C}_r$) are illustrated; (b) The relative error with respect to the right eye (left in Fig. (b)). A circle with a radius of 0.25 relative error is drawn around the eye center. Image courtesy [23].

a relative error of 0.25 equals a distance of half an eye width (see Fig. 5(b)). In this paper, an eye is considered detected if D_{eye} $0{:}25$, and is rejected otherwise.

In the second experimental scenario, visible face images (probes) were matched against visible images (gallery) for the purpose of establishing a baseline for comparison. Then, intra-spectral matching is performed, where we compared MWIR against MWIR face images acquired on the same session (indoors vs. indoors) and on a different session (indoors vs. outdoors). The same standoff distance of 6.5 feet (2m) was used for both sessions, and the faces were, finally, normalized after using our eye detection and face normalization steps. The identification performance of the system was evaluated through the cumulative match characteristic (CMC) curve. The CMC curve measures the $1 : m$ identification system performance, and judges the ranking capability of the identification system.

A. Eye Detection Results

For the eye detection experiments, only the full frontal still face images from the MWIR face dataset (of the WVU-VT database) were used. The detection accuracy of both the left and right eye was calculated by the number of accurately detected left or right eyes divided by the total number of left or right eyes. Both the generated average (left and right) eye templates, as well as the average (left and right) ocular templates (see Fig. 4) were employed for template-based matching, and two main scenarios were considered. First, all steps (from (a) to (h)) of the proposed eye detection methodology were used. Second, from the proposed eye detection methodology we excluded step (e), i.e., the usage of the integral projections method, and step (f), i.e., the application of integral projections that results in a limited search space where the template-based matching was applied. The purpose of investigating these experimental scenarios was to identify which scenario results in the highest eye detection performance.

Table I summarizes the results, where we can see that the average detection performance of our method is 90.95% (vs. 99.8% average eye detection performance when using the visible dataset of the WVU-VT database). We know that when operating in the thermal band, apart from the human eyelashes and eyebrows, there are no other clear features in the ocular region that can be extracted (see Fig. 1). Thus, when template matching was performed using the eye templates, the average eye detection results were very low (24%). By choosing an ocular template (that includes the eyebrow region) the average eye detection accuracy increased 2.35 times. Another issue that was investigated was whether the usage of integral projections can improve the eye detection accuracy of our method. The experimental results showed that integral projections improve eye detection accuracy by 2.75 times when using eye templates, and by 1.36 times when using ocular templates. This is because the usage of integral projections limits the search space where the template-based matching is applied, and thus, we have fewer occurrences of falsely detected eyes.

TABLE I

EYE DETECTION ACCURACY (%) AFTER APPLYING OUR PROPOSED APPROACH, WHEN USING DIFFERENT EYE TEMPLATES, BEFORE AND AFTER APPLYING THE INTEGRAL PROJECTIONS METHODOLOGICAL STEPS. L/R=LEFT/RIGHT; IP=INTEGRAL PROJECTIONS

Target	Eye/NoIP	Eye/IP	Eyebrow/NoIP	Eyebrow/IP
R Eye	23.33	65.71	67.62	92.38
L Eye	24.29	65.24	66.67	89.52

B. Face Identification Results

The results of the baseline (when operating in the visible spectrum) experiments are illustrated in Fig. 6. We can see that when G8 was used, the identification rate was 100% at Rank-1. When appearance-based techniques were used (e.g., PCA, LDA) a lower performance was achieved (e.g., PCA resulted in 71% identification rate at Rank-1).

The face identification results when we compared MWIR to MWIR (indoor vs. indoor; indoor vs. outdoor sessions) face images are summarized in Fig. 6. The main conclusion here is that FR in MWIR spectrum is comparable to traditional FR in the visible spectrum (e.g., for MWIR we achieved 100% identification rate at Rank-2). Thus, in theory, MWIR FR can be considered superior to visible FR due to the fact that similar performance can be obtained when using MWIR face images acquired in complete darkness. A problem is noticed when using outdoor thermal images as probes, i.e., we have a 25% performance drop across all used algorithms. This is mainly due to outdoor environmental effects on the appearance of the thermal faces. Another conclusion is that appearance-based FR approaches (e.g., PCA) perform much better in the MWIR spectrum, even in the indoor vs. outdoor session test. This was expected due the fact that MWIR imaging is illumination invariant. A similar conclusion was reported in [24] where PCA was employed when using visible and Long-Wave IR (thermal) face images.

V. CONCLUSIONS

In this paper we discussed the problem of eye detection and its effects in FR when operating in the MWIR spectrum. A set of experiments was performed on the WVU-VT dataset. The experimental results show that when using MWIR still