

Multispectral Eye Detection: A Preliminary Study

Cameron Whitelam Zain Jafri Thirimachos Bourlai

West Virginia University, LDCSEE, 6201, Morgantown, WV, 26505, U.S.A

Email: {cwhitela, zjafri}@mix.wvu.edu {Thiramachos.Bourlai}@mail.wvu.edu

Abstract—In this paper the problem of eye detection across three different bands, i.e., the visible, multispectral, and short wave infrared (SWIR), is studied in order to illustrate the advantages and limitations of multi-band eye localization. The contributions of this work are two-fold. First, a multi-band database of 30 subjects is assembled and used to illustrate the challenges associated with the problem. Second, a set of experiments is performed in order to demonstrate the possibility for multi-band eye detection. Experiments show that the eyes on face images captured under different bands can be detected with promising results. Finally, we illustrate that recognition performance in all studied bands is favorably affected by the geometric normalization of raw face images that is based on our proposed detection methodology. To the best of our knowledge this is the first time that this problem is being investigated in the open literature in the context of human eye localization across different bands.

Keywords-Multispectral Imaging; Eye Detection; Pupil Detection; Face Recognition

I. INTRODUCTION

Within the last two decades, we notice improvement in the performance of face recognition (FR) systems in controlled conditions characterized by suitable lighting and favorable acquisition distances. However, over the years the technology has steadily progressed to tackling increasingly more realistic conditions rather than adequately handling only well-controlled imagery. Most related research emphasizes maintenance of high recognition performance while coping with increased levels of image variability.

Among the most insidious problems for visible-spectrum-based FR algorithms are (1) the variation in level and nature of illumination, and (2) the fact that as the level of illumination decreases, the signal to noise ratio rises quickly, and thus automatic processing and recognition become impossible. In order to address these issues recent research has moved into the use of infrared imagery (e.g., intensified near-infrared (NIR) [1], Short Wave IR [2], Long Wave IR [3]). Table I summarizes the electromagnetic bands of interest, their wavelength range, and illumination sources.

The reflected IR spectral bands provide advantages for a solution to the face detection problem due to the fact that human skin shows unique reflectance characteristics and facial signatures are less variable in these spectral bands. Also the reflected IR light is, for the most part, invisible to the human eye so the system can remain unobtrusive and covert. For example when using SWIR spectrum for FR,

Table I: Imaging Ranges of Interest for Day and Night Environments.

| Band | Wavelength Range (nm) | Illumination | Viability |
|---------|-----------------------|--------------------------------------|--------------|
| Visible | 450 - 750 | Ambient light Broadband sources | High High |
| NIR | 700 - \approx 1000 | AlGaAs LEDs, LDs | High |
| SWIR | \approx 1000 - 2500 | InGaAs LEDs, LDs | High |
| LWIR | 7K-14K | Subject reflected ambient thermal | High |

the benefits are: (a) usefulness in a night time environment, (b) usefulness in detecting disguised faces due to unique and universal properties of the human skin in this sub-band (SWIR upper band, i.e., 1.4-2.5 μ m), and (c) SWIR imagery can be combined with visible-light imagery to generate a more complete image of the human face.

The scope of this study is motivated by the aforementioned issues. However, FR robustness should be further supported by correct eye localization as the fundamental step for the initialization of most of the commercial and academic FR techniques. The problem is that the localization error influences the baseline techniques in a non linear way, and as a result, their accuracy can rapidly decay as the localization quality decreases [4].

Many algorithms have been reported to efficiently detect the presence of an eye. In the visible spectrum, the algorithm proposed in [5] uses Circular Hough Transforms to detect the circular shape of the eye, resulting in accuracy no greater than 86%. In [6] rectangular as well as pixel-pattern-based texture features (PPBTF) are used for eye detection, achieving a detection accuracy of 97%. Also in [7], probabilistic classifiers are employed to separate eyes and non-eyes. Multiple classifiers are then combined in AdaBoost to form a robust and accurate eye detector that overall achieved a 94.5% accuracy.

In the infrared spectrum, Dowdall et al. [8] used the lower band of the IR illumination to find the eyes since it is darker than the rest of the face, and the upper IR band to find the eyebrows, considering that the eyebrows reflect light extremely well in this range. The algorithm proposed in [9] operates also in the IR spectrum, and uses Kalman filtering and the mean shift tracking algorithm to track and detect eyes in real time. All aforementioned methods operated in either the visible or the IR spectrum, but were not designed to efficiently operate in both spectra.

In this paper we propose a multi-spectral eye detection methodology that further detects the pupil center as well. It is tested on both the West Virginia University Multispectral (WVUM) database, as well as a subset of the FRGC database [10]. Three different experiments have been performed. The first one investigates the detection accuracy of our method on FRGC and on each of the three WVUM datasets (visible, multispectral, and SWIR). In the second experiment, our method is tested against a commercially available eye detection software¹. The third experiment investigates the effect of geometric normalization - triggered by our eye detection methodology - to FR accuracy.

The rest of this paper is organized as follows. Section 2 describes the experimental setup. Section 3 provides our detection methodology. Finally, Section 4 describes the experiments performed and discusses our results, before conclusion are made in Section 5.

II. DATABASES

(a) **WVUM**: Three different types of cameras have been used to create this database. All face images were captured in a controlled environment and adhered to the level 40 NIST “head and shoulder” acquisition scene constraint²:

- **Canon EOS 5D Mark II**: This digital SLR camera (www.canon.com) has a 21.1-megapixel full-frame CMOS sensor with DIGIC 4 Image Processor, and a vast ISO Range of 100-6400. It is used to obtain standard RGB, ultra-high resolution frontal pose face images (see Fig. 1(a)).

- **DuncanTech MS3100**: This camera (Fig. 1(c)) incorporates 3 CCD and three band-pass prisms behind the lens to simultaneously capture four different wavelength bands. The IR and Red (R) sensors of the multispectral camera have spectral response ranges from 400nm to 1000nm. Also, the Green (G) channel has a response from 400nm to 650nm, and the Blue (B) channel from 400nm to 550nm. Note that the *IR and Red* sensor outputs an image of size 1392x1040. The G and B images are recorded on a RGB Bayer pattern sensor and are, therefore, one-third the resolution of the other images. Then G/B images are interpolated to have the same resolution as the IR and R images.

- **XenICs**: This camera (Fig. 1(d)) has an Indium Gallium Arsenide (InGaAs) 320×256 Focal Plane Array (FPA) with $30\mu m$ pixel pitch, 98% pixel operability and three stage thermoelectric cooling. It has a relatively uniform spectral response from 950 - 1700 nm wavelength (lower SWIR band) across which the InGaAs FPA has largely uniform quantum efficiency. Response falls rapidly at wavelengths lower than 950 nm and near 1700 nm.

(b) **FRGC**: In order to extend our visible spectrum sample size, we used a subset of the Facial Recognition Grand Challenge database. It consists of frontal face images at constrained poses of 84 male and 63 female subjects, resulting

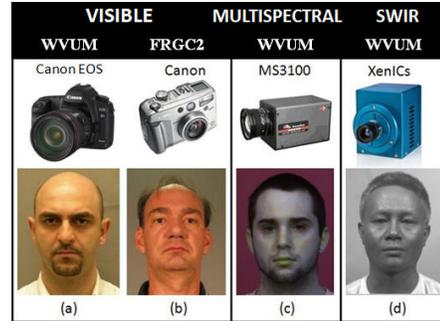


Figure 1: Face samples of the (a) WVUM visible, (b) FRGC visible, (c) WVUM Multispectral, and (d) WVUM SWIR datasets.

in a total of 458 images. The images were acquired by a 10 megapixel Canon PowerShot G2 camera (Fig. 1(b)), and they are in JPEG format.

III. METHODOLOGY

In this section we outline the technique we employ to perform eye and pupil localization (see Fig. 2). The salient stages of the proposed method are described below:

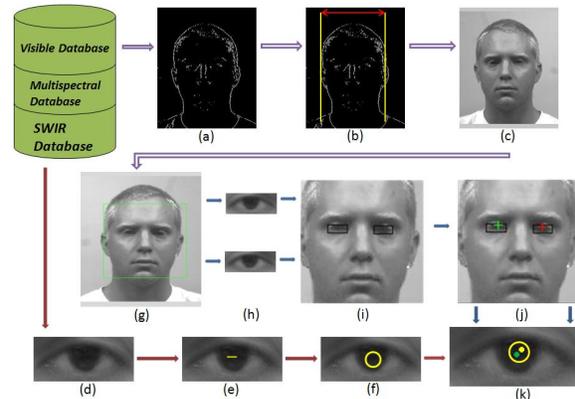


Figure 2: Overview of the methodology used to perform eye and pupil detection. (a) Sobel Edge Detection; (b) Average Face Width Estimation; (c) Face Resizing; (d) Eye Template Generation; (e) Av. Pupil Estimation; (f) Pupil Acceptance Circle; (g) Face Detection; (h) Left/Right Eye Scanning; (i) Eye Detection; (j) Pupil Detection; (k) Estimation of Pupil Detection Accuracy.

- **Face Width Estimation**: Before performing eye detection, all images must be resized to the same scale so that the face width (or the inter-ocular distance) is relatively constant across all subjects. The purpose of this step is to achieve the image normalization required for our algorithm to obtain consistent results. Thus, we first apply on each face the Sobel edge detection algorithm (Fig. 2(a)). Then, for each column we calculate the maximum number of white pixels (representing an edge) to identify the column positions on the x-axis where the left (X_l) and right (X_r) side of each face are. These two positions determine a face width estimation W_e (Fig. 2(b)), that can be calculated by subtracting the left from the right edge column value ($W_e = X_r - X_l$ pixels). This information is used to resize the face (Fig. 2(c)) to a

¹www.neurotechnology.com

²http://fingerprint.nist.gov/standard/Approved-Std-20070427.pdf

pre-determined width (W_p) (generated for the creation of an average face - as described in the next stage). If $W_e > W_p$ the image is scaled down. Otherwise it is scaled up.

- *Eye Template Generation*: From each dataset employed we randomly select a certain number of subjects and manually crop both right and left eyes. We then average the cropped eyes to form the right and left eye templates (Fig. 2(d,h)). We empirically identified that 10 subjects are enough to generate templates that can achieve satisfactory eye detection results.

- *Pupil Acceptance Circle*: For each generated eye template we estimate the pupil diameter (Fig. 2(e)), and create an accuracy circle centered at the pupil center that is a few pixels wider than the estimated diameter. This is to compensate for varying pupil diameters caused by exposure to light as well as to negate any human error there might have been when manually selecting the pupil (Fig. 2(f)), and (Fig.3).

- *Face Detection*: The Viola & Jones face detection algorithm (Fig. 2(g)) is used to localize the spatial extent of the face and determine its boundary. The algorithm was observed to perform reasonably well on the face images acquired in this work.

- *Eye Detection*: After face detection the four coordinates of the face boundary are located. These boundaries assist in the application of *template convolution* by first placing each of the generated eye templates to the top left corner of each face, and then calculate the *Pearson Product Moment correlation coefficient* - (r). This measure is illustrated in *equation 1*, 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.

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

We continue this throughout the face until the coordinates of the two highest correlation coefficient are found, i.e., one for each eye (Fig. 2(i)).

- *Pupil Detection*: Within each detected eye we search for the lowest intensity, i.e. the minimum gray scale value inside the found template, and record its x - y position as the detected pupil location. Then, the positions of the two detected pupils (one for each eye) are compared with that of the manually annotated ones. If the locations identified by our method are found within the *pupil acceptance circle* then they are considered as correctly detected (Fig. 2(j)). Otherwise they are rejected. Then, the number of identified pupils is summed up and the acceptance percentage is computed. Finally, for each we estimate the Euclidean pixel distance between the detected and the ground truth pupil location (Fig. 2(k)).

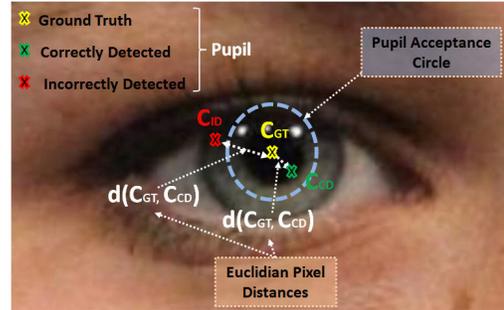


Figure 3: Example diagram of pupil detection: The manually annotated pupil (yellow cross) is compared to the correctly (green cross) and incorrectly (red cross) detected pupils in terms of the Euclidean pixel distance.

IV. EXPERIMENTS

Three different experiments have been performed. In the *first experiment* we investigate the eye and pupil detection accuracy of our method when using each of the WVUM datasets, viz., visible (400nm-700nm), multispectral (400nm-1000nm), and SWIR (950nm-1700nm), as well as the FRGC subset of images in the visible spectrum. Eye detection accuracy is computed for each eye, and is measured as the number of accurately detected eyes divided by the total number of eyes in each dataset employed (see *Table II*). Pupil detection is measured as the Euclidean pixel distance between the true positions of each pupil center (by manual annotation) and the detected pupil. *Fig. 3* illustrates a diagram where the manually annotated pupil (yellow cross) is compared to the correctly (green cross), and incorrectly (red cross) detected pupils in terms of the Euclidean pixel distance.

Table II: Eye detection accuracy after applying the Viola & Jones face detection algorithm, on the FRGC subset in the visible spectrum, as well as on each of the WVUM datasets.

| Target | FRGC | Visible | Multispectral | SWIR |
|-----------|---------|---------|---------------|-------|
| Left Eye | 444/458 | 142/142 | 59/60 | 58/60 |
| Right Eye | 425/458 | 142/142 | 59/60 | 59/60 |

In the *second experiment* the efficiency of our eye detection and pupil detection algorithms is tested against a commercially available software provided by Verilook on all of the aforementioned datasets. Once again, we use the manual annotated eye centers as a ground truth, and compare them to the detected eye centers by employing both our proposed methodology and the commercial software. Experimental results are summarized in *Table III*.

In the *third experiment* we investigate whether face alignment (in terms of scale and rotation) via geometric normalization, which is triggered by our eye detection methodology, affects FR accuracy. The study is performed on each of the WVUM datasets, before and after face normalization. The academic FR techniques we use are the *Principal Components Analysis* (PCA) [11], and the *Linear*

Table III: Comparison of our detection method to that of commercial's software. Detection accuracy (DetA%) is the number of correctly detected eyes. EPD is the Euclidean distance in pixels between the manually annotated pupil centers to that being automatically detected by the commercial and our proposed method. LE/RE=Left/Right Eye.

| Datasets | Criteria | Verilook | Proposed |
|----------------------|-----------|----------|----------|
| Visible (FRGC) | DetA (LE) | 97.14 | 97.58 |
| | DetA (RE) | 98.24 | 93.40 |
| | EPD (LE) | 2.62 | 2.41 |
| | EPD (RE) | 2.68 | 2.47 |
| Visible (WVUM) | DetA (LE) | 100 | 100 |
| | DetA (RE) | 100 | 100 |
| | EPD (LE) | 1.15 | 0.83 |
| | EPD (RE) | 1.50 | 0.86 |
| Multispectral (WVUM) | DetA (LE) | 93.33 | 98.33 |
| | DetA (RE) | 93.33 | 98.33 |
| | EPD (LE) | 4.21 | 4.84 |
| | EPD (RE) | 3.85 | 4.81 |
| SWIR (WVUM) | DetA (LE) | 100 | 95.00 |
| | DetA (RE) | 96.66 | 96.66 |
| | EPD (LE) | 1.44 | 0.98 |
| | EPD (RE) | 2.05 | 1.15 |

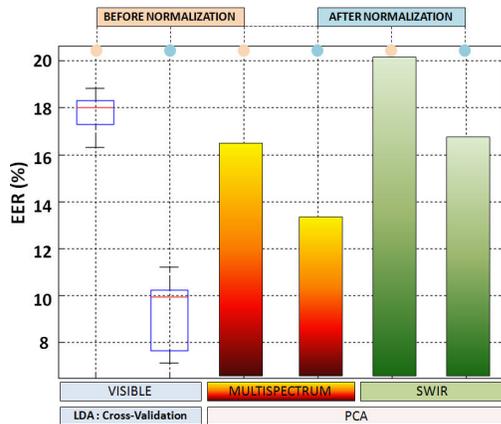


Figure 4: Recognition accuracy in terms of the Equal Error Rate (%), before and after applying face normalization. Note that we applied cross-validation only when using the visible datasets (multiple samples). In the other cases (one gallery/probe sample) we use PCA.

Discriminant Analysis (LDA) [12]. Both PCA/LDA are used in combination with the k-nearest neighbor algorithm (k-NN) [13]. Experimental results are summarized in Fig. 4.

V. CONCLUSIONS

We have presented a study on the problem of eye detection across three different bands, i.e., visible, multispectral, and short wave infrared. The experiments were performed using the WVU Multispectral database, and a subset of the FRGC database. Experimental results show that, unlike previous approaches in the literature that operate only on either the visible or IR spectrum, our eye detection method can be applied across different spectra with promising results (see Table III). Our method's accuracy is also comparable to that of the commercial eye detection software employed. These results are especially important when operating in either ideal (day-time) or non-ideal (night-time) environments.

Finally, related to the face recognition study performed, face alignment significantly increased system performance for both academic algorithms during intra-spectral matching, viz., visible-visible or IR-IR.

ACKNOWLEDGMENT

This work is sponsored in part through a grant from the Office of Naval Research, grant number N00014-08-1-0895, and support from the NSF CITEr, award number IIP-0641331. The authors are grateful to N. Kalka for his valuable assistance.

REFERENCES

- [1] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition—a review. *CVIU*, 97(1):103–135, 2005.
- [2] T. Bourlai, N. D. Kalka, A. Ross, B. Cukic, and L. Hornak. Cross-spectral face verification in the short wave infrared band. In *ICPR*, 2010.
- [3] D. Socolinsky and A. Selinger. A comparative analysis of face recognition performance with visible and thermal infrared imagery. In *ICPR*, pages 217–222, 2002.
- [4] P. Campadelli, R. Lanzarotti, and G. Lipori. Eye localization: a survey. *Biometrical Issue NATO Science Series*, 18:234–245, 2007.
- [5] W.M.K.W.M. Khairoufaizal and A.J. Nor'aini. Eye detection in facial images using circular hough transform. *SP&A*, 1:238–242, 2009.
- [6] H. Lu, W. Zhang, and D. Yang. Eye detection based on rectangle features and pixel-pattern-based texture features. *SISPCS*, 1:746–749, 2007.
- [7] P. Wang, M. Green, Q. Ji, and J. Wayman. Automatic eye detection and its validation. *CSCV*, 1:164–164, 2005.
- [8] J. Dowdall, I. Pavlidis, and G. Bebis. Face detection in the near-ir spectrum. *ICV*, 21:565–578, 2003.
- [9] Z. Zhu, Q. Ji, K. Fujimura, and Kuangchih Lee. Combing kalman filtering and mean shift for real time eye tracking under active ir illumination. *ICPR*, 4:318–324, 2002.
- [10] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Boyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *CSCV*, 1:947–954, 2005.
- [11] M. Turk and A. Pentland. Eigenfaces for recognition. *LCN*, 3(1):7186, 1991.
- [12] P.N. Belhumeur, J. Hespánha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *TPAMI*, 19:4558, 1996.
- [13] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *TIT*, 13(1):2127, 1967.