ON DESIGNING SWIR TO VISIBLE FACE MATCHING ALGORITHMS

Contributors

Cameron Whitelam
Multispectral Imagery Lab,
West Virginia University

Thirimachos Bourlai
Multispectral Imagery Lab,
West Virginia University

Recent advances in facial recognition have trended towards cross-spectrally matching visible gallery face images to probe face images captured under different wavelengths of the electromagnetic spectrum. In this article, we study the problem of matching visible images to images taken in the short-wavelength infrared (SWIR) spectrum, more specifically, the 1550-nm band. There are many benefits to using the SWIR spectrum for face recognition, including covert capturing in nighttime environments as well as imaging through certain environmental conditions such as fog and smoke. However, due to the fact that the moisture in the skin tends to absorb the 1550-nm wavelength, all subjects appear to have dark or black skin tone. Because of the stark contrast between 1550-nm and visible face images, standard face recognition protocols fail to accurately match images captured using sensors operating on different bands. While preliminary work in this area resulted in fairly good performance results, it was determined that a more sophisticated approach could be developed to further improve our original face recognition algorithm in terms of (i) accuracy, (ii) speed, and (iii) adaptability, that is, the proposed algorithm should achieve good results on a wider variety of testing scenarios (diverse face datasets).

More specifically, we study the advantages and limitations of our new proposed cross-spectral matching (visible to SWIR) technique when using an extended set of challenging FR scenarios. The proposed face matching algorithm is a significant improvement when compared to the original algorithm where fused texture-based scores of a large number of photometric normalization combinations between SWIR and visible images were used to achieve satisfactory recognition performance results. Our contributions are threefold. Firstly, multiple databases are considered, which represent different difficult environments, that is, multiband face images were acquired under different lighting conditions and behind different obscurants (multiple levels of tinted glass). Secondly, we demonstrate that the use of a random selection of intensity-based normalization techniques is not necessary. This is because a random combination of such techniques does not have a significant amount of discriminatory information to accurately match one subject's face to another, yielding undesirably low face-matching scores. Thirdly, we demonstrate that a smart selection of a subset of normalization techniques not only results in obtaining more accurate face recognition performance scores, but also drastically decreases the long processing time required to produce even a single face-to-face image match score. Our design also incorporates the usage of parallel processing to further boost the time needed to perform cross-spectral matching. Finally, our experiments indicate that the level of improvement in recognition accuracy is scenario dependent.

Introduction

The past decade's research efforts in the area of facial recognition resulted in a significant improvement in terms of recognition performance. This can be inferred from the results of the 2010 Multiple-Biometrics Evaluation (MBE) study organized by NIST.[1] In 1993, at a false match rate (FMR) of 0.001 percent, the best performing face matcher had a false non-match rate (FNMR) of 79 percent. According to NIST's MBE, the FNMR has significantly dropped to an impressive 0.003 percent at a false accept rate (FAR) of 0.001 percent. Typically, face recognition algorithms perform well in the visible band of the electromagnetic spectrum (380–750 nm). However, the problem of matching facial images remains a challenge when dealing with difficult and diverse scenarios, including the usage of different capturing sensors (for example 2D, 3D, visible, or IR), large datasets, or having to deal with facial obscurants, as well as pose, illumination, and expression variations. There are many inherent problems that come along with visible face recognition. First and foremost, the effect of illumination variation on visible band images is among the most insidious problems that face matching algorithms need to efficiently deal with. Because of this problem, recent FR trends are leading away from the visible spectrum and heading to different infrared bands: near infrared (NIR: 750–1100 nm)[2][3], short-waved infrared (SWIR: 900-1900 nm)[4], mid-waved infrared (MWIR: 3–5 μm)[5][6], and long-waved infrared (LWIR: 7–14 μm).[7] The main disadvantage of IR-based biometric identification at the present time is the high price of the majority of high end sensors. However, the cost of infrared security cameras has dropped considerably and is now comparable to high end digital single-lens reflex (DSLR) cameras (visible band). For example, in the thermal band, FLIR is now offering LWIR cameras starting at less than USD 3,000, making them more affordable and, thus, researchers can utilize them in several innovative ways. This scenario would have been inconceivable just a few years ago. Affordable IR sensors provide the opportunity to create more challenging databases (larger scale, different sensors operating on the same or different IR bands), allowing also for the development and testing of heterogeneous face matching algorithms where visible images are matched to NIR, SWIR, MWIR, and LWIR face images.

The focus of this work is matching visible to SWIR band face images. There are many benefits when using SWIR camera sensors for the purpose of designing and developing face recognition algorithms. First, the SWIR spectrum allows for covert capture of face images in nighttime environments considering that the illumination source is invisible to the human eye (due to the wavelength being well beyond the visible spectrum). Another advantage of the SWIR band is the capability to see through different types and levels of tinted glass as well as sunglasses.[11] SWIR has a longer wavelength range than NIR and is more tolerant to low levels of obscurants like fog and smoke. Finally, different facial features can be extracted in the SWIR band that can be combined with those extracted in the visible band to create a more accurate and complete representation of a subject's face.[4] Our previous studies determined that this

capability resulted in an increase of rank-one identification rates under variable face matching scenarios.[16]

While previous FR studies have mainly concentrated on the visible and NIR bands, FR in the SWIR band, more specifically the 1550 nm wavelength, has received limited attention. Prior work focused on dealing with face datasets assembled under controlled and challenging conditions.[4] However, in uncontrolled scenarios (long range recognition and imaging behind glass operational conditions), there is a need for efficient intelligence and surveillance reconnaissance (ISR) interoperability, that is, operational teams (for example, armed forces) are required to effectively manage, access, and use ISR to improve command and control, and enhance information sharing and situational understanding to improve the effectiveness of operations while minimizing collateral damage in a complex environment. One particular issue that we address in this article is the ability to capture a subject's face behind glass (that can be used in commercial buildings, homes, or vehicles), especially when the glass is tinted. Being able to image a subject behind different types of tinted glass and accurately match them with images from a database of visible images (such as a watch list, do-not-fly list, and so on) is an important step in improving human identification in operational environments.

Goals and Contributions

In this article, we propose a new cross-spectral face matching algorithm. It significantly enhances the capability of the original approach proposed by Kalka et al.[4] to match SWIR to visible face images in variable challenging scenarios, including scenarios where face images were captured behind different types of tinted glass. Firstly, in order to evaluate the efficiency of our proposed approach, a database of subjects was assembled behind multiple types of tinted glass and under different lighting conditions (that is, ambient lighting or the usage of SWIR active illumination). Secondly, we determined that our wavelength- and scenario-dependent eye detection algorithm performs very well on all datasets that it was tested on. In addition, experiments using our proposed face matching algorithm show that the use of randomly selected photometric normalization techniques (as proposed in Kalka et al.[4]) is not necessary to improve FR performance. This is due to the fact that certain normalization techniques do not yield enough discriminatory information in the face which, in turn, yields low face match (similarity) scores. Specifically, we demonstrate that the use of only a small subset of more than 45 normalization techniques (and their combinations) available and tested was necessary to increase the overall performance of our face matcher, while drastically reducing the computational time required to perform a single match (that is, using a small subset vs. all possible combinations). Our proposed design also includes the use of parallel processing, which further reduces the time needed to perform a single match. Finally, our experiments show that the level of improvement achieved when using our proposed face matching approach in variable challenging face datasets is scenario dependent.

The rest of the article is organized as follows. The following section, "Background Information," discusses some background work done in the field of heterogeneous face recognition. "Data Collection" describes our data collection process, while the section "Methodological Approach" offers insights into our methodology. Finally, the section "Experimental Setup" demonstrates experimental results, while the section "Conclusions and Future Work" draws conclusions and discusses our future work.

Background Information

The field of heterogeneous face matching can be broken down into four different categories: NIR-visible, SWIR-visible, MWIR-visible and LWIR-visible matching. Because each band of the IR spectrum reveals different facial characteristics, different face recognition algorithms (including face/eye detection, feature extraction, and matching) must be designed, developed, and used when working in any specific face matching category described above. In the NIR spectrum, Klare and Jain[2] learned discriminative projections using common feature-based representations (LBP and HOG features) as well as linear discriminant analysis (LDA) on both NIR and visible images. In the proposed approach, the authors matched NIR to visible images directly using random subspace projections as well as using sparse representation classification. Zhu et al.[3] proposed the transductive heterogeneous face matching (THFM) method that adapts the NIR-visible matching, learned from a training database, to target images. With the use of their version of a Log-DoG (difference of Gaussian) filtering, along with local encoding and feature normalization, they were able to alleviate the heterogeneous difference between the two spectral bands. The transduction-based approach simultaneously reduces the domain difference and learns the discriminative model for target subjects. This resulted in fairly accurate NIR-visible matching scores.

In the category of SWIR-visible face matching, Mendez et al.[7] use nonlinear dimensionality reduction approaches. Global nonlinear techniques, such as Kernel-PCA and Kernel-LDA, as well as local nonlinear techniques, such as local linear embedding and locality preserving projections were compared to their linear counterparts, PCA and LDA. Experiments showed that the use of local nonlinear dimensionality reduction techniques resulted in higher FR matching rates in the SWIR band on two controlled databases. Kalka et al.[4] used a photometric fusion technique that incorporated six different illumination normalization schemes. These techniques were combined in both the SWIR and visible bands to create 36 photometric combinations. A simple summation fusion scheme was then used to determine the final match score. The approach was tested on a set of datasets representing difficult challenging environments including SWIR images taken in an operational setting (that is, in the wild). Experimental results showed that this approach outperformed other texture-based approaches and was dependent on the scenario tested.

In the field of passive IR face matching, Bourlai et al.[5][17][18][19][20][21] use variable schemes, including a texture-based fusion scheme to match MWIR or LWIR

probe to visible gallery face images. Using different databases ranging from 50 to more than 100 subjects, different approaches, including texture-based ones such as local binary and ternary patterns (LBP/LTP), pyramid histogram of gradient (PHOG), and scale invariant feature transform (SIFT) were compared and fused after applying different photometric normalization techniques. Osia and Bourlai[6][17] also match MWIR probe images to visible gallery images. Face features, such as veins, scars, and wrinkles, are first extracted using multiple techniques including a standard fingerprint extraction method, the scale-invariant feature transform (SIFT), and the speeded up robust feature (SURF) method. A fingerprint matcher is then used to match the extracted features (from either the whole face or subregions of the face) from same band face images (visible-visible, MWIR-MWIR and so on).

In this work, we are discussing the problem of matching SWIR against visible face images captured under variable conditions. What follows is a description of the face datasets used for the purpose of our study.

Data Collection

To the best of our knowledge, there are no publicly available face databases in the research community that are composed of visible and SWIR face images captured behind tinted glass. In this work, we discuss the database collected at WVU for the purpose of our study and the data collection protocol designed and executed. To simulate an operational environment, where face images need to be captured through tinted glass, a three-sided booth, with a 1-ft. × 1-ft. window slot, was built for subjects to sit behind and have their faces collected by both visible and SWIR cameras. The window slot was set so different types of tinted glass could be easily switched. Both industrial and automotive tinted glass was used for the purpose of this study. The three types of glass that were used can be described as follows:

- Industrial clear glass panel (clear with 0-percent tint)
- Industrial clear glass panel with tinted automotive film applied (80-percent tinted film)
- Industrial tinted glass panel (Solarcool-2 Graylite)

The use of these panels was chosen in order to test the performance of heterogeneous face recognition across a wide spectrum of varying levels of tint. Each panel allows for the simulation of different operational scenarios such as imaging through normal glass (0-percent tint), through a vehicle's tinted side window (80-percent tinted film) or through a secure building with highly tinted windows (Solarcool-2 Graylite).

Two different light sources (scenarios) were used to illuminate the subjects' faces while sitting behind the booth and glass panels. In the first scenario, an interior light source (inside the booth with the subject present) was used to illuminate the subject's face. Two 250 W tungsten lightbulbs were positioned in the booth to optimize the illumination on the subject's face without hotspots

and produced ~3.87 kilolux of light. In the second scenario, an external active SWIR illumination source was used. A 1550-nm laser source with a 500 mW light diffuser was positioned outside of the booth and illuminated the glass externally. The SWIR illumination source was set up at an angle from the glass panel in order to minimize the reflections back into the camera. An image of each subject's face behind the glass panel was captured with the two different illumination sources using the following cameras:

- *Visible Camera:* A Canon EOS 5D Mark II camera was used to capture the visible images. This digital SLR camera has a 21.1-megapixel full-frame CMOS sensor with a DIGIC 4 image processor and a vast ISO range of 100–6400. It also has an auto lighting optimizer and peripheral illumination correction that enhances its capabilities. The Mark II was used to collect RGB images of the subjects at the ground truth level (no glass) as well as when using each glass panel with either of the two illumination sources tested.
- *SWIR Camera:* A Goodrich SU640 camera was used to capture SWIR face images. The SU640 is an indium gallium arsenide (InGaAs) video camera featuring high sensitivity and wide dynamic range. This model has a 640 × 512 FPA with 25 m pixel pitch and >99 percent pixel operability. The spectral sensitivity of the SU640 ranges uniformly from 700–1700 nm wavelength. The response falls rapidly at wavelengths lower than 700 nm and greater than 1700 nm.

A total of 140 subjects participated in our study. When a subject arrived for collection, ground truth images were first taken with no glass panels, when using either the visible or the SWIR camera sensors. Then, the following conditions were considered when using the aforementioned glass panels, the SWIR camera, and different illuminators. In addition, a 100-nm band pass filter, centered at 1550 nm, was placed in front of the SWIR camera to ensure that only SWIR light in that particular waveband was entering the camera.

- 0 percent tint
  - Internal visible illumination
  - External active SWIR illumination
- 80 percent tinted film
  - Internal visible illumination
  - External active SWIR illumination
- Solarcool-2 Graylite
  - Internal visible illumination
  - External active SWIR illumination

Overall, we assembled a total of seven databases under variable scenarios (including the ground truth database where no glass was used between a subject's face and the camera). The final set of databases assembled consisted of the data of 140 subjects, that is, 980 SWIR face images (140 subjects × 7 scenarios) and

140 visible (ground truth) face images, totaling 1,020 face images. These images were then used to conduct the heterogeneous face recognition experiments that are described later in the section "Experimental Setup." In the next section, we describe our methodological approach used to deal with the variable challenges of the cross-spectral face matching scenarios investigated in this study.

Methodological Approach

In this section, our methodological approach to perform heterogeneous face recognition is discussed. First, we outline an overview of our automatic face/eye detection algorithm. This allows for each subject's face to be geometrically normalized to the same plane and used for face recognition. Then, we describe our photometric normalization fusion face recognition algorithm. Our method uses a texture-based matching algorithm, local binary patterns (LBP) and local ternary patterns (LTP), as well as our new proposed cross-photometric normalization fusion scheme to accurately match SWIR (probe dataset) to visible face images (gallery dataset). An empirical study is conducted to help reduce the number of photometric normalization techniques used, which, in turn, helps reduce the time required to obtain a face match score. What we discuss in the following sections are the steps used in our proposed heterogeneous face recognition approach.

Automatic Face and Eye Detection

In general, commercial and academic facial recognition algorithms require that the face images of each individual be standardized (in terms of orientation, interocular distance, masking, and so forth). Typically, feature points of sub-facial regions, more specifically the locations of human eye centers, are used to rotate and translate a face to a standard representation. While this operation can be manually performed by an operator on a limited size dataset, when having to deal with larger databases (from thousands to millions of subjects), manually obtaining the eye locations of each subject, and for each sample per subject available, is not practical. Therefore, an accurate and robust eye detection method needs to be employed since it is expected to have positive impact on FR performance on any typical FR system dependent on eye locations. Our face and eye detection method comprises five main processes: preprocessing, automatic face detection, eye region localization, summation range filtering, and geometric normalization. This leads to an image that is suitable for a face recognition system.

Preprocessing

SWIR images tend to have low contrast in the facial region, especially in the 1550-nm band. Instead of the human skin reflecting those wavelengths back into the camera, the moisture from the skin tends to absorb higher SWIR wavelengths, causing the skin to appear very dark (even for very light-skinned subjects). In order to compensate for this, photometric normalization techniques bring out unique features that are beneficial for face and eye detection algorithms applied to wavelength-specific face datasets.

Because we use a *template matching* scheme to detect the face and eye regions, average templates are needed. Therefore, for the purpose of this study, seven subjects are randomly selected from each capturing scenario and their faces are geometrically normalized, cropped, and averaged together to create an average face template. Then, the eye regions from this template are cropped and used as *average eye templates*. These average templates are, finally, saved and used on all images in the database.

Automatic Face Detection

Because of the unique qualities that SWIR images have, typical face detection algorithms could not be used. Therefore, a template-based face detection algorithm was developed to spatially locate the face region. For each pixel in the query face image, the 2D normalized cross-correlation is computed between the region of that pixel and the average face template. Mathematically, the 2D normalized cross-correlation can be described as:

$$\delta(u,v) = \frac{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]\left[t(x-u, y-v) - \bar{t}\right]}{\left\{\sum_{x,y}\left[f(x,y) - \bar{f}_{u,v}\right]^2 \sum_{x,y}\left[t(x-u, y-v) - \bar{t}\right]^2\right\}^{1/2}} \quad (1)$$

where f is the image, $\bar{t}$ is the mean of the template, and $\bar{f}(u,v)$ is the mean of $f(x,y)$ in the region under the template. Then, the convolution of the image and the average template yields a correlation map. The highest location within the correlation map (the peak) is the location of the face. However, different average templates yield different eye detection results. Because of this issue, multiple average templates (in our case five) are created and used to increase the chance of finding the correct location. The final location of the face can be described with the following formula:

$$\hat{\delta}(u,v) = argmax\,(\delta_x(u,v)) \quad (2)$$

where $\delta_x(u,v)$ is the location of the highest correlation coefficient obtained from average template x (in our case $x = 1,...,5$). Then, $\hat{\delta}(u,v)$ corresponds to the upper left point of the face region, and finally, the face can be cropped to the size of the average templates used. By only using the face area determined by this approach, our subsequent eye detection method only has to search the face area (a much smaller region), instead of the entire image as a whole.

Eye Region Localization

Since the location of the face is now known, the location of the eye regions can be easily determined. In order to further reduce the search space, the face is split into four equal regions (top left, top right, bottom left, and bottom right). Assuming that the face region is found correctly by using the method described above, the right and left eyes should be located in the top right and top left regions respectively. Therefore, to obtain the left and right eye regions, the average eye templates are convolved with their respective quadrants using Equation 1. As stated above, different average eye templates yield different results. Therefore, the process is repeated

multiple times using unique templates to increase the chance of obtaining the correct region. Then, Equation 2 can be used to find the final location of the eye regions.

Summation Range Filter

Although the region of the eye can be easily found, the center of the eye cannot always be determined to be the center of the found region. Therefore, an accurate way of determining the correct center of the eye must be employed. Knowing that the pupil is typically much darker than an iris, and that, in certain conditions, light reflections from an illumination source are available within the eye (typically in the pupil), summation range filters can be used to more accurately determine the center of the eye. The summation range map *S (x,y)* can be described as follows:

$$S(x,y) = \sum_{x=-1}^{1} \sum_{y=-1}^{1} R(x,y) \tag{3}$$

where

$$R(x,y) = argmax\left(I(x-1{:}x+1, y-1{:}y+1)\right) - argmin\left(I(x-1{:}x+1, y-1{:}y+1)\right) \tag{4}$$

and where *I(x,y)* is the original cropped eye region. Then, the final eye center is determined to be

$$P(x,y) = argmax\left(S(x,y)\right) \tag{5}$$

This process is repeated for both the right and left eye regions to determine the final locations for the right and left eye respectively.

Geometric Normalization

In order to assist facial recognition systems, the left and right eye locations are used to geometrically normalize the image. By setting a standard interocular distance, the eye locations can be centered and aligned onto a single horizontal plane and resized so all images are similar to each other. This ensures that, if the eyes are found correctly, the left and right eyes are guaranteed to be in the same position every time, an assumption that is made by most facial recognition algorithms. Therefore, all face images are geometrically normalized based on the found locations to have an interocular distance of 60 pixels with a resolution of 130 × 130 pixels. The geometrically normalized images can then be used in our facial recognition system.

Face Matching Algorithm

In this work, both commercial and research software was employed to perform the face recognition experiments: (1) Commercial software, such as Identity Tools G8 provided by L1 Systems and (2) standard texture-based feature methods. Two different texture-based schemes were used to test our algorithms, namely local binary patterns (LBP) and local ternary patterns (LTP).

In the LBP operator, patterns in an image are computed by thresholding 3×3 neighborhoods based on the value of the center pixel. Then, the resulting binary pattern is converted to a decimal value. The local neighborhood is defined as a set of sampling points evenly spaced in a circle. The LBP operator used in our experiments is described as $LBP^{u^2}_{P,R}$, where P refers to the number of sampling points placed on a circle with radius R. The symbol u^2 represents the uniform pattern, which accounts for the most frequently occurring pattern in our experiments. The pattern is important because it is capable of characterizing local regions that contain edges and corners. The binary pattern for pixels, lying in a circle f_p, $p = 0,1,\ldots,P-1$ with the center pixel f_c, is mathematically computed as follows:

$$S(f_p - f_c) = \begin{cases} 1 \; if f_p - f_c \geq 0; \\ 0 \; if f_p - f_c < 0; \end{cases} \tag{6}$$

Following this a binomial weight 2^p is assigned to each sign $S(f_p - f_c)$ to compute the LBP code,

$$LBP_{P,R} = \sum_{P=0}^{P-1} S(f_p - f_c) 2^p \tag{7}$$

LBP is invariant to monotonic gray-level transformations. However, one disadvantage is that LBP tends to be sensitive to noise in homogeneous image regions since the binary code is computed by thresholding the center of the pixel region.

Consequently, LTP has been introduced to overcome such a limitation, where the quantization is performed as follows:

$$S(f_p - f_c) = \begin{cases} 1 \; if f_p - f_c \geq t \\ 0 \; if \; |f_p - f_c| \leq t \\ -1 \; if f_p - f_c \leq -t \end{cases} \tag{8}$$

The output of this operator is a 3-valued pattern, as opposed to a binary pattern. Furthermore, the threshold t, can be adjusted to produce different patterns. The user-specific threshold also makes the LTP code more resistant to noise.

An Empirical Study on Photometric Normalization

The problem of cross-spectral FR, matching visible to SWIR face images, is very challenging because of the interaction between the electromagnetic waves (visible and SWIR) and the material (in our case, human skin). This results in different reflectance, transmission, and scattering properties. Because of this, contrast, texture, and so on are different when dealing with visible and SWIR images, respectively. Photometric normalization algorithms traditionally have been employed in order to compensate for these changes in illumination, such as shadows and varying light conditions. In this work, we employ six different

photometric normalization techniques in order to facilitate cross-spectral matching. More specifically, we employ the following techniques: contrast-limited adaptive histogram equalization (CLAHE), tangent-based single-scale retinex (TBSSR), log-based single-scale retinex (LBSSR), TBSSR followed by CLAHE (C-TBSSR), LBSSR followed by CLAHE (C-LBSSR), and the Tan and Triggs[9] normalization (TT). Sample images of these photometric normalizations can be seen in Figure 1.

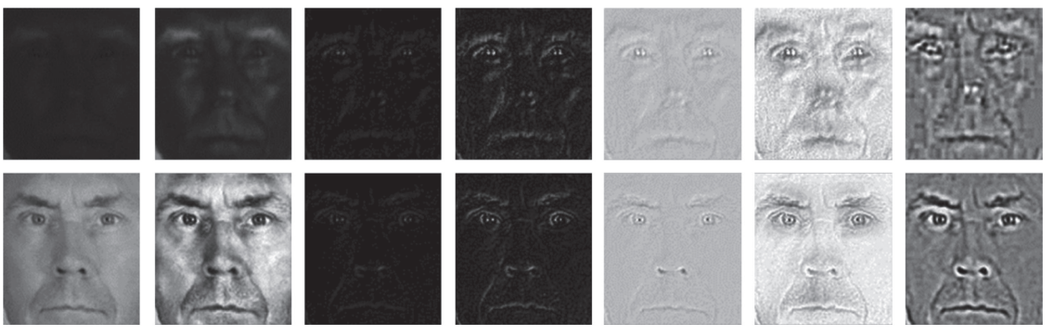

Figure 1: Sample with different photometric normalizations: a) Original data b) CLAHE c) LBSSR d) C-LBSST e) TBSSR f) C-TBSSR and g) TT
(Source: West Virginia University, 2014)

- CLAHE: This technique operates on small local regions (8×8 for our experiments) in the image and applies histogram equalization on each individual region (in contrast to the entire image in regular histogram equalization). In order to increase contrast while decreasing the amount of noise, CLAHE redistributes each histogram so that the height of each bin falls below a predetermined threshold (0.1 in our reported experiments).
- TBSSR: This decomposes the image into two components, illumination $L(x,y)$ (the amount of light falling on the targeted object) and reflectance $R(x,y)$ (the amount of light reflecting off the targeted object). The illumination component is estimated as a low-pass version of the original image, while the reflectance component is obtained by dividing the original image from other illumination images. Therefore, to calculate the TBSSR,

$$R(x,y) = atan\left(\frac{I(x,y)}{L(x,y)}\right) \tag{9}$$

- C-TBSSR: A common problem with TBSSR is that images tend to become oversaturated or "washed out." This can have negative effects on eye detection algorithms. Furthermore, "halo" artifacts may be introduced depending on the scene and scale of value chosen for the Gaussian smoothing function. To diminish the cost of processing speed, we applied the CLAHE approach listed above to TBSSR face images to help compensate for the aforementioned approaches and to increase the contrast of the image.

- LBSSR: By using different nonlinear transformations on the TBSSR, different image representations can be obtained. Therefore, the *atan* in Equation 1 is replaced with a logarithmic transformation, resulting in the following formula:

$$R(x,y) = log_{10}\left(\frac{I(x,y)}{L(x,y)}\right) \tag{10}$$

- C-LBSSR: As described above, the LBSSR can cause over saturation and haloing effects. Therefore, the CLAHE approach was also applied to the LBSSR image to correct the contrast issues mentioned above.
- TT: This photometric normalization[9] incorporates a series of algorithmic steps that allow for the reduction of illumination variations, local shadowing, and highlights, while still preserving the essential elements of visual appearance. These steps include gamma correction (raising each pixel value to a certain value, in this case 2), difference of Gaussian filtering (subtraction of an original image from the blurred version of the same image), and contrast equalization (suppressing larger intensities while maintaining lower intensities).

In this empirical study, we wanted to determine which combination of photometric normalization algorithms produces the best match scores between visible images and 1550-nm SWIR face images. In order to do this, a heterogeneous cross-spectral approach was used. First, all gallery and probe images are photometrically normalized using the techniques described above. Then, each normalization-per-probe image is matched with each normalization-per-gallery image. With the original face image and the six photometric normalizations used ($n = 7$), 49 different photometric combinations are created per match (7 probe representations × 7 gallery representations). This resulted in 49 different match scores for a single probe-gallery match. An overview of this process can be seen in Figure 2.

Once all probe images are matched to all gallery images, each photometric combination is broken down into their respective genuine (true positive) and imposter (true negative) scores. Then, the *receiver operator curve* (ROC) is computed for all 49 photometric normalizations. The ROC is used to examine the relationship between the true positive and the false positive rate. To quantify which ROC performs better than another, a measure must be taken. In this case, the *area under the curve* (AUC) is used as a measurement to determine which photometric combination performs better than the others. Higher AUCs show combinations that have a wider gap between true positives and false positives, which, in turn, results in higher performance.

After computing the AUCs for all 49 combinations, we can determine which photometric combinations result in higher performance. Sample ROCs and their respective AUCs can be seen in Figure 3.

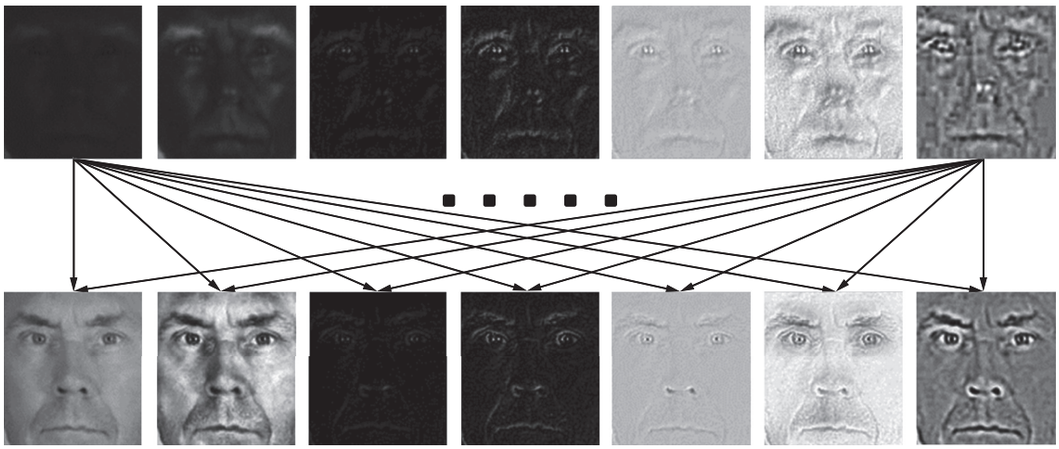

Figure 2: An overview of the cross-photometric empirical study done. Notice that each representation in the gallery set is matched against all representations in the probe set creating 49 combinations.
(Source: West Virginia University, 2014)

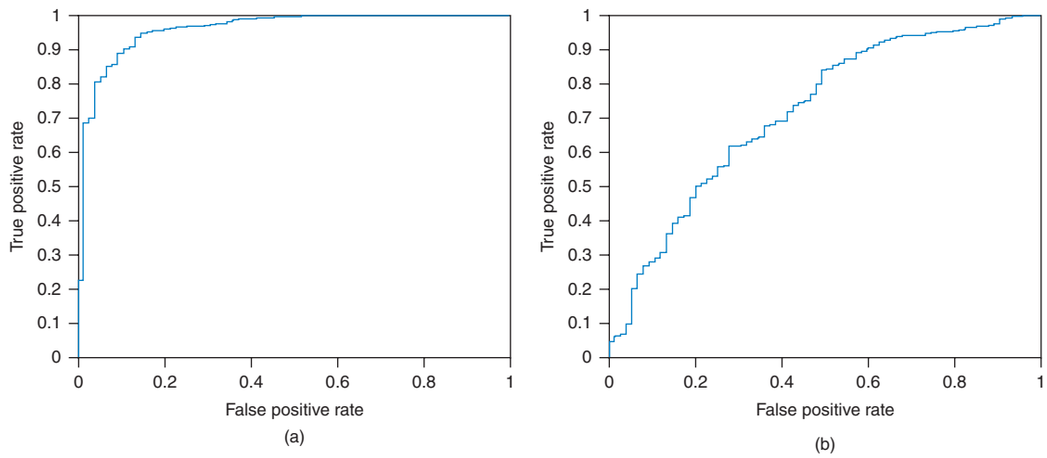

Figure 3: Sample ROCs for two different photometric combinations a) Combination 44: 95.85 percent AUC and b) Combination 2: 71.67 percent.
(Source: West Virginia University, 2014)

Score Level Fusion

Since we know which photometric combinations perform best and which combinations perform worse, based on their respective AUCs, we can take advantage of multiple combinations to further increase the final match score. Simple fusion of all 49 combinations can be performed, as also described by Kalka et al.[4] However, this is not feasible in practice due to the lengthy process of applying all photometric normalization schemes and matching all 49 combinations. Therefore, a second empirical study was conducted to determine

which three combinations, fused together from the top five photometric combinations observed, provide the best matching results for our study. Choosing only three combinations allows for a vast increase in processing time while still maintaining the level of accuracy desired. After the combinations were determined, the testing phase only required these three combinations be used. The final match score S for any probe image is computed by using the following formula:

$$S = \sum_{i=1}^{3} m(Pt_i, Gt_i) \tag{11}$$

Where Pt and Gt are the gallery and probe templates respectively and i represents the photometric normalization combination determined previously. Matching function $m(Pt,Gt)$ corresponds to the matching algorithms listed above, LBP and LTP. An overview of this process is illustrated in Figure 4.

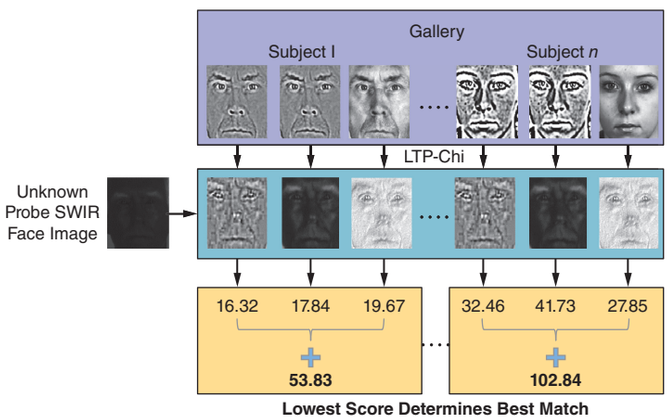

Figure 4: An overview of the score level fusion scheme used to improve cross-spectral face recognition.
(Source: West Virginia University, 2014)

After the completion of this empirical study, it was determined that the following photometric combinations yield the highest rank 1 identification rates:

1. *Gallery:* Tan and Triggs – *Probe:* Tan and Triggs
2. *Gallery:* Tan and Triggs – *Probe:* Contrast-limited adaptive histogram equalization
3. *Gallery:* Contrast-limited adaptive histogram equalization – *Probe:* Tangent-based single-scale retinex

Another advantage of our approach is that instead of performing all 49 photometric normalization combinations as was proposed by Kalka et al.[4], which uses a lot of processing time, our proposed approach only requires three

such combinations. As we will show in the following section, the advantage of this approach is that it manages to increase the performance rate as well as boost the processing speed when compared to the original algorithm described by Kalka et al.[4]

Experimental Setup

Two sets of experiments were conducted to demonstrate the efficiency of our algorithm. First face/eye detection tests were performed on all face datasets. Then, heterogeneous face recognition tests were performed. Our approach is compared with baseline texture-based approaches as well as commercial software (G8 provided by L1 Systems).

Eye Detection Validation

In order to test the accuracy of our eye detection algorithm, a validation scheme was used. First, we performed our face detection algorithm on each capturing scenario to determine the spatial location of the face region. If the face detection failed, the location was manually marked and subsequently used. To test the accuracy of the eye detection method, the normalized average error was used. This error, indicating the average error between both eyes, was used as the accuracy measure for the found eye locations and can be described as:

$$e = \frac{mean\ (d_{left}, d_{right})}{W} \tag{12}$$

Where d_{left} and d_{right} are the Euclidean distances between the found left and right eye centers with the manually annotated ground truth and w is the Euclidean distance between the eyes in the ground truth image. In the normalized error, $e < 0.25$ (or 25 percent of the interocular distance) roughly corresponds to the width of the eye (corner to corner), $e < 0.10$ roughly corresponds to the diameter of the iris, and $e < 0.05$ roughly corresponds to the diameter of the pupil.

Texture-Based Distance Metrics

In order to get the final match scores from the LBP and LTP feature vectors, two different distance metrics were used, the distance transform and the chi-squared metric. The distance transform (defined as the distance or similarity metric from image X to image Y) is defined as follows:

$$D(X,Y) = \sum_{Y(i,j)} w(d_x^{K_{Y(i,j)}}(i,j)) \tag{13}$$

Where $K_{Y(i,j)}$ is the code value of pixel *(i,j)* of image Y, and w is a user controlled penalty function.

The chi-squared distance is defined as follows:

$$\chi^2(n,m) = \frac{1}{2}\sum_{i}^{l} \frac{h_n(k) - h_m(k)}{h_n(k) + h_m(k)} \tag{14}$$

Where h_n and h_m are the two histogram feature vectors, l is the length of the feature vector and n and m are two sample vectors extracted from an image of the gallery and probe sets respectively.

Experimental Results

In this section we discuss the two main experiments we performed. The first one is on eye detection and heterogeneous face recognition, and the second one is discussing the time efficiency (computational complexity) of our proposed fusion approach when compared to original one, and after we incorporate in our design parallel processing.

Eye Detection and Heterogeneous Face Recognition

In order to show the efficiency of our algorithms, we performed both eye detection and heterogeneous face recognition tests on the seven different databases listed above: ground truth (no glass), 0-percent tint with active SWIR and visible illumination, 80-percent tint with active SWIR and visible illumination, and the Solarcool-2 Graylite glass with active SWIR and ambient lighting. One hundred forty subjects were used for one-to-one comparison of visible gallery images to SWIR probe images. We compared our proposed heterogeneous face recognition scheme with multiple algorithms, including the baseline texture-based approaches (LBP, LTP) with both distance metrics, a commercial face recognition algorithm (L1 Systems G8), and the original cross-photometric score level fusion approach described by Kalka et al.[4] An overview of the results of this experiment can be found in *Table 1*.

	Eye Detection (% @e<.25)	**LBP – χ^2**	**LBP – DT**	**LBP – χ^2**	**LTP – DT**	**L1's G8**	**Kalka et al [4]**	**Proposed Method**
Ground Truth	96.81	62.14	80.14	70.71	86.43	81.43	35.00	**94.26**
0% Visible Lighting	100.00	38.57	50.00	47.86	53.57	56.43	19.29	**67.86**
0% Active SWIR Lighting	97.86	12.86	17.86	19.29	24.29	12.14	5.71	**35.71**
80% Visible Lighting	99.29	26.43	**40.82**	20.41	30.61	6.12	13.57	34.69
80% Active SWIR Lighting	98.57	7.14	12.14	8.57	13.57	4.29	5.00	**23.57**
Solarcool Visible Lighting	36.96	39.29	48.57	52.14	55.00	**69.29**	23.57	61.43
Solarcool Active SWIR Lighting	57.61	0.71	0.71	0.71	0.71	1.43	0.71	**4.29**

Table 1: Experimental results for the proposed eye detection and heterogeneous face recognition algorithm. Eye detection (second column) uses the normalized average error at $e = 0.25$ while the face recognition results show the rank 1 percentage. (Source: West Virginia University, 2014)

In these results, the eye detection (second column) reports the percentage of eyes whose normalized average error is $e < 0.25$. In reference to the face recognition studies performed (columns 3 through 9), the percentage of subjects who obtained a rank 1 identification rate was reported. Note that in all

cases, except for the 80-percent visible lighting and Solarcool visible lighting, our proposed algorithm outperformed all other algorithms.

Time Efficiency

One of the main drawbacks to the algorithm proposed by Kalka et al.[4] is the length of time that it takes to complete a single probe-gallery match. Because the algorithm is essentially repeating the same process 49 times (just with different image representations), the time to match one probe SWIR image to a gallery of visible images is impractical, in an operational standpoint, and grows as the size of the gallery grows. Therefore, in order to increase speed, as well as increase matching accuracy, our empirical study, described earlier in the section "An Empirical Study of Photometric Normalization," was performed to narrow the photometric normalization combinations down from 49 to 3. Although this helps speed up the matching process by approximately 18.5 times, it's still too slow to have any practical matching ability. In order to decrease the process time further, parallel processing was used. Eight cores were used simultaneously to perform the matching algorithm described above. All experiments were conducted on a gallery of 140 subjects. The results for the time efficiency test can be found in Table 2, where we can see the time it takes (in seconds) for a single probe to match a gallery image. All experiments described above were performed on a 64-bit Windows 7 machine with 12 GB of RAM running Intel® Core™ i7 CPU at 3.2 GHz using MATLAB R2012b. The MATLAB Parallel Processing Toolbox was used to test the parallel processing speeds.

Algorithm	Kalka et al.[4]	Proposed	Proposed with Parallel Processing
Avg. Time (sec)	12.059	0.650	0.207

Table 2: Results of the time efficiency test. All times reported are in in seconds for a single probe to match a gallery face image.
(Source: West Virginia University, 2014)

As we can see in Table 2, the proposed method, when using parallel processing, further speeds up the time it takes to make a single gallery to probe match. Also it is clear that by reducing the number of photometric normalizations, our algorithm is much faster and more efficient than the algorithm proposed by Kalka et al.[4]

Conclusions and Future Work

In this article, we studied the advantages and limitations of performing cross-spectral face matching (visible against SWIR) in different challenging scenarios represented in a set of different face databases. We first showed that our eye detection approach performs extremely well on our assembled face databases. Specifically, we managed to achieve an eye detection rate of greater than 96 percent in the majority of the scenarios we investigated. This achievement is

critical in improving the efficiency of the automated face recognition system proposed. Secondly, we proposed an approach that enhances the capability of the original cross-photometric score level fusion proposed by Kalka et al.[4] Our experimental results showed that the use of a fairly small set of photometric normalization combinations is sufficient to yield desirable face recognition scores due to our empirical study that determined the efficiency in matching probe to gallery face images under specific pairs of photometric normalization algorithms. In other words, our study showed that a smaller number of combinations results in an increase of rank-1 identification rate, in addition to an improvement of the computational complexity of the proposed approach, that is, when using a subset vs. a complete set of photometric normalization techniques and their combinations. By using the best three photometric normalizations instead of all 49 combinations tested, the time required for a single gallery to probe face match increased by more than 18.5 times. In addition, by utilizing MATLAB's parallel processing toolbox, we were able to further increase the matching speed by 58 times when compared to the original matching algorithm.

Another benefit of our face matching algorithmic approach is that in all but two scenarios, it outperformed all other face matching algorithms tested. We obtained a rank 1 identification rate of 94.26 percent when using our ground truth data, which is about 2.7 times improvement over the original algorithm[4] and a more than 7-percent improvement over LTP-DT, the algorithm that achieved the second-best rank 1 identification rate. The only scenarios where our proposed algorithm did not outperform all others were when we used the 80-percent visible lighting and the Solarcool visible lighting face datasets.

For future work we are planning to extend our experiments in the field of heterogeneous face recognition. One of our focus areas will be to test our proposed *enhanced cross-photometric score level fusion algorithm* on face images captured on different bands than the SWIR and visible ones we used in this study. Studies in the near IR and passive IR bands would be beneficial to show the robustness of this algorithm, even when we have to deal with face images affected by different image degradation factors, such as camera and motion noise, or information loss due to the acquisition of face images at long ranges (close to or further than the capabilities of the camera used).

Acknowledgements

This work is sponsored through a grant from the Office of Naval Research (N00014-08-1-0895) Distribution A - Approved for Unlimited Distribution. The authors are grateful to the students and staff at West Virginia University, especially Dr. Bojan Cukic, Dr. Jeremy Dawson, as well as WVU students including Nnamdi Osia, Neeru Narang, and Jason Ice, for their assistance in this work.

References

[1] Grother, P., G. Quinn, P.J. Phillips, "Report on the Evaluation of 2D Still-Image Face Recognition Algorithms," *Nat'l Inst. Of Standards and Technology interagency/internal report (NISTIR) 7709*, 2010.

[2] Klare, B. and A. Jain. "Heterogeneous face recognition: Matching NIR to visible light images," *IEEE International Conference on Pattern Recognition (ICPR)*, pp. 1513–1516, 2010.

[3] Zhu, J. Y., W.S. Zheng, J.H.Lai, and S. Li, "Matching NIR Face to VIS Face using Transduction," *IEEE Transactions on Information Forensics and Security*, 2014.

[4] Kalka, N., T. Bourlai, B. Cukic, and L. Hornak, "Cross-spectral Face Recognition in Heterogeneous Environments: A Case Study on Matching Visible to Short-wave Infrared Imagery," *International Joint Conference on Biometrics (IEEE, IAPR)*, pp. 1–8, 2011.

[5] Bourlai, T., A. Ross, C. Chen, and L. Hornak, "A Study on using Middle-Wave Infrared Images for Face Recognition," *SPIE Biometric Technology for Human Identification IX*, pp. 83711K–83711K, 2012.

[6] Osia, N. and T. Bourlai, "Holistic and Partial Face Recognition in the MWIR Band Using Manual and Automatic Detection of Face-based Features," *IEEE Conf. on Technologies for Homeland Security*, pp. 273–279, 2012.

[7] Mendez, H., C. San Martin, J. Kittler, Y. Plasencia, and E. Garcia-Reyes, "Face recognition with LWIR imagery using local binary patterns," *Advances in Biometrics*, Springer, pp. 327–336, 2009.

[8] Akhloufi, M. A., A. Bendada, and J.C. Batsale, "Multispectral face recognition using nonlinear dimensionality reduction," *Proceedings of SPIE, Visual Information Processing XVIII*, volume 7341, 2009.

[9] Tan, X. and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, pp. 19:1635–1650, 2010.

[10] Bourlai, T., N. Kalka, D. Cao, B. Decann, Z. Jafri, C. Whitelam, J. Zuo, D. Adjeroh, B. Cukic, J. Dawson, L. Hornak, A. Ross, and N. A. Schmid, "Ascertaining human identity in night time environments," *Distributed Video Sensor Networks*, Springer, pp. 451–467, 2011.

[11] Ice, J., N. Narang, C. Whitelam, N. Kalka, L. Hornak, J. Dawson, and T. Bourlai, "SWIR imaging for facial image capture through tinted materials," *SPIE Defense, Security, and Sensing. International Society for Optics and Photonics*, pp. 83530S–83530S, 2012.

[12] Jobson, D., Z. Rahman, and G. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transaction on Image Processing*, pp. 6(3):451–462, 1997.

[13] Whitelam, C., Z. Jafri, and T. Bourlai, "Multispectral eye detection: A preliminary study," *IEEE International Conference on Pattern Recognition*, pp. 209–212, 2010.

[14] Whitelam, C., T. Bourlai, and I. Kakadiaris, "Pupil Detection under Lighting and Pose Variations in the Visible and Active Infrared Bands," *IEEE Workshop on Information and Forensics Security (WIFS)*, pp. 1–6, 2011

[15] Short, J., J. Kittler, and K. Messer, "A comparison of photometric normalization algorithms for face verification," *IEEE Conference on Automatic Face and Gesture Recognition*, pp. 254–259, 2004.

[16] Bourlai, T., "Short-Wave Infrared for Face-based Recognition Systems," *SPIE Newsroom Magazine - Defense & Security*, 2012 [Invited Article].

[17] Osia, N. and T. Bourlai, "A Spectral Independent Approach for Physiological and Geometric Based Face Recognition in the Visible, Middle-Wave and Long-Wave Infrared Bands," *Image and Vision Computing, Journal - Elsevier,* 2014.

[18] Bourlai, T. et al., "Applications of Passive Infrared Imaging to Forensic Facial Recognition," *InfraMation (Thermal Imaging Leading User's Conference - Organized by FLIR), Loews Royal Pacific*, 2013.

[19] Bourlai, T., "Mid-wave IR Face Recognition Systems," *SPIE Newsroom Magazine - Defense & Security*, pp. 1–3, 2013.

[20] Bourlai, T. and B. Cukic, "Multi-Spectral Face Recognition: Identification of People in Difficult Environments," *IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 196–201, 2012.

Author Biographies

Cameron Whitelam received the BS degree in Biometric Systems from West Virginia University. From January, 2011, he started pursuing a PhD in computer engineering in the multispectral imaging lab under the direction of Dr. Thirimachos Bourlai. His main areas of research include automatic eye detection in visible, shortwave infrared, and 3D images. He also focuses on facial recognition algorithms and their performances. In the summer of 2012, he was awarded a prestigious internship at the Office of Naval Research where less than 5 percent of applicants were accepted. He has coauthored numerous conference papers on automatic eye detection and has presented his

work internationally. He has also coauthored a book chapter in *Video Sensor Networks* (published by Springer).

Thirimachos Bourlai holds a BS (M.Eng. Equiv.) degree in Electrical and Computer Engineering from Aristotle University of Thessaloniki (Greece), an MS in Medical Imaging with Distinction and a PhD (Facial Recognition) degree from the University of Surrey (U.K.). He completed his first postdoctoral appointment in 2007 at the University of Surrey and his second in 2009 in a joint project between The Methodist Hospital and the University of Houston, TX (USA), in the fields of thermal imaging and human-based computational physiology. From 2008 to 2012 he worked at West Virginia University, first as a visiting research assistant professor and then as a research assistant professor. Since August of 2012 he has been an assistant professor at WVU, where he is the founder and director of the Multi-Spectral Imagery Lab. He is also an adjunct assistant professor at the School of Medicine, Department of Ophthalmology at WVU. He is actively involved in several applied projects, in the areas of biometrics, biomedical imaging, deception detection, and human computer interaction. The projects have been funded by DoD-ONR, DoD-DTRA, FBI, CITeR, a National Science Foundation (NSF) Industry/University Cooperative Research Center (I/UCRC), and TechConnect WV. He has published numerous book chapters, journals, magazines, and conference papers. T. Bourlai is also a Senior Member of IEEE and a member of the IEEE Communications Society, the IEEE Signal Processing Society, the IEEE Biometrics Compendium, the National Safe Skies Alliance, the Biometrics Institute Organization, the National Defense Industrial Association, the European Association of Biometrics, and the American Physical Society.