Face recognition in the SWIR band when using single sensor multi-wavelength imaging systems

N. Narang *, T. Bourlai *

MILab, West Virginia University, PO Box 6201, Morgantown, WV 26506, United States

Abstract

In this paper, we study the problem of Face Recognition (FR) when using Single Sensor Multi-Wavelength (SSMW) imaging systems that operate in the Short-Wave Infrared (SWIR) band. The contributions of our work are four fold: First, a SWIR database is collected when using our developed SSMW system under the following scenarios, i.e. Multi-Wavelength (MW) multi-pose images were captured when the camera was focused at either 1150, 1350 or 1550 nm. Second, an automated quality-based score level fusion scheme is proposed for the classification of input MW images. Third, a weighted quality-based score level fusion scheme is proposed for the automated classification of full frontal (FF) vs. nonfrontal (NFF) face images. Fourth, a set of experiments is performed indicating that our proposed algorithms, for the classification of (i) multiwavelength images and (ii) FF vs. NFF face images, are beneficial when designing different steps of multi-spectral face recognition (FR) systems, including face detection, eye detection and face recognition. We also determined that when our SWIR-based system is focused at 1350 nm, the identification performance increases compared to focusing the camera at any of the other SWIR wavelengths available. This outcome is particularly important for unconstrained FR scenarios, where imaging at 1550 nm, at long distances and when operating at night time environments, is preferable over different SWIR wavelengths.

Keywords: Multi-spectral imaging, Weighted score level fusion scheme, Face recognition, Classification, Pre-processing, SWIR band

1. Introduction

In a biometric-based recognition system, there are different modalities that can be used (fingerprints, face, iris, retina, voice etc.). Among those modalities, face is considered one of the top choices because, face recognition is one of the most common human experiences: it is easy to capture at a distance and in a non-cooperative manner and, finally, face recognition (FR) technology is fairly accurate. Depending on the application, face can be used either independently or in combination with other modalities in order to increase the performance of human recognition systems.

There are a number of practical issues that still need to be solved with FR systems. When designing such systems, one has to deal with a variety of problems that arise from each module of the overall architecture, i.e. data collection, transmission, data storage, signal processing and decision making. The data collection module has its own challenges. For example, FR systems perform well with frontal faces captured under controlled conditions (indoors, short standoff distance, controlled illumination). The problem becomes more complicated when face images are captured under variable illumination conditions, expressions and poses.

Another problem is when the collection of face images is performed using sensors that operate at different spectral bands (visible or infrared). The main question here is “depending on the application and the conditions (controlled or uncontrolled) under which we operate, which band should be selected (either independently or in combination with others) so that we can achieve high face/eye detection and face recognition accuracy?”. Since we are dealing with different camera systems, covering different bands (multi-spectral imaging) and different wavelengths within each band (hyperspectral imaging), lets move on to discuss about the IR spectrum and the focus of our work.

The infrared (IR) spectrum is comprised of the active IR band (near-IR or SWIR), and the thermal (passive) IR band. The passive IR band is further divided into the mid-wave (MWIR) and long-wave IR (LWIR) bands. The MWIR range is 3 μm to 5 μm, whereas the LWIR range is 7 μm to 14 μm. Both MWIR and LWIR cameras can sense temperature variations across human faces at a distance and produce thermograms in the form of 2D images. In this paper, our work is focused in the SWIR band. It is a part of the reflected IR (active) band (in our experiments, it ranges from 0.9 μm to 1.7 μm). SWIR has a longer wavelength range than Near Infrared (NIR) and is more tolerant to low levels of obscurants like fog and smoke. Differences in appearance between images sensed in the visible and the active IR band are due to the properties of
the object being imaged. The reflected spectral signatures in the SWIR band require an external light source. However, some of the benefits of SWIR-based imaging systems are that they can take advantage of sunlight, moonlight, or starlight [1]. Another benefit is that when images in the SWIR spectrum are fused with visible images, FR performance can increase [2].

There are three main disadvantages of using the SWIR band in comparison to the LWIR band. First, thermal imagery can be acquired without any external illumination during day or night environments, while regions in the active IR band might require an external light source. Second, vein patterns or other anatomical features not observable in the SWIR band, can be observable in the LWIR. Finally, background clutter in thermal images is not always visible. For example, the texture of a wall will not usually be visible if it is uniform and has the same surface temperature signature. Thus, under certain conditions, we can say that when operating in the thermal band, the tasks of face detection, localization, and segmentation (fundamental processes of a typical face recognition system) are comparatively easier and more reliable than when operating in the active IR and visible bands.

1.1. SSMW face recognition systems in the SWIR band

Conventional imaging systems use a specific sensor (e.g., an SWIR camera) that can be operated without an external hardware, and utilize their complete spectral range to capture images. The information is collected over the wide spectrum and the integration process is responsible for getting less qualitative information than multi-spectral systems. Moreover, it is difficult to separate the information related to the light distribution (amount of light absorbed and reflected back from an image) when operating at an arbitrary band. Multi-Imaging Systems (MIS) are either Multi-Sensor (MS), Single-Sensor Multi-Wavelength (SSMW), or a combination of the two, i.e., Multi-Sensor Multi-Wavelength (MSMW). MIS are composed of multiple sensors that operate in different bands. For example, we can use band-specific cameras to acquire images in the visible, NIR and SWIR bands. On the other hand, SSMW imaging systems utilize a single imaging sensor in combination with external hardware. Such systems, before applying the aforementioned processing steps, are capable of acquiring images at specific wavelengths within the same band, e.g., a camera system like that can have a set of wavelength-selective band pass filters placed in front of a camera.

The key drawback of current SWIR-based face image acquisition systems is that they lack the capability of real-time simultaneous acquisition of multiple, wavelength-specific face images. In our work, we have developed a SSMW face image acquisition system, where faces are acquired at a selected set of wavelengths. The proposed system supports the acquisition and usage of unique facial information per individual that can enhance the performance of our proposed FR system. However, when using a SSMW system, such as the one proposed for the purpose of this work, there are challenges that the system designer has to mitigate.

During data collection, in order to acquire good quality face images, different design steps need to be considered, including the selection of the appropriate hardware components and image acquisition parameters that the practitioner can select via the graphical user interface that handles image acquisition. Design steps include (i) the selection of light source (e.g., tungsten, fluorescent etc.), (ii) the selection of the wavelength to focus the camera during data collection, (iii) the setting up of the optical sensor response, (iv) the camera and filter wheel synchronization, and (v) the set up of the filter wheel speed. In our previous work [3], an empirical optimization of the experimental set up was performed to acquire good quality face images. In this work, we are focusing on developing the necessary algorithms to further pre-process the acquired good quality face images. This design step is very important because it can further contribute to the performance of our proposed SSMW FR system. What follows is a description of the three specific challenges that this paper is addressing.

- **Challenge 1**, classification of multi-wavelength face images to individual IR wavelengths: The purpose of our proposed classification approach is that, in practice, during or after data collection, an operator needs to deal with the processing of many face datasets (as we will show, the data were captured at five different wavelengths within the SWIR band). This can be a large pool of images that need to be categorized to the right wavelength. When this effort is performed manually, it can be very time consuming and can result in many errors, especially when dealing with large datasets [4]. Thus, classification needs to be performed automatically, using either supervised or unsupervised methods [5]. In this work, we propose an automated quality-based score level fusion scheme for the classification of multi-wavelength face images. Classification is performed using a combination of parametric and non-parametric statistical-based methods, namely an estimation-based Bayesian Classifier [6] and a multi-wavelength face image k-NN classifier [7]. The big challenge we originally had to deal with was the extraction and, then, selection of features. The way we mitigated this challenge will be discussed in detail in the Methodology section.

- **Challenge 2**, classification of frontal vs. non-frontal face images: FR systems perform well when dealing with full-frontal face images. In this work, our collected SSMW dataset consists also of face images captured at variable face poses. In order to keep only full-frontal, good quality, face images to be used for FR studies, we developed an automated method that classifies frontal vs. non-frontal face images based on a set of image quality scores. These scores were generated by estimating different image quality factors (such as blurriness).

- **Challenge 3**, wavelength-based selection of frontal face images for improved FR performance: Our SSMW FR system was used to first focus the camera at various wavelengths, and then, capture multiple face samples at various wavelengths and at various poses and light sources. Thus, it is pertinent to investigate under which conditions good quality, full frontal, face images can be selected out of the raw dataset for further processing (face matching). The goal is to have a fully automated system where accurate localization of faces and eyes as well as high recognition performance can be achieved. The system should be able to perform well when using the normalized face images generated from the selected full frontal, good quality dataset constructed after we successfully mitigated the previous two challenges.

1.2. Goals and contributions

In this paper, our main contributions are the following: (i) A SSMW system is designed and developed that operates within the SWIR band. By using our system, a database was collected under the following scenarios, i.e. Multi-Wavelength (MW) images were captured (by placing a tunable filter wheel in the optical path) when the camera was focused at either 1150, 1350 or 1550 nm. We also developed, a hybrid image quality assessment method (reference and no-reference-based) to compute different image quality scores from our collected face images. (ii) An automated image quality-based score level fusion scheme was also designed and developed. The purpose of this scheme was the automated classification of multi-wavelength face images to individual wavelengths. Note that the images were acquired when the camera was focused at three different SWIR wavelengths, namely the 1150, 1350 and 1550 nm. (iii) An image quality, weighted-based, score level fusion scheme, was designed and developed for the automated classification of full frontal vs. non-frontal SWIR face images. (iv) A set of experiments were conducted indicating that our proposed classification algorithms developed are beneficial when designing different steps of a multi-spectral face recognition (FR) system, including face detection, eye detection and face recognition. In the latter set of experiments, we also determined the most beneficial wavelength to focus the camera in order to acquire good quality SWIR face images and achieve increased...
recognition rates. Finally, we determined which SWIR wavelength should be selected to provide us with good quality face images and high recognition rates, independent of the wavelength the camera is focused at. As we will discuss below, the study was based on Image Quality Assessment (IQA) and Face-based Image Quality Assessment (FIQA) methods.

1.3. Paper organization

The rest of the paper is organized as follows. Section 2 briefly reviews related work in the literature. Section 3 describes our proposed methodologies, the developed multi-spectral face image acquisition system and the face datasets collected. Section 3 also provides the results from the experiments conducted, before conclusions are made in Section 4.

2. Related work

MIS systems are usually available in the visible and near infrared (NIR) regions. Munsell color laboratory developed an imaging system that can acquire images in the visible band, aiming to overcome the problems related to loss of information, using only three visible bands and color distortion information [8]. In their system, they combined the commercial trichromatic camera with two additional filters that were placed in the optical path, and the evaluation of color reproduction was then performed.

Fig. 1. Overview of the SSMW system operating in the SWIR band. Step 1: the camera is focused at one of the selected wavelengths (1150, 1350 or 1550 nm); Step 2: multiple images are acquired during wheel rotation for a selected set of wavelengths. The acquired images are shown at the bottom of the figure, when the camera is focused at 1150 nm (rows 1, 2), 1350 nm (rows 3, 4) and 1550 nm (rows 5, 6). The remaining steps that involve classification, pre-processing, evaluation of face images in terms of the image quality, and selection of the SWIR wavelength to focus the camera, will be discussed in the Methodology section.
Pan et al. [9], proposed a SSMW imaging system that can capture images in the NIR band by placing a liquid crystal tunable filter (LCTF) wheel in the optical path. The LCTF wheel is built of a birefringent liquid crystal plate and a set of polarizers. The time required for the wheel to move from one polarizer to another depends upon the relaxation time of the liquid crystal plate. The system was used for the collection of face images under variable facial expressions and poses.

Chang et al. [10], proposed an automated method that specifies the optimal spectral ranges to improve the performance of their recognition system in comparison to the conventional broad-band system. A liquid crystal tunable filter [11], operating in the visible spectrum, was used to acquire multiple images, which were generated after focusing the narrow band filter at different wavelengths between 400 nm and 720 nm. The authors illustrated that when using their system, FR performance in the visible spectrum can be improved. In that study, faces with frontal pose and neutral expression were mainly considered.

For the multiclass image classification [12], authors proposed the bag of features method to classify the images. This method generates a codebook or dictionary. They have implemented this method when using visible images to classify the objects based on a selected set of features. To implement this method in our experimental set, where we need to perform for multi-wavelength face image classification is a big challenge. The problem is that this method suffers from high complexity to discriminate among various face features extracted from SWIR band images. The limitations of this method, when we considered to apply it to our face images, are determined to be lack of geometry representation and high complexity in the number of classes [13]. Sun et al. [14], proposed a method regarding the classification of hyper-spectral images in the field of remote sensing area. The classification is performed based on the measured spectral response by using a spectrometer.

Namin et al. [15], proposed a method to classify images in visible and NIR band captured with a multispectral camera. The subject’s face images look similar when captured using a conventional camera but are more easy to discriminate when captured using a multispectral camera. The classification was performed using local features and an SVM classifier. Their band classification (visible vs. NIR) is not considered to be as challenging as our proposed wavelength (within band) classification that we will discuss below.

While band classification is a challenging problem, pose estimation for the purpose of improving FR performance is another problem. Researchers have adopted different methods to estimate the variation in

<table>
<thead>
<tr>
<th>Average quality scores</th>
<th>1150 nm</th>
<th>1250 nm</th>
<th>1350 nm</th>
<th>1450 nm</th>
<th>1550 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRIQA method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharpness (q1)</td>
<td>0.018</td>
<td>0.041</td>
<td>0.051</td>
<td>0.055</td>
<td>0.052</td>
</tr>
<tr>
<td>Blurriness (q2)</td>
<td>11.35</td>
<td>7.17</td>
<td>5.93</td>
<td>5.08</td>
<td>6.65</td>
</tr>
<tr>
<td>RIQA method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminance (q3)</td>
<td>0.90</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>Contrast (q4)</td>
<td>0.77</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td>AD (q5)</td>
<td>−70.98</td>
<td>−20.97</td>
<td>24.54</td>
<td>−6.0</td>
<td>−0.99</td>
</tr>
<tr>
<td>SC (q6)</td>
<td>0.45</td>
<td>0.72</td>
<td>1.30</td>
<td>0.88</td>
<td>0.97</td>
</tr>
<tr>
<td>PSNR (q7)</td>
<td>9.04</td>
<td>10.24</td>
<td>10.27</td>
<td>10.39</td>
<td>11.83</td>
</tr>
<tr>
<td>NK (q8)</td>
<td>1.37</td>
<td>1.02</td>
<td>0.71</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>NAE (q9)</td>
<td>0.61</td>
<td>0.50</td>
<td>0.51</td>
<td>0.47</td>
<td>0.41</td>
</tr>
<tr>
<td>MSE (q10)</td>
<td>8096</td>
<td>6148</td>
<td>6110</td>
<td>5943</td>
<td>4263</td>
</tr>
<tr>
<td>MD (q11)</td>
<td>167</td>
<td>237</td>
<td>247</td>
<td>250</td>
<td>222</td>
</tr>
</tbody>
</table>
pose and to further improve the performance of FR systems, Guillemaut et al. [16], proposed a pose-specific recognition method. They adopted the active appearance models, which were used toemap the non-frontal probe image into a frontal image. This method needs frontal images in the gallery set to train the face recognition system. Frontal facial pose recognition system [17] was proposed using a discriminate feature extraction method. Most of these methods are training based methods and perform well when using visible and NIR face images and there has been no study, to our knowledge, that the same methods can perform well in a SSMW system such as the one we propose in this work.

Most of the face recognition systems reported in the literature are focused in visible and NIR band [2]. Li et al. [18], developed a face recognition system in the NIR band to minimize the effect of environment illumination conditions. Bourlai et al. [19], investigated the problem of FR in the NIR band at long distance, when performing cross spectral and cross distance matching. In this paper, we move away from the NIR band and focus on the problem of FR in different wavelengths of the SWIR band. What follows, is a discussion of our proposed methodological approach that manages to efficiently deal with a variety of problems related to the automated pre-processing and matching of multi-wavelength SWIR-based face images.

3. Methodology

In this section, we outline the algorithms developed for the automated classification of multi-wavelength face images, the classification of frontal vs. non-frontal face images, and the evaluation of our SWIR-based SSMW system in terms of determining the effect on image quality and FR performance when selecting different wavelengths to focus the camera and then, capture face images. Below we describe each step of our proposed methodological approach.

3.1. Quality based score level fusion scheme for automated classification of multi-wavelength face images to individual IR wavelengths

The classification of multi-wavelength SWIR images to individual wavelengths (1150, 1350 and 1550 nm) is a challenging task. The key problem is that the computed image quality scores are close to each other [3,1]. In this work, an automated method is developed that takes as an input all face images of our database and classifies them to individual wavelengths. There are two main approaches to deal with multi-class classification problems: One-versus-All (OvA) and All-versus-All (AvA).

Our empirical study determined that when the classification is performed using the AvA approach, the computational cost and elapsed time is less in comparison to OvA. In AvA, each class is compared to all available classes and predicts the class label to which the test image belongs to, based on the selected models (i.e. maximum posterior probability for naïve class and minimum distance for k-NN classifier).

In our system, 1150 nm images are selected as class 1, 1250 nm as class 2, 1350 nm as class 3, 1450 nm as class 4 and 1550 nm as class 5. The main steps of our classification approach involve, (i) database collection, (ii) feature extraction, (iii) feature selection, (iv) model selection and (v) training and classifier evaluation.

- Database collection: The constructed wheel, placed in front of the SWIR camera, has five filters, i.e. 1150, 1250, 1350, 1450 and 1550 nm (see Fig. 1). During data collection, we focused the camera at 1150, 1350 and 1550 nm and collected the face images of 30 subjects. The data collection was performed in 2 sessions on different days. Each session took 25 min, while both sessions required 50 min in total. Unfortunately, not all subjects could come on the same day, so we had to collect Session 1 within a couple of days, and Session 2 within a couple of days too. This was a convenient solution for the subjects that had to appear two times within a few weeks. In each session we collected 275 frames per subject (Fig. 2(a) i.e. 5 set of selected wavelengths × 55 images per wavelength) [3]. The face images collected are left profile, right profile, and full frontal.

- Feature extraction: We work with a set of features typically used for determining (computing) the levels of different image quality factors such as sharpness, blurriness, structural content, and contrast. The features are measured using reference-based image quality assessment (RIQA) or non-reference based image quality assessment (NRIQA) methods (see Fig. 2(b)). In RIQA methods, the quality image is compared against a reference or target image that is considered to be an image of satisfactory quality (e.g. face image captured under controlled conditions of a highly cooperative subject). RIQA methods are fast and easy to compute. However, reference images may not always be available. In our previous work, reference based quality assessment methods provided us with good results (baseline) by defining specific conditions for the selection of the wavelength to focus the camera [3]. In this work, we managed to improve the performance of our baseline system, by considering both reference and no-reference based (no target image was used to establish the quality of the query image) quality assessment methods [20,21].

Next, we will describe the process we used to extract the necessary features for classification when using either the RIQA or the NRIQA methods.

- Feature extraction using RIQA methods: In our work, we use objective quality assessment methods based on (i) average pixel distance (AD), which is defined as the pixel by pixel difference between the reference and query images, (ii) maximum pixel distance (MD) as maximum difference exists, (iii) PSNR measure, which is defined as

<table>
<thead>
<tr>
<th>Selection of features</th>
<th>f1u</th>
<th>f2u</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1 (reference scores)</td>
<td>c3 + q5 + q7 + q9 + q11</td>
<td>q4 + q6 + q8 + q10</td>
</tr>
<tr>
<td>c2 (no-reference scores)</td>
<td>q1</td>
<td>q2</td>
</tr>
</tbody>
</table>

Reference and no-reference

- c3 q1 + q2
- c4 q1 + q2
- c5 q1 + q2
- c6 q1 + q2
- c7 q1 + q2
- c8 q1 + q2
- c9 q1 + q2
- c10 q1 + q2
- c11 q1 + q2
- c12 q1 + q2
- c13 q1 + q2

Selection of features for fusion by considering different combinations cn (where, n represents the selected combination number i.e. 1, 2, …, 13). Features acquired from NRIQA: q1 — Sharpness, q2 — Blurriness and from RIQA: q3 — Luminance, q4 — Contrast, q5 — SC, q6 — PSNR, q7 — NK, q8 — NAE, q9 — MSE, q10 — MD, q11 — AD.
the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is defined (in units of decibels) via the MSE. (iv) MSE and PSNR are measured as described in Eqs. (1) and (2). A higher PSNR would normally indicate that the reconstruction is of higher quality. (v) NAE is defined in Eq. (3). The large value of Normalized Absolute Error means that the image is of poor quality. (vi) Luminance is measured using Eq.(4) where \( \mu_x \) and \( \mu_y \) represent mean intensity, (vii) contrast using Eq. (5) where \( \sigma_x \) and \( \sigma_y \) represent base contrast, (viii) structural content from Eq. (6) \( \sigma_{xy} \) represents the correlation coefficient and (ix) normalized correlation (NK). The comparison between a reference (a good quality face image captured under controlled conditions for each subject in the database) and a query image (captured in the SWIR band) provides a numerical quality index value (output). The aforementioned quality measures are illustrated in Fig. 2(b) and coded as q3, q4, ..., q11 respectively.

\[
\text{MSE} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (x(i,j) - y(i,j))^2 \tag{1}
\]

\[
\text{PSNR} = 10 \times \log_{10} \frac{D^2}{\text{MSE}} \tag{2}
\]

\[
\text{NAE} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} |x(i,j) - y(i,j)|}{\sum_{i=1}^{M} \sum_{j=1}^{N} |x(i,j)|} \tag{3}
\]

- Feature extraction using NRIQA methods: The two quality factors computed are, spectral sharpness, and blurriness (see Fig. 2(b), coded as q1 and q2 respectively). In order to measure the first quality factor, we developed an approach based on a method that was originally developed to deal with visible images [20], but is tuned to operate with our SWIR face images. First, each SWIR face image is divided into blocks with size of 16 × 16 and then, image contrast is calculated as the ratio of spread of intensity values to mean value for each block (using the contrast Root Mean Square or RMS method). Next, the spectrum magnitude and its slope are calculated from the radial frequency component and orientation [22].

\[
l(x,y) = \frac{2\mu_x \mu_y}{\mu_x^2 + \mu_y^2} \tag{4}
\]

\[
c(x,y) = \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{5}
\]

\[
s(x,y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{6}
\]

\fig{3}{Variance plots for extracted set of features for the 1150 nm wavelength when the camera is focused at 1350 nm.}

\fig{5}{Variance plots for extracted set of features for the camera is focused at 1350 nm.}

\fig{4}{Variance plots for extracted set of features for the 1150 nm wavelength when the camera is focused at 1350 nm.}

\fig{4}{Variance plots for extracted set of features for the 1150 nm wavelength when the camera is focused at 1350 nm.}

\fig{5}{Estimation of distributions for fused scores (fu1 top figure and fu2 bottom figure) from combination c13.}
Spatial sharpness is measured based on maximum variation between the blocks. The total variation is computed by combining both the spectral and spatial sharpness using a weighted mean method [21]. Blurriness (i.e. the second quality factor) is computed based on the method proposed in [23]. We know that poor quality images can be the result of camera shake, motion, and de-focusing [24]. In this work, the blur parameter is calculated from the line spread function (LSF), which is constructed using the edge spread function [23]. The edge points correspond to the local maximum gradient in a blurred image. The point spread function (PSF) is identified using the edges in the arbitrary directions. First, we applied a Gaussian filter to perform image smoothing. Edges were detected using the Sobel operator.

**Fig. 6.** Variance plots for combinations, c1 (top left), c3 (top right), c2 (bottom left) and c4 (bottom right).

**Fig. 7.** Variance plots for combinations, c5 (top left), c7 (top right), c6 (bottom left) and c8 (bottom right).
After edge detection is applied in arbitrary directions, lines or straight edges are calculated from the radon transform [23]. Eventually, the blur parameter (PSF) is calculated from the determined line spread function (LSF).

A number of quality scores are calculated using the aforementioned image quality assessment methods. A total of eleven quality measures ($q_1, q_2, \ldots, q_{11}$) that we compute and use as features in our proposed SWIR wavelength-based classification scheme. Features $q_1$ and $q_2$ are computed using NRIQA methods, while $q_3$ to $q_{11}$ using RIQA methods (as illustrated in Table 1).

- **Features selection (fusion scheme):** Although the selection of features used in the fusion scheme was originally a big challenge, we dealt with this problem by considering 13 different combinations in total labeled as $c_1, c_2, \ldots, c_{13}$ (as illustrated in Table 2). Finally we selected the combination of features for fusion, which provided the best results. In Table 2, $fu_1$ and $fu_2$ represent the fused scores generated from the NRIQA methods and RIQA methods. The selected combination for features (the quality scores) are fused together (sum rule) as illustrated in Eq. (7) (see Fig. 2(c)) and converted to a 2D feature space [25]. The $S_i$ in Eq. (7) is defined as sum of features (where $i$ in the range of $[n_1, n_2, \ldots, n_n]$ the number of features selected for fusion, depends upon the combination selected from Table 2). For combination $c_1$, scores are fused and represented by $fu_1$ (the value of $n_n$ is $n_5$) and $fu_2$ (value of $n_n$ is $n_4$). For $c_{13}$, the value of $n_n$ is $n_2$ for $fu_1$ ($q_1 + q_2$) and $n_7$ for $fu_2$ ($q_3, q_4, q_5, q_6, q_7, q_8$ and $q_{11}$).

\[
Score = \sum_{i=n_1}^{n_2} S_i
\]  

- **Statistical analysis of features:** In order to examine the effectiveness of the extracted features and before fitting these extracted features into the model, the statistical analysis tests are performed. To see the variance between the selected set of features, we first plotted the box-plots for all the five wavelengths and the plots for 1150, 1350 and 1550 nm wavelength are shown in Figs. 3, 4(a) and (b) for all the selected set of features from RIQA and NRIQA methods.

From these plots, it is clear that we cannot use these scores to classify the images into individual wavelengths, as scores are close to each other (scores are normalized). To deal with this problem, we selected the fusion score method. We selected the sum rule fusion method with 13 different combinations and statistical analysis was performed to select the best combination.

The first test was performed to estimate the distribution of all the acquired scores. Then, based on the kernel density estimation method, distributions are not normal as shown in Fig. 5. Thus, based on the distribution estimation test, we have concluded that we cannot consider the parametric tests like ANOVA1 and ANOVA2 to find the relationships between and within features.

Finally, we considered the Kruskal–Wallis test (non-parametric test) and variance plots are used to select the best combination to fit the data into model. Variance plots from all the 13 combinations ($c_1, c_2, c_{13}$) are...
Figs. 6, 7, 8 and 9. From Fig. 6 we concluded that, for all the four combinations c1, c2, c3 and c4 fused score values fu1 and fu2 are very close to each other for five classes (1150, 1250, 1350, 1450 and 1550 nm) and we cannot fit these fused scores in the model to classify them into individual wavelengths and same is the case for combinations from c5 to c12 (see Figs. 7 and 8). There is overlap between the fused scores (fu1 and fu2) between the classes from the combination c1 to c12.

From combination c13 (see Fig. 9), fused score fu1 is close but the fused score fu2 has no overlap between the values between five classes. Based on these statistical analysis tests, we concluded that the combination c13 is selected as the best among all combinations.

Model selection: Model selection refers to the complexity selection or in other words flexibility to fit the data. We have applied Bayesian model and k-NN model to classify the images see Fig. 2(c). In the Bayesian classifier, the decision is made based on probability distribution of the training data and used to classify the test data. It depends upon the hyper-parameters like prior distribution and parameters such as the statistical mean and variance are computed to estimate the distribution of classes [26]. Our main goal is to maximize the model accuracy and to select the classifier with minimum classification error rate. In the case of the k-NN model, the complexity controlling parameter is the value of k (number of nearest neighbors under consideration). We select 15 different values of k and run

![Flow chart diagram representing the process involved in our proposed approach for classifying the frontal and non-frontal face images.](image)

![Steps followed that generate the canonical face images after SWIR face image preprocessing.](image)
the experiment ten times for each selected value. Finally, the value of \( k \) is selected, which provided us with the minimum classification error rate. The classification is performed based on the majority of \( k \)-nearest neighbors, and the distance is measured between an unknown class’s feature vector and \( k \) nearest neighbors using the Euclidean distance metric. The classifier is easy to use and provide us with the minimum or just a low but comparable classification error.

- **Training and testing of classifier:** To train the classifier for the selected set of features, the data is randomly divided, based on training set sizes using 20%, 30% and up to 80%, while rest of the data as testing set is used to evaluate the performance of the classifier. And, repeated this process ten different times for each set (i.e. training set sizes using 20%, 30% and up to 80%), using a random selection of training and testing data each time. To verify the validity of our results, during collection, we manually labeled all face images to their actual wavelength and pose and this was used as ground truth information. The selection of the best classifier (Bayesian or \( k \)-NN) was performed based on the classification error rate that was computed using the confusion matrix as illustrated in Eq. (8). True Positive (TP) and (TN) represent, the proportion of correctly classified images. False Positive (FP) and False Negative (FN) represent the proportion of images that are incorrectly classified.

\[
\text{Classification error} = \frac{FP + FN}{TP + TN + FN + FP} \quad (8)
\]

### 3.2. Proposed weighted quality-based score level fusion scheme for automated classification of FF vs. NFF images

The classification of FF vs. NFF face images can be done reasonably well by human observers, where the practitioner is requested to visibly distinguish different face poses. However, when having to deal with large datasets this process needs to be performed automatically. In our case, we had to solve a more complicated problem since we are dealing with multiple subjects with variable face poses captured at five different wavelengths when the camera is focused at three different wavelengths within the SWIR band.

Our proposed classification approach is based on the estimation of image quality measures but, in contrast to other reported approaches in the literature, is designed to operate in the SWIR band. Various quality scores are measured based on both RIQA and NRIQA methods (see Fig. 10). Then, the scores are fused using a scheme that is based on quality scores. Our approach is efficient, easy to operate, convenient to use, and there is no need for the users/operators to manually process data captured by multiple sensors. Although quality scores can be combined using a number of different rules, such as the simple sum, max, min, median and the weighted sum or by other approaches [25], we empirically determined that a fusion scheme, based on the weighted sum rule, provided the most promising results [27]. The proposed approach is evaluated by comparing the classification results to that of a practitioner’s, i.e. manual classification of multi-spectral face images to frontal and non-frontal. The five steps of the proposed approach, as illustrated in Fig. 10, are described below.

- **Step 1 — Normalization of quality scores computed by RIQA methods:** To normalize the scores, well known methods available for use are Z-Score, min–max, tanh etc. [25]. In this work, score normalization is done using the min–max rule, illustrated in Eqs. (9) and (10). The value of quality scores lies between 0 and 1, for both \( q_1 \) and \( q_2 \) (sharpness and blurriness) quality scores considered in this problem.

\[
qnr_1 = \frac{q_1 - \min(q_1)}{\max(q_1) - \min(q_1)} \quad (9)
\]

Before normalization of RIQA method based quality scores, fusion of quality scores is performed to overcome the dimensionality problem illustrated in Eqs. (11) and (12). Two sets of features are used before fusion. The first set consists of PSNR, SC, Lummi and Contrast, and the second set consists of the NK, MD, AD, MSE and NAE features.

\[
r_1 = (q_3 + q_4 + q_5 + q_6) \quad (11)
\]

\[
r_2 = (q_7 + q_8 + q_9 + q_{10} + q_{11}) \quad (12)
\]

After fusion, the normalized quality scores are computed based on min–max rule as illustrated in Eqs. (13) and (14). The scores \( s' \) and \( s'' \) are average scores computed from NRIQA and RIQA based methods (see Eqs. (15) and (16)).

\[
\text{rrn}_1 = \frac{r_1 - \min(r_1)}{\max(r_1) - \min(r_1)} \quad (13)
\]

\[
\text{rrn}_2 = \frac{r_2 - \min(r_2)}{\max(r_2) - \min(r_2)} \quad (14)
\]

\[
s' = \frac{\text{rrn}_1 + \text{rrn}_2}{2} \quad (15)
\]

### Table 3

<table>
<thead>
<tr>
<th>Camera focused at</th>
<th>1150 nm</th>
<th>1350 nm</th>
<th>1550 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150 nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k )-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1300 nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k )-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1550 nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k )-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Camera focused at</th>
<th>1150 nm</th>
<th>1350 nm</th>
<th>1550 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1150 nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k )-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1300 nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k )-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1550 nm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( k )-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of quality assessment methods used and, in our case, since we used two methods (RIQA and NRIQA) the value is 2. Also, \( m \) represents the individual assessment method used to measure the quality scores. Please note that the value of the weight decreases as the error rate increases.

\[
W^m = \left( \frac{1}{\sum_{m=1}^{M} \frac{1}{E_r^m}} \right)
\]

(21)

• Step 4 — Weighted sum fusion of quality scores computed by RIQA and NRIQA methods: In Eq. (22), \( n \) stands for the number of quality scores under consideration, \( s^i \) stands for the obtained scores from the RIQA and NRIQA methods. When the value of \( i = 1 \), it represents the NRIQA method. Also we have \( w^1 \) as the computed weight from Eq. (21) and \( s^1 \) as the quality scores computed from Eq. (15). When \( i = 2 \), it represents the RIQA method and \( w^2 \) from Eq. (21) is the computed weight while \( s^2 \) from Eq. (16) represents the computed scores.

\[
F = \sum_{i=1}^{n} W^i s^i \sum_{i=1}^{n} W^i
\]

(22)

• Step 5 — Classification of frontal vs. non-frontal face images: Our proposed approach for the classification of FF and NFF images is based on a weighted sum rule. To classify the images, a threshold value is selected based on the scores from the weighted fusion method.

Fig. 12. Results from selected combinations of features for fusion scheme.

Fig. 13. Results from selected combinations of features for fusion scheme using k-fold cross validation method.
Images are labeled as 1 for frontal faces and 0 for non-frontal faces as illustrated in Eq. (23b).

\[ IF \ w_1s_1 + w_2s_2 \leq \text{threshold} = \text{Frontal face} \tag{23a} \]

Else = Non-frontal face \tag{23b}

3.3. Pre-processing

We performed a set of experiments indicating that the information extracted by the score level fusion studies above is beneficial when designing different steps of a multi-spectral face recognition (FR) system, including face detection, eye detection and recognition.

3.3.1. Face detection

Face detection is a fundamental step of a FR system. It can affect the performance of the FR system (not accurately detected faces can result in false match scores). In our work we used the Viola & Jones face detection method [28] that is applied to all three of our SWIR multi-spectral datasets (as described at the top of Section 3.1).

3.3.2. Eye detection and geometric normalization

Eye detection was performed using a WVU designed and developed software [29]. The method is based on a template-based matching approach. The eye center positions are first located on the detected face images and, then, used to geometrically normalize the images. The normalization scheme compensates for slight perturbation in frontal pose and consists of an affine transformation that is based on eye center locations.

3.3.3. Photometric normalization and masking

Photometric normalization is also applied to the SWIR face images to compensate for changes in illumination conditions Fig. 11. We applied the contrast limited adaptive histogram equalization (CLAHE), to deal with ambient variations or strong shadows. Finally, pixel normalization is performed to adjust the pixel values to have zero mean and a standard deviation of 1. The version of the face image used in our follow up face recognition experiments is resized to 150 × 130 pixels where, 70 pixels is the fixed interocular distance [30].

3.4. Face recognition systems

In order to perform the face recognition experiments the standard FR methods provided by the CSU Face Identification Evaluation System [30] were employed, including Principle Components Analysis (PCA) [31], a combined Principle Components Analysis and Linear Discriminant Analysis (PCA + LDA) algorithms [32] and the Bayesian Intrapersonal/Extra-personal Classifier (BIC) [30] using either the Maximum likelihood (ML) or the Maximum a posteriori (MAP) hypothesis [33].

MW frontal face images with good quality, obtained from the proposed algorithms and after pre-processing were randomly broken down into groups based on training set sizes using 20%, 40%, 60% and 80%, while the rest of the images were selected for testing part. We finally selected the training set that achieved the best FR performance.

4. Empirical evaluation

A unique facial image database (see Fig. 1) was acquired to facilitate the proposed study at variable focus points for the camera. The database (DB) is composed of three datasets (DB1, DB2, DB3). SSMW images for the camera focused at 1150, 1350 and 1550 nm. Our SSMW system consists of the following components: a SWIR Goodrich camera, a 5-position rotating filter wheel, a servo-motor and a reflective sensor.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Selected value of k for k-NN classifier after repeating each combination ten times.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected value of k</td>
<td>Accuracy</td>
</tr>
<tr>
<td>3</td>
<td>96.24</td>
</tr>
<tr>
<td>5</td>
<td>95.85</td>
</tr>
<tr>
<td>7</td>
<td>95.18</td>
</tr>
<tr>
<td>11</td>
<td>94.42</td>
</tr>
<tr>
<td>13</td>
<td>93.98</td>
</tr>
<tr>
<td>15</td>
<td>93.67</td>
</tr>
<tr>
<td>17</td>
<td>93.30</td>
</tr>
<tr>
<td>21</td>
<td>93.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Results for k-NN and Bayesian classifier for randomly selected training sets.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera focused at classifier</td>
<td>1150 nm</td>
</tr>
<tr>
<td></td>
<td>k-NN</td>
</tr>
<tr>
<td>20%</td>
<td>96.74</td>
</tr>
<tr>
<td>30%</td>
<td>93.25</td>
</tr>
<tr>
<td>40%</td>
<td>94.37</td>
</tr>
<tr>
<td>50%</td>
<td>95.10</td>
</tr>
<tr>
<td>60%</td>
<td>95.35</td>
</tr>
<tr>
<td>70%</td>
<td>95.68</td>
</tr>
<tr>
<td>80%</td>
<td>95.89</td>
</tr>
</tbody>
</table>
The spectral response of the camera depends upon four main factors: the spectral distribution of the light source \( L(\lambda) \), the spectral reflectance from the target (face image of a subject) \( R(\lambda) \), the spectral transmittance \( T(\lambda) \) of the filter placed in front of the camera, and the camera sensitivity. The face of a subject under study can be illuminated by a light source. The reflected light from the subject's face passes through one of the filters of the wheel, before it finally reaches the camera sensor. The evaluation of the experimental scenarios investigated in this paper are described in the following sections.

4.1. SWIR wavelength-based classification evaluation

In this section, we illustrate the classification results when applying our proposed approach on sample face images and the associated quality scores used for classification. The main steps involved are, feature extraction, feature selection (for fusion scheme and to reduce the dimensionality problem for feature space) and finally the evaluation of best classifier for three different scenarios. Features are extracted based on RIQA and NRIQA methods as shown in Table 1 and for the feature selection part, we have considered 13 different combinations (as illustrated in Table 2). For the selection of training and testing data we have selected the random sampling method.

To get the more generalized results, the \( k \)-fold cross validation method was considered and compared to the results with random selection method. In \( k \)-fold cross validation, the value of \( k \)-fold was selected. The experiments are performed for three scenarios, when the camera is focused at 1150, 1350 and 1550 nm as shown in Table 4. And for each experiment in 10-fold cross validation, the data is divided into 10 equal size folds, trained on 9 folds and tested on 1 fold. The experiment was repeated 10 times and mean value is selected this way. We have applied this method on both selected classifiers (Bayesian and \( k \)-NN). The same procedure is repeated for all the 13 combinations. Results from cross validation method are comparable to the random selection method (see in Tables 3 and 4) and scores are very close to each other.

Based on the results, it is concluded that the selection of features for fusion effects the performance of the classifier. As shown in Figs. 12 and 13(a) and (b), combination c1 (based on reference feature scores) has an accuracy of more than 85% for random selection and cross validation method for \( k \)-NN classifier. For \( c_5 \), accuracy is more than 97% for Bayesian classifiers for both the selected methods (random and \( k \)-fold). Combination c1 works well for Bayesian but results are not satisfactory for \( k \)-NN and \( c_5 \) works well only for the Bayesian classifier. The best results we obtained are from \( c_{13} \) for both the classifiers, using both random and cross validation methods and accuracy is higher than 95% (see in Tables 3 and 4). We have selected the combination \( c_{13} \) (selection of both reference and no-reference features) out of 13 different combinations for both (Bayesian and \( k \)-NN classifiers).

For \( k \)-NN classifier, to select the value of \( k \), for the quality scores extracted from images in the SWIR band, 15 different values of \( k \) (see Fig. 14) are experimented with. Finally, the selected value of \( \text{protect } k \) is 27, the accuracy is higher than other selected values and is greater than 95% (Table 5 represents average value from each selected value of \( k \) after repeating 10 times). Based on the results, it is concluded that results are comparable from both the classifiers (Bayesian and \( k \)-NN). For the Bayesian classifier, best performance results are achieved from 80% and 60% training data for the camera focused at 1150 and 1530 nm respectively. The rest of data is used for testing (as shown in Table 6) and accuracy is greater than 96%. Whereas for \( k \)-NN, best results are obtained from 20% training data for the camera focused at 1150 and 1530 nm.

The selected \( k \)-NN classifier (where 20% data is used for training), provides us with more data for testing of MW images captured when the camera is focused at 1150 and 1530 nm. When the camera is focused at 1550 nm, Bayesian classifier is selected (for \( k \)-NN classifier the variance is large and mean value is smaller (box-plot representation) than Bayesian classifier see Fig. 15).

The performance is more than 95% for all the three scenarios when the camera is focused at 1150, 1350 and 1550 nm, for both classifiers (Bayesian and \( k \)-NN) and is higher than other distributions of datasets (see Fig. 15). In addition, based on the mean and variance of boxplots, the accuracy is higher when using images captured with the camera focused at 1350 nm (see Fig. 15) in comparison to the camera focused at 1150 nm and 1550 nm.

We have also compared the results from random selection method with the \( k \)-fold cross validation method. For the \( k \)-fold cross validation method, the data is divided into different sets of folds. For 20% training data, the data is divided into 5 folds (\( k=5 \) fold) where the first fold is...
used for training data and rest of the 4 folds combined together into one fold. This experiment is repeated 5 times. For 25% training data, the data is divided into four folds \((k = 4 \text{ fold})\), where the first fold is used for training and the rest of the folds are combined to get the test data. We have followed the same procedure for the rest of the training sets and applied the Bayesian classifier. Finally, we compared the results from the random selection method with \(k\)-fold cross validation method see Table 7. Based on the results, it is concluded that the results are comparable from both the methods.

### 4.2. Frontal vs. non-frontal classification evaluation

The quality scores used for the classification of FF vs. NFF face images are computed based on NRIQA and RIQA methods. The scores computed using the reference-based method (i.e. quality scores such as the luminance, contrast, SC and PSNR and NK, NAE, MSE, MD and AD) are fused together using the simple sum rule and converted into a 2D feature space. The fused scores are normalized using the simple min max rule. The weights are assigned to individual normalized scores for both reference and non-reference based methods according to the computed ER and the assigned weights are shown in Table 8. Fig. 16 represents the quality scores (spatial sharpness), for non-frontal and frontal faces at five different wavelengths. Here, face images depict frontal and non-frontal face masks obtained from spatial mapping. To check the accuracy of the developed automated method, classification of frontal and non-frontal faces is performed based on the human visual system. In terms of manual classification, faces are classified based on human observation and this is performed for the camera focused at three different wavelengths. The confusion matrix is generated to check the classification accuracy by making a comparison between the

![Graphical mapping based on quality scores. The vertical axis depicts quality scores (sharpness) from NRIQA. Face images at the top side of the figure depicts frontal and non-frontal face masks obtained from spatial mapping.](image)

**Fig. 16.** Graphical mapping based on quality scores. The vertical axis depicts quality scores (sharpness) from NRIQA. Face images at the top side of the figure depicts frontal and non-frontal face masks obtained from spatial mapping.

![Boxplots depict performance of face detection methods for SWIR multi-spectral images for the camera focused at three different wavelengths (1150 nm, 1350 nm and 1550 nm). The blue box is for the camera focused at 1150 nm, the green box is for the camera focused at 1350 nm, and the red box is for camera focused at 1550 nm. The table at the right of the figure represents the accuracy results.](image)

**Fig. 17.** Boxplots depict performance of face detection methods for SWIR multi-spectral images for the camera focused at three different wavelengths (1150 nm, 1350 nm and 1550 nm). The blue box is for the camera focused at 1150 nm, the green box is for the camera focused at 1350 nm, and the red box is for camera focused at 1550 nm. The table at the right of the figure represents the accuracy results.
automatic methods investigated with the manually classified images. In the confusion matrix, data is represented as a $2 \times 2$ matrix.

$TP$ represents the proportion of frontal face images that are correctly classified as frontal faces from the developed automated method. $TN$ represents the proportion of non-frontal faces that are correctly classified as non-frontal faces. $FP$ represents the proportion of face images incorrectly classified as frontal while they are non-frontal and $FN$ is the proportion of face images that are incorrectly classified as non-frontal while they are frontal. Based on the results using our confusion matrix, we have concluded that the accuracy for images at 1150 nm is more than 90% for the camera focused at 1350 nm filter. For the 1250 nm, 1350 nm and 1550 nm images, the system accuracy is more than 85% for all the three different focus points used. In the case of the 1450 nm, the classification accuracy is lower and approximately 81%.

In the case of 1350 nm the classification accuracy is higher than the other two focus points and is more than 90%, but the accuracy drops when using 1450 nm images. A valid reason for this result is the atmospheric absorption effect. The spectral response (amount of absorbed and reflected light) is unique for each image at 1150, 1250, 1350, 1450 and 1550 nm. The present moisture content in the environment has an impact on the face appearance. Water vapors are responsible for more light energy absorption at 1450 nm and images appear dark [34]. However, the reflectance is greater for images at 1150 nm in comparison to the other four selected wavelengths.

4.3. Face detection experiments

Viola & Jones face detection algorithm is applied to the three different conditions, camera focused at 1150 nm, 1350 nm and 1550 nm filter.

Based on the results, we have concluded that the face detection system performs well for the camera focused at 1350 nm (as shown in Fig. 17). In this scenario, 95% of the faces are detected and maximum accuracy reaches to 100%, for the images at 1150, 1250, 1350, 1450 and 1550 nm. Whereas, for the camera focused at 1150 nm and 1550 nm, maximum accuracy reaches to 95%.

Face detection performs well for the camera focused at 1350 nm and performance of system improves, and best results were achieved when we applied, only on the selected images, from our developed algorithms for classification of MW images and FF vs. NFF face images. Viola & Jones did not work well for images at 1450 nm and a very small number of faces are detected. When this wavelength was used, and in order to perform face recognition experiments, the 1450 nm faces are manually detected.

4.4. Eye detection experiments

Eye detection experiments were performed using both WVU’s method as well as the commercial eye detection technique (L1 eye detection).

The detection accuracy for MW images at 1150 nm, 1250 nm and 1350 nm is more than 90% using either method. For the camera focused at 1350 nm accuracy is more than 96% and is higher than other two sets for the camera focused at 1150 nm and 1550 nm. For MW images at 1450 nm and 1550 nm, accuracy, as expected due to the aforementioned reasons, is about 70% (as shown in Fig. 18).

To overcome the problem we have detected the eye centers using the WVU eye detection method [29]. When using WVU eye detection method, accuracy increases up to 97% for 1150 nm and 1350 nm images and up to 92% for 1250 nm. Accuracy was higher than the one achieved by using the available commercial detection method, and reached 71%. For images captured at 1450 nm and 1550 nm (Fig. 18), accuracy was 70%.

4.5. Face identification experiments

A number of identification experiments were performed for randomly selected training sets (i.e. 20%, 40%, 60% and 80% training set). For each FR algorithm used, each experiment was repeated five times (i.e. each time a different training set was randomly selected), based on the mean and variance of the boxplots. The boxplot that provided us with more accuracy for the system (see Fig. 19(a)) was selected.

Based on results (see Fig. 19(b)), it was determined that better performance results are achieved for both 40% and 60% training set data in comparison to the 20% training data. For 20% training, performance reaches greater than 95% for 1150 and 1350 nm images, but it did not provide us with good results for 1550 nm images. Whereas, the accuracy reaches 100% for all the selected algorithms (Bayesian Map, Bayesian ML and PCA Euclidean) using 40% and 60% training data (Table 9). Finally, a 40% training set is selected which provides a large testing set compared to 60% and has less computational cost.

Experimental results indicate that (Fig. 20), overall, the identification rate is the highest for 1150 nm images (Table 9). The accuracy reaches 100% when the camera is focused at 1150 nm (see Fig. 20). When the camera is focused at 1350 and 1550 nm, the accuracy is greater than 97%. For 1350 nm images, the identification rate is higher for the camera focused at 1350 nm (reaches 99%) than that focused at 1150 nm and 1550 nm (as shown in Table 9). Finally, the identification rate for 1550 nm images is higher for the camera focused at the 1550 nm filter than the other two wavelengths, with the rank-1 rates reaching 100% accuracy. Moreover, the identification rate is greater than 97% for 1550 nm images, when the camera is focused at 1150 and 1550 nm wavelengths.

The overall best results are attained for three wavelengths (1150 nm, 1350 nm, and 1550 nm), when the camera is focused at
1350 nm (see Fig. 20). Also, better quality images can be acquired when the camera is focused at 1350 nm followed by 1550 nm. In either case the identification rate is higher than when operating at the rest of the available SWIR wavelengths. The results are also independent of which wavelength the camera is focused at.

5. Conclusions

We have studied the problem of FR when using SSMW imaging systems in the SWIR band. Our study demonstrated that our developed SSMW face acquisition system and proposed automated algorithms are beneficial when designing different steps of multi-spectral face recognition systems including: (i) The capability to acquire a sequence of co-registered images with good quality in a very short duration of time. (ii) The capability to perform automated classification of the multi-wavelength images that takes input face images from the database and classifies them into individual wavelengths. In this scenario, we applied two classifiers, in which the quality scores were computed based on RIQA and NRIQA methods and were eventually used as feature vectors. The study concluded that, in terms of classification accuracy, the k-NN classifier provides the most promising results when the camera is focused at 1150 and 1350 nm while the Bayesian classifier performs better when the camera is focused at 1550 nm. (iii) The capability to classify FF vs. NFF images. The proposed method is not only efficient, but is also easy to operate and convenient to use. The evaluation of the proposed method was performed by comparing the automated computed results to that of a practitioner’s (that manually classified frontal and non-frontal faces). Our study concluded that the classification accuracy is more than 90% for 1150 nm face images and consistently about 90% for face images captured at all other wavelengths within the SWIR band. (iv) The capability to achieve accurate localization of faces and eyes as well as high recognition performance. Based on the experimental results, we concluded that face detection, eye detection and face recognition systems performed well when the camera is focused at 1350 nm. We also determined that 95% of face images were detected using the Viola & Jones algorithm on images captured at different SWIR wavelengths, excluding the 1450 nm. In terms of eye detection we achieved more than 96% accuracy when using face images captured at 1150 and 1350 nm while the camera was focused at 1350 nm. Finally, when using the standard FR methods provided by the CSU face identification evaluation system, the FR accuracy was almost 100% for all three scenarios we investigated.

In terms of FR performance, we determined that the identification rate is almost 100%, (a) for 1150 nm images when the camera is focused at 1150 nm (see Fig. 20). Also, better quality images can be acquired when the camera is focused at 1350 nm followed by 1550 nm. In either case the identification rate is higher than when operating at the rest of the available SWIR wavelengths. The results are also independent of which wavelength the camera is focused at.
at 1150 nm, (b) for 1350 nm images when the camera is focused at 1350 nm and (c) for 1550 nm images when the camera is focused at 1550 nm. The overall best FR results are attained for all the three wavelengths (1150, 1350 and 1550 nm), when the camera is focused at 1350 nm.

Finally, we demonstrated that when using face images captured at 1550 nm, high identification rates are obtained that are also independent of which wavelength the camera was focused. This conclusion is particularly important for unconstrained FR scenarios, where we may need to capture face images at 1550 nm (eye safe wavelength), at long distances and when operating at night time environments (preferable over other SWIR wavelengths).

For our designed SSMW system, multi-wavelength classification is performed using the baseline classifiers. In the future, we are planning to improve the quality of our work by testing more classification algorithms such as CVM or SVM. Another improvement will be in the development of better face detection algorithms. In this work, the academic face detection algorithm we used did not work as well as expected for all SWIR wavelengths, especially for the 1450 nm face images. Thus we will develop more efficient methods that will be capable of detecting multi-spectral faces captured under both controlled and uncontrolled conditions.

**Acknowledgments**

This work was sponsored in part through a grant from the Office of Naval Research (ONR), contract N00014-09-C-0495, “Distribution A –
Approved or Unlimited Distribution”. We are also grateful to all faculty and students who assisted us with this work as well as the reviewers for the kind and constructive feedback. The views expressed in this manuscript are those of the authors and do not reflect the official policy or position of the Department of Defense, or the U.S. Government.

References