

2016

## Eye Detection and Face Recognition Across the Electromagnetic Spectrum

Cameron F. Whitelam

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Whitelam, Cameron F., "Eye Detection and Face Recognition Across the Electromagnetic Spectrum" (2016). *Graduate Theses, Dissertations, and Problem Reports*. 6944.  
<https://researchrepository.wvu.edu/etd/6944>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# Eye Detection and Face Recognition Across the Electromagnetic Spectrum

Cameron F. Whitelam

Dissertation submitted to the  
Benjamin M. Statler College of Engineering and Mineral Resources  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of

Doctor of Philosophy  
in  
Computer Engineering

Thirimachos Bourlai, Ph.D., Chair  
Bojan Cukic, Ph.D.  
Lawrence Hornak, Ph.D.  
Arun Ross, Ph.D.  
Kostas Sierros, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia  
2012

**Keywords:** Eye Detection, Face Recognition, Summation Range Filter, Template Matching, Heterogeneous Face Recognition, Document to Live Face Recognition

Copyright ©2016 Cameron Whitelam

# Abstract

## Eye Detection and Face Recognition Across the Electromagnetic Spectrum

Cameron F. Whitelam

Biometrics, or the science of identifying individuals based on their physiological or behavioral traits, has increasingly been used to replace typical identifying markers such as passwords, PIN numbers, passports, etc. Different modalities, such as face, fingerprint, iris, gait, etc. can be used for this purpose. One of the most studied forms of biometrics is face recognition (FR). Due to a number of advantages over typical visible to visible FR, recent trends have been pushing the FR community to perform cross-spectral matching of visible images to face images from higher spectra in the electromagnetic spectrum.

In this work, the SWIR band of the EM spectrum is the primary focus. Four main contributions relating to automatic eye detection and cross-spectral FR are discussed. First, a novel eye localization algorithm for the purpose of geometrically normalizing a face across multiple SWIR bands for FR algorithms is introduced. Using a template based scheme and a novel summation range filter, an extensive experimental analysis show that this algorithm is fast, robust, and highly accurate when compared to other available eye detection methods. Also, the eye locations produced by this algorithm provides higher FR results than all other tested approaches. This algorithm is then augmented and updated to quickly and accurately detect eyes in more challenging unconstrained datasets, spanning the EM spectrum. Additionally, a novel cross-spectral matching algorithm is introduced that attempts to bridge the gap between the visible and SWIR spectra. By fusing multiple photometric normalization combinations, the proposed algorithm is not only more efficient than other visible-SWIR matching algorithms, but more accurate in multiple challenging datasets. Finally, a novel pre-processing algorithm is discussed that bridges the gap between document (passport) and live face images. It is shown that the pre-processing scheme proposed, using inpainting and denoising techniques, significantly increases the cross-document face recognition performance.

*This work is dedicated to my family, whom without I would never have been able to complete this work. Their unwavering encouragement, understanding, and advice kept my head looking forward when all I wanted do was turn around.*

## Acknowledgments

First, and foremost, I would like to give my deepest gratitude to my committee chair and advisor Dr. Thirimachos Bourlai. Through your unique mentoring abilities, I have not only learned the technical and research skills one needs to complete this doctorate program, but have advanced my education on integrity and morality. I will never forget the lessons you have taught me and will strive to exude them in all of my personal and professional relationships. I am truly thankful for everything that you have done for me and will always be indebted to you.

ευχαριστώ.

To my graduate committee members: Dr. Larry Hornak, Dr. Arun Ross, Dr. Bojan Cukic, and Dr. Kostas Sierros. The commentary, direction, and discussions you all have provided me have been invaluable for strengthening this work. I have been honored to have worked with and been your student in my undergraduate as well as post graduate studies.

Finally, I would like to acknowledge my friends and colleagues: Dr. Jeremy Dawson, Dr. Nathan Kalka, Nnamdi Osia, Neeru Narang, Kyle Smith, Amanda Holbert, Jason Ice, Brianna Maze, and Christo Robison. Without your support and friendships, none of this work would have been possible.

# Contents

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction to Biometrics . . . . .	1
1.2 Advantages of the IR Spectrum . . . . .	4
1.3 Research Objective: Eye Localization . . . . .	8
1.3.1 Contributions . . . . .	9
1.4 Research Objective: Cross-spectral Matching . . . . .	12
1.4.1 Contributions . . . . .	13
1.5 Proposal Organization . . . . .	14
<b>2 Face-based Biometrics Under Challenging Conditions</b>	<b>19</b>
2.1 Eye Detection across the EM . . . . .	19
2.1.1 Visible Eye Detection . . . . .	20
2.1.2 NIR Eye Detection . . . . .	26
2.2 Face Recognition across the EM . . . . .	28
2.2.1 Cross-spectral FR . . . . .	28
2.2.2 Document to Live FR . . . . .	30
<b>3 West Virginia University Multispectral Face Database</b>	<b>37</b>
3.1 Multi-Spectral Face Collection for Cross-Spectral Identification Applications	37
3.1.1 Challenges . . . . .	38
3.2 Standard Visible Mugshot Capture . . . . .	40
3.3 SWIR Face Capture . . . . .	42
3.3.1 Tinted Material Characterization . . . . .	42
3.3.2 SWIR Face Image Collection . . . . .	43
3.4 Conclusion . . . . .	53

<b>4</b>	<b>Accurate Eye Localization in the Short Waved Infrared Spectrum through Summation Range Filters</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Methodology . . . . .	59
4.2.1	Preprocessing . . . . .	61
4.2.2	Automatic Face Detection . . . . .	61
4.2.3	Eye Region Localization . . . . .	63
4.2.4	Summation Range Filter . . . . .	63
4.2.5	Geometric Normalization . . . . .	65
4.3	Evaluation . . . . .	67
4.3.1	Face Detection Accuracy and its Effect on Automatic Eye Detection .	68
4.3.2	Effect of Using Summation Range Filter for Automatic Eye Detection	70
4.3.3	Effect of Photometric Normalization on Automatic Eye Detection . .	70
4.3.4	Effects of Variable Average Templates on Automatic Eye Detection .	76
4.3.5	Robustness of Automatic Eye Detection Approach . . . . .	79
4.3.6	Comparison with the State of the Art . . . . .	88
4.3.7	Face Recognition Study . . . . .	96
4.4	Conclusion . . . . .	97
<b>5</b>	<b>Unconstrained Eye Detection across the Electromagnetic Spectrum</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.1.1	Contributions . . . . .	108
5.2	Methodology . . . . .	112
5.2.1	Spectrum Classification . . . . .	112
5.2.2	Active Illumination Behind Glass and Long Distance SWIR Face Datasets	116
5.2.3	IR Face Datasets with and without Eye Glasses . . . . .	117
5.2.4	Visible Face Dataset with Extreme Pose and Illumination Variation .	122
5.3	Evaluation . . . . .	125
5.3.1	Databases . . . . .	125
5.3.2	Experimental Results . . . . .	130
5.3.3	Limitations . . . . .	153
5.4	Conclusion and Future Work . . . . .	155
<b>6</b>	<b>Designing SWIR to Visible Face Matching Algorithms</b>	<b>159</b>
6.1	Introduction . . . . .	159
6.2	Methodology . . . . .	161
6.2.1	FR Matchers . . . . .	161

---

6.2.2	Distance Metrics . . . . .	163
6.2.3	Photometric Normalization Empirical Study . . . . .	164
6.2.4	Score Level Fusion . . . . .	166
6.3	Evaluation . . . . .	169
6.3.1	Heterogeneous Face Recognition . . . . .	169
6.3.2	Time Efficiency . . . . .	171
6.4	Conclusion . . . . .	172
<b>7</b>	<b>Document to Live Facial Identification</b>	<b>177</b>
7.1	Introduction . . . . .	177
7.1.1	Analogy to Cross-Spectral FR . . . . .	179
7.1.2	Passport Standardization . . . . .	180
7.2	Pre-processing Techniques . . . . .	182
7.2.1	Threshold-Based (TB) Denoising . . . . .	182
7.2.2	Total Variation Wavelet Inpainting . . . . .	185
7.2.3	Exemplar-Based Inpainting . . . . .	187
7.3	Experimental Evaluation . . . . .	188
7.3.1	Evaluating Image Quality . . . . .	189
7.3.2	Evaluating Texture Based Approaches . . . . .	191
7.3.3	Assessing the Effects of Mask Dilation . . . . .	194
7.3.4	Assessing the Effects of Threshold-Based Nonlinear Denoising . . . . .	200
7.3.5	Performance Assessment using a Commercial Matcher . . . . .	205
7.4	Conclusion and Future Work . . . . .	205
<b>8</b>	<b>Conclusions</b>	<b>213</b>
8.1	Conclusions . . . . .	213
8.2	Future Work . . . . .	216





# List of Figures

1.1	An overview of a typical biometric pattern recognition system broken into its main parts, i.e. sensors, feature extraction, database, matching algorithm, score, and decision. . . . .	2
1.2	The electromagnetic spectrum depicting the active (NIR, SWIR) and passive (MWIR, LWIR) infrared bands [25]. . . . .	6
3.1	Liquid water absorption levels across the electromagnetic spectrum. Note that the absorption levels in the SWIR spectrum are significantly higher than that of the visible spectrum. Because of this, The moisture in the human skin absorbs these wavelengths and causes the face to look black. . . . .	39
3.2	Photo station layout for the visible mugshot collection aspect of the WVUMFD	41
3.3	Overview of the WVUMFD SWIR capture setup. (a) birds eye view of the setup (notice the angle at which the camera was setup to reduce reflection) and (b) picture of actual capturing setup. . . . .	44
3.4	Sample images from the WVUMFD captured using different wavelengths (1150-1550nm) and lighting conditions when the subject was behind clear glass with 80% film tint. . . . .	46
3.5	Sample images from the WVUMFD captured using different wavelengths (1150-1550nm) and lighting conditions, including the ground truth data, when the subject was wearing Oakley Flak Jackets. . . . .	51
3.6	Demographic breakdowns of subjects collected under the WVUMFD. (a) Number of participants collected weekly; (b) Participants by age group; (c) Participants by ethnicity group; and (d) demographics by gender and ethnicity.	52
4.1	The electromagnetic spectrum shows the research gap in SWIR and sample images from a) 1150nm b) 1250nm c) 1350nm d) 1450nm and e) 1550nm. Notice how eye centers from both academic (green and blue) and commercial (red) algorithms are inaccurate and produce normalized face images with both rotational and scale issues (bottom row). . . . .	58

4.2	Overview of the proposed fully automated pre-processing methodology. a) Query Image b) Photometric Normalization c) Sample Average Face and Eye Templates d) Face Detection e) Eye Detection f) Eye Center Localization using Summation Range Filtering g) Geometric Normalization . . . . .	60
4.3	Correlation map of query face image for automatic face detection. The highest peak of the correlation map corresponds to the location of the face. . . . .	62
4.4	Sample detected left and right eye regions and their corresponding summation range images. . . . .	64
4.5	Sample summation range maps and their corresponding detected eye regions. Note that the peaks are not in the center of the found regions. . . . .	66
4.6	Results of running the proposed algorithm with (red) and without (blue) the summation range filter. When the summation range filter is not used (blue), the center of the found eye region is used. At $e = 0.05$ , the summation range filter increases the accuracy of the original approach by 25.3%, 28.5% and 16.4% for the 1150, 1350, and 1550nm bands respectively. . . . .	69
4.7	Examples of photometric normalizations used: (a) Original Image (b) CLAHE (c) LBSSR (d) CLAHE LBSSR (e) TBSSR (f) CLAHE TBSSR (g) TT . . .	71
4.8	Sample success (top row) and failed (bottom row) eye detection locations on the 1150nm database. Red dots are manual locations while blue dots are found locations. . . . .	75
4.9	Different average eye templates for (a) the right eye and (b) the left eye in the 1550nm database . . . . .	76
4.10	Effects different templates have on eye detection accuracy on the 1150nm database. . . . .	77
4.11	Effects different templates have on eye detection accuracy on the 1550nm database. . . . .	78
4.12	Effect of eye detection on downsampling images from 100% to 10% in 10% increments in the 1150nm database. . . . .	80
4.13	Effect of eye detection on downsampling images from 100% to 10% in 10% increments in the 1550nm database. . . . .	81
4.14	Sample JPEG compressed images from 1150nm with a) 100% b) 50% and c) 10% compression rates. . . . .	82
4.15	Effect of eye detection on JPEG compression from 100% to 10% in 10% increments in the 1150nm database. . . . .	83
4.16	Effect of eye detection on JPEG compression from 100% to 10% in 10% increments in 1550nm database. . . . .	84
4.17	Sample images with Gaussian blur from 1150nm with a) $\sigma = 0.5$ b) $\sigma = 1.0$ and c) $\sigma = 1.5$ . . . . .	85

4.18	Effect of eye detection on Gaussian blur from $\sigma = 0.5$ to $\sigma = 1.5$ in 0.1 increments in the 1150nm database. . . . .	86
4.19	Effect of eye detection on Gaussian blur from $\sigma = 0.5$ to $\sigma = 1.5$ in 0.1 increments in the 1550nm database. . . . .	87
4.20	Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (blue), and the state of the art (black) for 1150nm assuming the faces are known up to $e \leq 0.25$ . . . . .	89
4.21	Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (blue), and the state of the art (black) for 1150nm assuming the faces are known up to $e \leq 0.10$ . . . . .	90
4.22	Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (blue), and the state of the art (black) for 1350nm assuming the faces are known up to $e \leq 0.25$ . . . . .	91
4.23	Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (cyan), Viola and Jones with photometric normalization (blue), and the state of the art (black) for 1350nm assuming the faces are known up to $e \leq 0.10$ . . . . .	92
4.24	Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (cyan), Viola and Jones with photometric normalization (blue), and the state of the art (black) for 1550nm assuming the faces are known up to $e \leq 0.25$ . . . . .	93
4.25	Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (cyan), Viola and Jones with photometric normalization (blue), and the state of the art (black) for 1550nm assuming the faces are known up to $e \leq 0.10$ . . . . .	94
4.26	ROC curve showing the results of the LBP matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red). . .	99
4.27	CMC curve showing the results of the LBP matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red). . .	100
4.28	ROC curve showing the results of the PCA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red). . .	101

4.29	CMC curve showing the results of the PCA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red). . .	102
4.30	ROC curve showing the results of the LDA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red). . .	103
4.31	CMC curve showing the results of the LDA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red). . .	104
5.1	The electromagnetic spectrum depicting the active (NIR, SWIR) and passive (MWIR, LWIR) infrared bands. In this chapter, a scenario adaptable pupil detection approach that can successfully localize eye regions (a), and eye centers (pupils) (b) when using challenging face images captured in visible and different active IR bands, including Near Infrared (NIR) and Short Waved Infrared (SWIR) is proposed. . . . .	108
5.2	Overview of a typical face recognition system. The main contributions of this paper (green) include spectrum classification, eye detection, and eye center localization. Multiple studies were conducted to show the effect of the eye detection and center localization scheme (blue). The data collection protocol as well as the face detection algorithm used for the purpose of this work (red) will also be discussed. . . . .	110
5.3	Overview of the methodological approach described in this paper. First, a classification model is built to determine which spectrum (i.e. Visible, NIR, or SWIR) a face image is subject to. Then, the eye detection approaches are employed based on the models decision. V1 describes the baseline approach used in challenging SWIR scenarios (i.e. active illumination and long distance). V2 describes the approach used in another challenging scenario (i.e. facial obstruction) in the NIR band and V3 describes the approach tested on extreme pose and illumination variation in the visible band. . . . .	111
5.4	Sample HoG images for the three spectral scenarios: a) Visible b) Near Infrared and c) Short Waved Infrared. . . . .	113
5.5	Overview of the spectrum classification algorithm. HoG descriptors are extracted for all images in the three different spectra. Then, two different SVM models are built, i.e. visible or NIR/SWIR (model 1) and NIR or SWIR (model 2). When testing, HoG features are extracted from a query image and sent into SVM model 1 to determine if it is visible. If not, it is passed into SVM model 2 to determine if it is NIR or SWIR. . . . .	114

5.6	Flowchart of the proposed eye detection in images with facial obstructions, in this case eye glasses. a) original image b) feature extraction c) SVM classifier d) eye detection/location prediction, e) pupil detection and f) geometric normalization. . . . .	117
5.7	Examples of images with and without eye glasses. a) Original Image b) Summation Range Image and c) features used for SVM. Notice that the image with the glasses produce higher reflectance pixels therefore increasing the number of SVM features. . . . .	118
5.8	Example images with eye glasses in which the original algorithm managed to either not find accurately enough, or not find the eye region at all (see red boxes). Notice how the eyebrows and large areas of reflectance causes the eye regions to be inaccurate when only using the baseline approach. Then, see how the eye localization prediction approach manages to correctly find the eye regions (see green boxes). . . . .	120
5.9	Overview of the methodology used to detect eyes under extreme pose and illumination variation in the visible dataset. . . . .	123
5.10	Sample face images, after face detection, of 2 randomly selected subjects for each of the four face databases used for the purpose of this study, i.e. Active Illumination in SWIR, Long Distance in SWIR, Face Obstruction in NIR, and Extreme pose and illumination variance in Visible . . . . .	126
5.11	Sample images in which the proposed approaches accurately finds the pupil in the SWIR, NIR, and Visible spectrum. Ground Truth = Red; Found Location = Blue. . . . .	131
5.12	WVUM: Eye detection accuracy when the subjects' face images are under active illumination (a) and at long distances (b). The proposed method is compared to L1's G8 (commercial), a state-of-the-art academic software, and three differently trained Viola and Jones algorithms. . . . .	133
5.13	CASIA: Eye detection accuracy comparing the original method (green) with the extended method (red) described in Section 3.3. . . . .	135
5.14	CASIA: Eye detection accuracy when the subjects are wearing eye glasses. The proposed method is compared to L1's G8 (commercial), a state-of-the-art academic software, and three differently trained Viola and Jones algorithms. . . . .	136
5.15	UHDB11: Eye detection accuracy comparing the original method (green) with the extended method (red) described in Section 3.4. . . . .	138
5.16	UHDB11: Eye detection accuracy when the subjects have extreme pose and illumination variation. The proposed method is compared to L1's G8 (commercial), a state-of-the-art academic software, and three differently trained Viola and Jones algorithms. . . . .	139

5.17	Top: Sample face image from the UHDB11 database after applying face detection. Bottom: The effects that accurate (b) and inaccurate (a,c) eye localization has on geometric normalization. As expected, the accurate eye localization (green) produces the best face image for for recognition . . . . .	142
5.18	Verification results on the Active Illumination database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	145
5.19	Verification results on the Long Distance database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	146
5.20	Verification results on the Facial Obstruction database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	147
5.21	Verification results on the Extreme Pose and Illumination Variation database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	148
5.22	Identification results on the Active Illumination database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	149
5.23	Identification results on the Long Distance database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	150
5.24	Identification results on the Facial Obstruction database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	151
5.25	Identification results on the Extreme Pose and Illumination Variation database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against. . . . .	152
5.26	Sample images where the proposed approach fails to accurately detect the eyes in the SWIR, NIR, and Visible spectrum. Issues such as poor contrast, eye closure, and extreme face rotation lead to falsely detected pupil locations. Ground Truth = red; Found Locations = blue. . . . .	154
6.1	Overview of the LBP operator . . . . .	162
6.2	Overview of the LTP operator. Notice how the LTP operator splits the code into two separate LBP feature vectors. . . . .	162
6.3	Sample images from the visible gallery (bottom row) and SWIR probe (top row) after applying six different photometric normalization: (a) CLAHE (b) LBSSR (c) C-LBSSR (d) TBSSR (e) C-TBSSR (f) TT . . . . .	164

6.4	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. . . . .	166
6.5	Sample ROC curves for two different photometric normalization combinations (a) Combination 44: 95.85% AUC and (b) Combination 2: 71.67% AUC. . .	167
6.6	An overview of the score level fusion scheme used to improve cross-spectral face recognition. . . . .	168
7.1	Flow chart representation illustrating the resemblance to matching across the electromagnetic (EM) spectrum where cross-restoration comparison is analogous to cross-spectral comparison throughout the EM domain. . . . .	180
7.2	A collection of passports ranging from 1884 to 2009. All are real passports and are available in the public domain [92]. . . . .	181
7.3	Sample images used from the WVU Identity Document Database: a) ground truth gallery images, b) original document images, c) manually annotated passport masks in red, and d) the resulting masks marking the regions of interest (ROI). . . . .	190
7.4	An illustration of the evaluation of the improvement in image quality where the UIQ metric is used to evaluate the restoration performance using a) TV wavelet and b) EB inpainting where each inpainting method was also coupled with threshold-based (TB) denoising. Note that the inpainting/denoising combination is the most effective regardless of the inpainting method that was employed. . . . .	192
7.5	Identification results that solely use TV wavelet inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	195
7.6	Identification results that solely use TV wavelet inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	196
7.7	Identification results that solely use Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	197



7.8	Identification results that solely use Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	198
7.9	Identification results that use denoising and Total Variation based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	201
7.10	Identification results that use denoising and Total Variation based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	202
7.11	Identification results that use denoising and Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	203
7.12	Identification results that use denoising and Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of $\chi^2$ and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases. . . . .	204
7.13	CMC curves illustrating commercial performance comparisons between TV wavelet and EB restoration strategies. . . . .	206
8.1	Facial landmarks detected on image with extreme pose and illumination variation. The open sourced algorithm DLIB was used. Notice the inaccuracy in landmark detection in this challenging image. Further testing determined that this algorithm could not detect any facial landmarks on SWIR face imagery.	217





# List of Tables

2.1	Different methods for eye localization in the literature . . . . .	21
3.1	Eye safe power specifications at different wavelengths for multiple durations [76]. The blue shading depicts the SWIR wavelengths used in this collection for an illumination time of greater than 10 seconds. Empirical testing determined the illumination source used for this collection was well below the eye safety thresholds. . . . .	49
4.1	Face detection rates and normalized errors ( $e = 0.15$ ) for the proposed eye detection method using the face detection approach as well as using the manually annotated face regions. . . . .	70
4.2	Normalized errors (maximum error for both left and right eye) across six different photometric normalizations to determine the highest overall accuracy per wavelength. 5% normalized error is approximately within the diameter of the pupil, 10% diameter of the iris, and 25% the diameter of the eye. The bold results represent the highest 5% normalized error and were used for further experimentation. . . . .	74
5.1	Confusion Matrix displaying the performance of the SVM classifier. Note that only 12 out of 1,902 sample images was mis-classified. . . . .	134
5.2	Numerical results for each algorithms performance on the four tested databases. Notice that the proposed algorithm outperforms all other algorithms in all experiments except for the rank-1 score in the Active Illumination database. . .	153
6.1	Experimental results for the proposed eye detection and heterogeneous face recognition algorithm. Face recognition results show the rank-1 percentage .	170
6.2	Results of the time efficiency test. All times are in seconds for a single probe to match a gallery face image. . . . .	172

7.1	Top five ranks and which width mask generated the scores for both inpainting algorithms. Results are generated from the LTP-DTS texture based approach as well as the commercial algorithm G8. . . . .	207
-----	--	-----





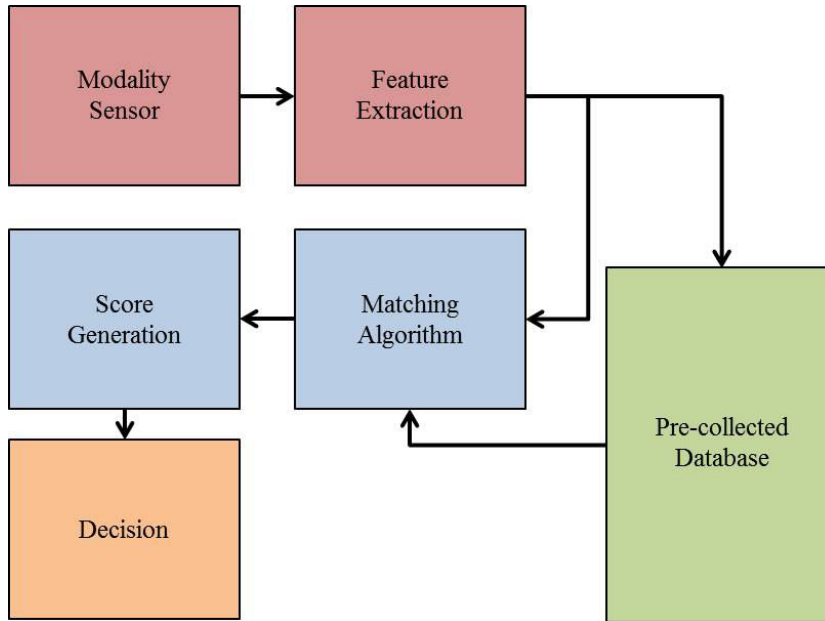
# Chapter 1

## Introduction

### 1.1 Introduction to Biometrics

Over the past century, humans have always had the need to identify themselves for a multitude of reasons. Historically, tokens, pseudonyms, passports, legal names, etc. have been used when asked to present something that is unique to that individual [1] [2]. With an ever more connected world, the need to be able to securely and accurately identify someone is of paramount importance. Whether used for encrypting personal files, protecting banking records, or securing a nation's borders, biometrics can be used to address some of the security concerns that plague these industries. Biometrics, or the ability to uniquely identify an individual based on their physical or behavioral traits, have emerged in the last two decades as a sound alternative to the classical means of identity. Physical and behavioral traits, such as face, fingerprint, iris, voice and gait [3] [4] [5] [6], allow for an alternative means of identification compared to the knowledge (passwords) and token (smart-cards) based approaches. Because biometrics are convenient (i.e. users do not have to remember carrying identifying





**Figure 1.1:** An overview of a typical biometric pattern recognition system broken into its main parts, i.e. sensors, feature extraction, database, matching algorithm, score, and decision.

items on their person or have to remember ever changing passwords) and secure (i.e. imposters are easily detected, making it difficult to falsely impersonate someone’s identity), a great deal of research has gone into the design and development of such algorithms to create viable security products. Because of this, biometric systems have been employed not only in both commercial and private industries, but also in the forensic and military sectors.

A typical biometric-based recognition system strictly relies on comparing two digital representations, or feature vectors, of a persons physical or behavioral trait. Usually, a newly collected feature vector is compared with one that has been previously collected and is stored on file. A typical biometric pattern recognition system can be broken down into six main components, and can be seen in Figure 1.1.

- **Sensor:** A sensor is used to collect a digital representation of the biometric trait being

captured. For example, a digital camera must capture a face image of an individual or an optical scanner can be used to obtain a palm or finger print.

- **Feature Extraction:** A set of algorithmic steps are used to break down the digital representation of the biometrics into a set of salient characteristics, creating a template of features, also known as a feature vector.
- **Database:** Any recognition system must have a database to be compared to that has been previously collected and whose feature templates have already been extracted.
- **Matching:** A newly collected feature vector is matched with some or all feature vectors in the database using a specified algorithm designed to work strictly for that biometric modality.
- **Score:** Two feature vectors are compared to each other, producing a single numerical match score. In some cases, such as similarity scores, a higher number indicates a better match, where as in distance scores, a lower number indicates a better match.
- **Decision:** Based on the matching score between two feature vectors and a predetermined threshold, a decision can be made to determine if the newly collected sample is a genuine match (i.e. the same identity) or an imposter match (i.e. different identities).

Depending on the application, there are two main types of biometric systems, i.e. an identification and a verification system. An identification system can be classified as a system in which a newly collected sample is compared and classified to all samples in a pre-collected database. Also known as a 1-N system, this type of system has the capability to inform the operator if a collected sample is on a watch list or anywhere in the system.

On the other hand, a verification system, also known as a 1-1 match, performs a comparison between the newly collected template and one matching the template the user identifies with. If the score is above a set threshold, the user is verified. If it is below, the user is denied access. Although similar, these two systems have their own advantages and disadvantages, which should be taken into consideration when implementing either.

In this work, the modality being studied is that of a human face. Facial recognition (FR) systems are becoming ever more popular due to the advantages of face over other biometric modalities (such as iris, voice, fingerprints, etc.): it is non-intrusive, understandable, and can be captured at standoff distances. The past decade has seen significant progress in the field of automated face recognition as is borne out by results of the 2010 Multiple-Biometric Evaluation (MBE) organized by NIST [7]. For example, at a false match rate (FMR) of 0.001%, the false-nonmatch rate (FNMR) of the best performing face recognition (FR) system has decreased from 79% in 1993 to 0.003% in 2010. Countless industries, from casinos in Las Vegas to military and governmental watchlists, use face recognition to keep their assets secure. With an estimated market worth of \$6.5 billion dollars by 2018\*, research into increasing its overall performance and exploring new options in the field of facial recognition is paramount.

## 1.2 Advantages of the IR Spectrum

Conventional FR systems depend on capturing face images in the visible range of the electromagnetic spectrum (EM), i.e. 380-750nm. However, in operational scenarios, such as military and law enforcement, harsh environmental conditions are common and dealt with

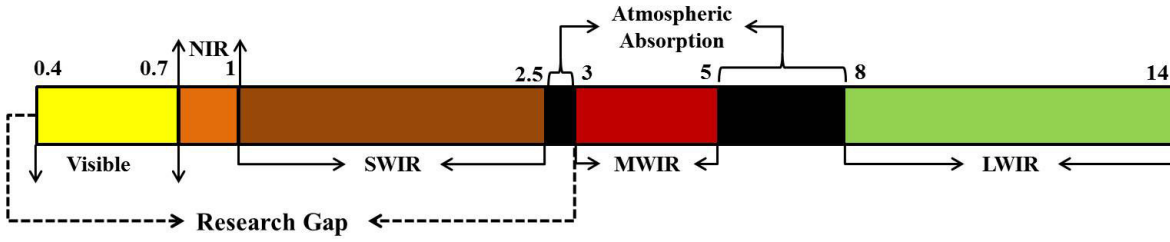
---

\*<http://www.marketsandmarkets.com/PressReleases/facial-recognition.asp>

in a sufficient and advantageous manner. These conditions can be characterized mainly by difficult lighting, which can cause unfavorable shadows and thus, limit the effectiveness of conventional FR systems in terms of their human identification accuracy. For example, in low light environments, human recognition, based solely on visible band face images, may show a significant decrease in FR performance and may not even be achievable [8]. Because of this, the problem of matching facial images, when considering diverse scenarios, remains to be a challenge.

In order to deal with such difficult environments, camera sensors operating in both the visible and infrared bands can be used. This is due in part by the fact that multispectral camera sensors have the advantage of imaging both day and night. By utilizing different sensors, it is possible to capture face images under varying representations (e.g. 2D, 3D, visible, IR), creating large datasets that vary not only in pose and expression, but also in varying illumination. The varying levels and nature of illumination is also among visible FR's most insidious problems. Recent FR trends are pointing to interest in higher spectra, i.e. Short Waved Infrared (SWIR) [9], Mid-Waved Infrared (MWIR) [10] [11] and Long-Waved Infrared (LWIR) [12]. With the prices of these camera sensors decreasing rapidly, the practical benefit of using different spectra for automatic FR is increasing. The potential to create more challenging (larger scale, different sensors) databases is becoming more of a reality for testing and developing face recognition algorithms with hopes of accurately matching visible gallery images to SWIR, MWIR, or LWIR probe images (namely the heterogeneous problem in FR, also known as cross-spectral matching) for the use of nighttime biometrics.

Nighttime biometrics, or the ability to perform biometric identification and/or verification during the night, is operationally relevant with limited research conducted. Operationally,



**Figure 1.2:** The electromagnetic spectrum depicting the active (NIR, SWIR) and passive (MWIR, LWIR) infrared bands [25].

having the ability to identify a subject during the night has its advantages and disadvantages. Advantageously, the act of identification is more covert than it would be during the day. The targeted subject would have a more difficult time determining that they are being studied, allowing operational entities greater security when conducting their work. On the other hand, the operational entities may have to deal with different challenges than they would operating during the day. Nighttime biometrics should not be confused with “low-light biometrics”. Although darkness, in and of itself, is a hindrance and a significant challenge of nighttime biometrics, it may be obfuscated by actively illuminating a subjects face in wavelengths above the visible spectrum (e.g. NIR, SWIR, etc.). However, darkness is not always present in nighttime biometric scenarios. Subjects may be standing under a street lamp or walking under full moonlight. These passive illumination sources can be a disadvantage to nighttime biometrics, causing saturation or reflections on the subjects face. Due to these advantages and disadvantages, this work describes face recognition advancements in the field of nighttime biometrics, specifically by using the IR spectrum.

The infrared spectrum (IR), as can be seen in Figure 1.2, is divided into different spectral bands, however the exact boundaries between these bands can vary depending on the scientific field and application. The IR bands, discussed in this work, are based on the response

of various detectors. Specifically, the IR spectrum is composed of the active IR (focus of this work) and thermal (passive) IR band. The active band ( $0.7\text{-}2.5\mu\text{m}$ ), which must be actively illuminated in order to be imaged, is further divided in the NIR ( $0.7 - 0.9\mu\text{m}$ ) and the SWIR ( $0.9 - 1.7\mu\text{m}$ ) bands. The NIR band is advantageous in the respect that face images acquired are close enough to the visible light spectrum. This allows for the capturing device to still obtain the correct structure of the face, however, it is far enough removed to not change the facial appearance due to visible light illumination changes [15]. The SWIR band has a longer wavelength range than the NIR and is more tolerant to low levels of obscurants like fog and smoke, allowing to image straight through. Because the wavelengths used are larger than the water vapor (fog) and carbon (smoke) particles, the optical energy passes through the fog and smoke, allowing it to not be imaged. However, in the visible spectrum, the wavelengths are significantly smaller and scatter/diffract off of the water vapor and carbon particles, obscuring the imaging scenario behind it.

The SWIR band also provides the capability to image through certain types of tinted glass [16], such as automotive tint and different forms of sunglasses, due to the low absorption of the tinting materials at these wavelengths. The passive IR band, which does not need to be illuminated in order to be imaged, is further divided into the Mid-wave (MWIR) and Long-wave (LWIR) infrared bands. MWIR ranges from  $3\text{-}5\mu\text{m}$ , while the LWIR ranges from  $7\text{-}14\mu\text{m}$ . In this work, the visible and SWIR bands are the main focus.

Multi-spectral and hyper-spectral (multiple wavelengths within a band) image understanding is an important capability of a biometrics system, not only in terms of recognition performance [17] but also in terms of operational efficiency. For example, it is impractical for a forensic tool operator to manually annotate thousands to millions of eye centers before

he/she can further apply any set of FR pre-processing, feature extraction, and matching algorithms. More specifically, this is one problem forensic operators and researchers have to deal with on a regular basis. Facial databases are enlarging to a point where millions of individuals are enrolled into biometric systems, where the data collected is for either research purposes or not, e.g. they can be collected under operational conditions for military and law enforcement applications.

### 1.3 Research Objective: Eye Localization

As facial recognition algorithms become ever more popular, the size of test databases must increase to accommodate testing structures. Most academic and commercial FR systems typically use normalized face images for their algorithms. A normalized face image is an image in which a face has been geometrically shifted and resized in order for most or all facial features to be located in the same pixel area. This allows for a fair comparison between two different sets of facial features. Typically, the face normalization process requires both the left and right eye locations. By using these locations, an algorithm can be written to rotate the face so both eyes are on the same plane, have the same inter-ocular distance apart (i.e. have the same number of pixels between both eyes), and have the same spatial resolution. Usually, when smaller databases are used for testing purposes, these eye locations are manually annotated. However, as discussed above, databases are reaching into the thousands and millions of face samples. Having an operator manually annotate all of these images is unfeasible. It is an extremely time-consuming task that is not practical in operational scenarios. Because a face recognition algorithm relies so heavily on the normalized face, i.e. accurate eye locations, a faster and more time-effective way must be used to determine the

locations of both the left and right eyes.

Reported work in the literature shows that automated and robust eye localization can have a positive impact on FR performance [18]. However, while available eye detection approaches perform well in the visible band, they perform poorly in the SWIR spectrum. This happens mainly due to the fact that these algorithms were not designed to work well in bands other than the visible. The eye locations found when using three well known eye detection methods, i.e. both commercial and academic (Viola and Jones and state of the art [19]) on the hyper-spectral SWIR face dataset (face images are captured in multiple wavelengths within the SWIR band) do not perform well. More importantly, when face images in a dataset are geometrically normalized based on the found eye locations, the resultant image cannot be used by face recognition algorithms due to the effect that rotations and scale have on face images. Therefore, an efficient eye localization algorithm needs to be designed, developed, and applied on both the gallery (conditioning of enrolled individuals) and probe wavelength specific SWIR datasets.

### 1.3.1 Contributions

In this work, a novel eye localization method that is fast and robust when operating in different bands of the SWIR spectrum (e.g. 1150, 1250, 1350, 1450, and 1550nm) will be presented. The following contributions to this can be described as followed:

- An extensive dataset is designed and collected that captures a large number of face images in the SWIR spectrum. Ground truth data, along with different challenging scenarios, including sunglasses and different tinted pieces of glass, are collected for the use of testing eye detection methods and cross-spectral face recognition algorithms.



- A novel eye detection method is proposed and described that incorporates face and eye localization using the normalized cross correlations and pupil detection using summation range filters.
- An evaluation on which pre-processing photometric normalization techniques obtain the highest eye detection accuracy in all bands of the SWIR spectrum.
- A comparison study that shows that the proposed method advances the state of the art in SWIR eye detection and outperforms all other tested algorithms, including both commercial (L1 Identity Systems G8 algorithm) as well as academic algorithms (Viola and Jones method as well as the current state-of-the-art [19]).
- A study on the robustness of the eye detection algorithm including testing the performance on JPEG-compression, spatial resolution of the face image, and Gaussian blur.
- An extensive study on the effect that each of the tested eye detection methods have on multiple facial recognition systems is conducted. Both feature and texture based FR approaches, including PCA, LDA, and LBP, are tested and results for identification and verification tests are shown using the Cumulative Match Characteristic (CMC) curve as well as the Receiver Operating Characteristic (ROC) curve.

It will be shown that the proposed method not only performs well in multiple bands of the SWIR spectrum but also advances the state-of-the-art. The proposed method is also robust to many different image degradation factors and produces the best face recognition performance compared to all other tested algorithms.

Although this method is fast and robust, it was only designed to work on ground truth straight forward images in the SWIR spectrum. In operational scenarios, unconstrained images are more likely to be collected. Because of this, a fast and accurate eye detection that works across the EM spectrum and on challenging unconstrained databases is documented. The contributions of this work is listed as follows:

- Use of multiple databases across the electromagnetic spectrum to simulate different types of unconstrained scenarios including SWIR face images using active illumination and at long distances, NIR images in which a subject could be wearing eye glasses, and visible images in which the participants are subject to extreme pose and illumination variations.
- Developed an image classifier that automatically determines which unconstrained scenario a query image belongs to.
- Augmented and enhanced the eye detection methods to be able to accurately and quickly determine the location of the eyes in the unconstrained images.
- Compared the proposed eye detection methods with both commercial and state-of-the-art eye detection algorithms.
- Determined if the proposed eye detection method produces better FR results than all the compared algorithms.

By altering the proposed approach to work in more unconstrained environments, it is shown that the proposed eye detection approach works across the entire electromagnetic spectrum as well as under many different constrained and unconstrained capture scenarios.

## 1.4 Research Objective: Cross-spectral Matching

As described earlier in the chapter, there are many reasons as to why FR algorithms are trending to the upper wavelengths of the electromagnetic spectrum. Due to its covert ability to capture face images without the subject knowing and its ability to penetrate low level obscurants like fog and smoke, using the SWIR wavelength is ideal for operational teams. 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 [20]. However, in uncontrolled scenarios (e.g. long range recognition and imaging behind glass), there is a need for efficient intelligence and surveillance reconnaissance (ISR) interoperability. Often, 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 will be addressed 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.

### 1.4.1 Contributions

In this work, a novel approach to the cross-spectral matching problem is proposed. Specifically, the  $1550nm$  wavelength is studied in depth. The following contributions are made:

- The cross-spectral algorithm using photometric normalization fusion, based off of the work done by Kalka et. al. [20], is described in detail. This method shows that the number of photometric normalizations used in [20] is unnecessary.
- An empirical study is done to determine which photometric normalizations should be used to achieve the highest FR scores possible
- A comparison of the approach with other FR algorithms, including baseline texture based approaches, the algorithm described in [20] as well as a commercial algorithm, is performed on multiple datasets, including challenging scenarios.
- A time test is done to show that the cross-spectral approach is faster and more time efficient than the method described in [20].

It will be shown that not only does the proposed algorithm outperform all other tested algorithms in most datasets, but that this approach is much faster than the algorithm proposed in [20]. By implementing the algorithm using parallel processing, the time efficiency is increased even more.

Although cross-spectral FR is typically designed for comparing two face images from different spectra of the electromagnetic spectrum, similar concepts can be used when trying to match images across different image representations. For example, matching visible images to passport images have many of the same challenges that cross-spectral matching has. Because of this, the contributions to document to live face recognition is as follows:

- Each passport image is manually annotated to determine the locations of watermarks and to build binary masks.
- A pre-processing technique is performed involving two different inpainting algorithms, i.e. total variation inpainting and exemplar based inpainting, along with wavelet based denoising.
- The Universal Image Quality (UIQ) measurement is used to determine if the pre-processing techniques are working as expected.
- FR algorithms, including a commercial software, is used to determine if the pre-processing steps increase the baseline visible to passport FR results.

By performing these studies, a greater understanding of the problems with cross-spectral or cross-document FR will be known. The algorithms proposed will show an advancement to the state-of-the-art in cross-spectral and cross-document matching, an area where little work has been done in the literature.

## 1.5 Proposal Organization

In this section, an outline of the rest of the dissertation proposal will be described. In chapter 2, an extensive summary of work that has previously been done will be discussed. Methods used for automatic eye detection in both the visible and NIR bands will be explained as well as different approaches to the problem of automatic eye detection. Also, cross-spectral matching algorithms will be described that effectively match both visible-NIR and visible-SWIR images.

In chapter 3, an extensive overview of the data collection protocol is outlined. The West Virginia University Multispectral Face Database (WVUMFD), a database of constrained and unconstrained images collected in the visible and SWIR spectrum, is described in detail. First a short summary of different glass characteristics is described followed by the collection protocol of the visible and SWIR images. Finally, different demographics, such as gender and race, are outlined.

Chapter 4 presents automatic eye detection in the SWIR band. First, the motivation behind automatic eye detection is discussed. Then, the methodology behind the novel approach is outlined in detail including face/eye localization, pupil detection, and geometric normalization. Later, an extensive evaluation of the proposed approach is described including comparing the approach with multiple available eye detection algorithms. Finally, results are discussed and summarized.

Chapter 5 extends the methodology discussed in Chapter 4 to handle images that are far more challenging. These face images span the EM spectrum and range in challenges from images with/without glasses and extreme pose and illumination variation. The methodology for the approach is outlined in detail including the augments for each challenge. Later, an evaluation into the accuracy of the proposed algorithms compared to other available eye detection algorithms is conducted as well as a test on how the eye detection results affect face recognition performance. Finally, the results are discussed and summarized.

In Chapter 6, the work on cross-spectral face recognition is outlined. First, a discussion on the importance of cross-spectral FR is offered followed by the methodological approach, based on the work of [20]. Then, the experiments conducted are outlined and show that the proposed approach is not only faster, but is more accurate in most datasets tested. Finally,

a summary of the work is offered.

Chapter 7 outlines the proposed approach to document to live face matching. First, the motivation for this work as well as a discussion on the analogy between visible and SWIR and visible and passport photos is made. Secondly, an extensive description of the proposed work is offered. Later, evaluations on the proposed method is offered followed by conclusions and future work

Finally, in Chapter 8, overall conclusions are drawn and the main contributions of the dissertation are outlined, as well as areas of future research opportunities.







# Chapter 2

## Face-based Biometrics Under Challenging Conditions

### 2.1 Eye Detection across the EM

In recent years, as the size of face databases increases, there has been a lot of work done in the field of automatic eye detection. As reported in the literature, there are three main methodologies:

- *Feature Based methods* [19] [21] [22]: such methods use a variety of eye characteristics to detect the eyes. To identify distinctive features of the eyes, characteristics such as the color distribution of the sclera, the flesh of the eye, and the edge and intensity of the iris are used. Although these methods are typically efficient, they are not as accurate as other eye detection methods (appearance and template based methods), especially for low contrast images.

- *Appearance Based methods* [18] [23] [24]: in such methods, eye detection is considered the problem of classifying each scanned sub window as one of two classes (i.e., eye and non-eye). Appearance-based methods avoid difficulties in modeling 3D (depth) structure of the human eye by focusing on possible eye appearances under various conditions. In these methods, a classifier that determines an eye vs. a non-eye location may be learned from a training set composed of eye and non eye sample images. Different classification schemes, such as the AdaBoost algorithm, can be used to train the eye detector.
- *Template Based methods* [25] [26] [27] [28]: such methods use a generic eye model, typically based on eye shapes, to determine the location of the eye. Then, the eye model is moved across the image of a detected face, computing a correlation score. The location with the highest score is then determined to be the location of an eye.

All of these methods can be used to accurately detect the locations of the eyes in a face region. In this paper, we follow a *template based matching* approach and show that our method achieves a higher performance rate than previous work.

### 2.1.1 Visible Eye Detection

Because there is little to no research done in the field of SWIR automatic eye detection, it is difficult to compare methodologies. However, there is a multitude of work done in automatic eye detection performed in the visible ( $350nm - 850nm$ ) and Near Infrared (NIR -  $850nm - 1050nm$ ) spectra. With regards to the state of the art, Valenti et al [19] use isophote properties in the eye region to accurately detect the center eye location in visible

Method	Pre-requirements	Uses Learning	Used Feature	Band
Valenti [19]	Face/Eye Region	X	Isophotes	Visible
Asteriadis [23]	Face Region	-	Edges	Visible
Wang [18]	Face Region	X	FDA Descriptor	Visible
Song [29]	None	-	Binary Edge Images	Visible
Shafi [21]	Face Region	-	Hybrid	Visible
Khairrosfaizal [30]	Face Region	-	Circular Hough Transform	Visible
Dowdall [22]	None	-	Integral Projections	NIR
Zhu [27]	None	X	Kalman Filtering	NIR
Li [46]	None	X	AdaBoost Classifier	NIR
Previous Work [25]	Face Region	-	Normalized Correlation	SWIR
Proposed	None	-	Summation Range Filter	SWIR

**Table 2.1:** Different methods for eye localization in the literature

images. However, this approach assumes the location of the face is known, and within that location, anthropometric measures are used to estimate the eye region. Then, this estimated eye region is used to test the approach. This is problematic in the sense that different poses cause different eye locations. Therefore, assuming the eye region from an assumed face region is impractical in terms of an automated system standpoint. In contrast, our proposed algorithm empirically determines the location of the face and eye regions by using normalized correlation coefficients. Essentially, the input into our approach is a raw image, with an upper body, in which a face could be anywhere in the image. Then, we determine from this image where the face region is. If the face region is found incorrectly, it reflects in the accuracy of our eye detection as no eye can be found. However, in the compared methods, the input into their approach is the raw image and the spatial location of the face within that image. This guarantees that the eyes will be in every face image, making the point of eye detection trivial. It is impractical to assume knowing the face regions in a face image when large scale databases are being used. Because of the assumptions made,

the authors report a high accuracy rate at a low normalized error ( $\sim 80\%$  accuracy across databases at a normalized error of  $5\%$ ).

Asteriadis et al. [23] uses the standard PCA method, originally applied in intensity images, for eye region detection. First, they obtained the Canny edge map of an eye image and, for each pixel, a vector pointing to the closest edge pixel is calculated. Magnitude and slope are assigned to each pixel creating a vector length map and an angle map. Then, PCA is applied on a set of training eye images to derive eigenvectors for each map. In this work, 58 face images found on the web were used for training. In order to detect the eyes on an image, the length and angle maps of the candidate regions are subsequently projected on the spaces spanned by the training eigenvectors. The similarity of projection weights with those of model eyes is used to determine if an eye is present or not. If the algorithm determines that an eye is present, an eye center localization method is implemented, leaving the final location of the found eye. The authors tested their approach on two databases, namely the XM2VTSDB [34] and BioID [35]. These databases had 1,521 and 600 face images respectively. For the XM2VTSDB, the authors reports an eye detection accuracy of  $96.7\%$  while for the BioID database, the reported a performance of  $98.6\%$ . Although this eye detection is accurate, it requires an extensive training stage to learn the locations of the eyes.

Wang et al. [18] proposed a novel appearance based method for accurately determining the location of the eyes. In this work, the authors proposed to statistically learn discriminative features (by using Fischer Discriminant Analysis) to characterize eye patterns. Probabilistic classifiers are then learned to separate eyes from non-eyes. Multiple classifiers are combined in AdaBoost to form an accurate and robust eye detector. In this work, 500 pairs of eye patches were collected from both face image databases as well as the internet to include

more variance from the real world. Second, non-eye images were collected from background images. Also, falsely detected eye images were fed back into the non-eye category for re-training of the algorithm. This algorithm proved to be very effective, correctly detecting the eyes in 99% of found face images with an interocular distance error 2.67%. Although this algorithm is very accurate, it requires an extensive amount of training with thousands of eye and non-eye images. Also, time to retrain the classifier must be taken into consideration. The authors offered no insight into how many times the classifiers were re-trained before obtaining the 99% accuracy that they reported.

Song et al [29] used a robust Binary Edge Image (BEI) and light spots to accurately determine the locations of the eyes. In order to obtain the optimal BEI, multiple BEI's are created from different scales of a wavelet transform. These BEI's are merged together at every scale by simple pixel addition. Then, a foreground thresholding method is used to eliminate background noise caused from the merging. Following this, a Binary Edge Map (BEM) is created to alleviate the effects of noise. The final BEI is then obtained from this BEM. By using horizontal and vertical connection lengths, most of the components in the BEI can be eliminated. This leaves only areas of the face that could possibly be the eye. Finally, vertical projections are used. According to the authors, reflected light spots often exist in face images that are captured with flash photography and are typically located near the center of the iris. Exploiting this fact, localization of the eye centers is determined to be the region where the most light reflectance is located in the eye region. The authors used two databases that have a total of 777 images and achieved a detection rate of 98.7%.

Shafi et al [21] combines the work of [31], [32], and [33] to create a hybrid eye detection algorithm. In this work, three different methods are fused together to accurately detect the

locations of the eyes. The three methods used are as followed:

- **Illumination Based:** Two separate eye maps are created from both the chrominance and luminance components, respectively. The eye map from the chroma is based on the fact that the high  $C_b$  and low  $C_r$  values are present around the eyes, where  $C_b$  and  $C_r$  are the normalized blue and red chroma components respectively. By  $\oplus$  (logical AND) the two chrominance and luminance images, the locations of the eyes can be determined. This method resulted in 90% accuracy.
- **Color-based:** This method is a simple thresholding of pixel values. The authors assume that locations of the eyes are the darkest regions of the face. Therefore, by thresholding the images (in this case keeping all pixels below 20) and removing unwanted regions by component verification, the locations of the eyes can be obtained. This method resulted in 83.75% accuracy.
- **Edge-Density-Based:** Based on the observation that the eye region has the most edge density in a face region, a Sobel edge detector is used, along with numerous morphological operators, to obtain the areas with the largest density. Then, shape- and geometry-based rules are applied to the connected regions to extract the eyes. This method results in 80% accuracy.

By combining these three algorithms, the authors achieve an increase in performance between 3.75%-13.75% to obtain a final accuracy of 93.75%. One limitation to this paper, however, is that the algorithm was only applied to 60 images. More images would need to be considered in order to validate the performance of their method.

Khairafazail et al. [30] use a circular Hough transform to locate the pupil areas. First, a preprocessing step is done that first determines if the proper parameters are met, i.e. 2-D images, spatial resolution minimum (32x32 pixels), and gray scale image. Then, a face detection step is applied to spatially locate the area of interest. Following this, a median filter is used to rid the image of noise and enhance the brightness and color of the image. Afterwards, an accumulator array is built using the gradient magnitude and the linear indices. By using a local maxima filter, a mask is created to determine the eye pairs area of interest. Then, a general circular Hough transform is applied to the region of interest only. The algorithm was applied to 50 face images from the Illinois at Urbana-Champaign Face Database [36]. Using this method, the authors report a 86% rate of accurately determining the location of the subjects eyes. The authors attribute the low accuracy rate to the median filter not being able to successfully eliminate all illumination variations.

Other eye detection and landmark localization methods in the visible band include the following:

- Lu *et al.* [37] uses rectangular as well as pixel-pattern-based texture features.
- Feng *et al.* [38] uses multiple cues, such as intensity, the estimated direction of the line joining the centers of the eyes, and the proposed eye variance filter, to accurately determine the eye locations 92.5% of the time.
- Efraty *et al.* [40] uses an agglomerate of fern regressors along with component-landmark detectors, multi-scale analysis and learning of point cloud dynamics to find facial landmarks including the eye corners.
- Zhao *et al.* [41] uses a 3-D statistical facial feature model, which learns both the global



variations in configurational relationships between landmarks and the local variations of texture and geometry around each landmark, to find facial landmarks in 3-d facial data that has large facial expression variation and occlusions.

- Perakis *et al.* [42] uses a continuous map of principal curvature values of a 3-D objects surface and local descriptors of the objects 3-D point distribution to determine facial landmarks on 3-D images under large yaw and expression variations.
- Yu *et al.* [43] [44] [45] uses multiple approaches, including optimized part mixtures and cascaded deformable shape models, consensus of regression, and sparse shape representations, to accurately determine facial landmarks on images that have pose and expression variations.

### 2.1.2 NIR Eye Detection

As can be seen in the previous section, there has been a multitude of work done in the field of eye detection in the visible band. However, not much work has been done in the IR bands, especially in the SWIR spectrum. Few authors have reported work completed in the NIR band. For example, to assist in their accurate face detection method, Dowdall et al. [22] first detects the human skin and applies an integral projection to find the eyebrow regions. By using the upper and lower NIR bands, the eyes can be denoted as peaks and valleys. An inversion of the upper NIR band turns the peaks into valleys, which allows for a watershed algorithm to extract the regions. Finally, a template matching scheme is used to finalize a subjects eye locations. A database of 845 images was used with a reported 92.01% accuracy.

Zhu et al. [27] built an IR illuminator system and used kalman filtering to help assist

in the detection and tracking of the eyes in NIR spectrum. First, an illumination system consisting of two distinct rings was used when capturing the dataset. By using these two illumination rings and a video decoder, two de-interlaced face images are obtained, i.e. an even field face image and an odd field face image. By subtracting these two images, regions of possible eye locations are obtained. Then a trained SVM classifier using 558 eye images and 560 non eye images was used to determine if these potential areas were in fact eyes. Using this SVM classifier, the authors report a 95.5% accuracy rating.

Li et al [46] extensively train an NIR eye detector using the AdaBoost classifier. 86,000 NIR eye and 3,000,000 non eye examples are used where the non-eye examples are collected mainly from the upper-part of NIR face images. Each sample is a 21x15px region and is grouped into four distinct categories, i.e. left and right eyes and with and without glasses. By using a coarse-to-fine classifier tree, the authors report that this method can detect all possible eye candidate locations while rejecting more than 95% of non-eye examples.

As can be seen, there has been an extensive amount of work done in the field of automatic eye detection in the Visible band with some work having been done in the NIR band. However, there has been no work reported in the literature of automatic eye detection in the SWIR band of the electromagnetic spectrum. In this work, the author proposes a novel approach to quickly and accurately find the eyes in face images in the SWIR spectrum.

## 2.2 Face Recognition across the EM

### 2.2.1 Cross-spectral FR

As discussed in the introduction, there is a lot of interest in moving away from the visible spectrum and into the upper bands of the electromagnetic spectrum. Because face images in these wavelengths can be captured in a more covert fashion and can image through different levels of obscurants, there is a need for the design and development of FR algorithms that match visible gallery images to different spectra probe images. Although there has been a lot of work done in the visible spectrum for face recognition [47] [48] [49], there has been limited studies in the field of cross-spectral face matching. However, recently, researchers have been giving the problem more consideration. For example, Kalka et al. [20] [50] uses a photometric normalization fusion algorithm to minimize the differences between the two spectra. First, each visible gallery and SWIR probe image are altered using five different photometric normalization. Then, classical texture based FR algorithms are used to cross match each image representation in the gallery to each image representation in the probe, creating 36 unique scores per gallery to probe match. Then, these 36 scores are fused together to get the final match score. Using multiple SWIR-Visible databases, the authors report that cross-spectral matching in this form can achieve fairly high accuracy ratings when using a commercial matcher.

Nicolo et al [13] uses a combination of different feature vectors to cross-spectrally match SWIR probe to visible gallery images at short and long distances. After a pre-processing step to normalize the images pixels, the magnitude and phase response of the images is obtained using Gabor filters. For the magnitude response, two feature descriptors are used,

namely the Simplified Weber Local Descriptor and the Local Binary Pattern. For the phase response of the image, the General Local Binary Pattern descriptor is used. These features are then concatenated together to form one large feature vector. Then, when comparing two feature vectors from each spectra, the symmetric I-divergence distance is used. 48 subjects from the TINDERS [14] database is used for testing. At short distances, the authors report a 97.79% rank-1 score while at longer distances, rank-1 rates drop below 90%.

In the work conducted by Klare et al [15], the authors use two different image descriptors, namely Local Binary Patterns (LBP) an Histogram of Gradients (HOG) to create the feature vector for both the visible gallery and NIR probe. Then, using Linear Discriminant Analysis (LDA), random subspace classifiers are used to combine the mean vectors of the NIR and visible images. By altering the within scatter matrices for both spectra, the gap between the two spectra is minimized. When matching the visible and NIR images using nearest neighbor matching, sparse representation matching, and a commercial product on the CASIA database [52], the authors achieve a 93.45% true positive rate at a 0.1% false positive rate. These results were obtained when using 102 training subjects and 100 testing samples.

In the work conducted by Zhu et al [51], Transductive Heterogeneous Face Matching (THFM) is used to adapt the VIS-NIR matching. First, a Log-DoG (Difference of  $g$ =Gaussian) normalization is applied to both the visible and NIR images. Then, LBP and HOG features are extracted as in [15]. Finally, VIS-NIR face matching using transduction is applied and tested against other FR methods. The authors show that using their feature extraction method and matching algorithm, at a FAR of 0.1%, they achieve a classification accuracy of 98.42% when training on 150 subjects and testing on 50 subjects of the CASIA [52] database.

Finally, in the work performed by Akhloufi et al. [53], global and local non-linear techniques are used to bridge the gap between visible images and multiple spectra from the electromagnetic spectrum. Multiple different dimensionality reduction algorithms are tested on 2 different multispectral face databases. Advanced dimensionality reductions techniques, including local linear embedding and locality preserving projection, are tested against basic dimensionality reduction techniques, such as PCA, LDA, kernel PCA, and kernel LDA. Results show that different spectra react differently to different dimensionality reduction techniques in both datasets tested. However, good performance rates were obtained when performing multispectral face recognition.

These methods were all designed to work in the active band of the IR spectrum and had the goal of bridging the gap between the IR band studied and the visible band. There have been a number of publications that delve into the problem of matching visible images to the passive band of the IR spectrum [17] [54] [55] [56] [57], however, these studies are out of the scope of this proposal.

### 2.2.2 Document to Live FR

Document facial identification was first, explored by Staroviiodov *et al.* [58,59] who presented an automated system for matching face images, present in documents, against camera images. In that study, the authors constrained their work to earlier versions of passports (circa 1990) that were issued from a single country. The facial images used were reasonably clear and not “contaminated” by any security markings. Hence, the system’s ability to automatically identify the face photograph was not severely compromised.

Ramanathan and Chellappa [60] focused their attention in this area by addressing the

issue of age disparity prior to identification. Hence, they introduced a Bayesian classifier that approximates age estimation by estimating the differences between pairs of facial images.

Bourlai *et al.* [61] addressed the problem of facial matching over a variety of international passports. In their work, the authors introduced the following process:

1. *Face Detection* – Used to localize the spatial content of the face and determine its boundary. To accomplish this step, the algorithm proposed by Viola and Jones [24] was employed.
2. *Channel Selection* – Used to perform the appropriate color space normalization. Typically, document images are color images composed of the red, green, blue (RGB) channels and color space normalization is applied to prepare for further processing.
3. *Normalization* – Photometric and geometric normalization schemes, respectively, are needed, first, to compensate for illumination variation and next, for slight perturbations in the frontal pose. Geometric normalization is composed of two major steps: eye detection and affine transformation, where the eye detection is needed to create a global perspective in reference to all faces of the subjects within the database at hand. Photometric normalization is performed by employing histogram equalization and contrast adjustment. Histogram equalization is a nonlinear image enhancement method that transforms image brightness, which can (under certain conditions) improve verification performance [63].
4. *Wavelet Denoising* – Wavelet-based image denoising is needed to remove the additive noise present in documented facial images [64]. This supplemental noise is caused by variations in security markings as well as paper defects. For this step, the Translation-

Invariant Wavelet Transform (TI-WT) [65] was used to average out the translation dependence of the wavelet basis functions.

5. *Feature Extraction and Classification* – In this step, the appropriate facial features are extracted and, then, face matching is performed, i.e. matching of the preprocessed document facial photos against their live photo counterparts.

Empirical evaluations shown in [61], confirmed that document facial matching is a difficult problem due to challenges associated with person, document, and device-related factors. Consequently, applying the preprocessing methodology, described in Steps 1 - 4 above, improves overall recognition performance.

Bourlai et al. extended their original work discussed in [61] by incorporating an image restoration methodology that improved the quality of severely degraded facial images that are digitally acquired from printed or faxed documents [62]. The acute degradation types considered were: (a) fax image compression\*, (b) fax compression, then print, and finally scan, and (c) fax compression, then fax transmission, and finally scan. The authors' approach involved an iterative image restoration scheme to improve the textural content of the face images while removing noisy artifacts. Their observational results determined that the proposed image restoration scheme improved image quality as well as recognition performance. The works of [61, 62] also helped inspire investigations in automated image quality measurements to meet the needs of ISO/ICAO standards, where an evaluation benchmark was introduced [66].

The work of [61, 62] also motivated the need to design and develop more sophisticated

---

\*In that work *Fax image compression* is defined as the process where data (e.g., face images on a document) are transferred via a fax machine using the T.6 data compression, which is performed by a fax software on a controlling computer.

approaches for facial image restoration. Such approaches are not only limited to image denoising but where digital image inpainting can also be used as an additional tool that better deals with local structures, such as watermarks. Digital image inpainting was first introduced by Bertalmio *et al* [67], where a nonlinear third order partial differential equation (PDE) was used to fill in the selected region of interest. The results of this work spearheaded an interest in geometric interpolation and inpainting problems that include variational Partial Differential Equations (PDE) methods [68, 69], fluid dynamics inpainting [70], landmark based inpainting [71], inpainting by vector fields [72], and inpainting by corresponding maps [73].

Up to that point, inpainting and interpolation were mainly focused in the pixel domain; however, it was the work of Chan *et al* [74], who extended this practice to the wavelet domain. With this knowledge of wavelet interpolation, Bourlail *et al.* [75] extended the works of [61, 62] to perform total variation minimization in the wavelet domain and targeted wavelet coefficients associated with various security markings. By coupling this minimization with the general denoising algorithm [61, 62], the proposed restoration scheme showed much promise in the improvement of overall image quality as well as rank-1 identification accuracy.









# Chapter 3

## West Virginia University

## Multispectral Face Database

### 3.1 Multi-Spectral Face Collection for Cross-Spectral Identification Applications

Operational scenarios, such as that within law enforcement surveillance applications and government military defense situations, typically encounter scenarios in which face images are taken under non-ideal situations. Some of these non-ideal scenarios include pose, illumination, expression (PIE) variations, un-cooperative subjects, natural obscurants, such as fog and smoke, or physical obscurants such as windows, eye glasses, etc. One of the key aspects to cross-spectral FR is to be able to consistently and accurately match non-ideal face images captured in the SWIR band of the electromagnetic spectrum to ideal mugshot like images captured in the visible spectrum. However, to develop and test these types of

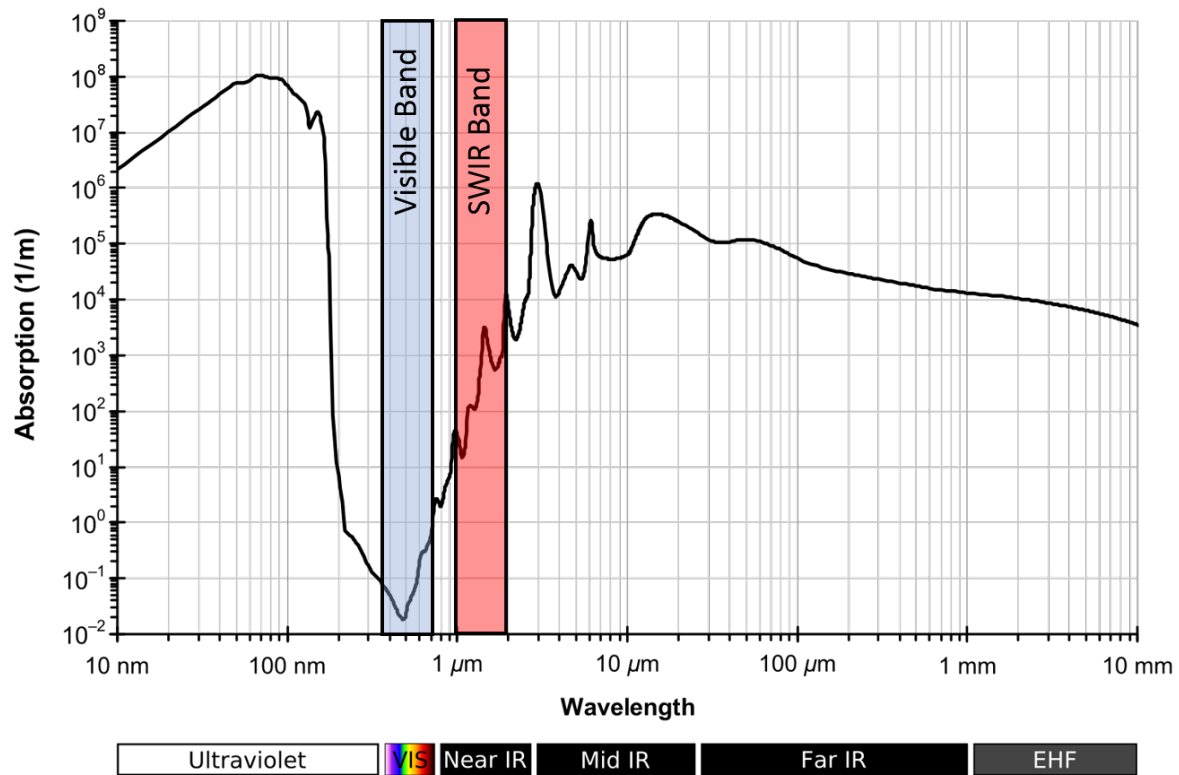
cross-spectral matching algorithms, a large amount of representative data, in both the visible and SWIR spectra, must be used. Because data such as this is not available to the general population, especially in the SWIR spectrum, an extensive and thorough collection protocol must be devised in order to collect data that is representative of some of the issues that are described. In order to have a fuller understanding of the challenges that arise when performing cross spectral face matching, as many of the non-ideal scenarios must be taken into consideration when collecting such data. In the collection protocol described below, many non-ideal scenarios are captured in the SWIR spectrum, such as varying illumination levels, different illumination sources, and multiple physical obscuring factors.

### 3.1.1 Challenges

Though collecting non-ideal SWIR data for developing and testing cross-spectral FR has its many advantages, there are still challenges that must be overcome when performing a new SWIR data collection. As can be seen in Figure 3.1\*, the level of water absorption in the SWIR spectrum is significantly higher than that of the visible band (at certain points a 5-6 order of magnitude increase). Because of this, when a human face is imaged at wavelengths above  $\sim 1400nm$ , the moisture found in the skin begins to absorb the infrared wavelengths. This gives the skin a dark or black pigment, even for light or fair skinned subjects. In situations where the ambient illumination is minimal, and an illumination source is necessary, the issue becomes further complicated in that this source must be eye-safe for the captured subject. Because the most eye-safe SWIR wavelengths are above  $1400nm$ , when illumination is needed to capture the subjects face, the black skin effect is unavoidable. Another challenge

---

\*[https://en.wikipedia.org/wiki/Electromagnetic\\_absorption\\_by\\_water](https://en.wikipedia.org/wiki/Electromagnetic_absorption_by_water)



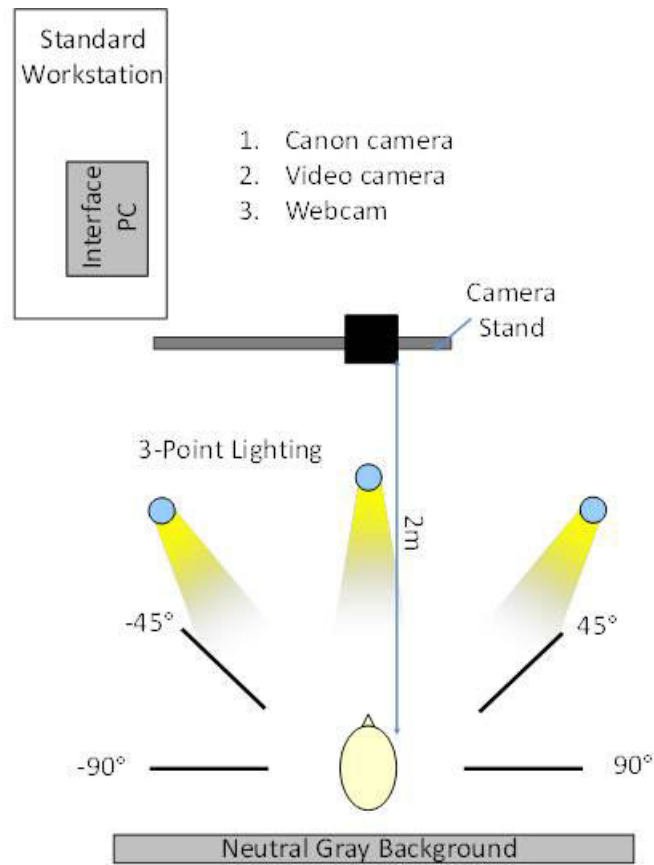
**Figure 3.1:** Liquid water absorption levels across the electromagnetic spectrum. Note that the absorption levels in the SWIR spectrum are significantly higher than that of the visible spectrum. Because of this, The moisture in the human skin absorbs these wavelengths and causes the face to look black.

associated with SWIR data collections is that membrane tissues, such as the sclera and iris of the eye, become darker as the SWIR wavelength increases. Because the pupil becomes obscure in these upper wavelengths, it poses problems when performing eye detection for automatic face recognition. Finally, certain oils produced by the skin reflect infrared light, causing a saturation effect in images with high intensity illumination sources and direct sunlight. Although these effects pose issues, many of them can be mitigated through filtering the images to specific wavelengths.

The benefits to multi-spectral face imagery outweigh the current associated issues found within. Because of this, West Virginia University has conducted several data collections involving capturing face images under different SWIR wavelengths. Through different biometric face collections, a West Virginia University Multispectral Face Database (WVUMFD) has been built. These images highlight specific bands across the SWIR spectrum and illustrate some of the non-ideal scenarios discussed above. Along with collecting the ideal visible spectrum mugshot face images, cross-spectral eye detection and face recognition algorithms can be developed and tested. This chapter will outline the methods and procedures used to collect and build the WVUMFD.

## 3.2 Standard Visible Mugshot Capture

To collect the visible ground truth gallery photos for the WVUMFD, a standard DSLR camera, more specifically the Canon 5D Mark II was used. This camera was equipped with a telephoto lens (e.g. Canon EF 800 mm f/5.6L IS USM) and captured high resolution face images. A strict capture protocol was built to capture each subject at 5 different poses, i.e.  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ , and  $90^\circ$ . In accordance with ANSI/NIST-ITL 1-2007



**Figure 3.2:** Photo station layout for the visible mugshot collection aspect of the WVUMFD



Best Practice Recommendation for the Capture of Mugshots<sup>†</sup>, 3-point lighting, comprised of one 250W fixture and two 500W fixtures, was used. Plastic diffusers in front of the reflector-mounted light bulbs are used to prevent “hot-spots” on the subjects face. Using a neutral gray backdrop, the light fixtures were positioned, with respect to the participant, in a slightly asymmetric position to avoid shadows in the background. The schematic of the capture protocol for the visible mugshots can be seen in Figure 3.2. The camera settings that typically produces the best face capture in terms of focal depth and image quality are as follows: White Balance: Tungsten, ISO: 1000, F/2.6: 1/10, Exposure: 1/60.

## 3.3 SWIR Face Capture

### 3.3.1 Tinted Material Characterization

Before the collection of the WVUMFD, two studies were performed to understand how different environmental impacts, such as lighting and temperature, affect the ability to image through the glass. First, a study was conducted to understand how changes in lighting (i.e. internal and external lighting), when coupled with a specific type of glass tint, affected the ability to image faces through glass. Secondly, a study on how temperature alters spectral transmission of different materials was conducted. The materials that were used in this study are as follows:

- Different common architectural and automotive glass with tint embedded in the material<sup>‡</sup>.

---

<sup>†</sup><http://www.nist.gov/itl/ansi/upload/Approved-Std-20070427-2.pdf>

<sup>‡</sup>All glass provided by Pittsburgh Plate Glass (PPG)

- 2 architectural samples with mirror coating applied.
- Clear plate glass covered with tinted plastic film<sup>§</sup>.
- Various types of plastic lenses in eye wear (i.e. sunglasses) from multiple manufacturers.

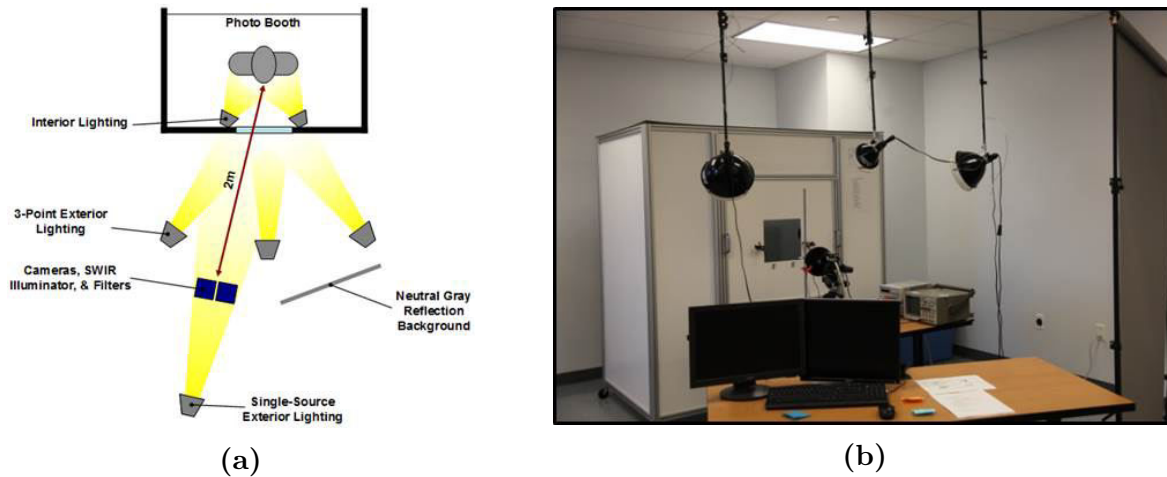
To summarize the findings of these studies, which are published in more detail in [16], temperature did not significantly impact the emission properties of the tinted materials. The largest effect on capturing face images behind the different tinted material was the lighting scenarios, i.e. different interior and exterior lighting. In order to down-select the number of glass materials used in the collection process, contrast quality metrics were applied to sample images under each lighting and tint condition. Then, each material was ranked according to transparency and image quality. From this, three samples were chosen to represent low, medium, and dark tint.

### 3.3.2 SWIR Face Image Collection

To provide operationally relevant data samples for development of eye/face detection techniques as well as to develop cross-spectral FR algorithms, the following data collection was performed. To collect the face images under the SWIR band, an Indium Gallium Arsenide (InGaAs) based Goodrich SU640KTSX-1.7RT SWIR camera was used for image acquisition, based on previous work done by [77] [78]. The solid-state InGaAs array, operating at a pixel depth of 640x512 and an operability of  $> 99\%$ , has a high sensitivity in 900-1700nm spectrum. For further specifications on the Goodrich SU640KTSX-1.7RT SWIR camera can be found here.

---

<sup>§</sup>Johnson Window Film



**Figure 3.3:** Overview of the WVUMFD SWIR capture setup. (a) birds eye view of the setup (notice the angle at which the camera was setup to reduce reflection) and (b) picture of actual capturing setup.

A photo booth and lighting setup, which can be seen in Figure 3.3 was built specifically for this capture protocol and possessed the ability to independently control both the interior and exterior lighting. The booth was designed with a removable front panel in which different tinted pieces of glass with varying transparency could be slid down into. For more details into these materials, please see [16]. Because test images were being captured with interference and obstructing reflections on the glass, the camera setup was slightly angled ( $\sim 12^\circ$ ) to reduce the impact of reflections on images under high exterior lighting conditions at a short distance (2m). Also, a neutral gray background was placed in the reflection path to provide a constant background. This allowed for further reduction of image reflectance.

An image capture protocol was developed and designed to collect face images for different materials under variable conditions at wavelengths ranging from  $1150nm$  to  $1550nm$  in increments of  $100nm$  using  $50nm$  band gap filters. For the scenario in which the booth panel was on (i.e. glass panel with tint slid into position), the facial images were collected

under the following scenarios, with all lux measurements taken at the interior booth seating location:

















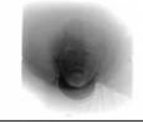







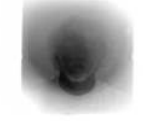

– **Glass Panels**

1. Clear w/ 0% Film Tint Glass Panel
2. Clear w/ 80% Film Tint Glass Panel
3. Solarcool (2) Graylite Glass Panel

– **Lighting Conditions**

1. Full interior ( $\sim 2600$  lux), 0 lux exterior
2. Minimum interior ( $\sim 60$  lux), 0 lux exterior
3. 0 lux interior, 3-point exterior ( $\sim 350$  lux)
4. Single external source ( $\sim 5$  lux)
5. 500mW 1550nm active illumination

Each glass panel was collected under all lighting conditions, creating a total of fifteen different tint/lighting scenarios. During initial testing, the 500mW 1550nm active illumination source was causing a singular illumination point. Therefore, to spread out the area of illumination, a Gaussian to square uniform point diffuser was used. When illuminating subjects with a SWIR illumination source, it is important to take into consideration the eye safety of that subject. As can be seen in Table 3.1 [76], for SWIR illumination between 1050nm and 1400nm, illuminating a subject between 10 and 1,000 seconds, the power must be below  $390\mu W$  and between 1400nm and 1550nm, the power must be below 10mW. After

<i>Variable Lighting – Clear glass with 80% Tint</i>					
	Full Internal, No External	Minimal Interior, No Exterior	No Internal, Full External	Single-Source External	Active Illumination
Visible					
OPR	6	8	7	9	9
ENH	OFF	OFF	ON	ON	OFF
1150 nm					
1250 nm					
1350 nm					
1450 nm					
1550 nm					

**Figure 3.4:** Sample images from the WVUMFD captured using different wavelengths (1150-1550nm) and lighting conditions when the subject was behind clear glass with 80% film tint.

applying the Gaussian to square diffuser, and empirically testing the illumination required to image a subjects face, the power of the illumination source was tested and well below the limits described in Table 3.1. Sample images taken under this capture protocol for the clear glass with 80% film tint can be seen in Figure 3.4.

Accessible emission limits for Class 1 and Class 1M laser products and $C_6 = 1$ <sup>a, b</sup>											
Wave-length $\lambda$ nm	Emission duration $t$ s										
	10 <sup>-13</sup> to 10 <sup>-11</sup>	10 <sup>-11</sup> to 10 <sup>-9</sup>	10 <sup>-9</sup> to 10 <sup>-7</sup>	10 <sup>-7</sup> to 1,8 × 10 <sup>-5</sup>	1,8 × 10 <sup>-5</sup> to 5 × 10 <sup>-5</sup>	5 × 10 <sup>-5</sup> to 1 × 10 <sup>-3</sup>	1 × 10 <sup>-3</sup> to 0,35	0,35 to 10	10 to 10 <sup>2</sup>	10 <sup>2</sup> to 10 <sup>3</sup>	10 <sup>3</sup> to 3 × 10 <sup>4</sup>
180 to 302,5	3 × 10 <sup>10</sup> W.m <sup>-2</sup>		30 J.m <sup>-2</sup>								
302,5 to 315	2,4 × 10 <sup>4</sup> W		Thermal hazard ( $t \leq T_1$ ) 7,9 × 10 <sup>-7</sup> C <sub>1</sub> J		Photochemical hazard 7,9 × 10 <sup>-7</sup> C <sub>2</sub> J ( $t > T_1$ )				7,9 × 10 <sup>-7</sup> C <sub>2</sub> J		
315 to 400											
400 to 450			7,9 × 10 <sup>-7</sup> C <sub>1</sub> J						7,9 × 10 <sup>-3</sup> J		7,9 × 10 <sup>-6</sup> W
450 to 500	5,8 × 10 <sup>-9</sup> J	1,0 $t^{0,75}$ J	2 × 10 <sup>-7</sup> J	7 × 10 <sup>-4</sup> $t^{0,75}$ J				3,9 × 10 <sup>-3</sup> J		3,9 × 10 <sup>-5</sup> C W	
500 to 700								3,9 × 10 <sup>-3</sup> C <sub>3</sub> J and <sup>c</sup> 3,9 × 10 <sup>-4</sup> W			
700 to 1 050	5,8 × 10 <sup>-9</sup> C <sub>4</sub> J	1,0 $t^{0,75}$ C <sub>4</sub> J	2 × 10 <sup>-7</sup> C <sub>4</sub> J	7 × 10 <sup>-4</sup> $t^{0,75}$ C <sub>4</sub> J				3,9 × 10 <sup>-4</sup> C <sub>4</sub> C <sub>7</sub> W			
1 050 to 1 400	5,8 × 10 <sup>-8</sup> C <sub>7</sub> J	10,4 $t^{0,75}$ C <sub>7</sub> J	2 × 10 <sup>-6</sup> C <sub>7</sub> J		3,5 × 10 <sup>-3</sup> $t^{0,75}$ C <sub>7</sub> J						
1 400 to 1 500	8 × 10 <sup>5</sup> W		8 × 10 <sup>-4</sup> J				4,4 × 10 <sup>-3</sup> $t^{0,25}$ J	10 <sup>-2</sup> $t$ J	1,0 × 10 <sup>-2</sup> W		
1 500 to 1 800	8 × 10 <sup>6</sup> W		8 × 10 <sup>-3</sup> J				1,8 × 10 <sup>-2</sup> $t^{0,75}$ J				
1 800 to 2 600	8 × 10 <sup>5</sup> W		8 × 10 <sup>-4</sup> J				4,4 × 10 <sup>-3</sup> $t^{0,25}$ J	10 <sup>-2</sup> $t$ J			
2 600 to 4 000	8 × 10 <sup>4</sup> W		8 × 10 <sup>-5</sup> J	4,4 × 10 <sup>-3</sup> $t^{0,25}$ J							
4 000 to 10 <sup>6</sup>	10 <sup>11</sup> W.m <sup>-2</sup>		100 J.m <sup>-2</sup>	5 600 $t^{0,25}$ J.m <sup>-2</sup>				1 000 W.m <sup>-2</sup>			
NOTE Laser products that meet the requirements for classification as Class 1 by satisfying measurement conditions 1 and 2 may be hazardous when used with viewing optics having greater than ×7 magnification or objective diameters greater than those specified in Table 11.											
<sup>a</sup> For correction factors and units, see Table 10											
<sup>b</sup> The AELs for emission durations less than 10 <sup>-13</sup> s are set to be equal to the equivalent power or irradiance values of the AEL at 10 <sup>-13</sup> s.											
<sup>c</sup> In the wavelength range between 450 nm and 500 nm, dual limits apply and a product's emission must not exceed either limit applicable to the class assigned.											

**Table 3.1:** Eye safe power specifications at different wavelengths for multiple durations [76]. The blue shading depicts the SWIR wavelengths used in this collection for an illumination time of greater than 10 seconds. Empirical testing determined the illumination source used for this collection was well below the eye safety thresholds.

In order to capture face images in which the subjects were wearing sunglasses with varying amounts of tint from multiple manufacturers, the photo booth panel was removed from the box. Therefore, facial images were collected under the following scenarios, with all lux measurements taken at the interior booth seating location:

– **Sunglasses**

1. Ground Truth - No Sunglasses
2. Oakley Flak Jacket Sunglasses
3. Oakley Straight Jacket Sunglasses
4. RB3449 59 Sunglasses
5. RB3025 58 Sunglasses

– **Lighting Conditions**

1. 0 lux interior, 3-point exterior ( $\sim 350$  lux)
2. Single external source ( $\sim 5$  lux)
3. 500mW 1550nm active illumination

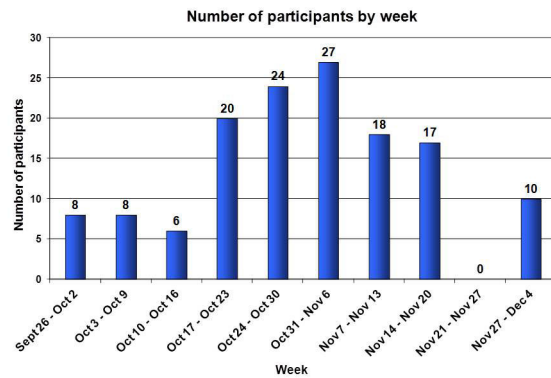
Each sunglass and ground truth image was captured under all lighting conditions, creating a total of fifteen different tint/lighting scenarios. Sample images, including the ground truth images, taken under this capture protocol for the Oakley Flak Jacket can be seen in Figure 3.5.

A total of 140 participants provided data for this collection between September 26, 2011 and December 4, 2011. An overview of the demographics of the subjects is broken down in

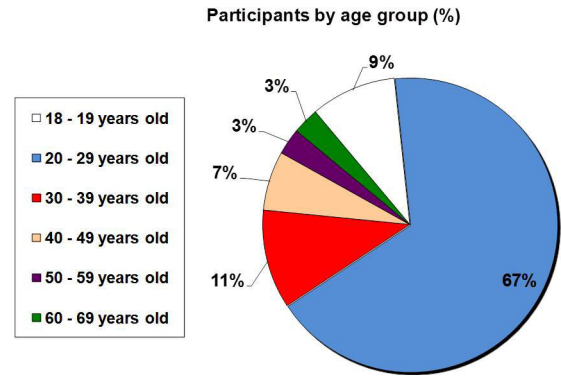


Variable Lighting – Oakley Flak Jacket				
	Ground Truth	No Internal, Full External	Single-Source External	Active Illumination
Visible				
OPR	5	6	9	7
ENH	OFF	OFF	OFF	OFF
1150 nm				
1250 nm				
1350 nm				
1450 nm				
1550 nm				

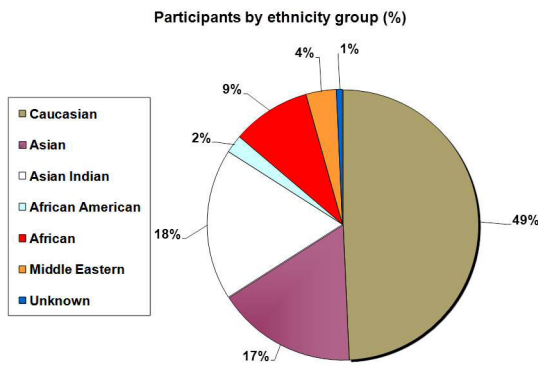
**Figure 3.5:** Sample images from the WVUMFD captured using different wavelengths (1150-1550nm) and lighting conditions, including the ground truth data, when the subject was wearing Oakley Flak Jackets.



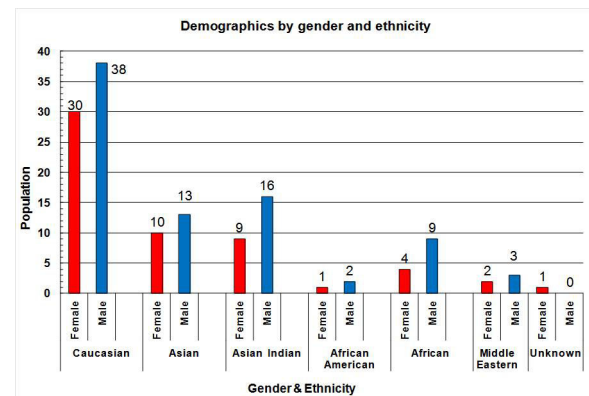
(a)



(b)



(c)



(d)

**Figure 3.6:** Demographic breakdowns of subjects collected under the WVUMFD. (a) Number of participants collected weekly; (b) Participants by age group; (c) Participants by ethnicity group; and (d) demographics by gender and ethnicity.

Figure 3.6. Figure 3.6(a) demonstrates the number of participants collected on a weekly basis during the period of collection. Figure 3.6(b) show the participants by age group. Notice that the majority of subjects collected were in the 20-29 year old age group due to the fact that the majority of subjects collected were students at West Virginia University. Figure 3.6(c) breaks down the participants by their ethnicity while Figure 3.6(d) demonstrates the distribution between gender and ethnicity.

## 3.4 Conclusion

Operational scenarios, such as law enforcement surveillance applications and government military defense situations, are limited in their ability to capture high quality and ideal imagery. Thus, representative data must be collected and used in order to further enhance face/eye detection methods and cross spectral FR algorithms. In this chapter, an overview of the West Virginia University Multispectral Face Database was provided. Discussed was an overview of the challenges that arise when collecting SWIR face imagery as well as the capture setups for both the ideal visible face images as well as non-ideal SWIR images. A glass characteristic study was performed to determine what type of tinted glass would best represent high, medium, and low tint and data was collected under 30 different lighting and tinting scenarios. Both tinted glass and sunglasses as well as multiple lighting scenarios were used to collect a significant number of subjects. Finally, demographics for the subjects collected show that there is a great diversity between the faces that are captured under non-ideal SWIR imaging. This collection allows for representative data to be used when designing, developing and testing face/eye detection and cross spectral FR algorithms.



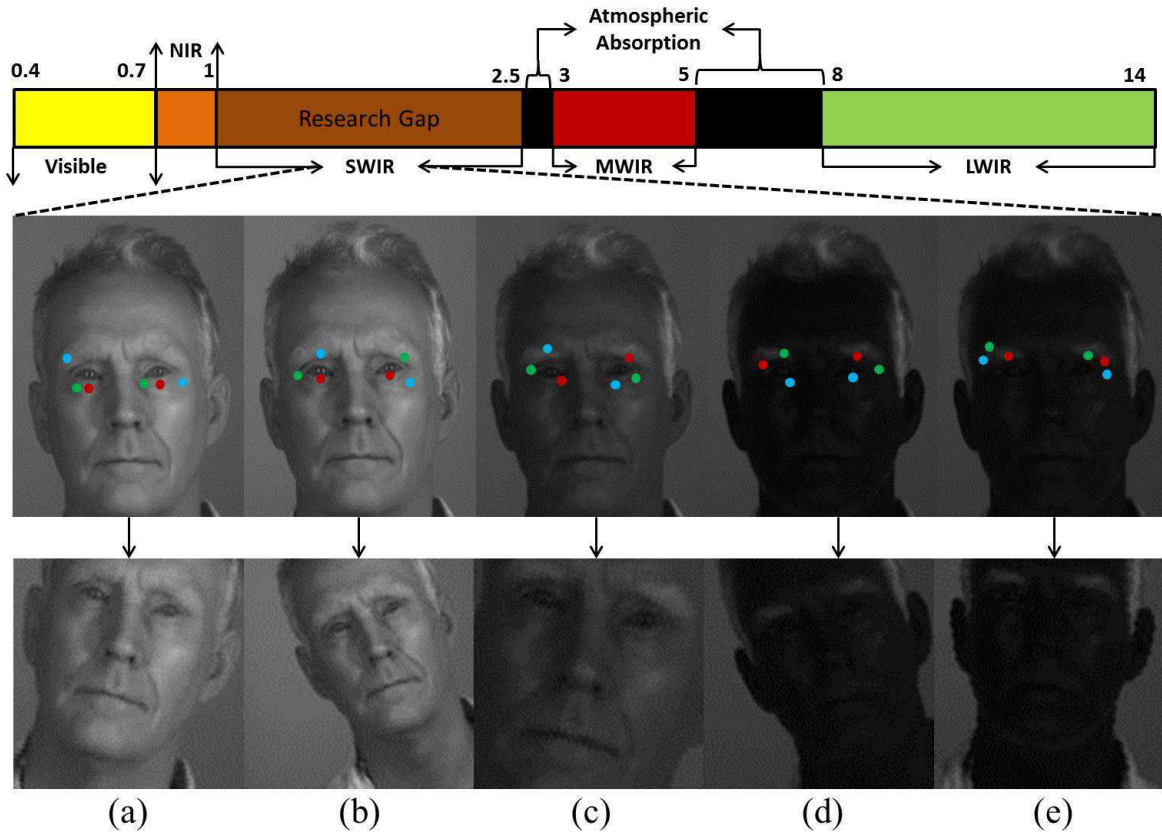


# Chapter 4

## Accurate Eye Localization in the Short Waved Infrared Spectrum through Summation Range Filters

### 4.1 Introduction

In a typical FR system, one of the main challenges that must be overcome in order to achieve high FR identification rates is to successfully determine the face and eye locations of all face images in a database that are being used for matching. Eye detection, in particular, is considered to be one of the fundamental steps in the majority of FR algorithms. Because FR trends have been moving into higher bands of the electromagnetic spectrum, as discussed in Chapter 1, it is important to have an eye detection method that is quick and accurate in the SWIR band. Reported work in the literature shows that automated and robust eye localization can have a positive impact on FR performance [18]. However, while available



**Figure 4.1:** The electromagnetic spectrum shows the research gap in SWIR and sample images from a) 1150nm b) 1250nm c) 1350nm d) 1450nm and e) 1550nm. Notice how eye centers from both academic (green and blue) and commercial (red) algorithms are inaccurate and produce normalized face images with both rotational and scale issues (bottom row).

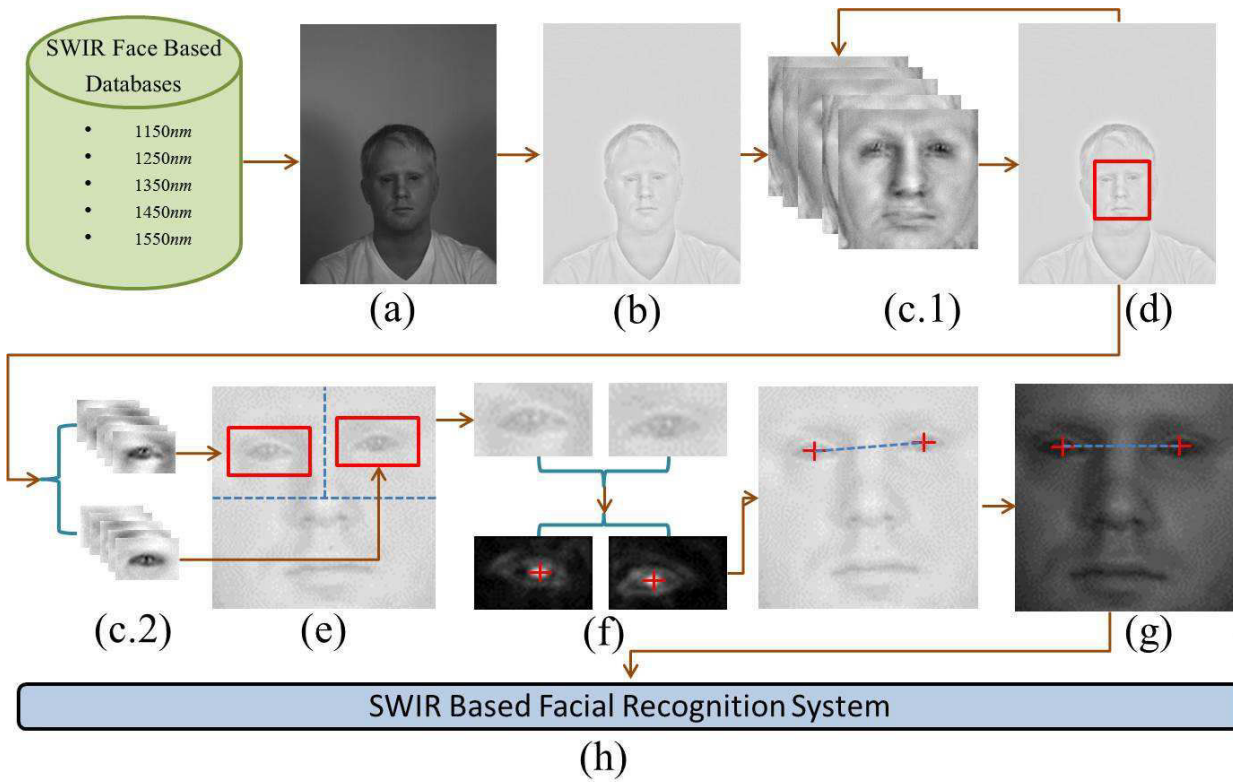
eye detection approaches perform well in the visible band, they perform poorly in the SWIR spectrum. Preliminary tests show that the state of the art algorithms, including commercial and academic approaches, do not output sufficiently accurate eye locations, which in turn, produces badly geometric normalized face images, as can be seen in Figure 4.1. This has the potential to significantly decrease the performance of an FR matcher.

In this chapter, a fast and robust eye detection method will be described that can accurately detect not only the face and eye regions, but determine the center of the subjects pupil. This allows for more accurate geometric normalization which leads to a more accurate FR matchers performance. The rest of the chapter is organized as followed. Section 2 describes the methodology of the proposed approach. Section 3 extensively validates the accuracy and robustness of the proposed approach while Section 4 summarizes the chapter.

## 4.2 Methodology

In this section, the methodology of performing automatic face/eye detection in multiple SWIR wavelengths (i.e. 1150, 1250, 1350, 1450, and 1550nm) will be described. The authors assumed the data used was full frontal images of an upper body area with ideal lighting conditions and did not consider more challenging cases, with head pose variation and eye rotations. An overview of the proposed approach can be seen in Figure 4.2. Given a query image (Figure 4.2(a)), the methodology consists of five main processes, i.e. 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. Sample images of subjects from each individual wavelength can be seen in Figure 4.1.





**Figure 4.2:** Overview of the proposed fully automated pre-processing methodology. a) Query Image b) Photometric Normalization c) Sample Average Face and Eye Templates d) Face Detection e) Eye Detection f) Eye Center Localization using Summation Range Filtering g) Geometric Normalization

### 4.2.1 Preprocessing

SWIR images tend to have low contrast in the face region, especially in the upper wavelength bands (i.e. 1350, 1450, and 1550 $nm$ ). Instead of the human skin reflecting those wavelengths back into the camera, the moisture in 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 are used to increase the contrast. However, different photometric techniques bring out unique features that are beneficial for face/eye detection in different wavelengths. For each wavelength, a specific photometric normalization is used (Figure 4.2(b)). For a detailed description and justification on which photometric normalization is used on each individual wavelength, refer to Section 4.3.3

Because the proposed technique is a *template based* method, further described in Chapter 2, average templates are needed. Therefore, for each wavelength, 7 subjects are randomly selected. The faces are geometrically normalized, cropped, and averaged together to create an *average face template* (Figure 4.2(c.1)). Then, the eye regions from this template are cropped and used as *average eye templates* (Figure 4.2(c.2)). These average templates are then saved and used on all images in the database.

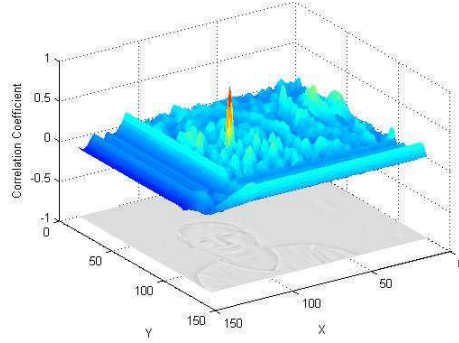
### 4.2.2 Automatic Face Detection

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. For each pixel in the query 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:

$$\delta(u, v) = \frac{\sum_{x,y} [f(x,y) - \bar{f}_{u,v}] [t(x-u, y-v) - \bar{t}]}{\{\sum_{x,y} [f(x,y) - \bar{f}_{u,v}]^2 \sum_{x,y} [t(x-u, y-v) - \bar{t}]^2\}^{0.5}} \quad (4.1)$$

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. The convolution of the image and average template then yields a correlation map.



**Figure 4.3:** Correlation map of query face image for automatic face detection. The highest peak of the correlation map corresponds to the location of the face.

Figure 4.3 shows the correlation map in correspondence with the query image. The highest location within the correlation map (i.e. the peak) is the location of the face. However, different average templates yield different results. Because of this, multiple average templates (in this case 5) are created and used to increase the chance of finding the correct location. Mathematically, the final location of the face can be described as:

$$\hat{\delta}(u, v) = \underset{x}{\operatorname{argmax}}(\delta_x(u, v)) \quad (4.2)$$

where  $\delta_x(u, v)$  is the location of the highest correlation coefficient obtained from average

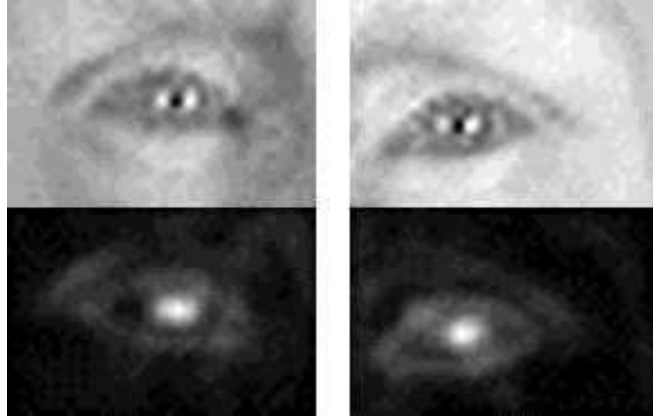
template  $x$  (in this case  $x = 1, \dots, 5$ ).  $\hat{\delta}(u, v)$  then corresponds to the upper left hand point of the face region and the face can be cropped to the size of the average templates (Figure 4.2(d)). This approach helps narrow down the search space when locating the eye regions.

### 4.2.3 Eye Region Localization

Since the location of the face is now known, the location of the eye regions can be determined. In order to further reduce the search space, the face is split into four equal regions (i.e. top left, top right, bottom left and bottom right). Assuming that the face region is found correctly through the method described above, the right and left eye 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 4.1. As stated above, different average eye templates yield different results. Therefore, the process is repeated multiple times (in this case 5) using unique templates to increase the chance of obtaining the correct region (Figure 4.2(e)). Justification for this practice is described in Section 4.3.4. Then, Equation 4.2 can be used to find the final location of the eye regions. Note that no assumptions as to where the eye regions are is made and if the face detection fails, the eyes will not be in the proper region and the effect is shown in the overall results, as can be seen in 4.3.1.

### 4.2.4 Summation Range Filter

Although the region of the eye can be empirically found, the center of the eye cannot always be said to be the center of the found region. Therefore, an accurate way of determining the correct center of the eye must be employed. During the data collection process, subjects



**Figure 4.4:** Sample detected left and right eye regions and their corresponding summation range images.

were asked to face and look directly at the camera as the samples were taken. Because an illumination source was used for an ideal capturing scenario, corneal reflections were assumed to be present and located in the center of the subjects eyes. Taking advantage of this, summation range filters can be used to more accurately determine the center of the eye by determining the location of those corneal reflections. The summation range map ( $S(x, y)$ ) can mathematically be described as:

$$S(x, y) = \sum_{x=-1}^1 \sum_{y=-1}^1 R(x, y) \quad (4.3)$$

where

$$\begin{aligned} R(x, y) = \\ \text{argmax}(I(x - 1 : x + 1, y - 1 : y + 1)) - \\ \text{argmin}(I(x - 1 : x + 1, y - 1 : y + 1)) \end{aligned} \quad (4.4)$$

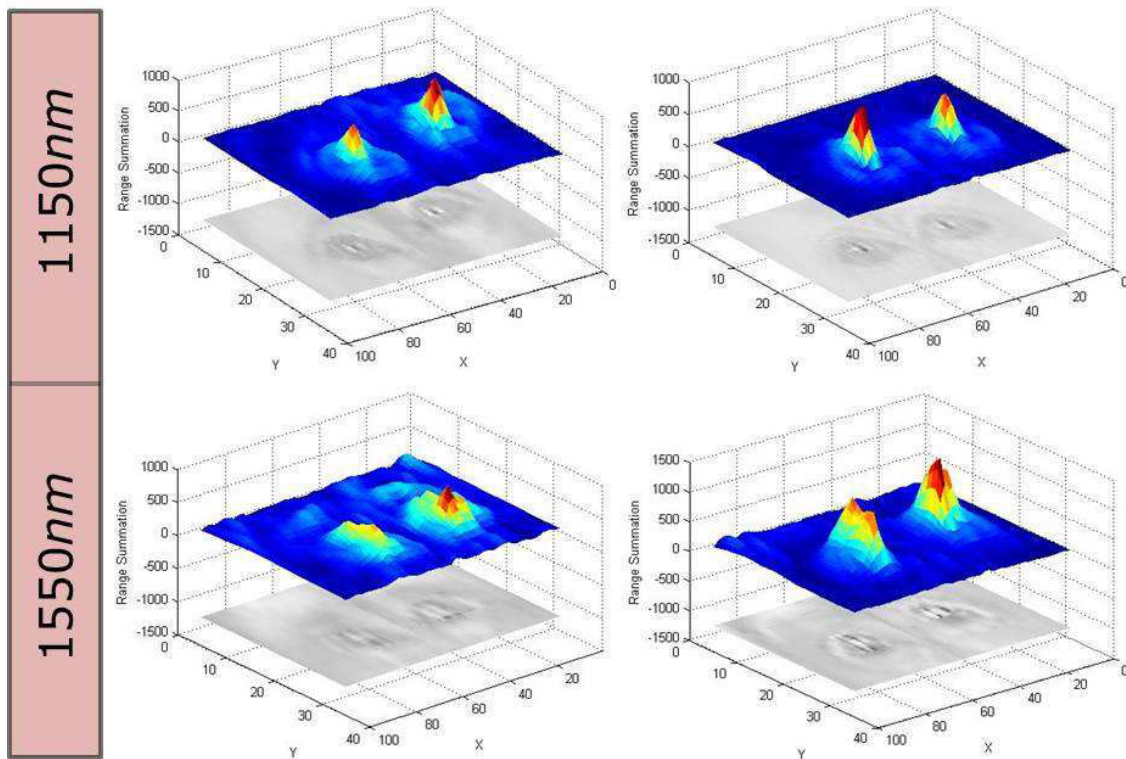
where  $I(x, y)$  is the original cropped eye region. Then, the eye final eye center is said to be:

$$P(x, y) = \operatorname{argmax}(S(x, y)) \quad (4.5)$$

This process is done for both the right and left eye regions to determine the final locations for the right and left eye respectively (Figure 4.2(f)). Figure 4.4 shows a sample found right and left eye region and its corresponding range filter while Figure 4.5 illustrates  $S(x, y)$  in both the 1150nm and 1550nm. Notice that the peaks in Figure 4.5 are not directly located in the middle of the found regions.

### 4.2.5 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 to fit said distance. 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 crucial to facial recognition algorithms. Therefore, all face images were geometrically normalized based on the found locations to have an interocular distance of 60px with a resolution of 130x130px (Figure 4.2(g)). The geometrically normalized images can then be used in a facial recognition system (Figure 4.2(f)).



**Figure 4.5:** Sample summation range maps and their corresponding detected eye regions. Note that the peaks are not in the center of the found regions.

## 4.3 Evaluation

In this section, the proposed method is tested on multiple bands of the SWIR spectrum. Additionally, the method is tested for robustness to change in scale, compression, and blur. Because there are no publicly available datasets collected in the SWIR band, the ground truth data from the WVUMFD, described in Chapter 3, was used. Due to some errors in the data, 135 subjects were collected with 10 samples per subject per wavelength, resulting in 6,750 images. Each subject was asked to look forward with no variability in pose, such as in mugshots, and was captured under ideal lighting conditions for each wavelength. All pupils were assumed to be in the middle of the eye, allowing for the pupil points to be used as geometric normalization points. After collection, both the left and right eye were then manually marked for their ground truth locations to use later for evaluation purposes. All experiments were performed on a 64-bit Windows 7 machine with 12GB of RAM running Intel Core i7 CPU @ 3.2Ghz using MATLAB R2012b.

In all experiments, the normalized error is used. This error, indicating the error obtained by the worse eye estimation, is used as the accuracy measure for the found eye locations. Proposed by [81], the normalized error is described as:

$$e = \frac{\max(d_{left}, d_{right})}{w} \quad (4.6)$$

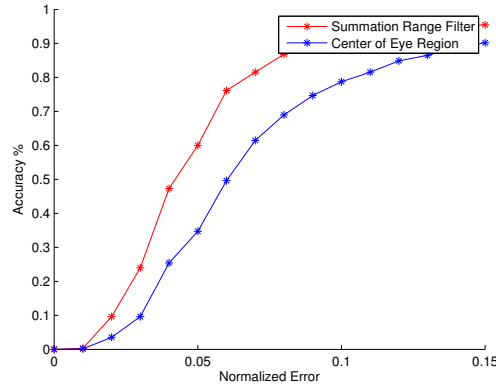
where  $d_{left}$  and  $d_{right}$  is the Euclidean distance between the found left and right eye centers with the manually annotated ground truth and  $w$  is the Euclidean distance between the eyes in the ground truth. In the normalized error,  $e \leq 0.25$  (or 25% of the interocular distance) roughly corresponds to the width of the eye (i.e. corner to corner),  $e \leq 0.10$  roughly



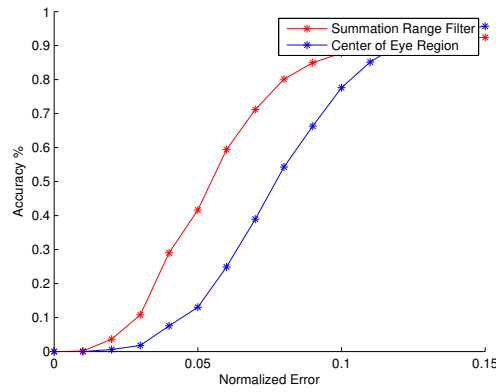
corresponds to the diameter of the iris, and  $e \leq 0.05$  roughly corresponds to the diameter of the pupil.

### 4.3.1 Face Detection Accuracy and its Effect on Automatic Eye Detection

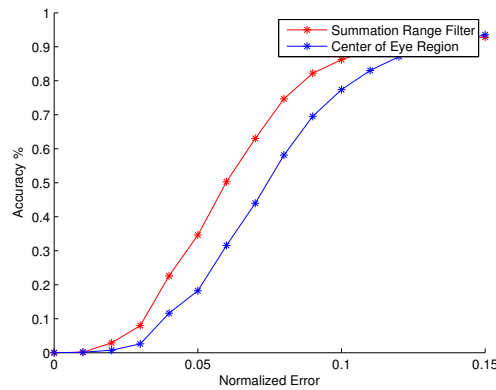
To show that the proposed algorithm is a complete eye localization algorithm from raw input image to final eye locations, an additional face detection step to the approach is added. In order to evaluate the accuracy of this face detection, the middle pixel of the found face is compared to the middle pixel of the manually annotated face. If the found face was  $e \leq 0.25$ , then it was deemed to be a found face. Whether the face was within the  $e \leq 0.25$  or not, the face was cropped accordingly and performed the performed eye localization algorithm. The eye localization algorithm using the found face locations as well as the manually annotated faces were compared and reported results at  $e = 0.15$ . Results for this experiment can be found in Table 4.1. It can be seen that in 1150, 1250, and 1350, where face detection has a high success rate, eye localization accuracy does not significantly decrease. In some cases (e.g. 1150 and 1250), accuracy improves if only very slightly. However, when the face detection accuracy is low, as in 1450 and 1550, eye localization accuracy is effected more greatly, dropping the accuracy by a significant amount (i.e. more than 11%).



(a) 1150nm



(b) 1350nm



(c) 1550nm

**Figure 4.6:** Results of running the proposed algorithm with (red) and without (blue) the summation range filter. When the summation range filter is not used (blue), the center of the found eye region is used. At  $e = 0.05$ , the summation range filter increases the accuracy of the original approach by 25.3%, 28.5% and 16.4% for the 1150, 1350, and 1550nm bands respectively.

	1150	1250	1350	1450	1550
Face Detection Rate	99.26	99.70	98.67	83.93	91.78
Eye Detection w/ Manual Faces	94.44	93.78	92.22	89.41	92.07
Eye Detection w/ Detected Faces	94.52	94.07	91.33	78.37	86.37

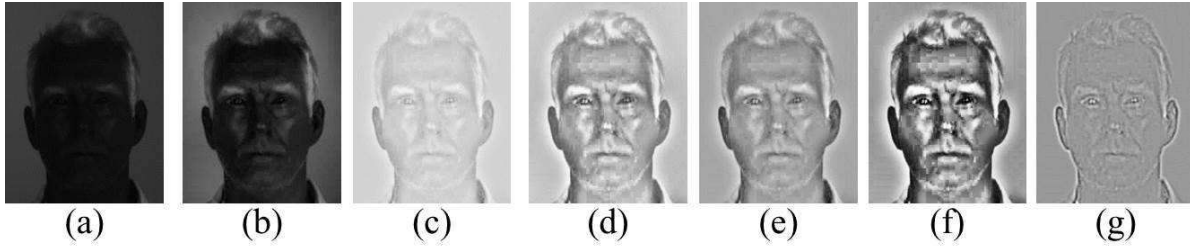
**Table 4.1:** Face detection rates and normalized errors ( $e = 0.15$ ) for the proposed eye detection method using the face detection approach as well as using the manually annotated face regions.

### 4.3.2 Effect of Using Summation Range Filter for Automatic Eye Detection

To show the effectiveness of adopting the summation range filter after eye region localization, the proposed algorithm was performed with and without the use of the summation range filter for three different wavelengths of the SWIR band. When the summation range filter was not used, the center of the localized eye region was used as the final location. The results of this experiment can be seen in Figure 4.6. It can be seen that the use of the summation range filter significantly increases the accuracy of the proposed algorithm across all three bands. At  $e = 0.05$ , the summation range filter increases the accuracy of the original approach by 25.3%, 28.5% and 16.4% for the 1150, 1350, and 1550nm bands respectively.

### 4.3.3 Effect of Photometric Normalization on Automatic Eye Detection

Because each individual wavelength has different contrasts and illumination properties, a study on multiple photometric normalization techniques was performed. In this study, seven



**Figure 4.7:** Examples of photometric normalizations used: (a) Original Image (b) CLAHE (c) LBSSR (d) CLAHE LBSSR (e) TBSSR (f) CLAHE TBSSR (g) TT

scenarios were considered and tested against each other to determine the photometric normalization that has the highest eye detection performance overall. The following photometric normalization techniques were used:

### Contrast Limited Adaptive Histogram Equalization (CLAHE)

CLAHE [82] operates on small local regions (8x8 for these experiments) in the image and applies histogram equalization on each individual region (in contrast to the entire image in regular histogram equalization). Mathematically, it is described as:

$$f(n) = \frac{N-1}{M} \times \sum_{k=0}^n h(k), \quad (4.7)$$

where  $M$  and  $N$  are the number of pixels and grey level bins in each sub-region, respectively, and  $h$  is the histogram of each sub-region. In order to increase the 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 reported experiments). Specifically, grey level counts above the threshold are uniformly distributed among the grey levels below it. Finally, the patches are subsequently combined using bilinear interpolation.

### Log Based Single Scale Retinex (LBSSR)

LBSSR [83] decomposes the image into two components, i.e. 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 the illumination image. Mathematically, this can be described as:

$$I(x, y) = L(x, y) \times R(x, y) \quad (4.8)$$

$$L(x, y) = I(x, y) * G_\sigma(x, y) \quad (4.9)$$

where  $G_\sigma$  is a Gaussian of scale  $\sigma$  and  $*$  denotes the convolution between the image and the kernel. Finally, the reflectance image is estimated as:

$$R(x, y) = \log_{10} \left( \frac{I(x, y)}{L(x, y)} \right) \quad (4.10)$$

### CLAHE LBSSR

A common problem with LBSSR is that images tend to become over saturated or “washed out”. This can have negative effects on eye detection algorithms. Furthermore, “halo” artifacts may be introduced depending on the scene and scale value chosen for the Gaussian smoothing function. Certain modifications to the LBSSR have been introduced, such as the multi-scale retinex approach [87], but at the cost of increased processing speed. Therefore, in this work, the CLAHE approach listed above is applied to the LBSSR image to help

compensate for the aforementioned approaches and to increase the contrast of the image.

### **Tangent Based Single Scale Retinex (TBSSR)**

By using different non-linear transformations on the LBSSR, different image representations can be obtained, e.g. by using TBSSR [83]. Therefore, the log in Equation 4.10 is replaced with an arc-tangent transformation, resulting in the following:

$$R(x, y) = \text{atan} \left( \frac{I(x, y)}{L(x, y)} \right) \quad (4.11)$$

### **CLAHE TBSSR**

As described above, the TBSSR can cause over saturation and haloing effects. Therefore, the CLAHE approach was also applied to the TBSSR image to correct the contrast issues mentioned.

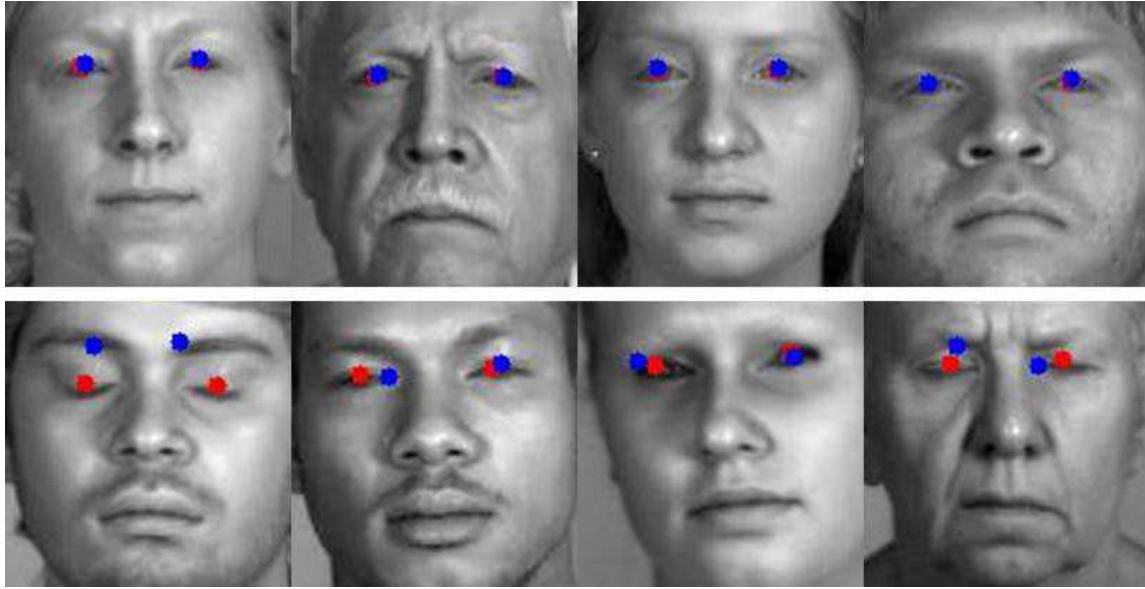
### **Tan and Triggs (TT)**

This photometric normalization, proposed by [84], 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 a blurred version of the same image), and contrast equalization (suppressing larger intensities while maintaining lower intensities).

Sample images for each photometric normalization can be seen in Figure 4.7. To evaluate the performance, the entire system (including face and eye detection) was tested on all pho-

		Photometric Normalization						
		Original	CLAHE	LBSSR	CLAHE LBSSR	TBSSR	CLAHE TBSSR	TT
1150	5%	42.74	27.70	<b>59.63</b>	41.78	56.52	33.33	51.93
	10%	80.67	65.70	<b>90.81</b>	81.78	90.00	72.96	83.26
	25%	93.19	90.37	<b>98.00</b>	96.89	98.22	95.85	96.96
1250	5%	36.89	25.63	<b>55.93</b>	44.15	53.56	35.85	42.52
	10%	81.48	69.63	<b>91.85</b>	86.30	91.33	79.70	77.33
	25%	90.15	90.81	<b>98.00</b>	95.33	98.37	95.85	86.44
1350	5%	20.37	15.93	<b>41.41</b>	27.63	53.56	23.41	32.30
	10%	52.44	46.81	<b>87.04</b>	71.78	91.33	60.67	75.26
	25%	68.37	81.11	<b>96.07</b>	93.63	98.37	89.70	91.56
1450	5%	8.89	10.00	22.81	<b>28.83</b>	26.81	28.52	12.00
	10%	24.81	30.52	70.30	<b>73.85</b>	70.67	72.37	31.33
	25%	44.07	46.15	84.74	<b>87.85</b>	83.70	86.67	38.15
1550	5%	10.00	11.78	32.15	31.63	<b>35.33</b>	31.48	24.74
	10%	22.96	29.19	80.67	74.89	<b>80.30</b>	73.41	53.56
	25%	46.52	49.85	90.74	89.93	<b>88.81</b>	90.89	59.56

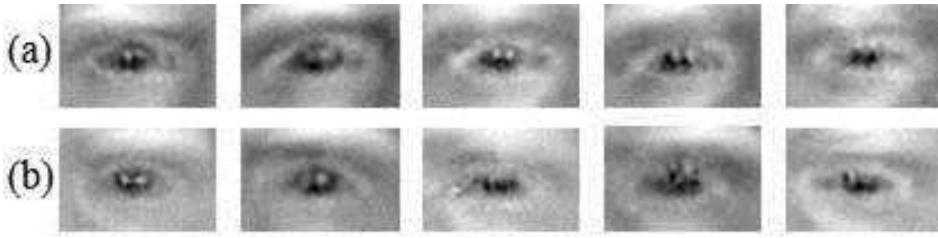
**Table 4.2:** Normalized errors (maximum error for both left and right eye) across six different photometric normalizations to determine the highest overall accuracy per wavelength. 5% normalized error is approximately within the diameter of the pupil, 10% diameter of the iris, and 25% the diameter of the eye. The bold results represent the highest 5% normalized error and were used for further experimentation.



**Figure 4.8:** Sample success (top row) and failed (bottom row) eye detection locations on the 1150nm database. Red dots are manual locations while blue dots are found locations.

tometric normalizations as well as the original image. Average face and eye templates were created for each specific photometric normalization and used to determine the location of the eye. Table 4.2 shows the results of the study while Figure 4.8 has example success and failure cases. It can be seen that the LBSSR has the best normalized error for 1150, 1250, and 1350 while CLAHE LBSSR and TBSSR performs best for 1450 and 1550, respectively. All further experiments on each database are conducted on the respective photometric normalizations that resulted in the highest normalized error.



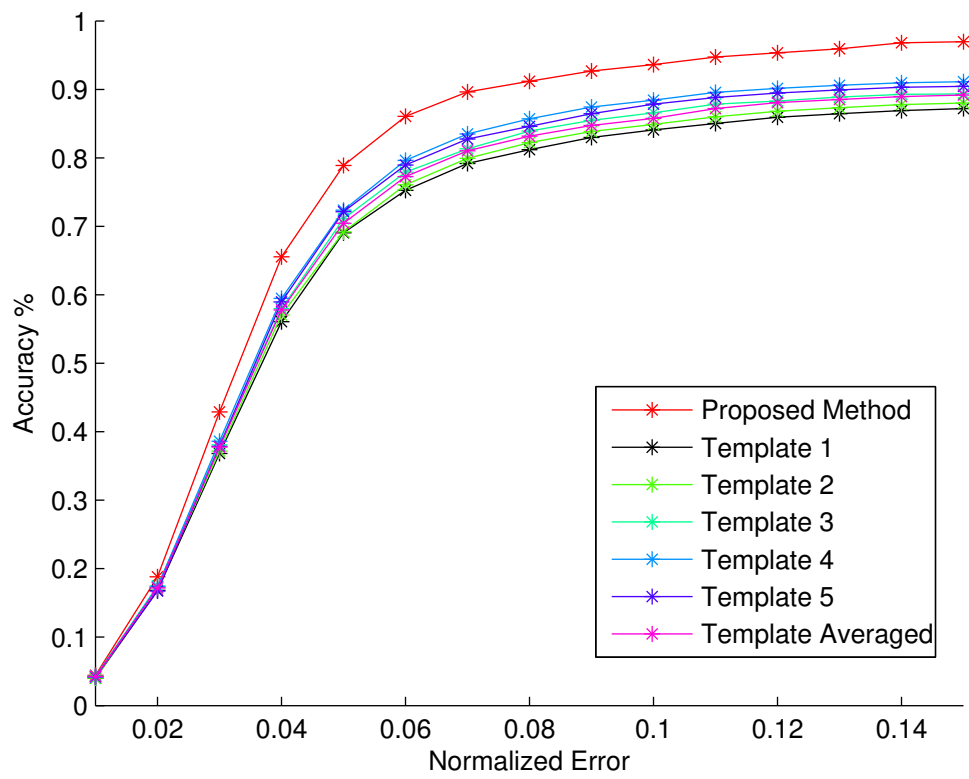


**Figure 4.9:** Different average eye templates for (a) the right eye and (b) the left eye in the 1550nm database

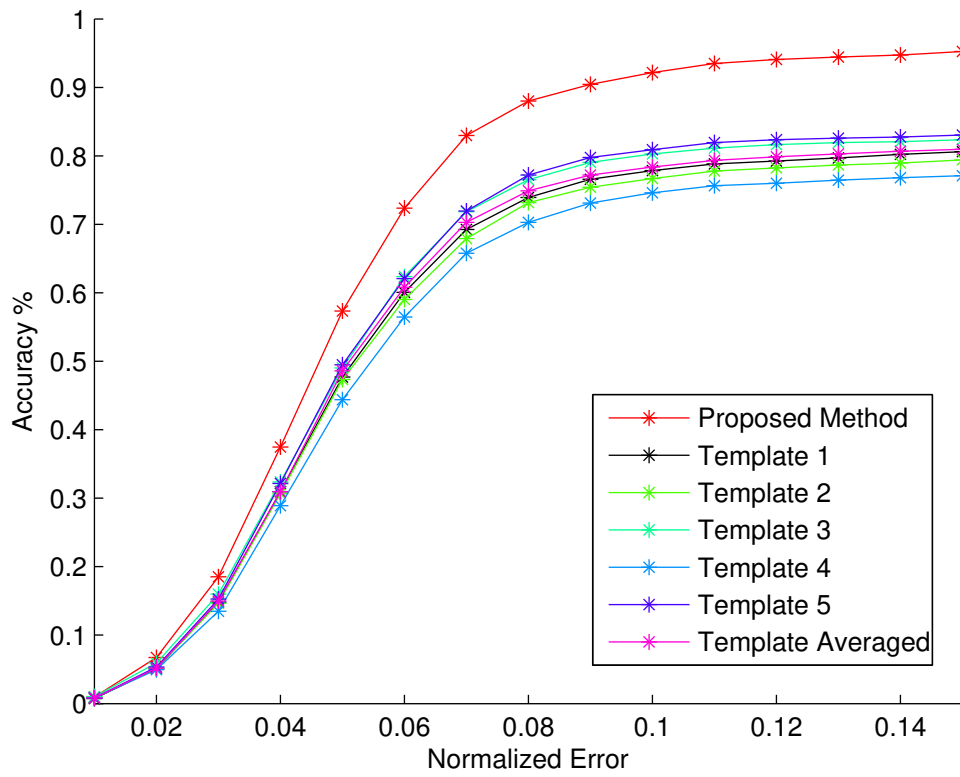
#### 4.3.4 Effects of Variable Average Templates on Automatic Eye Detection

In order to justify the use of multiple templates while performing the proposed eye detection, a study was conducted to show the effects different templates have on accurately detecting eyes. Here, 5 average templates were created from different groups of 7 subjects chosen at random, creating unique templates for each eye, as shown in Figure 4.9. Assuming the face is known, the proposed method (taking the best match from all five templates) is compared to the performance of each individual template as well as the average of all 5 templates on both the 1150 and 1550 database. The normalized error for all 5 individual methods as well as the proposed method can be seen in Figures 4.10 and 4.11.

Notice that each average template has varying accuracies, especially when  $e \geq 0.10$ . Also notice that the proposed method, which uses the best match from all 5 templates, achieves much higher accuracy than any one individual template and the average of all 5 templates.



**Figure 4.10:** Effects different templates have on eye detection accuracy on the 1150nm database.



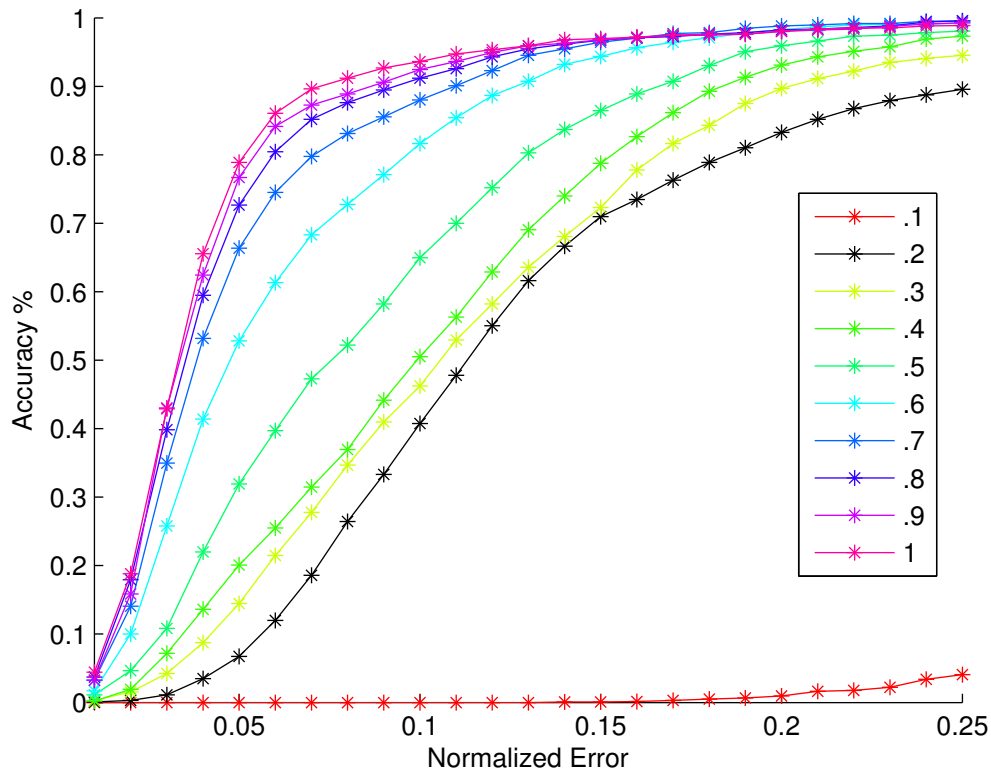
**Figure 4.11:** Effects different templates have on eye detection accuracy on the 1550nm database.

### 4.3.5 Robustness of Automatic Eye Detection Approach

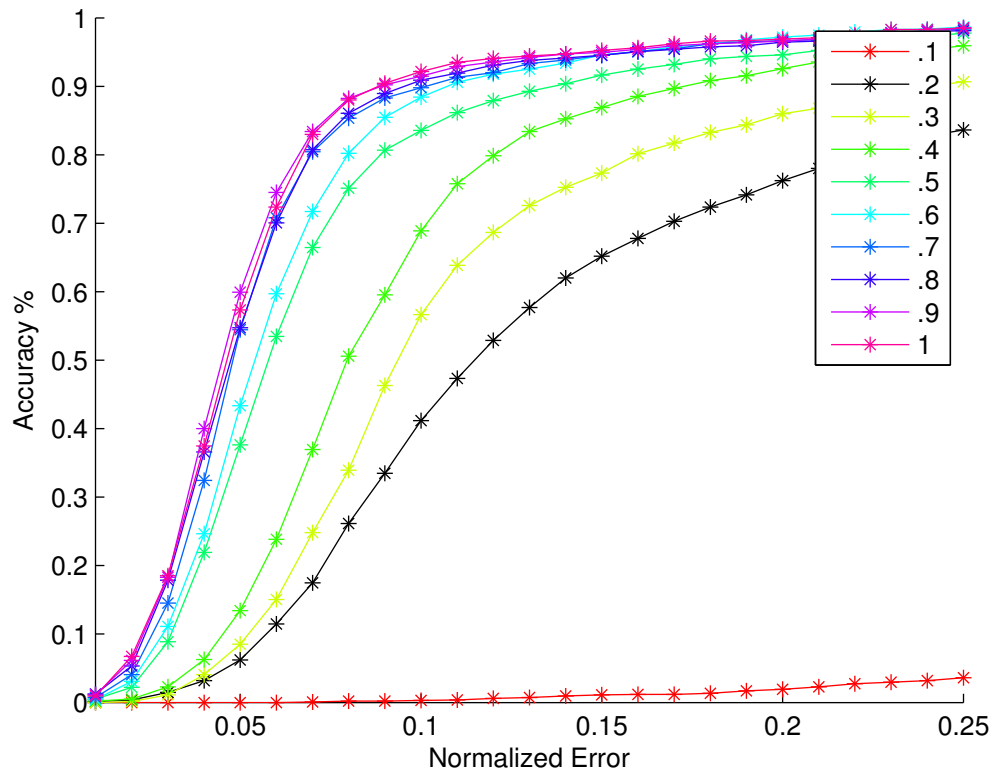
In order to test the robustness of the proposed approach, three studies were conducted, i.e. robustness to scale, robustness to image compression, and robustness to image blurring. Each experiment was conducted on all images from *1150nm* and *1550nm* databases assuming the face is known.

#### Size

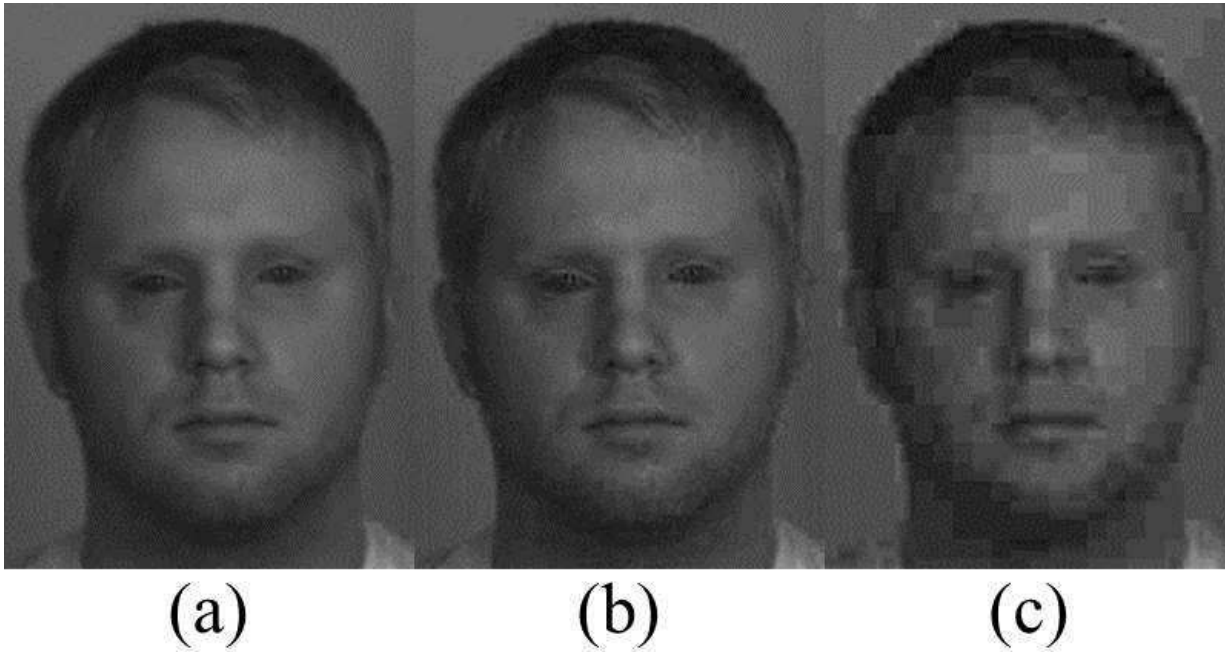
In order to study the effect of changing the scale (i.e. spatial resolution), the test images and average templates were downsampled from 100% ( $130 \times 130px$ ) to 10% ( $13 \times 13px$ ) in 10% increments. The results from this experiment can be seen in Figures 4.12 and 4.13. Note that in *1150nm*, downscaling from 100% to 80% does not significantly affect the results, while in *1550nm*, the images can be downsampled to more than 60% without a significant change.



**Figure 4.12:** Effect of eye detection on downsampling images from 100% to 10% in 10% increments in the 1150nm database.



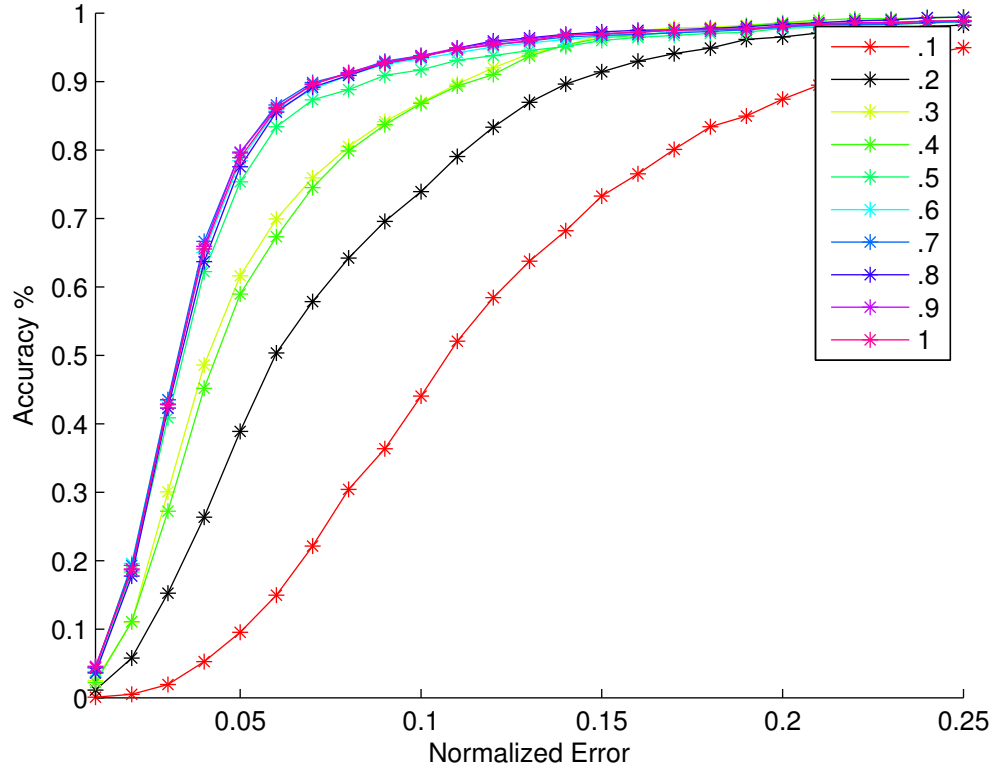
**Figure 4.13:** Effect of eye detection on downsampling images from 100% to 10% in 10% increments in the 1550nm database.



**Figure 4.14:** Sample JPEG compressed images from  $1150nm$  with a) 100% b) 50% and c) 10% compression rates.

### JPEG Compression

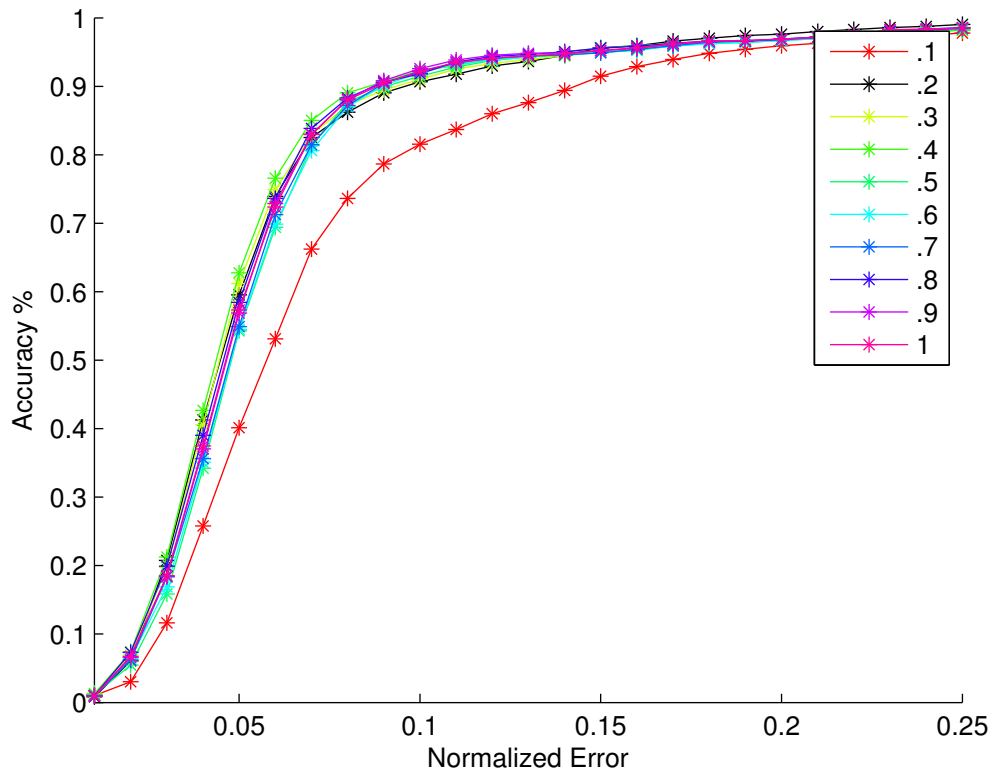
Due to the high number of pixels many cameras have available today, storage size of a single image can be an issue. In order to decrease the amount of storage needed to save an image, compression can be used. However, different levels of compression can cause different artifacts in the images. Therefore, a study on the effect that image compression has on the proposed approach was studied. Here, JPEG compression was applied to all images ranging from 100% (11.3 KB) to 10% (5.22 KB) in 10% intervals. Example images with 100%, 50% and 10% JPEG compression can be seen in Figure 4.14 while the results for the study can be seen in Figures 4.15 and 4.16. Note that JPEG compression of up to 50% does not



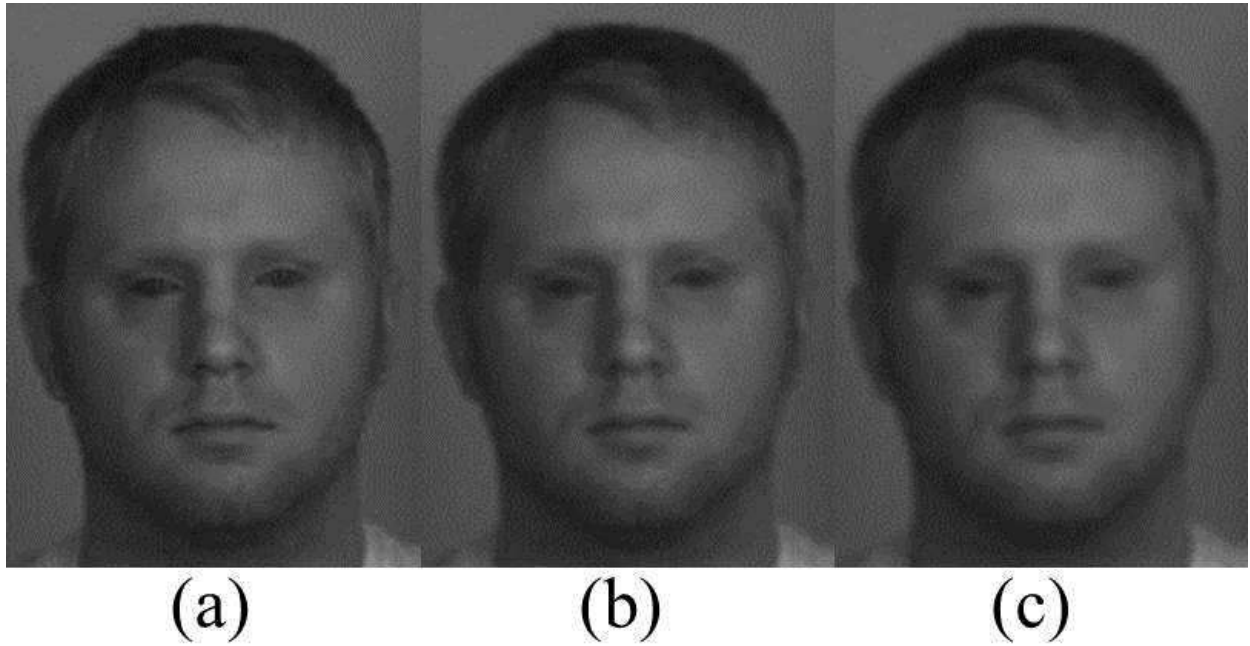
**Figure 4.15:** Effect of eye detection on JPEG compression from 100% to 10% in 10% increments in the 1150nm database.

significantly effect the results in 1150nm while in 1550nm compression of up to 20% has minimal effect.





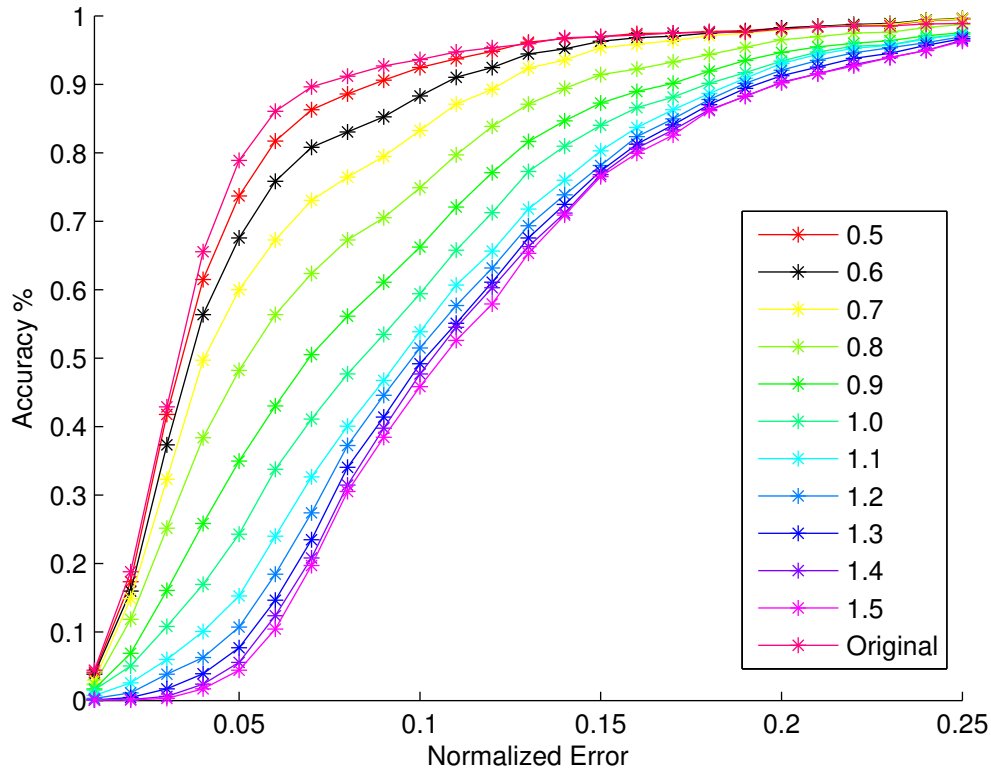
**Figure 4.16:** Effect of eye detection on JPEG compression from 100% to 10% in 10% increments in 1550nm database.



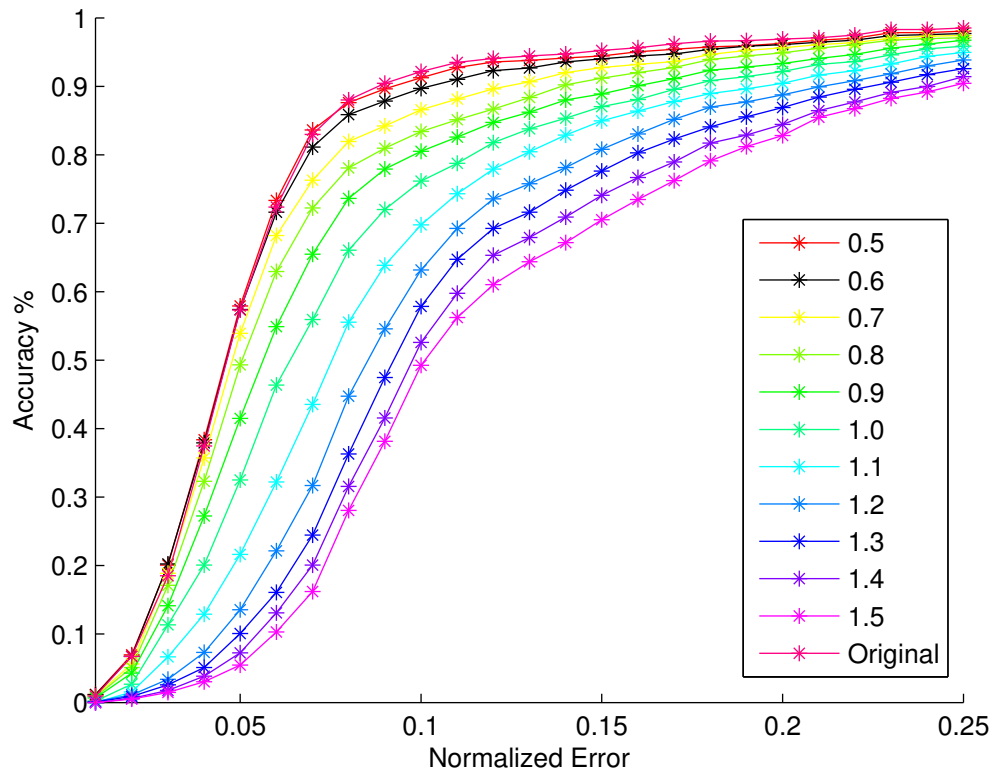
**Figure 4.17:** Sample images with Gaussian blur from  $1150nm$  with a)  $\sigma = 0.5$  b)  $\sigma = 1.0$  and c)  $\sigma = 1.5$ .

### Blur

Because images can easily be taken when the capturing lens is out of focus, a study on image blurring was considered. In order to simulate a camera being out of focus, Gaussian blur was added to all images with a window size of 7 and  $\sigma$  ranging from 0.5 to 1.5 in 0.1 increments. Sample images with Gaussian blur can be seen in Figure 4.17 while the results of the study can be seen in Figures 4.18 and 4.19. Note that as  $\sigma$  increases, the accuracy has a vertical shift downwards in both  $1150nm$  and  $1550nm$ . This can be explained by the fact that as  $\sigma$  increases, the blurring effect increases, which in turn decreases the range within the eye region. This causes higher errors when determining the center of the eye.



**Figure 4.18:** Effect of eye detection on Gaussian blur from  $\sigma = 0.5$  to  $\sigma = 1.5$  in 0.1 increments in the 1150nm database.



**Figure 4.19:** Effect of eye detection on Gaussian blur from  $\sigma = 0.5$  to  $\sigma = 1.5$  in 0.1 increments in the 1550nm database.

### 4.3.6 Comparison with the State of the Art

Because there is little to no work done in SWIR eye detection, it is hard to compare the obtained results with the state of the art. In order to do this, algorithms that are commercially or academically available for testing must be used. Therefore, 3 different methods have been tested for comparison. First, the Cascaded Adaboost eye detection classifier proposed by Viola and Jones [24] is available through the Computer Vision System Toolbox in MATLAB R2012b\*. This method uses weak classifiers and Haar features to encode details and uses a decision stump to determine the eye region. If an eye was detected, the final location used for accuracy was deemed to be the middle of the detected region. If no eye was detected, the location was set to [0,0]. Finally, the state of the art eye detection algorithm, proposed by Valenti et al [19] and published in PAMI in 2012, was obtained through the authors website† and tested using all optimized parameters listed in the publication. Finally, a commercial algorithm, namely G8‡ is used. G8 is a black box algorithm that takes as input a raw image and outputs both left and right eye locations. All methods tested were performed assuming that the face region of each subject was known, except in the commercial algorithm where the option was not given. The normalized errors for 1150nm, 1350nm, and 1550nm can be seen in Figures 4.20, 4.21, 4.22, 4.23, 4.24, and 4.25.

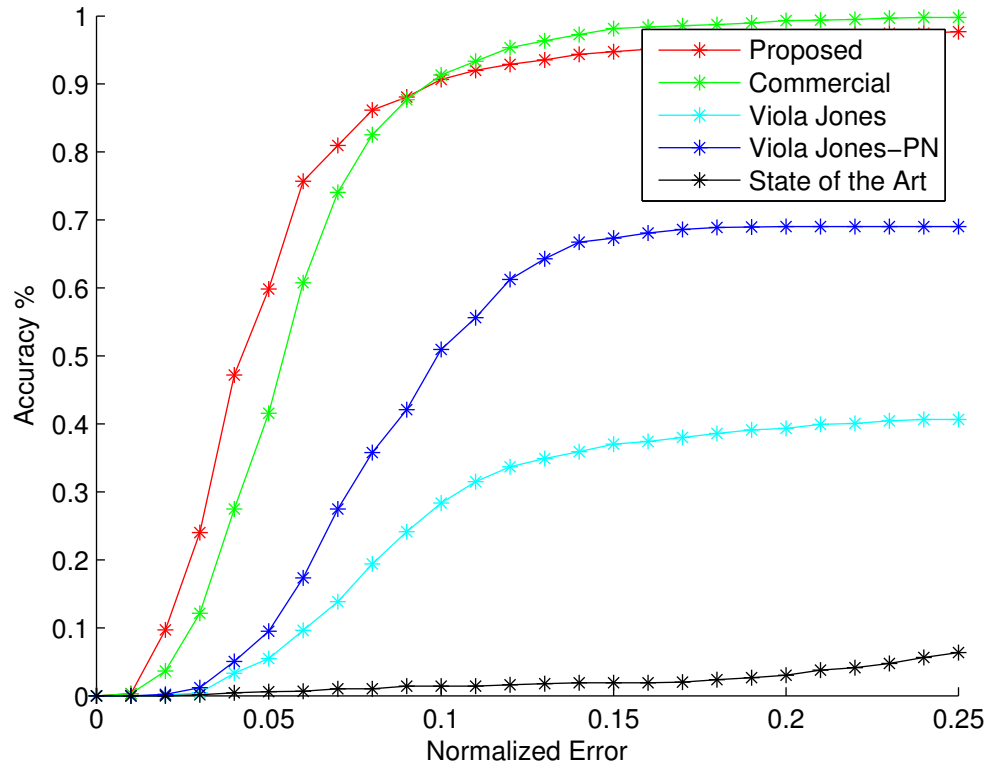
Notice that only in 1150nm does any of the algorithms outperform the proposed method, however, only at an error of  $e \geq 0.10$ . In order to have a fair comparison between the proposed algorithm and the Viola and Jones algorithm, retraining the classifier with SWIR images was necessary. Therefore, OpenCV's Cascaded Adaboost training algorithm, using

---

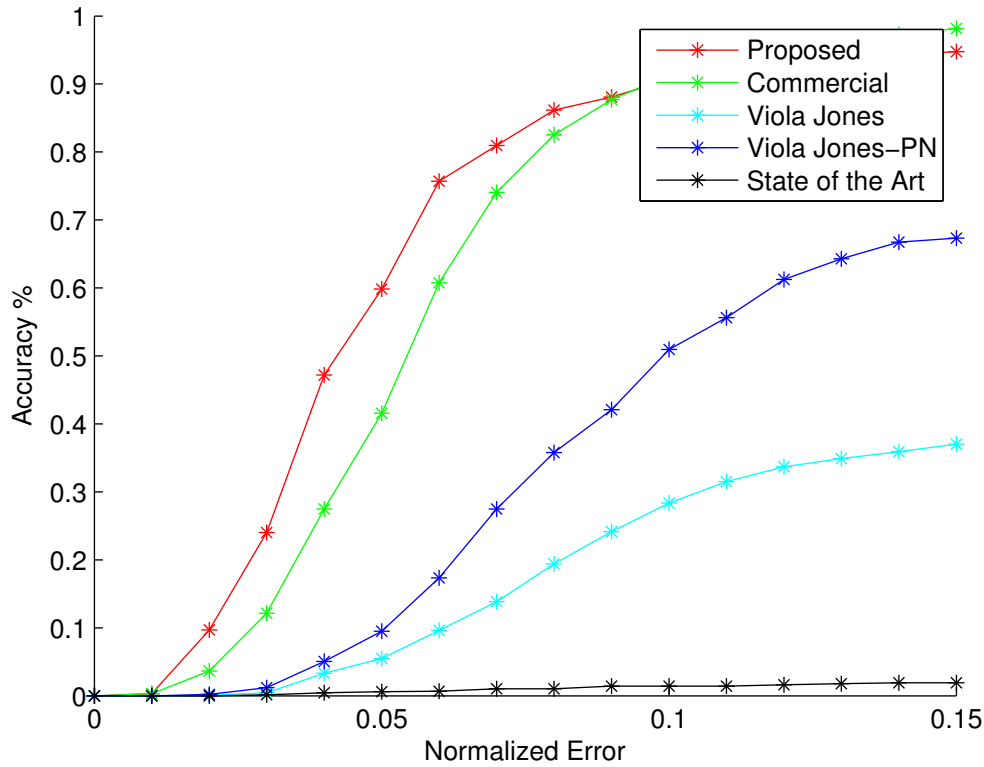
\*<http://www.mathworks.com/help/vision/ref/vision.cascadeobjectdetectorclass.html>

†<http://staff.science.uva.nl/~rvalenti/index.php?content=EyeAPI>

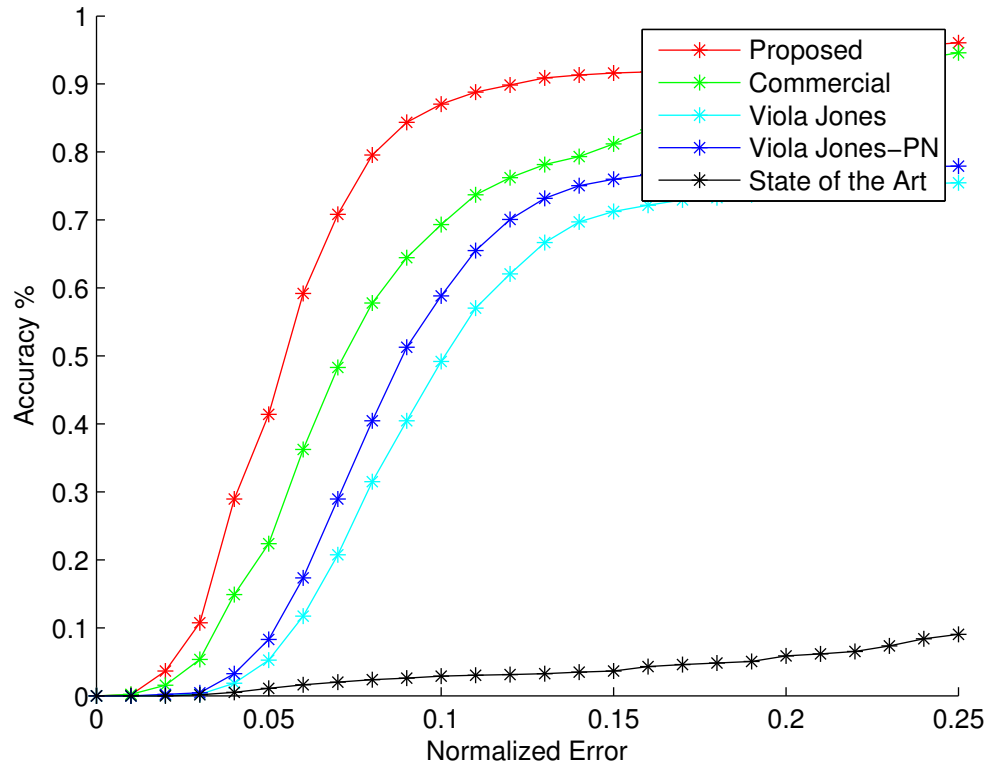
‡Provided by L1 Industries



**Figure 4.20:** Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (blue), and the state of the art (black) for 1150nm assuming the faces are known up to  $e \leq 0.25$ .

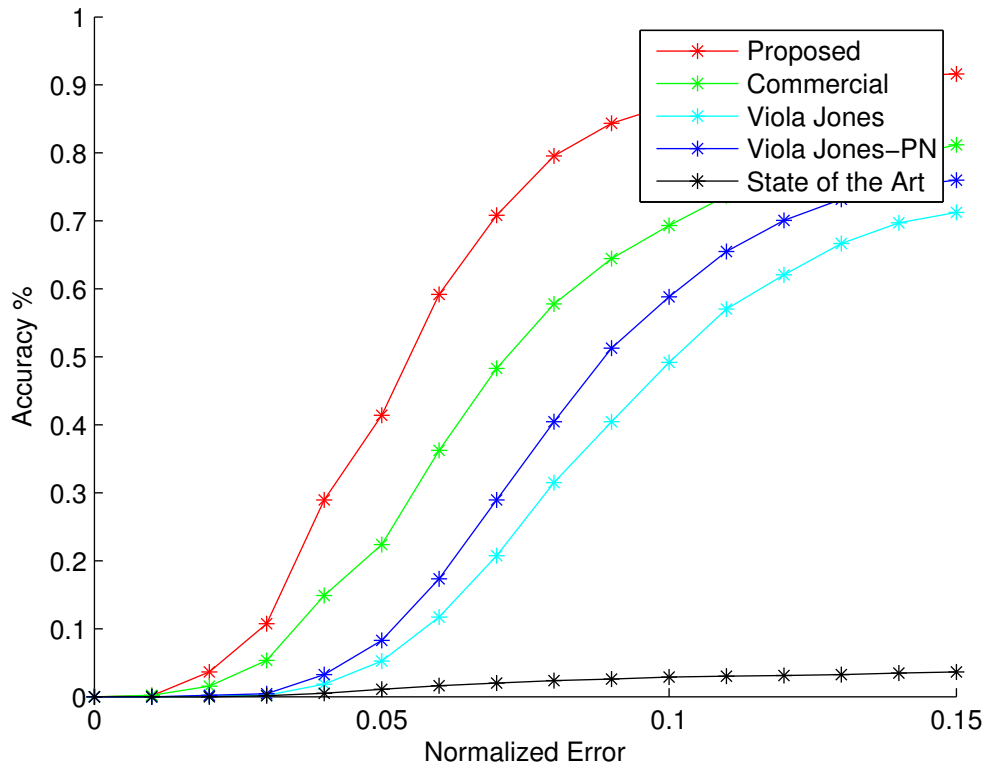


**Figure 4.21:** Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (blue), and the state of the art (black) for  $1150nm$  assuming the faces are known up to  $e \leq 0.10$ .

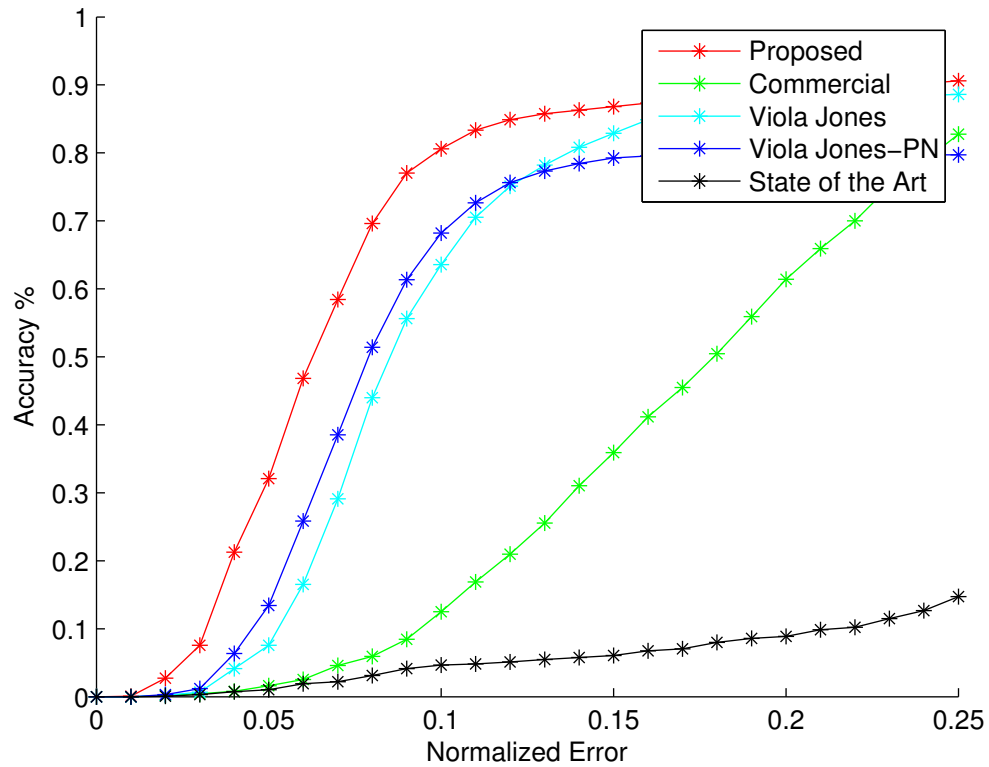


**Figure 4.22:** Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (blue), and the state of the art (black) for 1350nm assuming the faces are known up to  $e \leq 0.25$ .

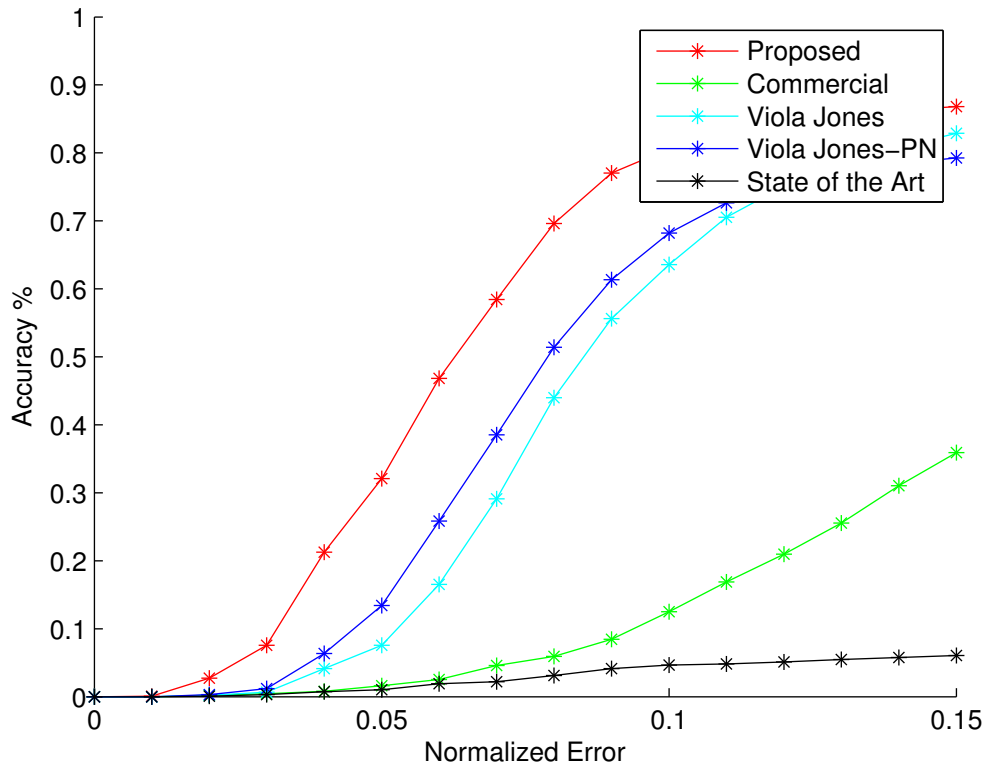




**Figure 4.23:** Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (cyan), Viola and Jones with photometric normalization (blue), and the state of the art (black) for  $1350nm$  assuming the faces are known up to  $e \leq 0.10$ .



**Figure 4.24:** Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (cyan), Viola and Jones with photometric normalization (blue), and the state of the art (black) for  $1550nm$  assuming the faces are known up to  $e \leq 0.25$ .



**Figure 4.25:** Our proposed method (red) compared to multiple available eye detection algorithms, including a commercial software (green), Viola and Jones (cyan), Viola and Jones with photometric normalization (blue), and the state of the art (black) for  $1550nm$  assuming the faces are known up to  $e \leq 0.10$ .

default Haar feature parameters, was used to build new classifiers using strictly 1550nm SWIR images as well as 1550 SWIR images photometrically normalized with the TBSSR approach (Viola and Jones - PN). 1,425 positive training samples (for both right and left eyes) and 13,590 negative samples, including 2,990 “hard” negative samples, were used for the training process. To further increase the fairness of the comparison, the tested faces were broken into four quadrants and the left eye classifier was only applied to the top left quadrant while the right eye classifier was applied to the top right quadrant, as is done in the proposed methodology. If multiple eye candidates were found, the one candidate that was closest to the ground truth was kept as the final eye location. Our method, however, empirically determines where the face is and produces a single final location while the Viola and Jones operator assumes the location of the face and can output multiple locations. It can be seen that even after retraining the classifiers, the proposed method still outperforms the Viola and Jones in all wavelengths. The authors recognize that to obtain a better understanding of how the Viola and Jones algorithm performs in lower wavelengths, specific classifiers should be made for the comparison, however, lack of data and time restraints made producing these results challenging.

The performance of the state of the art eye detection algorithm obtained from the authors website performs extremely poorly in comparison to the proposed method. With  $e \leq 0.25$ , the performance rate for 1150nm, 1350nm, and 1550nm is 6.37%, 9.04% and 14.74% respectively when given the face region only. After empirically testing the state of the art software, higher accuracies were achieved when limiting the search space to only the eye regions, however these results were not reported due to the authors publication and software in which it was stated that only the face region is necessary.

The proposed approach makes no assumptions as to where the eye regions are, and empirically determines where in the face each eye region is. This allows for a higher accuracy when only given a face region. Therefore, when decreasing the normalized error to  $e \leq 0.09$  (or less than the estimated width of the iris), the proposed algorithm outperforms all tested algorithms by more than 0.67%, 22.44%, and 73.63% for 1150nm, 1350nm, and 1550nm respectively.

### **4.3.7 Face Recognition Study**

The main purpose of automatic eye detection is to assist in a FR's systems ability to geometrically normalize the face. If the locations of the eyes are found incorrectly, geometric normalization will not be performed correctly (i.e. lining up the eyes in the same location for all subjects) and consequently facial recognition performance will suffer. This is why correct and accurate eye locations are needed. In order to test the accuracy of the proposed method, three different face recognition algorithms were used, i.e. Local Binary Patterns (LBP), Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). In these experiments, 1 sample from every subject (gallery image) was geometrically normalized using the manually annotated positions. This guaranteed that the gallery had one geometrically correct aligned face image. Then, this image was matched to 1 sample from each subject (probe image) that was geometrically normalized using the found eye locations from each respective method (i.e. manual annotation, proposed method, commercial, state of the art, and Viola and Jones). For both the LDA and PCA algorithms, 30 subjects with three samples per subject were chosen at random from the manually annotated geometrically aligned face images to perform training. Then the remaining 105 subjects were used for testing in

LDA, PCA, and LBP. In order to obtain a fair comparison, the subjects used for training the LDA and PCA were left out during the testing phase of LBP, even though LBP requires no training. The Receiver Operating Characteristic (ROC) curve and the Cumulative Match Characteristic (CMC) curve are used to describe the performance of the methods. Figures 4.26, 4.27, 4.28, 4.29, 4.30, and 4.31 shows the performance of each method using the LDA, PCA, and LBP algorithms in the  $1550nm$  databases (the most challenging database). Note that in all cases (LDA, PCA, and LBP), the proposed method produces better ROC curves and higher Rank-1 identification rates over all the other tested eye detection algorithms, except for the LBP algorithm where the proposed method has almost identical rates to the Viola and Jones method.

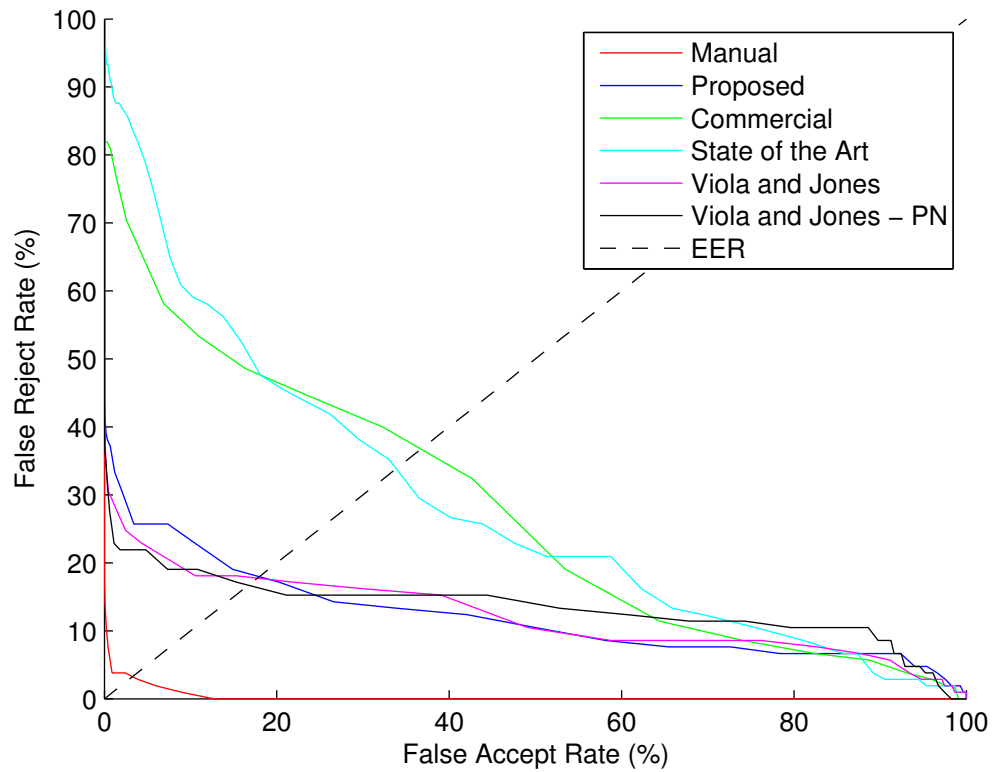
## 4.4 Conclusion

In this chapter, a novel unified eye detection method that can operate in multiple bands (i.e.  $1150nm$  to  $1550nm$ ) using normalized correlation coefficients and summation range filters has been proposed. After determining the location of the face, the eye regions are empirically found and summation range filters are used to accurately determine the center of the eye. The use of these simple filters yields low computational time (for 1,350 images,  $\mu = 75.7ms/image$  and  $std = 7.1ms/image$  using MATLAB R2012b) while allowing for robustness to multiple forms of image degradations, all while still achieving high face recognition rates.

An extensive evaluation of the proposed method was performed, testing it for accurate eye locations across multiple SWIR bands and for robustness to scale, image compression, and image blurring. The comparison with the commercial and academic algorithms and the state of the art suggests that under the general assumptions (i.e. the location of the face is known),

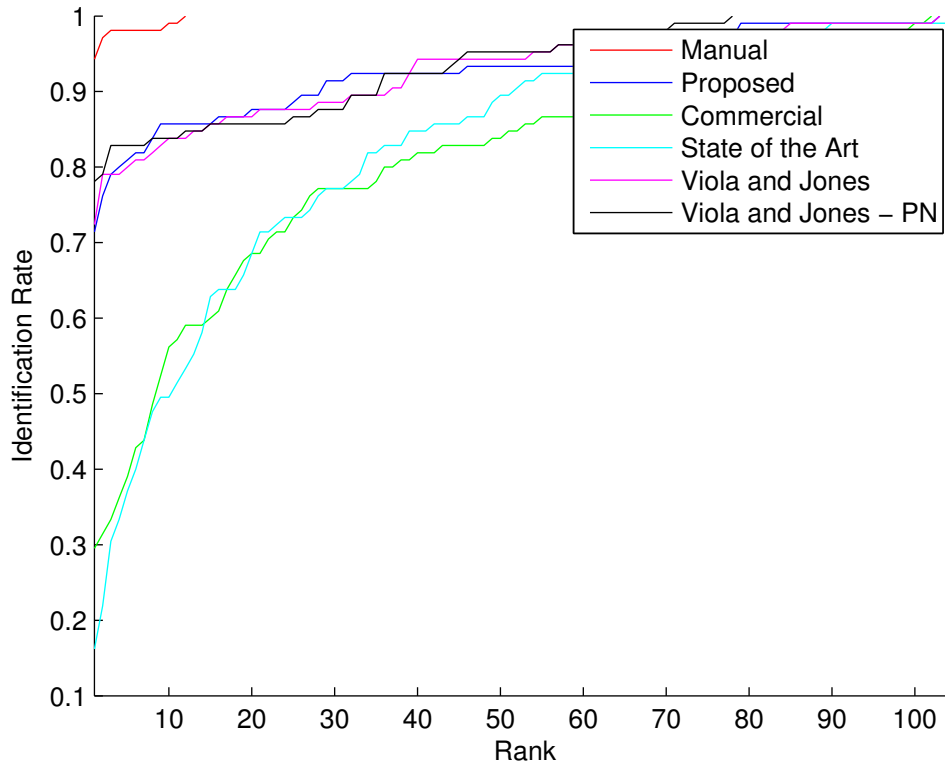
the proposed eye detection method is able to achieve higher accuracy and can be successfully applied to multiple SWIR bands. With  $e \leq 0.9$ , the proposed method outperforms every other method. Also, using the locations found through the method obtained higher face recognition results than almost all other methods across the spectra when geometrically normalizing the face. Given the reported accuracy of the overall system, it is believed that the proposed method provides a reliable way to not only localize the face region in SWIR images, but to accurately determine the correct location of the eyes.

As with any system, the proposed approach has drawbacks that can cause failed eye localization. If a person is wearing eye glasses with a large reflection from the illuminating source, this can cause an uncharacteristically high value within the summation range filter, causing the system to believe the eye location is somewhere on the rim of the glasses. In the described experiments, however, only 1 subject (or 10 samples per wavelength) was wearing eye glasses. Another drawback of the proposed system is dealing with uncooperative subjects who heavily squint or have their eyes totally closed. This causes very low range amounts in the range map. Consequently, the proposed approach has a tendency to believe that the eye corners or the eye brows is the correct location. In this system, this subsample equates to  $\sim 4\%$  of the database.

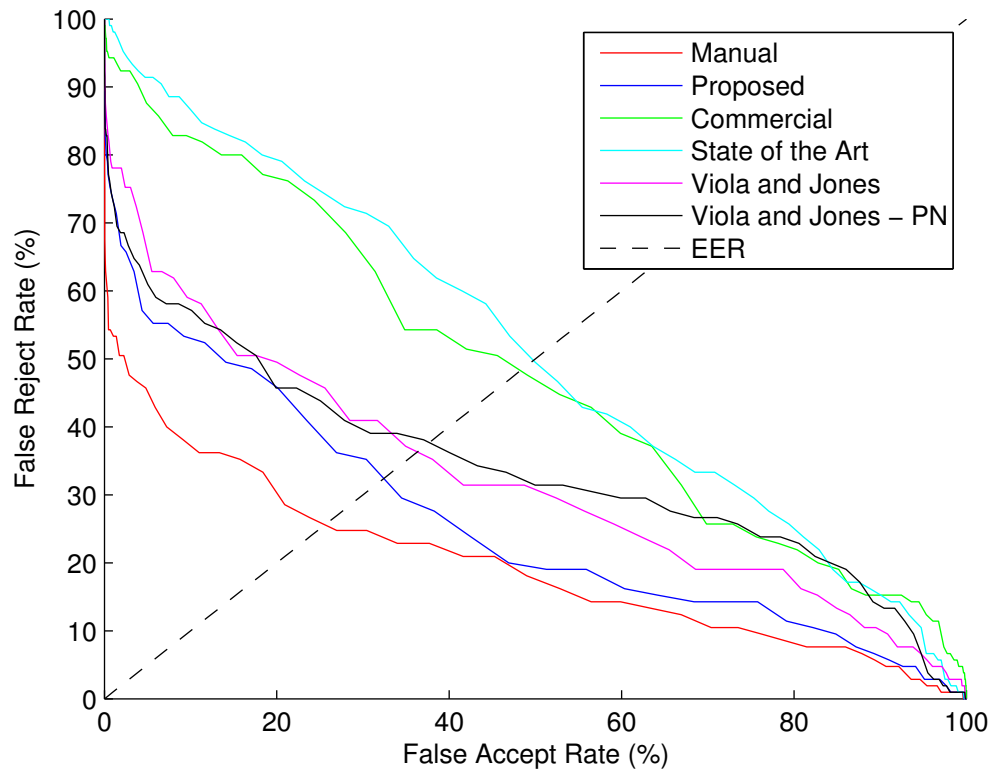


**Figure 4.26:** ROC curve showing the results of the LBP matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red).

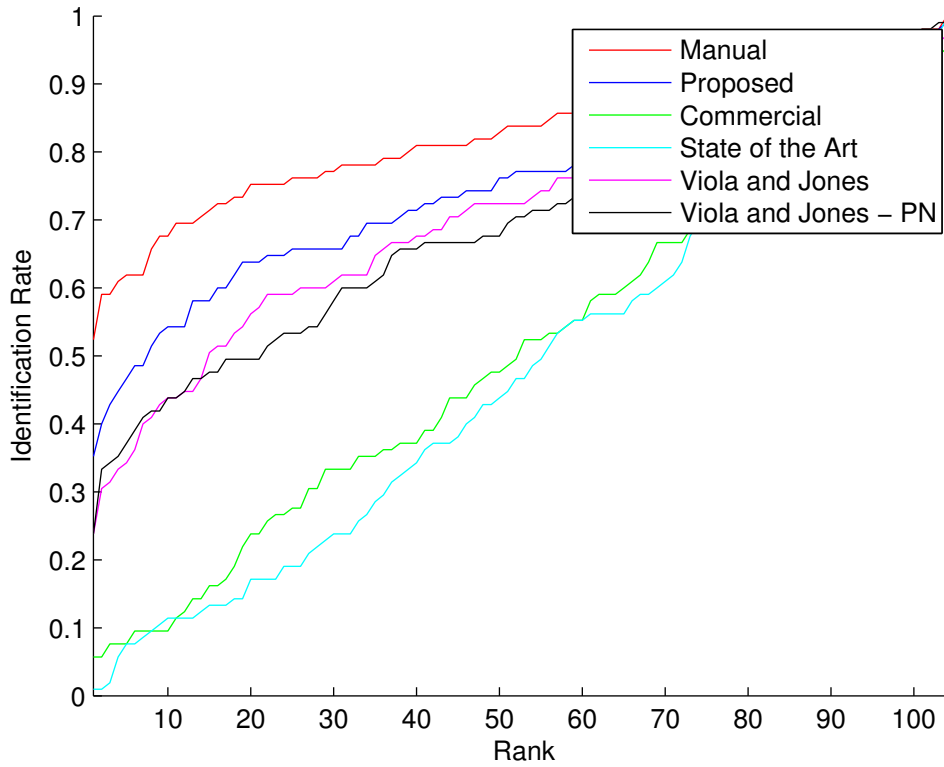




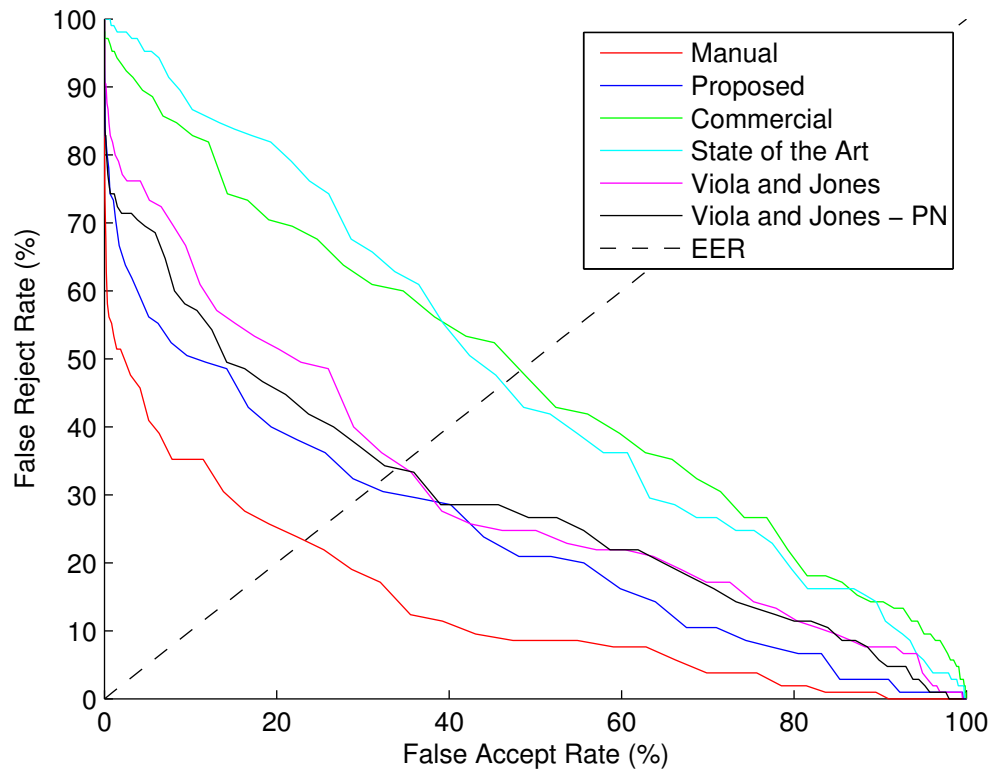
**Figure 4.27:** CMC curve showing the results of the LBP matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red).



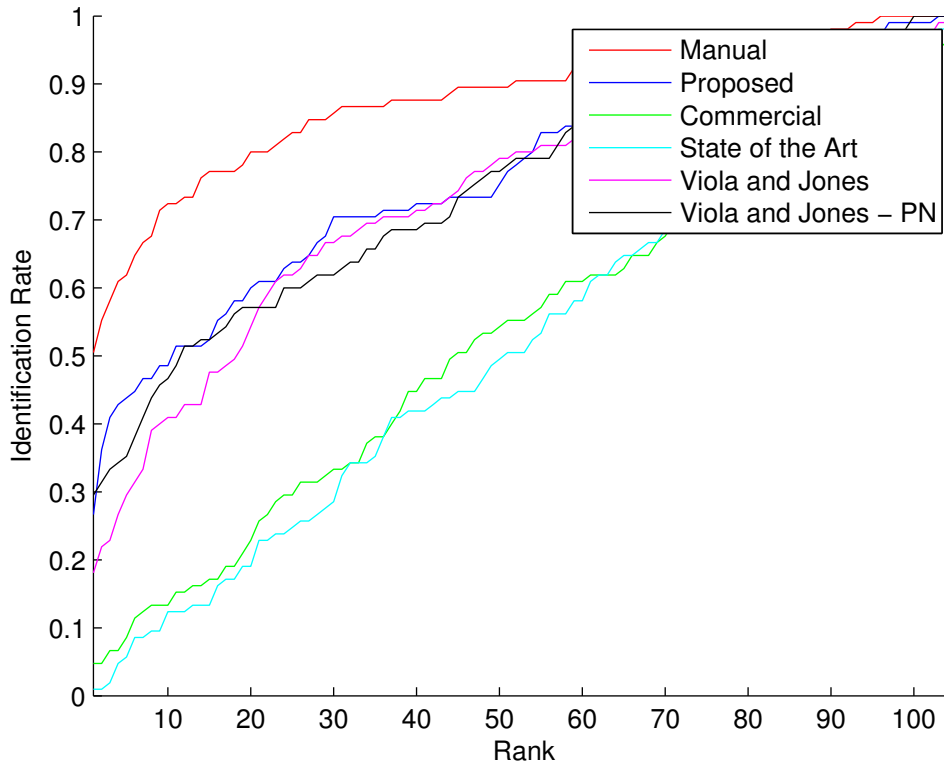
**Figure 4.28:** ROC curve showing the results of the PCA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red).



**Figure 4.29:** CMC curve showing the results of the PCA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red).



**Figure 4.30:** ROC curve showing the results of the LDA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red).



**Figure 4.31:** CMC curve showing the results of the LDA matcher when geometrically normalizing the face images with the proposed method (blue) with the other compared methods as well as the manually annotated ground truth (red).





## Chapter 5

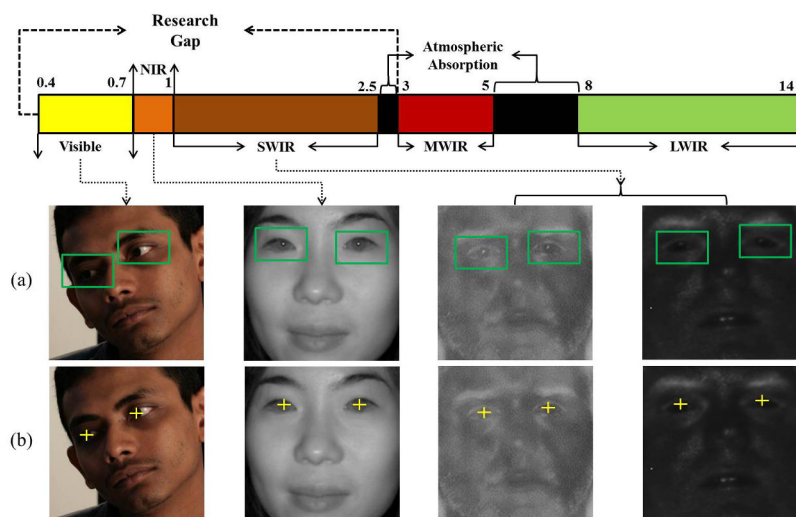
# Unconstrained Eye Detection across the Electromagnetic Spectrum

### 5.1 Introduction

As described in the previous chapter, face recognition systems are trending towards the mid and upper wavelengths of the electromagnetic (EM) spectrum due to the covert operational benefits that these bands provide. With eye detection being an integral part in any automated face recognition system, it is important to have an eye detection method that is cohesive and robust to all of the diverse attributes seen in each band of the EM spectrum. One of the limitations to the approach that is described in the previous chapter pertains to the fact that it was designed and developed to work on constrained (i.e frontal faces with ideal expression and illumination) data in the SWIR band.

As ideal full frontal face recognition has seen great advances in recognition performance, recent FR trends have been leaning towards striving to obtain high recognition rates in the





**Figure 5.1:** The electromagnetic spectrum depicting the active (NIR, SWIR) and passive (MWIR, LWIR) infrared bands. In this chapter, a scenario adaptable pupil detection approach that can successfully localize eye regions (a), and eye centers (pupils) (b) when using challenging face images captured in visible and different active IR bands, including Near Infrared (NIR) and Short Waved Infrared (SWIR) is proposed.

unconstrained (i.e. pose, illumination, and expression variant face imagery) domain [85, 86]. Furthermore, it is operationally infeasible to expect all face imagery captured to be full frontal under ideal illumination conditions. Due to the operational benefits of performing FR in the NIR and SWIR bands, it is important to have an eye detection algorithm that not only performs well across all EM spectra but also is robust to unconstrained face imagery.

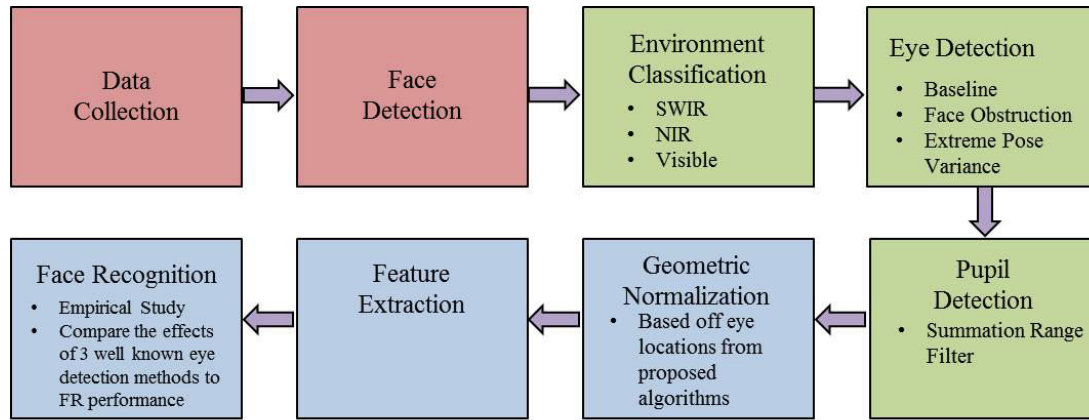
### 5.1.1 Contributions

In this chapter, a novel, scenario adaptable eye detection approach that efficiently locates human eye centers on face images captured using visible and infrared sensors when operating under challenging conditions, including the usage of (i) long range and night time face images, (ii) face images affected by partial face obstruction (eye glasses), (iii) face images affected by

pose and illumination variation and (iv) active IR illumination (Figure 5.1) is proposed. In order to mitigate all challenges pertaining to the unconstrained eye detection problem, a set of efficient eye detection approaches is designed and developed. The specific contributions are:

- Usage of challenging face datasets (multispectral, short and long range, faces behind tinted glass, etc.)
- Situation classification (determining under which scenario any raw input image was captured)
- Eye region localization on any face image of the challenging datasets tested
- Pupil detection using summation range filters for accurate eye center localization and geometric normalization
- Eye glasses classifier (on multispectral facial images, where the subjects are wearing or not wearing eye glasses)

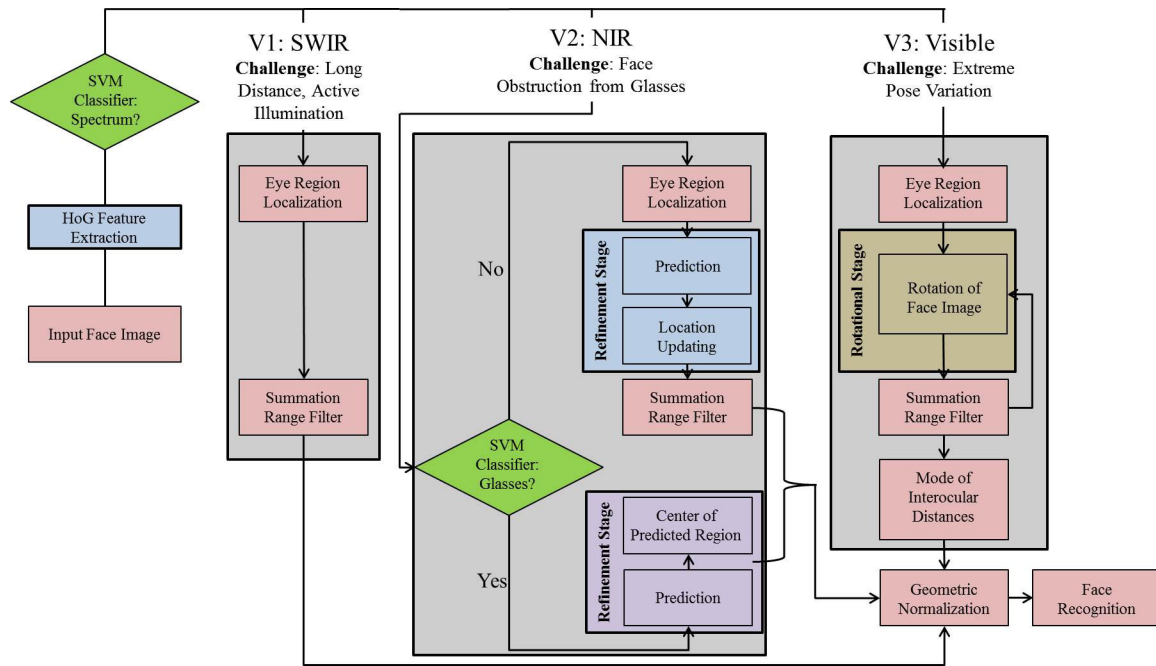
Our algorithmic approach developed is tested on the following challenging face datasets: (i) SWIR face images captured at night time using active IR illumination and at long distances, (ii) NIR face images with partial face occlusion, and, finally, (iii) visible face images when captured under extreme pose and illumination variation conditions. The experimental results demonstrate that the proposed situation classification model is highly accurate, while also, the proposed eye detection approach outperforms commercial and academic, state of the art, eye detection algorithms, including the Valenti and Gevers TPAMI eye detection algorithm [19], the Viola & Jones Adaboost approach [24] trained under multiple conditions,



**Figure 5.2:** Overview of a typical face recognition system. The main contributions of this paper (green) include spectrum classification, eye detection, and eye center localization. Multiple studies were conducted to show the effect of the eye detection and center localization scheme (blue). The data collection protocol as well as the face detection algorithm used for the purpose of this work (red) will also be discussed.

and the G8 commercial algorithm provided by L1 Systems. Furthermore, via a set of experiments, it is shown that the eye locations obtained from the approach result in higher face recognition scores across all eye detection algorithms tested. The experimental results demonstrate the robustness and accuracy of the proposed methodology when operating under difficult environments and multiple spectra.

Figure 5.2 shows the main contributions in this work (green) in regards to each step of a typical face recognition process and the studies that were conducted (blue). In this chapter, the outside databases considered, and the face detection algorithm used before the pupil detection approach is employed will be discussed.



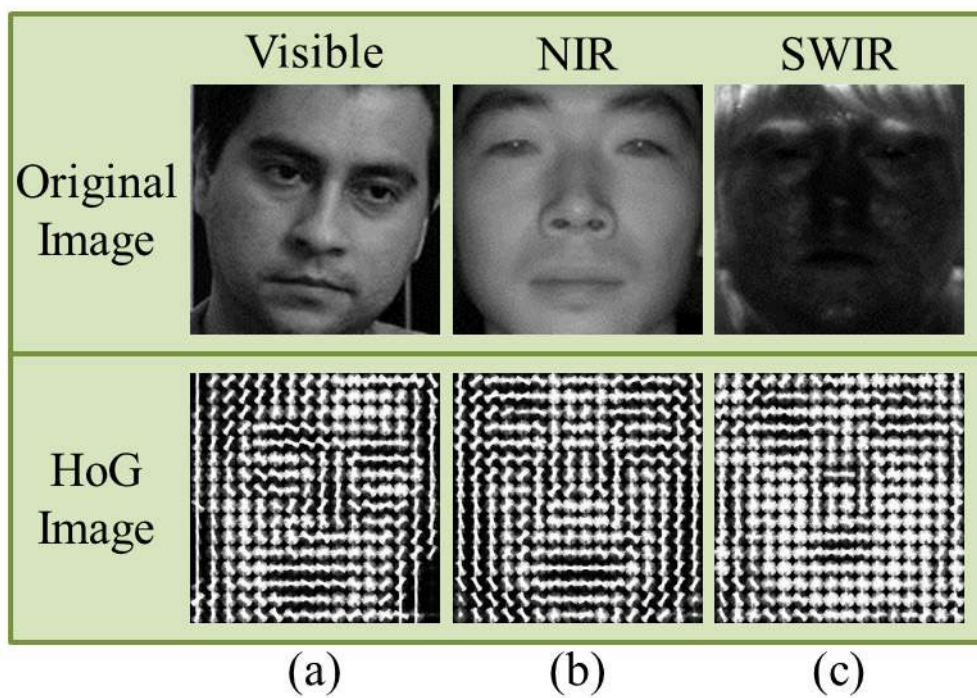
**Figure 5.3:** Overview of the methodological approach described in this paper. First, a classification model is built to determine which spectrum (i.e. Visible, NIR, or SWIR) a face image is subject to. Then, the eye detection approaches are employed based on the model's decision. V1 describes the baseline approach used in challenging SWIR scenarios (i.e. active illumination and long distance). V2 describes the approach used in another challenging scenario (i.e. facial obstruction) in the NIR band and V3 describes the approach tested on extreme pose and illumination variation in the visible band.

## 5.2 Methodology

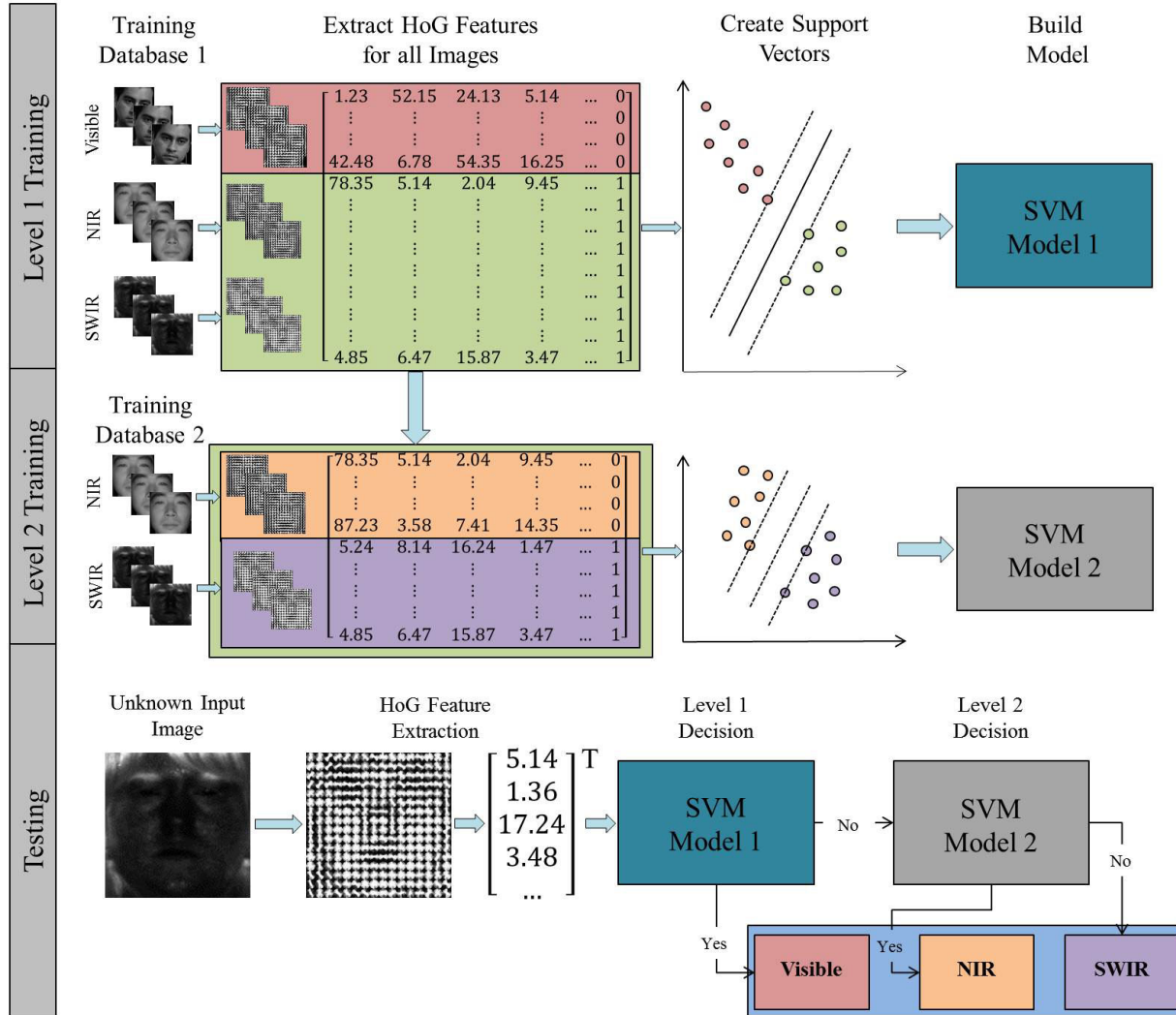
In this section, the techniques used to perform multi-wavelength eye detection in difficult environments, including extreme pose variation, facial obstructions, use of active illumination to capture face images behind tinted glass, and face images captured at long distances (i.e.  $106m$  or  $\sim 350ft$ ) are outlined. The overview of the methodological approaches can be seen in Figure 5.3 and is broken down into three main challenging scenarios, i.e. the baseline approach applied on SWIR face datasets and the adaptations made for more challenging scenarios (V2 and V3). The baseline method is presented in Section 3.2, while adaptations to this method for more challenging scenarios are considered in Sections 3.3 and 3.4 respectively.

### 5.2.1 Spectrum Classification

In this work, there are three different algorithms that compose the main eye detection approach. Each algorithm is band-specific, i.e. it is tailored to operate for each specific band. Thus, it is operationally important to have a spectrum classification algorithm that automatically determines the spectrum of each input image and, then, automatically launches the appropriate band-specific eye detection algorithm. This capability simplifies also the work (in terms of both time and operational efficiency) of the biometric operator in the loop. For this purpose, an algorithm that used Histogram of Oriented Gradients (HoG) features [79] in a multi-layer Support Vector Machine (SVM) to create two separate classification models is designed and developed. The models are used in the training as well as testing phase (see Figure 5.5). The first model is used to classify between visible and non visible (NIR and SWIR) face images, while the second model classifies NIR from SWIR face images. HoG



**Figure 5.4:** Sample HoG images for the three spectral scenarios: a) Visible b) Near Infrared and c) Short Waved Infrared.



**Figure 5.5:** Overview of the spectrum classification algorithm. HoG descriptors are extracted for all images in the three different spectra. Then, two different SVM models are built, i.e. visible or NIR/SWIR (model 1) and NIR or SWIR (model 2). When testing, HoG features are extracted from a query image and sent into SVM model 1 to determine if it is visible. If not, it is passed into SVM model 2 to determine if it is NIR or SWIR.

features, also known as a SIFT feature representation, is a compact vector representation of an image patch based on the orientation, magnitude, and spatial vicinity of the image gradients. After computing the gradient image from the original intensity image weighed by a Gaussian kernel using  $\alpha = 1/2$ , the HoG feature vector is calculated as follows. First, for a  $s \times s$  patch size (in this case  $s = 16$ ), the spatial coordinates are quantized into  $m \times n$  values, usually in which  $m = n$  (this case  $m, n = 2$ ). Then, the values of the gradient image pixel, ranging between  $[0, \pi)$ , are quantized into one of  $k$  orientations (in this case  $k = 9$ ). At each of the  $m \times n$  coordinates, the sum of the Gaussian weighted gradient magnitude values is computed. This yields a  $m \times n \times k$  feature descriptor for a single patch. This descriptor is initially normalized to unit length, although a second normalization is commonplace wherein the locations at which the feature vector is larger than 0.2 is suppressed to 0.2 and renormalized to unit length. Therefore, using face images of  $128 \times 128$ , with an overlap of 8 pixels per patch, the final feature vector becomes 9,216 dimensional. Sample HoG feature for the three different scenarios can be seen in Figure 5.4.

Once the HoG descriptors are obtained from the training set for each of the three spectra, they can be used in a multi-layer Support Vector Machine (SVM) scheme. First, the top layer SVM model was built to determine if a face image belonged to the visible spectrum or not. Using the HoG features, the SVM is trained with both positive (Visible) images and negative (NIR and SWIR) images using the Sequential Minimal Optimization algorithm that identifies the support vectors  $s_i$ , weights  $\alpha_i$  and bias  $b$ . Next, the second layer SVM model is built using positive (NIR) images and negative (SWIR) images in the same fashion. After the two models are built, a query image can be classified as either visible or not by



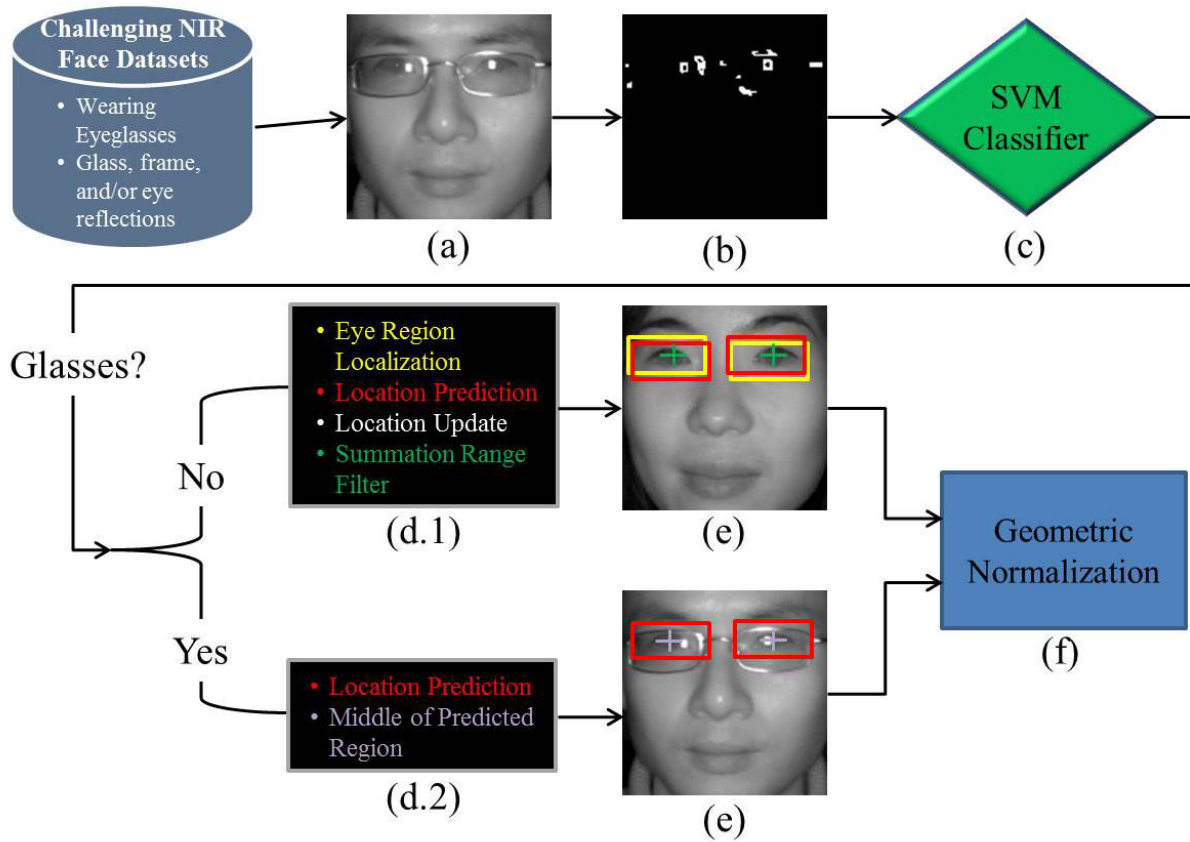
first calculating the HoG feature vector by using the following equation:

$$c = \sum_i \alpha_i k(s_i, x) + b_i \quad (5.1)$$

where  $s_i$ ,  $\alpha_i$ ,  $b_i$  are described above and  $k$  is the kernel function. Because this is a linear kernel,  $k$  can be computed by using the dot product. If  $c \geq 0$ , the sample is deemed to be a visible image. If  $c < 0$ , the query image is fed into the second layer SVM model in the same manner as above. Then, if  $c \geq 0$  in the second layer, the sample is classed as an NIR image. If not, it is classed as a SWIR image. An overview of this process can be seen in Figure 5.5. In order to assess the measure of accuracy, the bootstrap (with replacement) method was used tenfold: 25% training (i.e. 47 SWIR images, 751 NIR images, and 401 Visible images) and 75% (i.e. 142 SWIR images, 2,251 NIR images, and 1,201 Visible images) testing, only 2 false positives and 1 false negative were observed in the top layer of the classifier. The second layer model showed no errors when tested. This shows that the HoG features used are extremely efficient in classifying images from different spectra. This is key to be able to determine which version of the proposed eye detection algorithms is to be used.

### 5.2.2 Active Illumination Behind Glass and Long Distance SWIR Face Datasets

Although these datasets provide unique challenges on their own, such as being imaged in SWIR at long distances (106m) and behind tinted glass (80% light reduction), the baseline eye detection algorithm, described in Chapter 4.2 proved to be an efficient means of quickly and accurately determining the locations of the pupils. No augments to the algorithm were

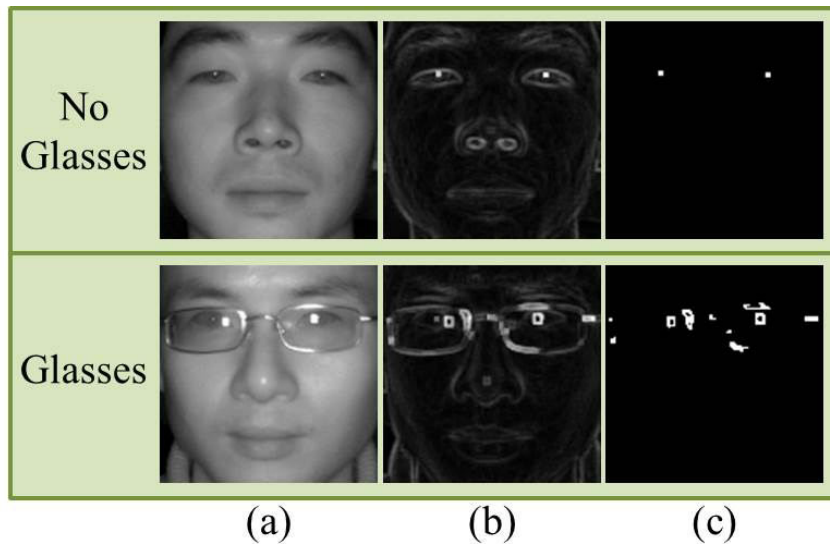


**Figure 5.6:** Flowchart of the proposed eye detection in images with facial obstructions, in this case eye glasses. a) original image b) feature extraction c) SVM classifier d) eye detection/location prediction, e) pupil detection and f) geometric normalization.

necessary for these datasets.

### 5.2.3 IR Face Datasets with and without Eye Glasses

Another challenging face dataset that was used in our study was the NIR face dataset where the subjects were not asked to remove their eye glasses (as was the case in the SWIR datasets). In order to accurately detect the eyes in such a challenging dataset, two additional steps to the aforementioned approach are implemented, namely eye glass classification and



**Figure 5.7:** Examples of images with and without eye glasses. a) Original Image b) Summation Range Image and c) features used for SVM. Notice that the image with the glasses produce higher reflectance pixels therefore increasing the number of SVM features.

eye location prediction.

### Eye Glass Classification:

In order to take the appropriate eye detection measures in the NIR database, it is important to know which face images have eye glasses and which don't. To do this, a Support Vector Machine (SVM) classifier to categorize each face image in the database as either an image with eye glasses or without was designed. As can be seen in Figure 5.7(a) and Figure 5.7(b), images with eye glasses tend to have more reflectance points than images without eye glasses. To determine the points in the image with high reflectance, the summation range filter (described in Chapter 4) was used. Using the assumption that images with glasses obtain more reflectance points above a certain threshold than images without glasses, the

SVM feature can be described as the following:

$$SVM_{feat} = \sum_{x=1}^a \sum_{y=1}^b S(x, y) > \lambda \quad (5.2)$$

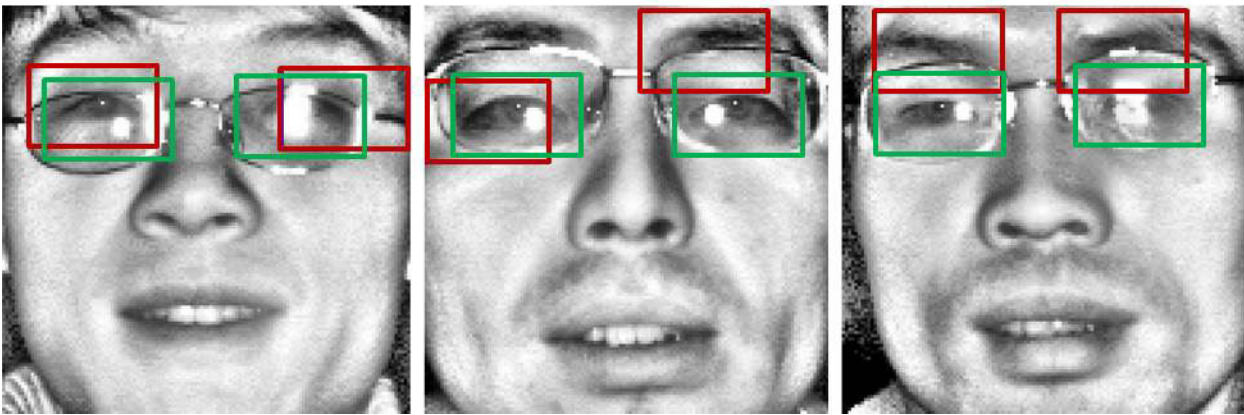
where  $S(x, y)$  is the summation range image of the entire face,  $\lambda$  is a threshold empirically determined before hand (in this case  $\lambda = 100$ ), and  $a$  and  $b$  are the width and height of the query image, respectively. These points can be seen in Figure 5.7(c). Using these features, the SVM is trained with both positive (glasses) images along with negative (no glasses) images in the same manner as described above Section 3.1. Therefore, if  $c \geq 0$ , the sample is deemed to be an image with glasses.

### Eye Location Prediction:

When having to deal with facial obstructions (i.e. glasses with reflectance points), the 2D normalized cross correlation (Equation 4.1) has a difficult time determining with a high degree of confidence where the eye region is, as can be seen in Figure 5.8.

In order to deal with this problem, Kalman filtering is used to predict the location of the eye region instead of empirically determining it using Equation 4.1. Because the face region is assumed, eye locations between subjects (frames) are expected to be in similar positions, however not in the exact same location. Therefore, the motion of the eye region from frame to frame can be described by both its position and its velocity. Let  $(x_t, y_t)$  represent the upper left hand coordinate of the initial eye region at time  $t$  and let  $(u_t, v_t)$  represent its velocity at time  $t$  in both  $x$  and  $y$  directions. Therefore,  $\delta_t = (x_t, y_t, u_t, v_t)^T$  can represent

— Prediction  
— No Prediction



**Figure 5.8:** Example images with eye glasses in which the original algorithm managed to either not find accurately enough, or not find the eye region at all (see red boxes). Notice how the eyebrows and large areas of reflectance causes the eye regions to be inaccurate when only using the baseline approach. Then, see how the eye localization prediction approach manages to correctly find the eye regions (see green boxes).

the state vector at time  $t$ . Using this, the system can then be modeled as:

$$\delta_{t+1} = \phi \cdot \delta_t + w_t \quad (5.3)$$

where  $w_t$  represents system perturbations and  $\phi$  describes how the system moves from one state to the next. We can further assume that a feature extractor estimates  $z_t = (\hat{x}_t, \hat{y}_t)$  where  $z_t$  is the estimated location of the upper left hand corner of the eye region at time  $t$ . Using this, the eye localization prediction model can be described as:

$$z_t = H \cdot \delta_t + v_t \quad (5.4)$$

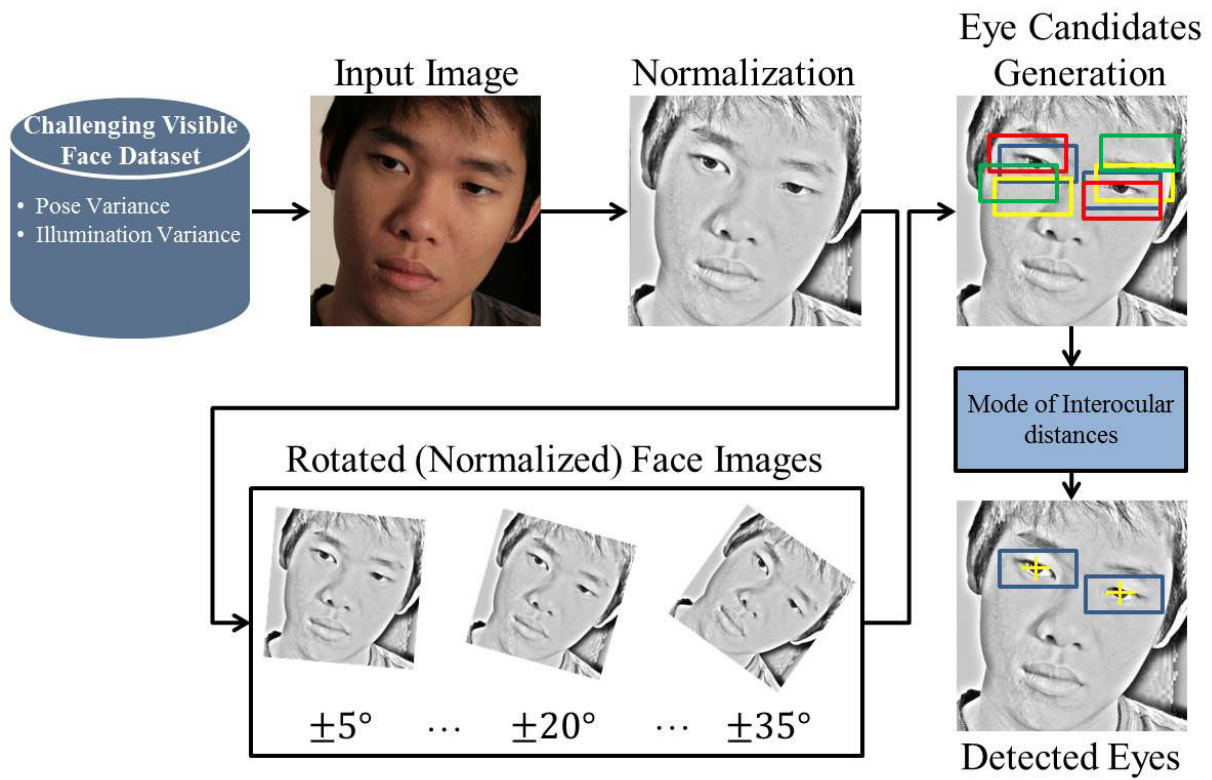
where  $v_t$  represents measurement uncertainty and  $H$  defines the mapping from the state vector  $\delta_t$  to the measurements vector  $z_t$ . More specifically, the location of the upper left hand corner of the eye region at current time (frame)  $t$  is estimated by thresholding a local neighborhood of predicted positions. Using some initial conditions as well as Equation 5.3 and Equation 5.4, the state vector  $\delta_{t+1}$ , along with covariance matrix  $\Sigma_{t+1}$ , can be updated using both the system model and the measurement model for location prediction and updating, respectively.

By using both the SVM classifier as well as the Kalman filter, performing eye detection on face images in the NIR with eye glasses can be described as follows and seen in Figure 5.6. Average eye templates are created for both the right and left eye as stated previously. Then the baseline eye localization algorithm described above is applied to the first image and then updates the eye region position. When subsequent images are processed, feature extraction is performed and the image is classified using the SVM model (Figure 5.6(b),(c))

and determined to either contain glasses or not. If the image does not have glasses, the eye localization is performed as normal. This location as well as the predicted location updates the model accordingly. Then the summation range filter is used on the found eye region to determine the final location of the subjects pupils (Figure 5.6(d.1),(e)). If the subject is wearing eye glasses, the eye localization scheme described above is not used and only the predicted eye region location is used as the found eye region (Figure 5.6(d.2),(e)). Note that when the image has glasses, the model is not updated with its location, essentially creating a Bayesian prior estimation scheme. This allows for a string of images with glasses to be predicted with no unfavorable updates to the filter. Because the images that have glasses typically have a much larger area of reflectance in them, as can be seen in Figure 5.7, the summation range filter fails to find the center of the pupil with much accuracy. Therefore, the location of the pupil is said to be the direct center of the predicted eye region. This scheme is performed on both the right and left eye regions to determine the locations of the left and right pupils, respectively. Finally, using these locations, the face images can be geometrically normalized (Figure 5.6(f)) and be used in a face recognition algorithm.

#### **5.2.4 Visible Face Dataset with Extreme Pose and Illumination Variation**

Another difficult scenario that was considered was data in the visible spectrum that has extreme pose and illumination variations. Because the face images are rotated from right to left ( $\pm 35^\circ$ ), additional improvements to the original methodology (as described in Section 3.2) had to be used to accurately detect the eyes. As before, photometric normalization was used to minimize illumination effects and five different sets of average eye templates were



**Figure 5.9:** Overview of the methodology used to detect eyes under extreme pose and illumination variation in the visible dataset.



created for the database. After this, the following methodology can be used to find eyes when there is extreme pose and illumination variance.

### Rotation of Faces:

In previous methodologies, it was assumed that the face area was known and narrowed down the search space by breaking the face into four equal quadrants and searching the upper right and left quadrants for both the right and left eyes respectively. Because the images that are considered in this dataset have extreme pose positions, i.e. extreme pitch, yaw, and rotation, those assumptions cannot be made anymore. Examples of these faces can be seen in Figure 5.10(d). Therefore, in order to compensate for this, each image was rotated at varying angles, from  $-35^\circ$  to  $35^\circ$  in  $5^\circ$  increments. At each of these increments, Equation 4.1 was used to get the correlation coefficient for all 5 average eyes for both the right and left eye respectively. At each increment, the summation range filter (described in Chapter 4) was used to find the eye centers for all five average eye pairs. This created, in total, 75 candidate eye pairs for each image.

After empirically determining that the most accurate locations of the eyes were typically found multiple times within these 75 candidates, it was observed that the same interocular distance would be repeated. Therefore, the Euclidean distance  $E_i$  was calculated for all 75 candidates and rounded off to the nearest pixel with  $\pm 5px$  standard deviation. To determine the final locations of the eyes, the following equation can be used:

$$\Gamma_i(x, y) = \delta_i(mode(E_i)) \quad (5.5)$$

where  $\Gamma_i(x, y)$  is the correlation coefficients of the candidates that have the most similar

(mode) interocular distances. Then, in order to determine a final pair of eye locations, the following equation can be used:

$$\hat{\Gamma}(u, v) = \underset{i}{\operatorname{argmax}}(\Gamma_i(u, v)) \quad (5.6)$$

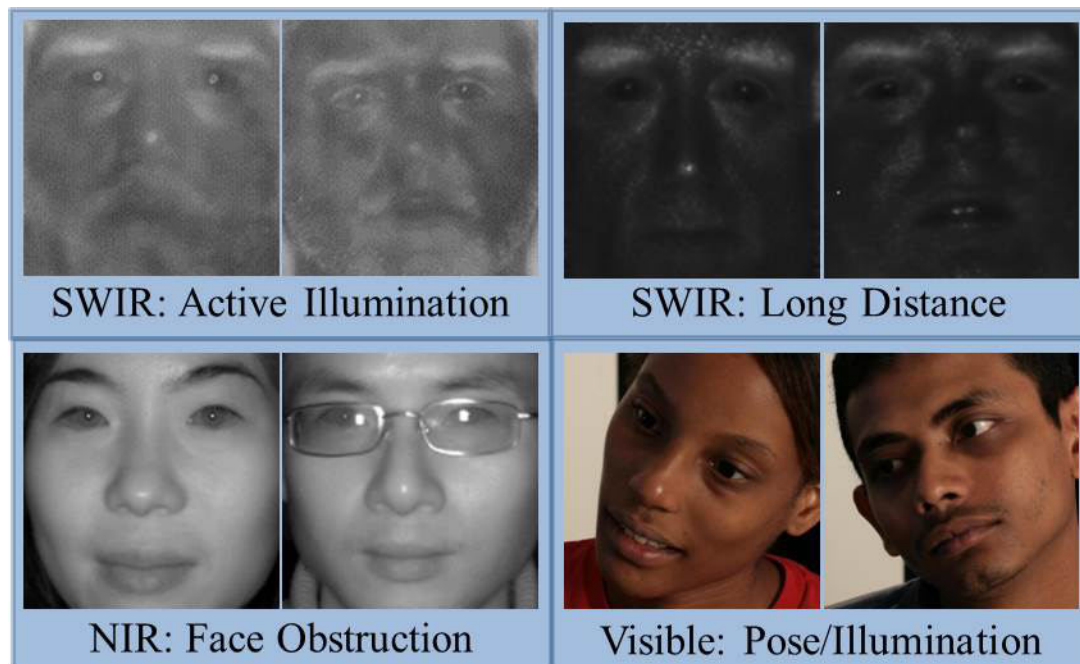
where  $\hat{\Gamma}(u, v)$  is the final location of the right and left eyes, respectively. This process can be summarized in Figure 5.9. Using the main methodology and its adaptations for more challenging scenarios, eye detection can be performed under multiple wavelengths as well as during difficult environments.

## 5.3 Evaluation

### 5.3.1 Databases

In this section, the approaches on multiple databases that represent the scenarios that have been outlined in the previous sections are tested. Additionally, the method is tested against other well-known eye detection methods, including the state-of-the-art by Valenti and Gevers [19], academic approaches (i.e. Viola & Jones Adaboost Classifier [24] retrained under various conditions), as well as a commercial algorithm (L1's G8). Finally, face recognition tests based on the found eye locations are performed to show the impact that the proposed algorithm has compared to others. To test the algorithms as well as others, the following databases were used.

- **SWIR: Active Illumination:** Because there are no publicly available datasets collected in the SWIR band behind glass with active illumination, the West Virginia



**Figure 5.10:** Sample face images, after face detection, of 2 randomly selected subjects for each of the four face databases used for the purpose of this study, i.e. Active Illumination in SWIR, Long Distance in SWIR, Face Obstruction in NIR, and Extreme pose and illumination variance in Visible

University Multispectral (WVUM) database, described in better detail in Chapter 3, was captured using a Goodrich SU640 camera. The SU640 is an Indium Gallium Arsenide (InGaAs) video camera featuring high-sensitivity and wide dynamic range. The model used has a 640x512 FPA with 25 $\mu$ m pixel pitch, and >99% pixel operability. The spectral sensitivity of the SU640 ranges uniformly from 700-1700nm wavelength. The response falls rapidly before 700nm and after 1700nm. For this work, subjects were captured behind automotive glass with 80% tint using 1550nm SWIR active illumination. For the database used, 140 subjects were collected under controlled conditions. Samples can be seen in Figure 5.10(a).

- **SWIR: Long Distance:** A proprietary optics and laser illumination system was created to covertly capture SWIR (i.e. 1550nm) images in nighttime at long distances. This system was used to capture images from 50 subjects at a standoff length of 106m and captured under semi-controlled conditions. Sample images can be seen in Figure 5.10(b).
- **Near IR: With/without Eyeglasses:** To demonstrate the effect that face obstructions has on the algorithm, a subsection of the CASIA Visible-NIR database was used [52]. This is a heterogeneous face database that contains both visible and NIR images. A subsection of 3002 NIR images from 202 subjects was used. They are manually aligned according to the eye coordinates, cropped, and resized to 128x128 pixels. In 9.2% of the images (or 275 total images), subjects were wearing eye glasses as can be seen in Figure 5.10(c).

- **Visible: Extreme Pose and Illumination Variance:** To demonstrate extreme pose and illumination variance, the University of Houston Database [80] (UHDB11) was used. This database consists of 1,602 face images from multiple subjects under variable pose and illumination conditions in the Visible spectrum. For each illumination condition, the subjects faced four different points inside the room simulating yaw (their face was rotated on the Y axis that is the vertical axis through the subjects' head). For each Y-axis rotation, three images were also acquired with rotations on the Z axis, simulating roll (that extends from the back of the head to the nose). Thus, the face images of the database were acquired under six illumination conditions, with four Y and three Z rotations. This caused extreme face poses in terms of pitch, yaw, and roll. Images were not labeled by their rotations or illuminations, therefore the level of illumination and rotation was unknown when the approach was applied. Samples can be seen in Figure 5.10(d).

In all cases, both the left and right eye were manually marked for their ground truth locations. All experiments were performed on a 64-bit Windows 7 machine with 12GB of RAM running Intel Core i7 CPU @ 3.2Ghz using MATLAB R2012b. Three different eye detection algorithms that are commercially or academically available were used for comparison to the proposed algorithm.

- A well-known commercial facial recognition algorithm (i.e. L1's Systems G8), whose SDK allows for eye location retrieval was used.
- The Cascaded Adaboost eye detection classifier proposed by Viola & Jones [24] is available through the Computer Vision System Toolbox in MATLAB R2012b\*. This

---

\*<http://www.mathworks.com/help/vision/ref/vision.cascadeobjectdetectorclass.html>

method uses weak classifiers and Haar features to encode details and uses a decision stump to determine the eye region. The original model (VJ-Visible), trained on visible images, was used as well as retrained models using both SWIR face images (VJ-SWIR) and SWIR face images after applying the TBSSR photometric normalization algorithm (VJ-SWIR-PN). Using all default parameters, new classifiers were built in order to more accurately test the NIR and SWIR databases. Due to time constraints, only 1550nm images were used for training. 1,425 positive training samples (for both the right and left eyes) and 13,590 negative samples, including 2,990 “hard” negative samples, were used for the training process. “Hard” negative samples include eye images that have extreme tilt from +45 deg to −45 deg in increments of 5 deg. After this trained on original SWIR images, the process was repeated after applying the TBSSR photometric normalization to all training samples. In order to further increase the fairness of the comparison, the tested faces were broken into four quadrants (except the case of the Visible database) and the left eye classifier was only applied to the top left quadrant while the right eye classifier was applied to the top right quadrant. Because the UHDB11 database has extreme pose variations, the eyes will not be guaranteed to be upper right and left quadrants. Therefore, the face image was broken only into left and right halves instead of quadrants and applied accordingly. If multiple eye candidates were found, the one candidate that was closest to the ground truth was kept as the final eye location.

- The state-of-the-art eye detection algorithm, proposed by Valenti & Gevers [19] and published in IEEE TPAMI in 2012, was obtained through the authors website<sup>†</sup> and

---

<sup>†</sup><http://staff.science.uva.nl/~rvalenti/index.php?content=EyeAPI>

tested using all optimized parameters listed in the publication.

All methods tested were performed assuming that the face region of each subject was known, except in the commercial algorithm where the option was not given.

### 5.3.2 Experimental Results

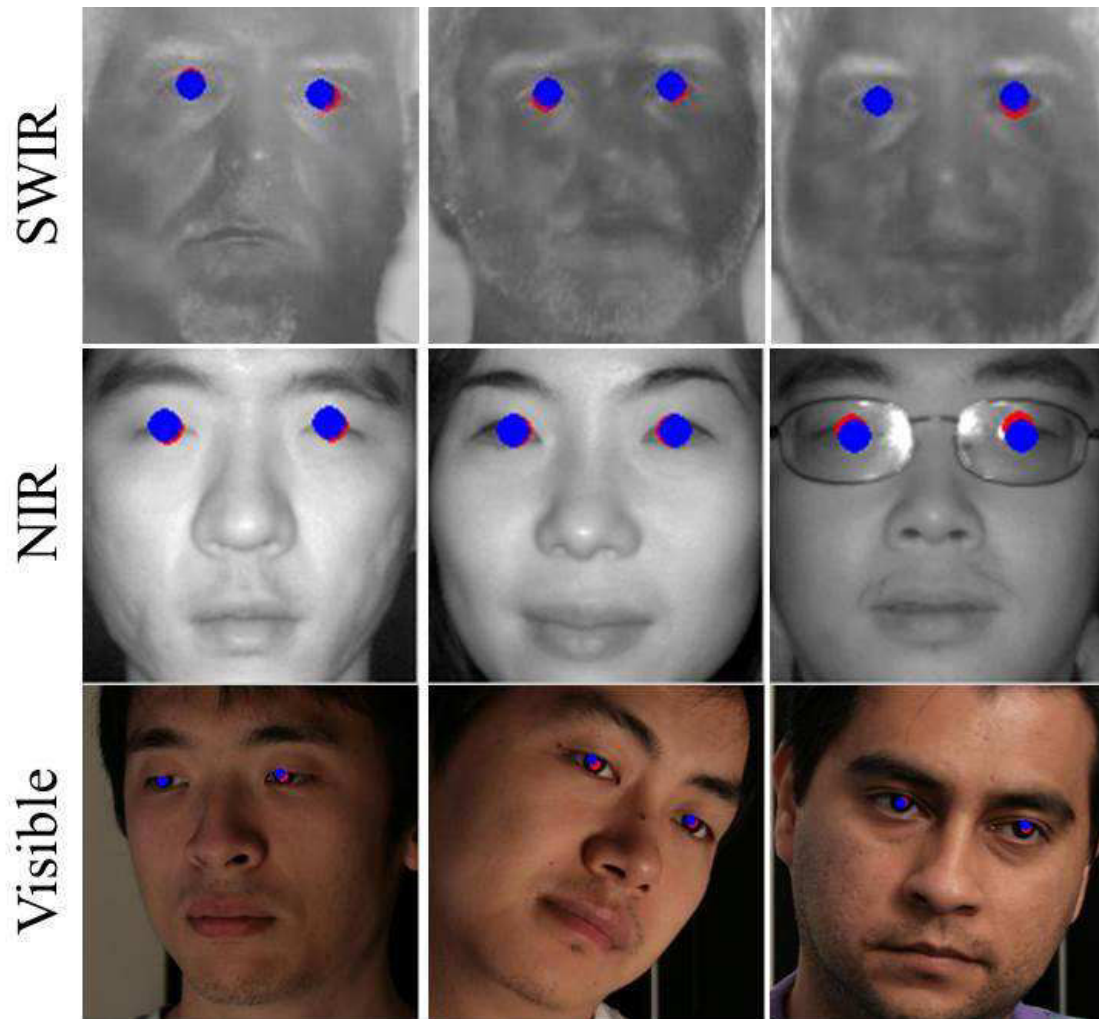
In all experiments, the normalized error is used. This error, indicating the error obtained by the worse eye estimation, is used as the accuracy measure for the found eye locations. Proposed by [81], the normalized error is described as:

$$e = \frac{\max(d_{left}, d_{right})}{w} \quad (5.7)$$

where  $d_{left}$  and  $d_{right}$  is the Euclidean distance 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. In the normalized error,  $e \leq 0.25$  (or 25% of the interocular distance) roughly corresponds to the width of the eye (i.e. corner to corner),  $e \leq 0.10$  roughly corresponds to the diameter of the iris, and  $e \leq 0.05$  roughly corresponds to the diameter of the pupil. Sample images in which the proposed approaches accurately finds the center of the pupil can be seen in Figure 5.11.

#### Active Illumination and Long Distance:

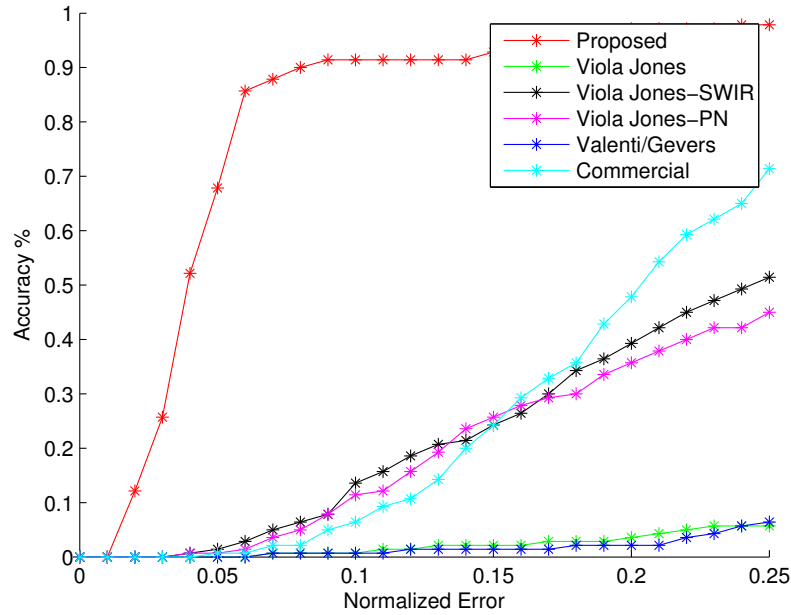
To show the effectiveness of the proposed approach on less challenging scenarios, we applied the methodology described in Section on both the Active Illumination and Long Distance databases. Figure 5.12 shows the results compared to all other tested algorithms including



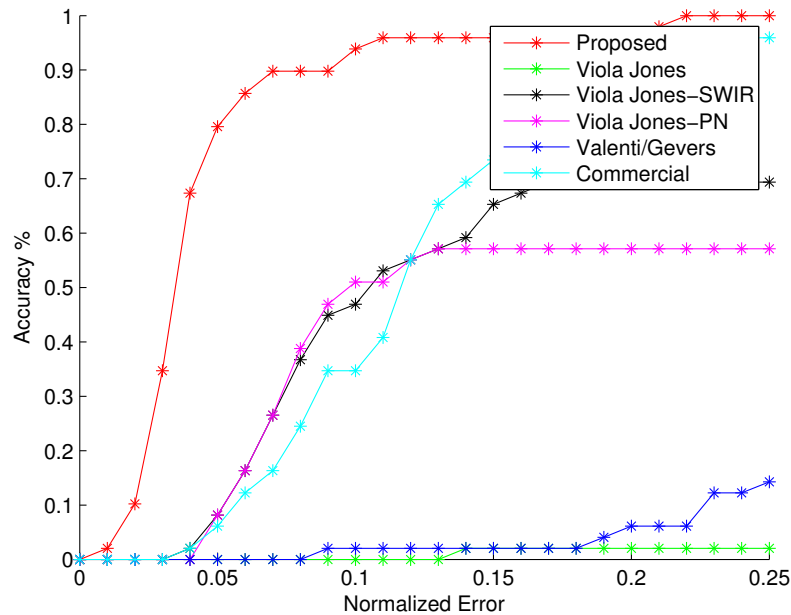
**Figure 5.11:** Sample images in which the proposed approaches accurately finds the pupil in the SWIR, NIR, and Visible spectrum. Ground Truth = Red; Found Location = Blue.



Valenti & Gevers, G8, and multiple trained instances of Viola & Jones. The proposed approach (red line) vastly outperforms all other approaches by a significant amount in both cases. In the Active Illumination database, at an interocular distance of 10%, the proposed approach yields 91.5%, while the state-of-the-art and the commercial algorithm yield 1% and 6.5% respectively. Also, for the Viola-Jones study, VJ-Visible yielded 0.7%, VJ-SWIR yielded 13.5% and VJ-SWIR-PN yielded 11.4%. In the Long Distance database, at an interocular distance of 10%, the proposed approach yields 93.8%, while the state-of-the-art and commercial yields only 2% and 34.5% respectively. Also, for the Viola-Jones study, VJ-Visible yielded 0%, VJ-SWIR yielded 46.9% and VJ-SWIR-PN yielded 51.0%.



(a)



(b)

**Figure 5.12:** WVUM: Eye detection accuracy when the subjects' face images are under active illumination (a) and at long distances (b). The proposed method is compared to L1's G8 (commercial), a state-of-the-art academic software, and three differently trained Viola and Jones algorithms.

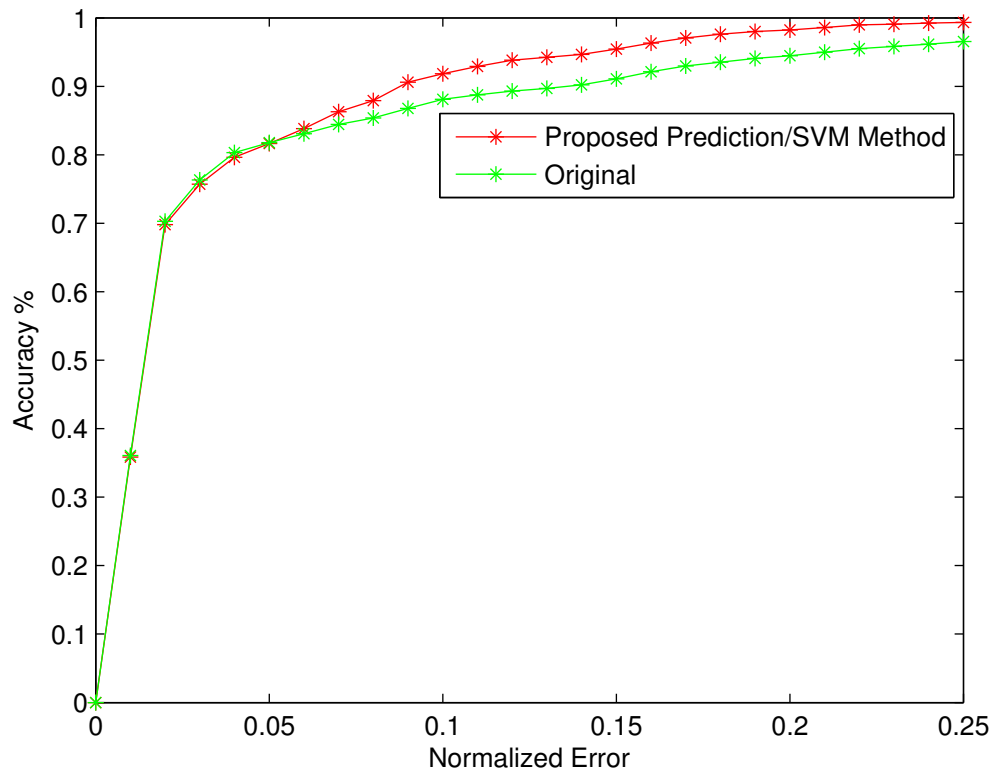
		Prediction	
		Glasses	No Glasses
Actual	Glasses	175	12
	No Glasses	0	1,715

**Table 5.1:** Confusion Matrix displaying the performance of the SVM classifier. Note that only 12 out of 1,902 sample images was mis-classified.

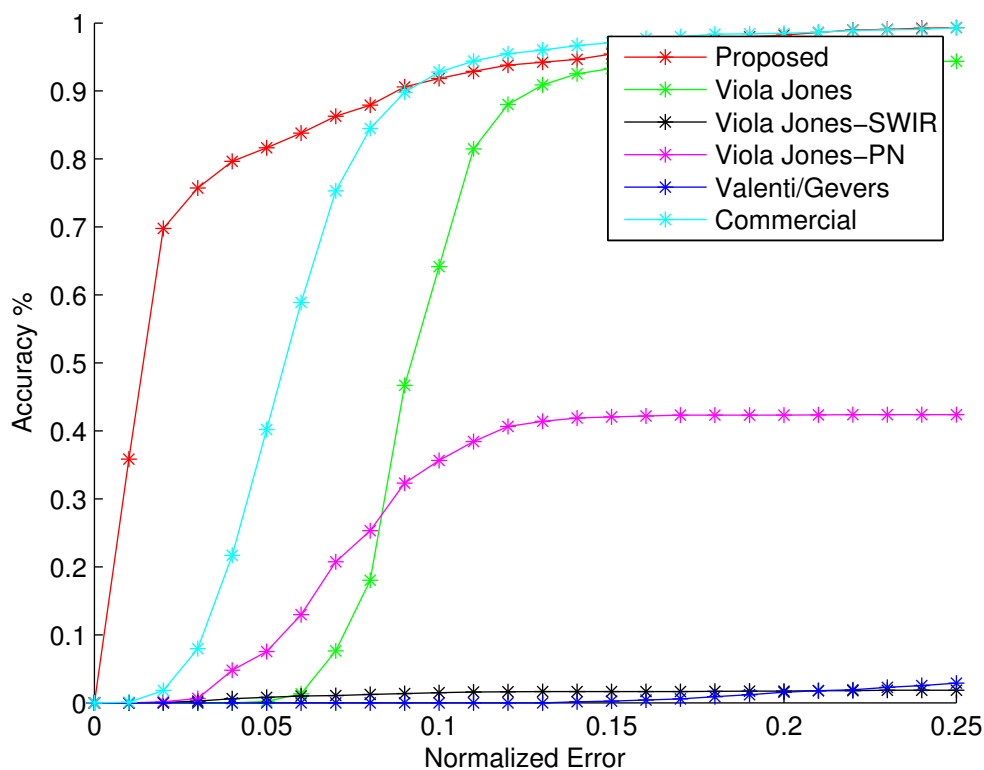
#### Wearing Eye Glasses:

To show the effectiveness of the proposed approach on images that have eyeglasses, the extended methodology proposed in Section 3.3 was considered. To train the SVM classifier, 100 images with glasses and 1000 images without glasses was used. Then, the remaining images were tested against to determine the accuracy of the classifier. Table 5.1 shows the confusion matrix. It can be seen that the accuracy of the SVM classifier is very high with a 99.36% true positive rate. This shows that the feature used for the SVM (Equation 5.2) is a highly discriminate feature for classifying images with large amounts of reflection due to eyeglasses. Secondly, we compared the original method (Section 3.2) with the extended methodology discussed in Section 3.3. The results for this test can be seen in Figure 5.13. Note that at small interocular distances, the accuracy is not much affected however when the interocular distance increases, the accuracy increases as well due to the location prediction scheme proposed.

Finally, the proposed approach is compared to the state-of-the-art and commercial algo-

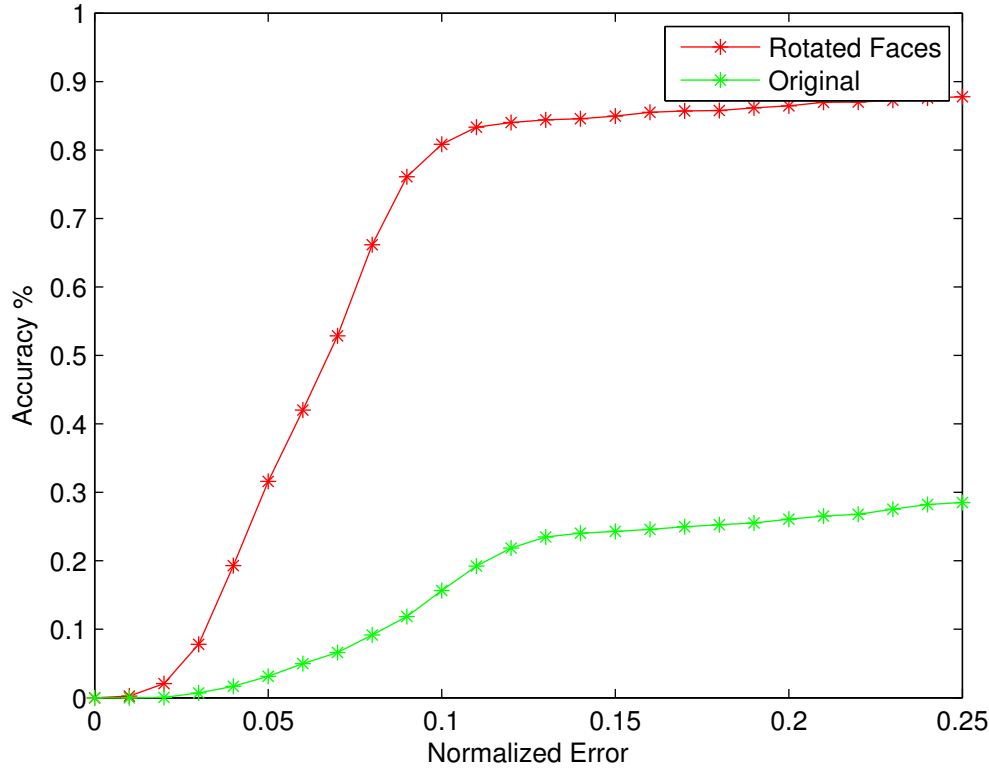


**Figure 5.13:** CASIA: Eye detection accuracy comparing the original method (green) with the extended method (red) described in Section 3.3.



**Figure 5.14:** CASIA: Eye detection accuracy when the subjects are wearing eye glasses. The proposed method is compared to L1's G8 (commercial), a state-of-the-art academic software, and three differently trained Viola and Jones algorithms.

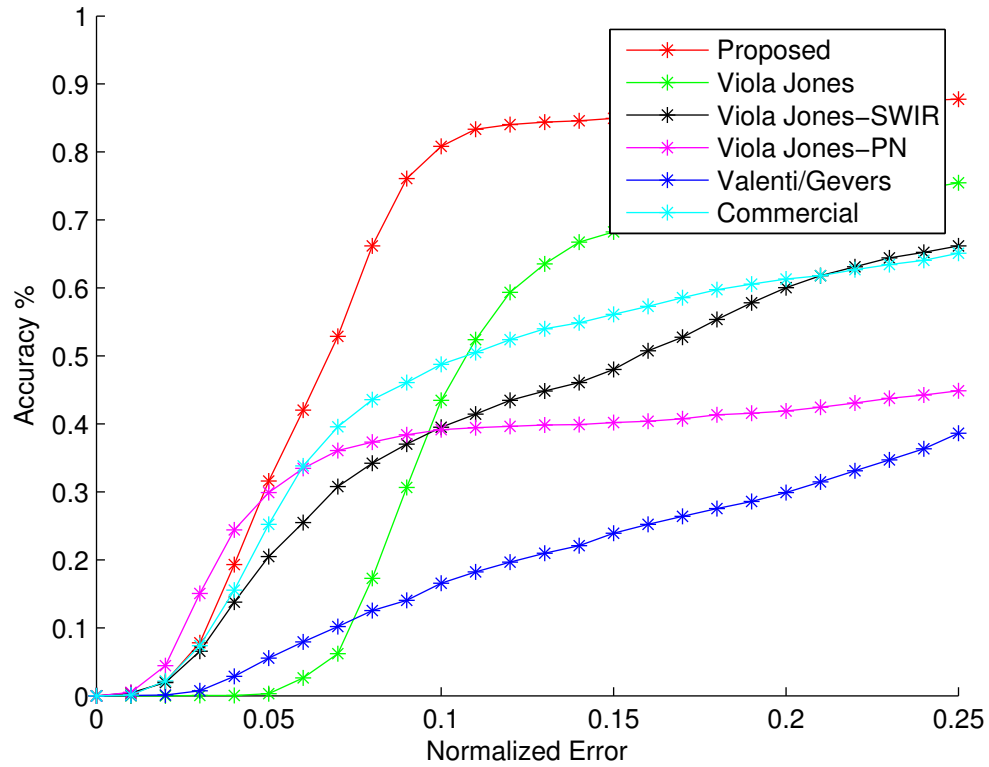
rithm. These results can be seen in Figure 5.14. It can be seen that the proposed approach (red line) yields a 81.5% accuracy at a 5% interocular distance while the state-of-the-art and commercial algorithm yield 0% and 40.5% accuracy. VJ-Visible yielded 0.2%, VJ-SWIR yielded 0.8% and VJ-SWIR-PN yielded 7.5%.



**Figure 5.15:** UHDB11: Eye detection accuracy comparing the original method (green) with the extended method (red) described in Section 3.4.

### Extreme Pose and Illumination Variance:

The approach that was described in Section 3.4 was used to test eye detection on images that have extreme pose and illumination variation. First, the approach was compared to the original approach mentioned in Section 3.2. The results can be seen in Figure 5.15. Notice that at an interocular distance of 10%, the proposed approach gains approximately 65% increasing from 16% to 81%. Finally, the proposed approach was compared to that of the state-of-the-art and the commercial algorithm. These results can be seen in Figure 5.16.



**Figure 5.16:** UHDB11: Eye detection accuracy when the subjects have extreme pose and illumination variation. The proposed method is compared to L1's G8 (commercial), a state-of-the-art academic software, and three differently trained Viola and Jones algorithms.



Note that at an interocular distance of 10%, the proposed approach achieves an accuracy of 81% compared to the 16.5% and 48.5% accuracy of the state-of-the-art and the commercial algorithm. VJ-Visible yielded 30.6%, VJ-SWIR yielded 37.0% and VJ-SWIR-PN yielded 38.4%.

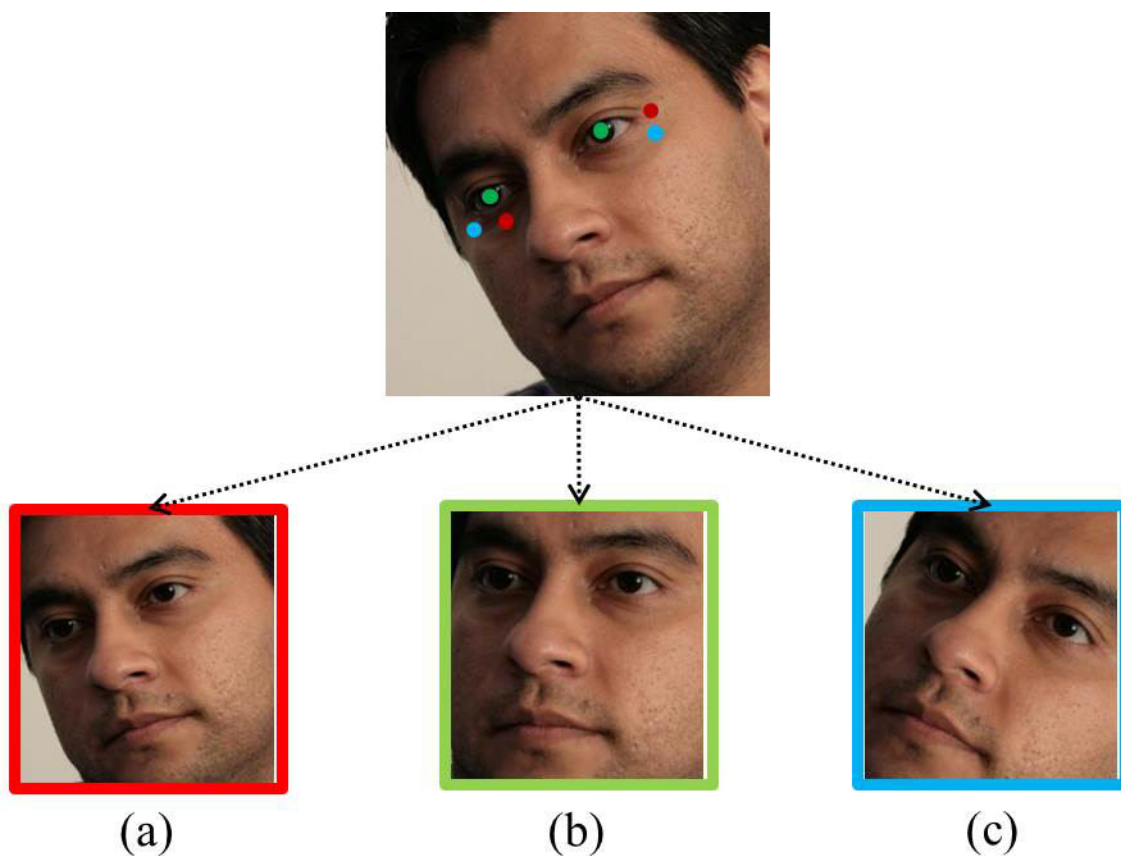
### Face Recognition Study:

The main purpose of automatic eye detection is to assist in a FR's systems ability to geometrically normalize the face. If the locations of the eyes are found incorrectly, geometric normalization will not be performed correctly (i.e. lining up the eyes in the same location for all subjects) as can be seen in Figure 5.17. Consequently, facial recognition performance will suffer. This is why accurate eye locations are needed. In order to test the accuracy of the proposed method, a texture based facial recognition algorithm, i.e. Local Binary Patterns (LBP), was used. Patterns in an image are computed by thresholding 3x3 neighborhoods based on the value of the center pixel. Then, the resulting binary pattern is converted to a decimal value. This local neighborhood is defined as a set of sampling points evenly spaced on a circle and is described in the experiments as  $LBP_{P,R}^{u^2}$  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. The binary pattern for pixels that are lying in a circle  $f_p$ ,  $P = 0, 1, \dots, P - 1$  with the center pixel  $f_c$ , is computed by the following:

$$f(n) = \begin{cases} 1 & \text{if } f_p - f_c \geq 0; \\ 0 & \text{if } f_p - f_c < 0; \end{cases} \quad (5.8)$$

After this, a binomial weight of  $2^P$  is assigned to each sign  $S(f_p - f_c)$  to compute the LBP code:

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c) 2^p \quad (5.9)$$



**Figure 5.17:** Top: Sample face image from the UHDB11 database after applying face detection. Bottom: The effects that accurate (b) and inaccurate (a,c) eye localization has on geometric normalization. As expected, the accurate eye localization (green) produces the best face image for for recognition

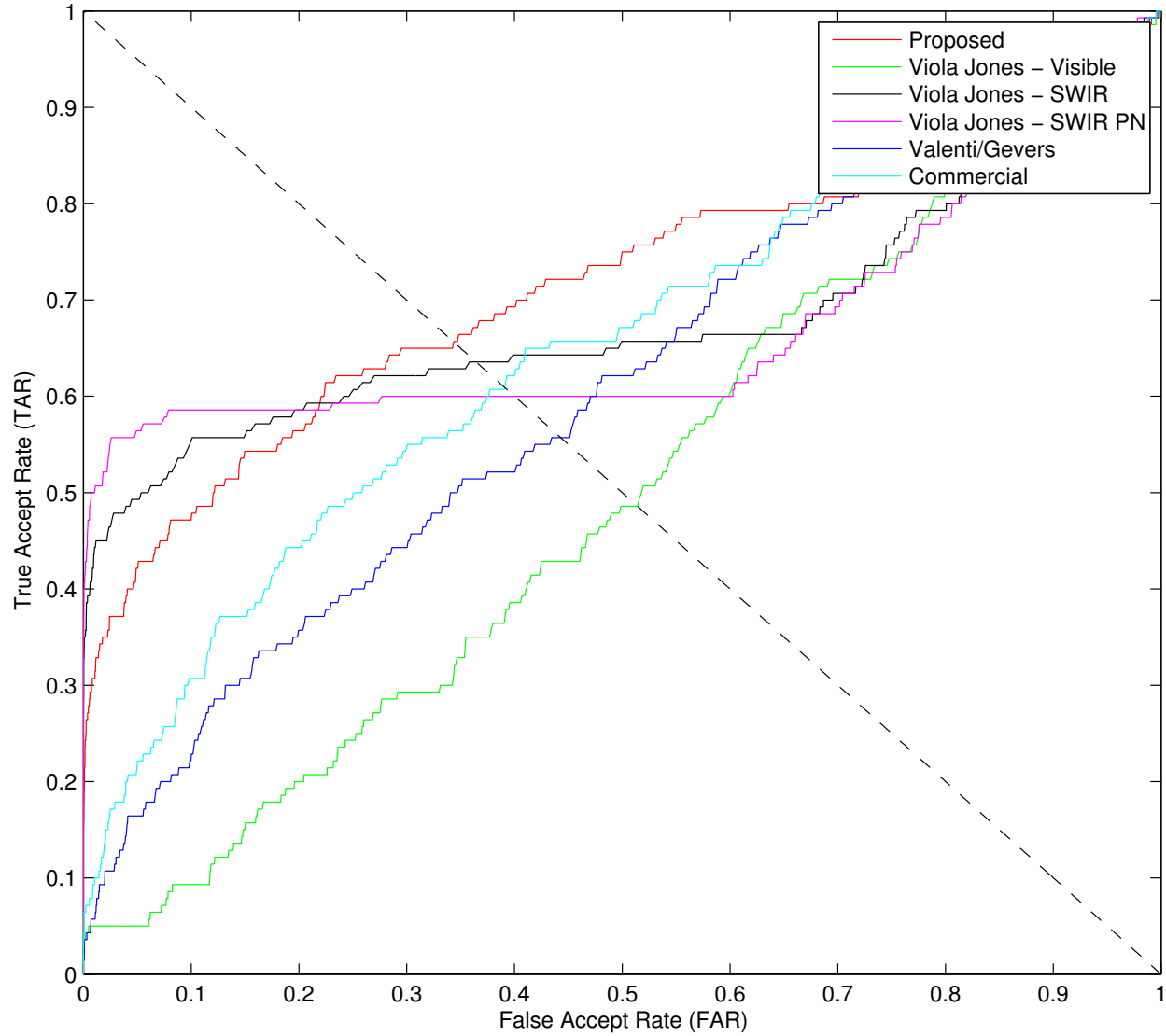
In these experiments, 1 sample from every subject (gallery image) was geometrically normalized using manually annotated eye center positions. Then, this image was matched to 1 sample from each subject (probe image) that was geometrically normalized using the found eye locations from each respective method (i.e. proposed method, Valenti & Gevers, G8, and Viola and Jones (VJ-Visible, VJ-SWIR, and VJ-PN). The LBP operator is applied all images of the gallery and probe face datasets and a distance score is computed per matching pair. After the process is completed for all matching pairs, the Receiver Operating Characteristic (ROC) curve and Cumulative Match Curve (CMC) is generated that shows the accuracy of the face recognition algorithm in both verification and identification scenarios. Using the Equal Error Rate (EER) and rank-1 and rank-10 scores, each algorithm can be numerically compared to each other.

The number of gallery and probe images per database is listed as follows:

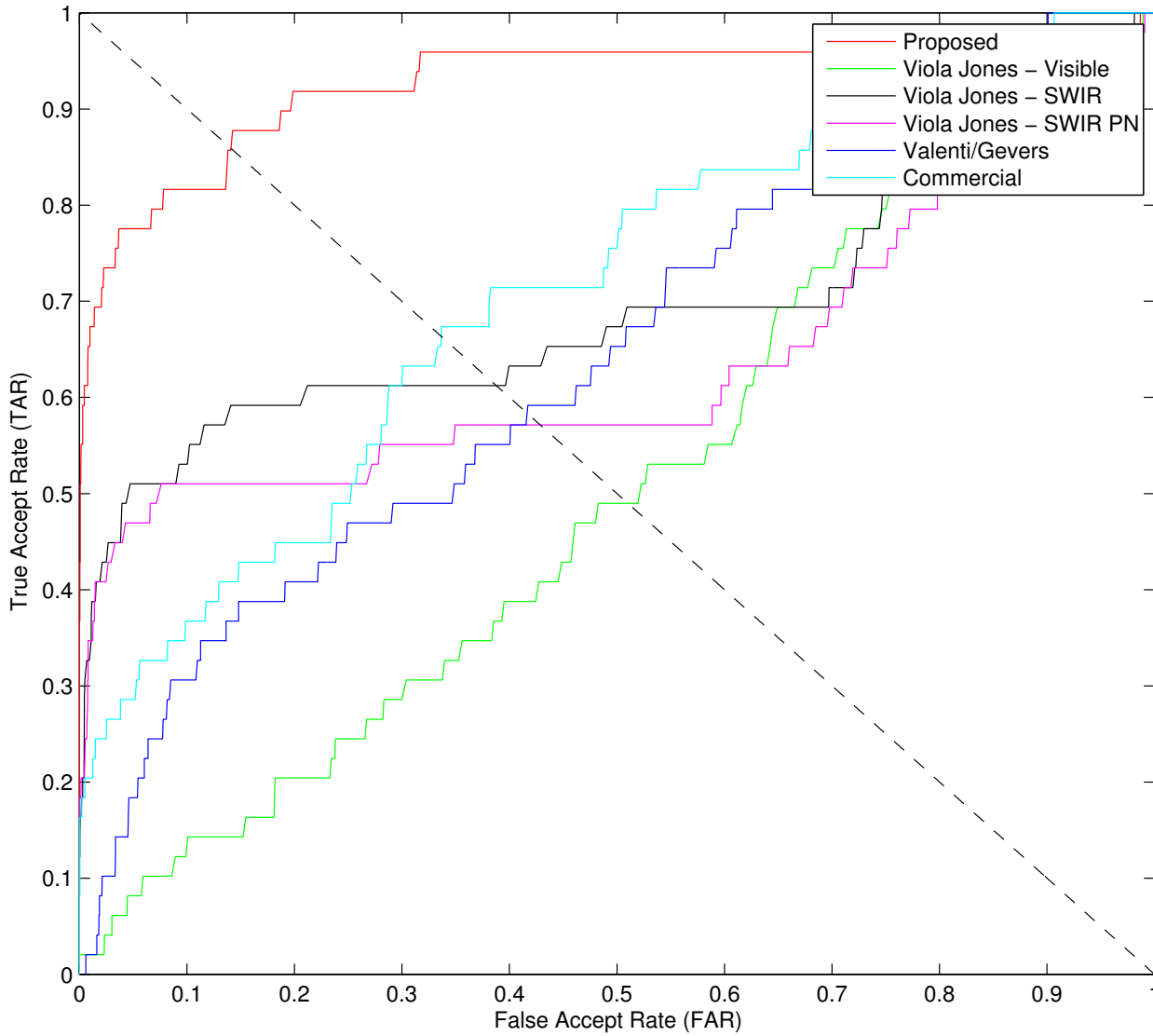
- Active Illumination:
  - Probe: 140 subjects, 1 sample per subject
  - Gallery: 140 subjects, 1 sample per subject
- Long Distance:
  - Probe: 49 subjects, 1 sample per subject
  - Gallery: 49 subjects, 1 sample per subject
- Face Obstruction:
  - Probe: 100 subjects, 5 sample per subject
  - Gallery: 100 subjects, 1 sample per subject

- Extreme Pose and Illumination Variation:
  - Probe: 23 subjects, 5 sample per subject
  - Gallery: 23 subjects, 1 sample per subject

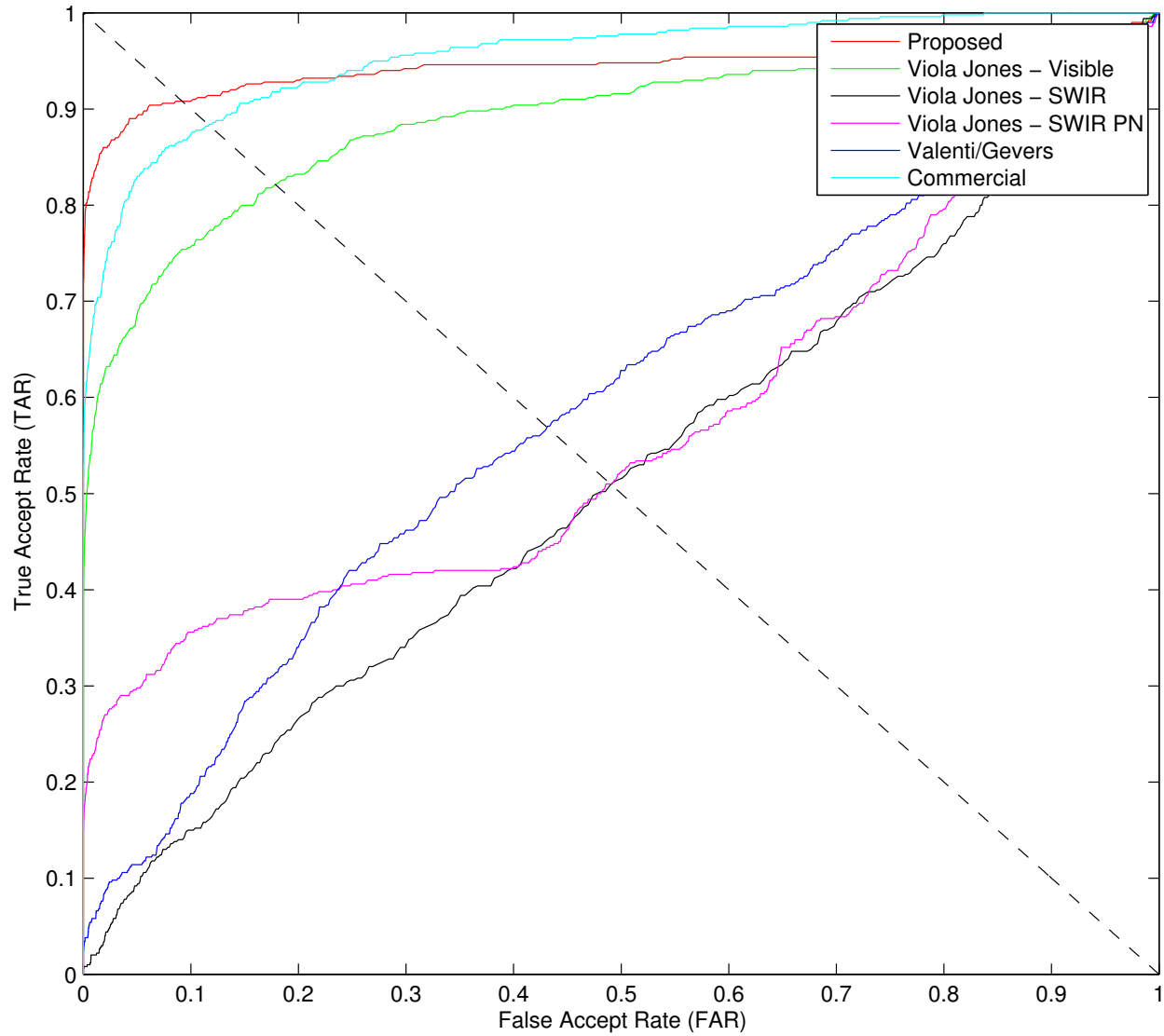
The numerical results for all face recognition experiments can be seen in Table 5.2. Verification results in the form of ROC curves can be seen in Figures 5.18, 5.19, 5.20 and 5.21, while identification results in the form of CMC curves can be seen in Figures 5.22, 5.23, 5.24 and 5.25. Note that in every face matching scenario studied, the proposed approach results in better face recognition rank-1 identification (except in Active Illumination database) rate as well as rank-10 rate and EER when compared to all other eye detection algorithms tested.



**Figure 5.18:** Verification results on the Active Illumination database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.

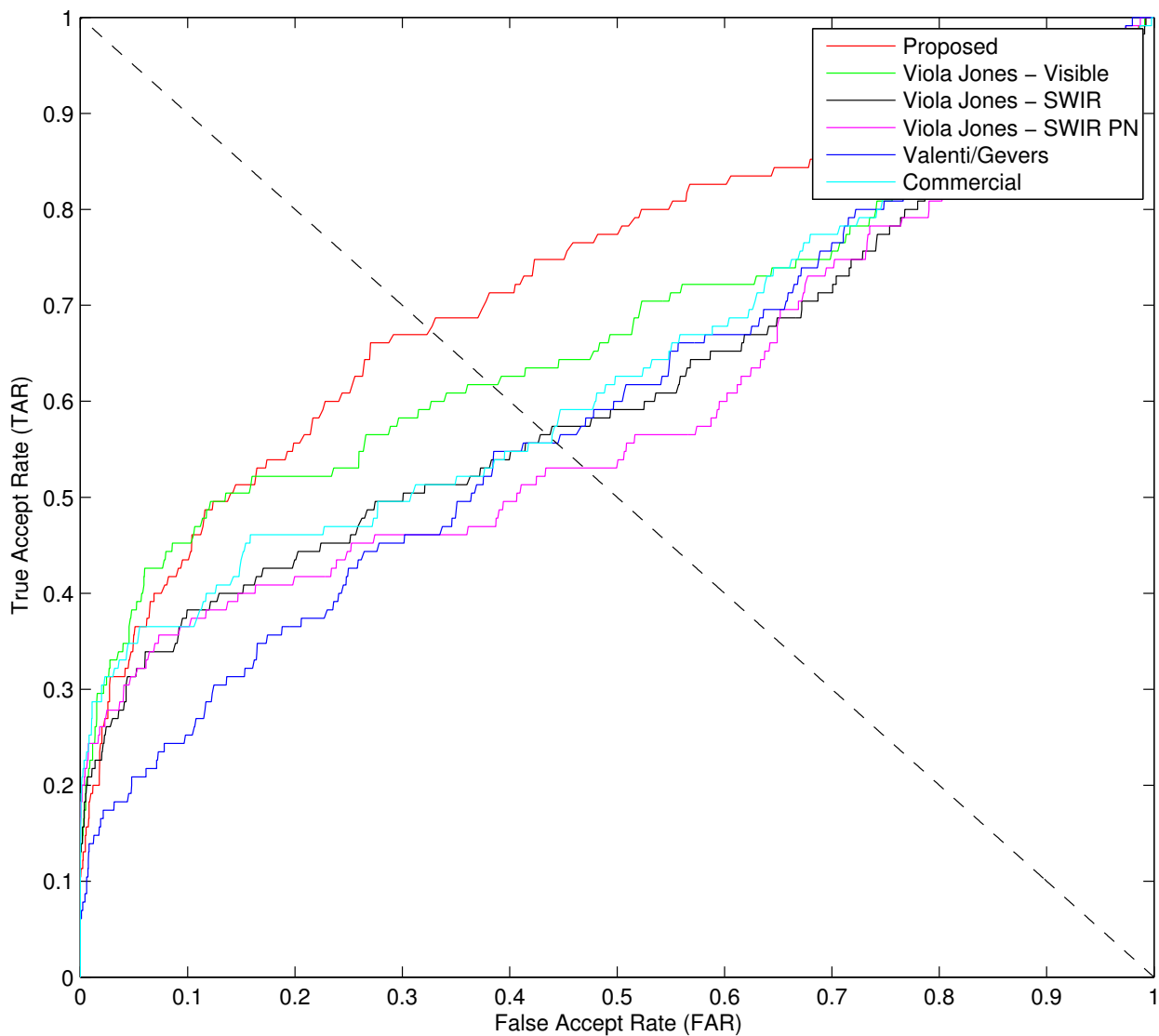


**Figure 5.19:** Verification results on the Long Distance database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.

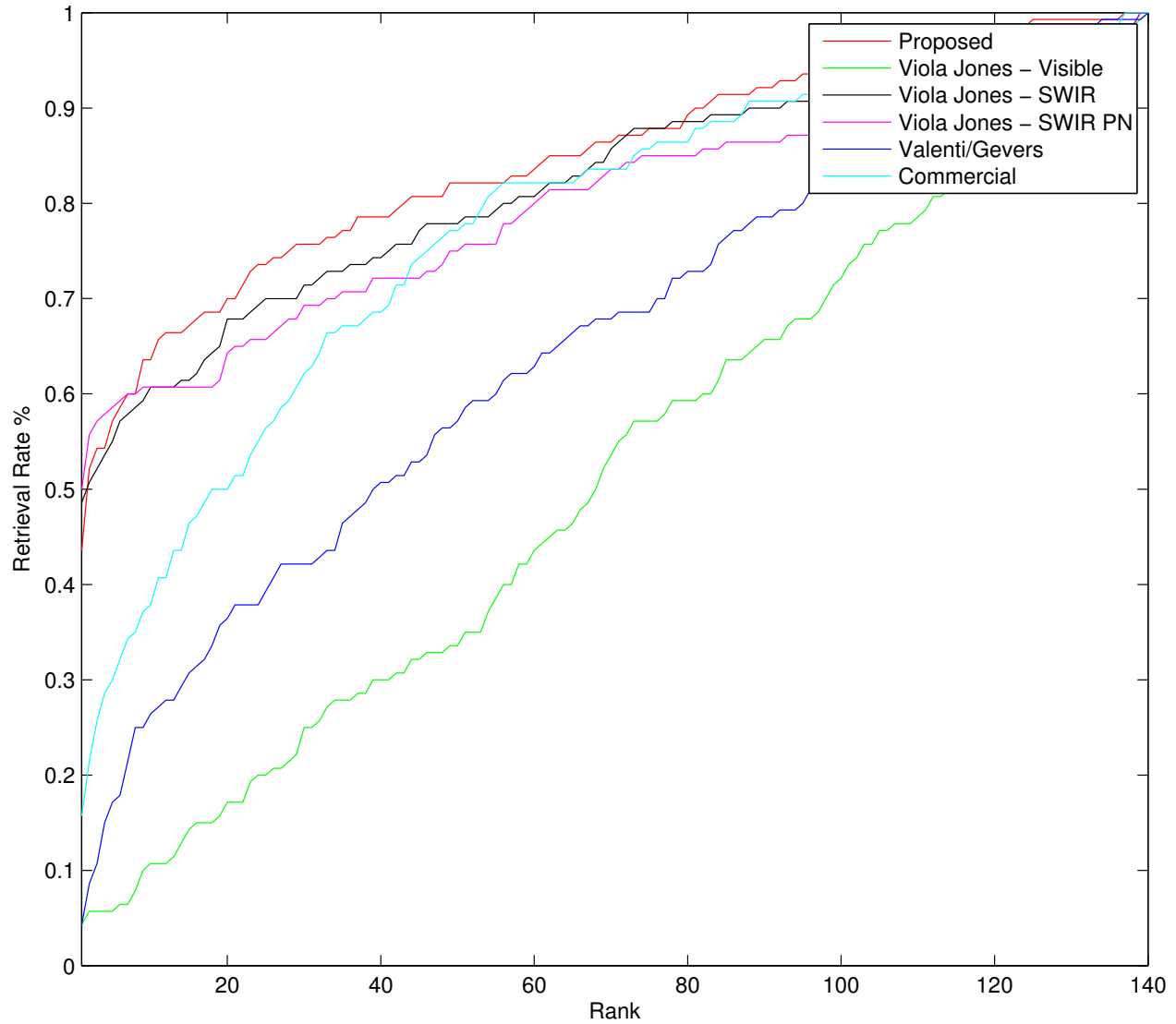


**Figure 5.20:** Verification results on the Facial Obstruction database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.

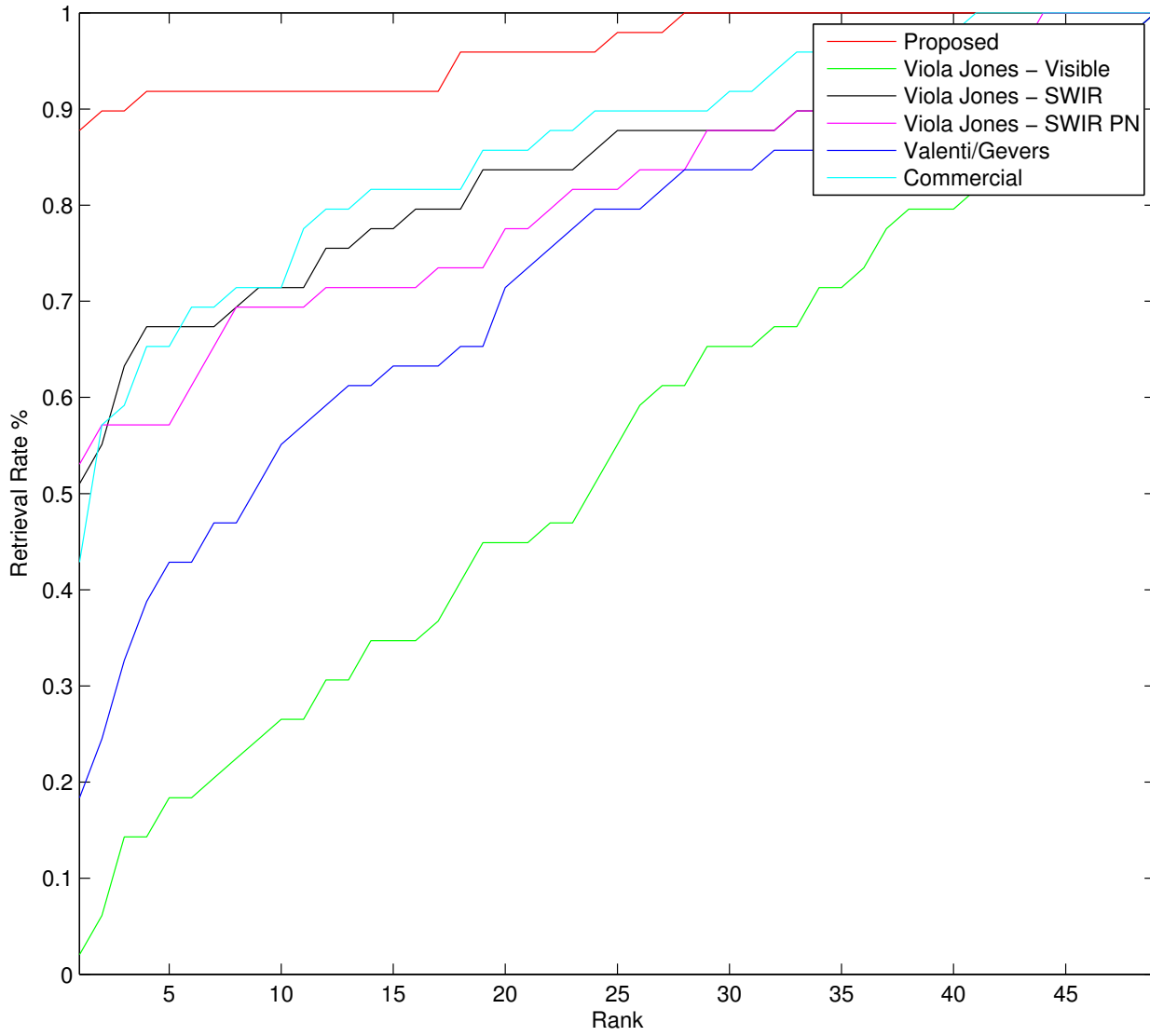




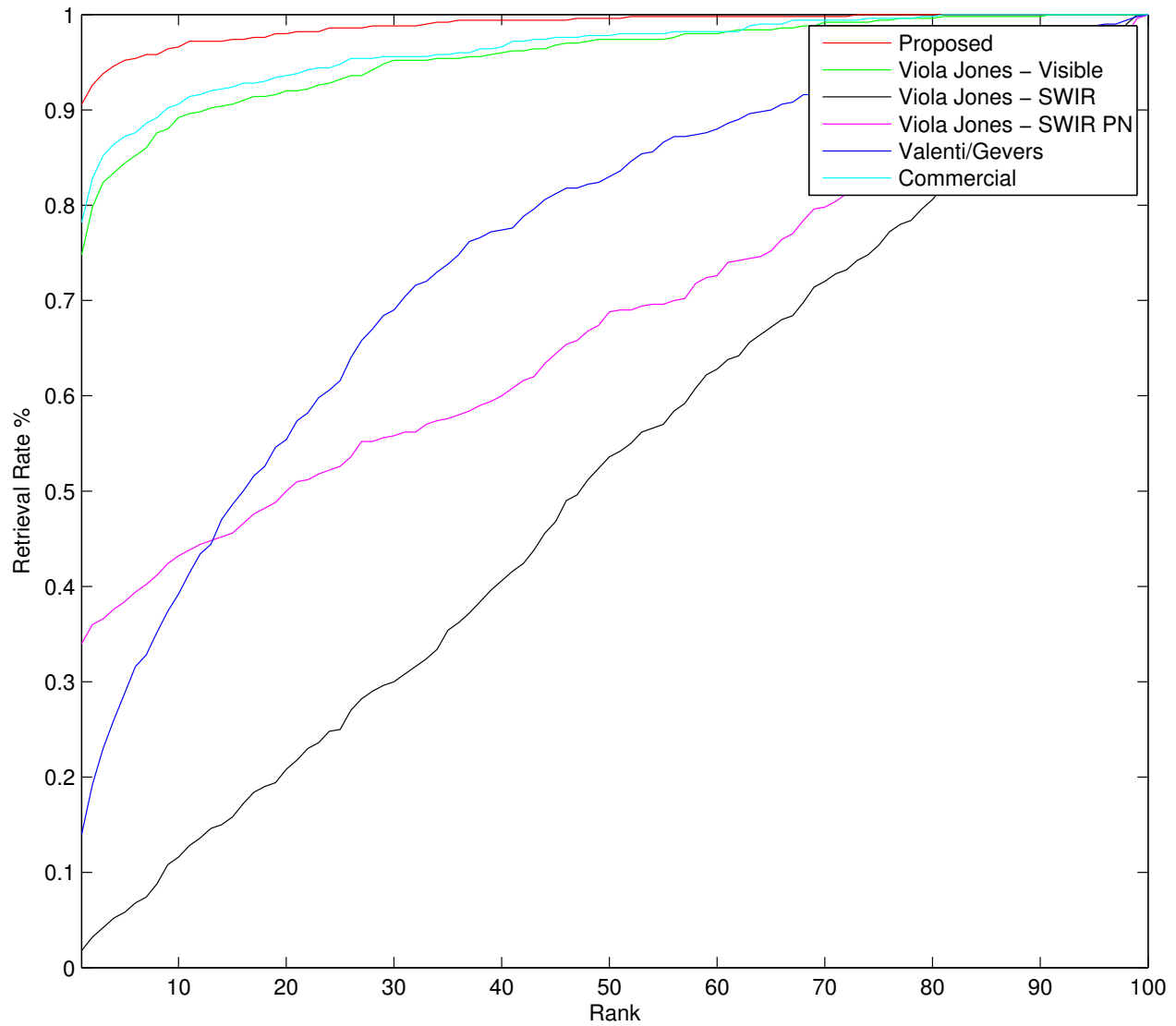
**Figure 5.21:** Verification results on the Extreme Pose and Illumination Variation database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.



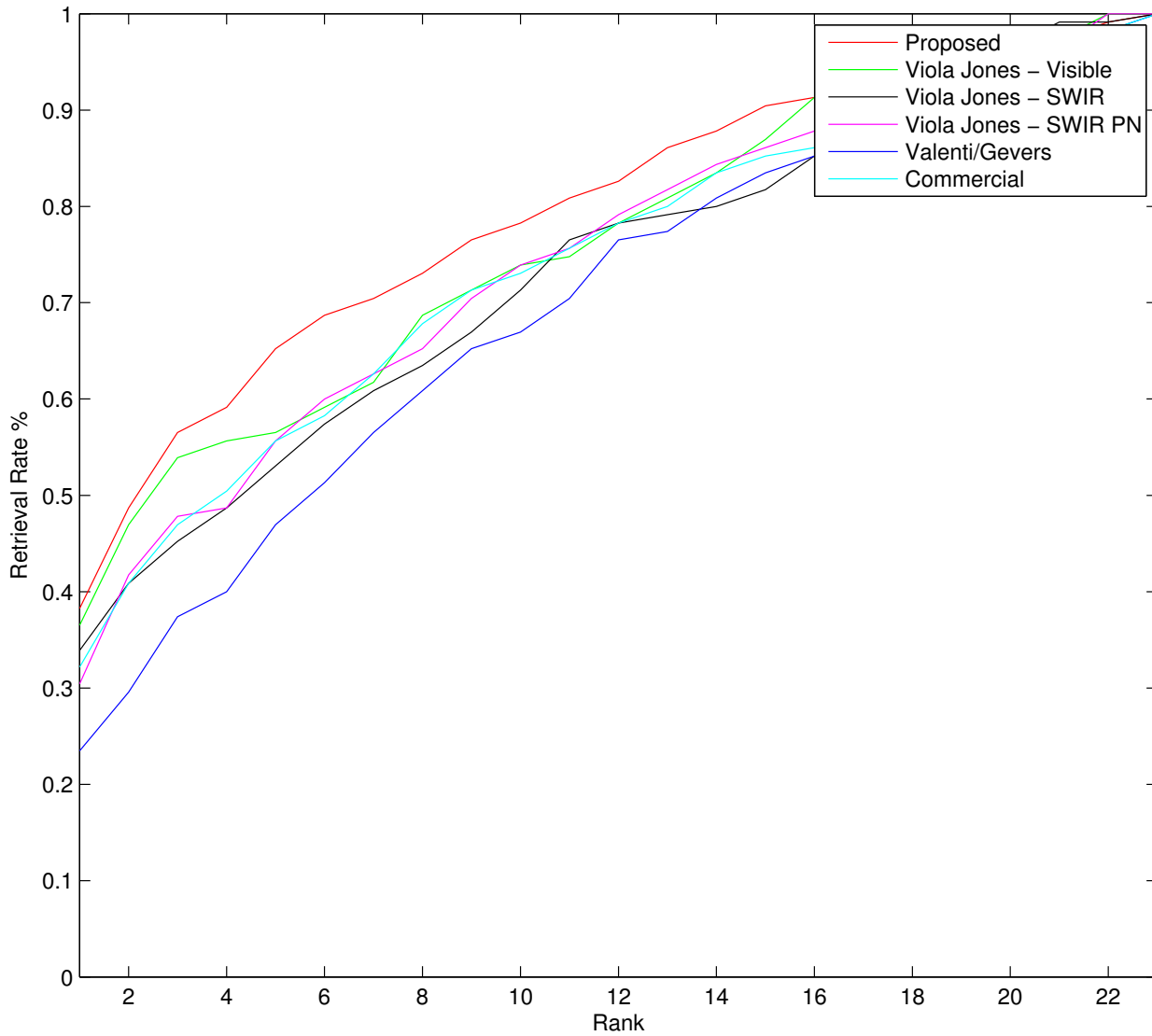
**Figure 5.22:** Identification results on the Active Illumination database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.



**Figure 5.23:** Identification results on the Long Distance database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.



**Figure 5.24:** Identification results on the Facial Obstruction database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.



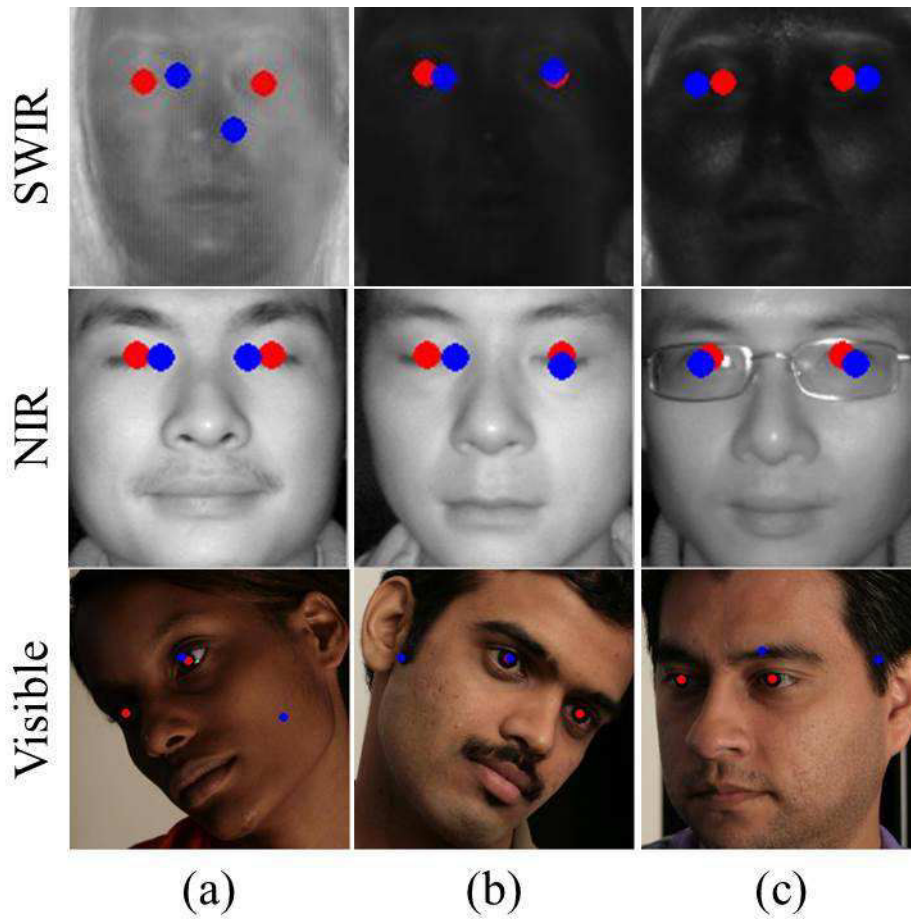
**Figure 5.25:** Identification results on the Extreme Pose and Illumination Variation database. Note that the results obtained from the proposed methodology (red) outperforms all other algorithms tested against.

	Active Illumination			Long Distance		
	EER	Rank-1	Rank-10	EER	Rank-1	Rank-10
Proposed	<b>0.344</b>	0.435	<b>0.635</b>	<b>0.142</b>	<b>0.877</b>	<b>0.918</b>
VJ-Visible	0.515	0.042	0.107	0.510	0.020	0.265
VJ-SWIR	0.364	0.485	0.607	0.388	0.510	0.714
VJ-PN	0.400	<b>0.500</b>	0.607	0.430	0.530	0.693
Valenti/Gevers	0.444	0.042	0.264	0.413	0.183	0.551
Commercial	0.392	0.157	0.378	0.332	0.428	0.714
	Facial Obstruction			Extreme Pose and Illumination Variation		
	EER	Rank-1	Rank-10	EER	Rank-1	Rank-10
Proposed	<b>0.093</b>	<b>0.906</b>	<b>0.966</b>	<b>0.325</b>	<b>0.382</b>	<b>0.782</b>
VJ-Visible	0.179	0.748	0.892	0.384	0.365	0.739
VJ-SWIR	0.491	0.018	0.116	0.435	0.339	0.713
VJ-PN	0.490	0.340	0.432	0.470	0.304	0.739
Valenti/Gevers	0.431	0.140	0.392	0.444	0.234	0.669
Commercial	0.117	0.782	0.906	0.442	0.321	0.730

**Table 5.2:** Numerical results for each algorithms performance on the four tested databases. Notice that the proposed algorithm outperforms all other algorithms in all experiments except for the rank-1 score in the Active Illumination database.

### 5.3.3 Limitations

As with any algorithm, the proposed approach fails under certain circumstances. For each band, there are different reasons in which the approach cannot accurately find the pupil center. For example, eye closure from non cooperative subjects (Figure 5.26 SWIR (a), NIR (a,b)) as well as illumination issues with poor contrast (Figure 5.26 SWIR (b,c)) cause the pupil detection algorithm to fail. In the NIR band, the reflection in certain sets of eye-glasses cause false eye localization (Figure 5.26 NIR (c)). Finally, in the Visible band, many failures



**Figure 5.26:** Sample images where the proposed approach fails to accurately detect the eyes in the SWIR, NIR, and Visible spectrum. Issues such as poor contrast, eye closure, and extreme face rotation lead to falsely detected pupil locations. Ground Truth = red; Found Locations = blue.

are obtained from extreme pitch, yaw, and roll issues (Figure 5.26 Visible (a,b)) as well as when faces are not in the center of the image (Figure 5.26 Visible (c)). Sample images that failed from our proposed approach can be seen in Figure 5.26. Because of these issues, other methods may perform better than the proposed approach, such as Active Appearance Models (AAM) [39] and Active Shape Models ASM [40]. By using these methods in conjunction with the approaches listed above, the proposed algorithms could be adapted to work with even more challenging scenarios, such as face in the wild. However, testing these approaches would require more collection of data for training and testing, which is extremely time consuming and costly due to IRB approval and sponsor funding.

## 5.4 Conclusion and Future Work

In this chapter, a novel set of methodologies to accurately locate the eye positions in multiple difficult environments and spectra is proposed. Multiple databases were collected and processed that represent different challenging scenarios. Then, each input face image was automatically classified into their respective scenarios. After locating the face positions and automatically determining if the image contained eye glasses or not (e.g. see experiments with CASIA face database), eye and pupil detection was performed. Each image was geometrically normalized and used for face recognition studies, before, finally, results were obtained. Experimental results showed that the proposed approach outperforms all other tested approaches by more than 85% and 55% when using the active illumination and long distance IR face databases respectively. Secondly, another challenging scenario was considered in the NIR spectra in which images were influenced by face obstruction (eyeglasses). It is shown that the use of the proposed eyeglass classifier as well as the pupil localization



prediction scheme significantly increases the performance of the algorithm allowing for an increase of 40% to L1's G8. Finally, another challenging scenario was explored in the visible spectrum in which subjects' faces are affected by pose and illumination variation. Our approach once again yields the highest accuracy over all other algorithms, outperforming them by more than 30% at an interocular distance of 10%. It is also demonstrated the benefit of the approach to face recognition performance. In all cases, the algorithm produces the highest rank-1 identification rates when geometrically normalizing the face images with the eye locations obtained from the approach verse the other tested approaches.

The authors recognize that in some databases (i.e. active illumination and long distance), the number of samples and subjects is small due to the fact that there are very few publicly available SWIR databases and that it is also very difficult to perform a new data collection under challenging conditions (due to university data collection policies, cost, time, etc.). Hence, we only managed to use the data that was either provided to us, or collected in past projects. In the future, the authors would like to explore applying the approach in larger face datasets collected using different IR camera sensors to further provide evidence of the efficiency of the algorithms. Also, the authors would like to optimize the algorithms to perform band-independent eye center localization in real time video scenarios. Finally, the authors are planning to explore even more challenging scenarios. By using alternative approaches, such as Active Appearance Models and Active Shape Models (under the condition that we redesign these models using IR band images), in conjunction with the approaches, eye detection could be performed in more challenging conditions, such as faces in the wild (i.e. completely uncontrolled scenarios).





# Chapter 6

## Designing SWIR to Visible Face Matching Algorithms

### 6.1 Introduction

The focus of this chapter is matching visible gallery images to SWIR band face probe images. As discussed in Chapter 1, 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 [16]. SWIR has a longer wavelength range than the 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 subjects face [20]. Previous studies [57], discussed in Chapter 2, determined that this capability resulted in an increase of rank-one identification rates under variable face matching scenarios.

While previous FR studies have mainly concentrated on the visible and NIR bands, FR in the SWIR band, more specifically the  $1550nm$  wavelength, has received limited attention. In this chapter, we propose a new cross-spectral face matching algorithm. It significantly enhances the capability of the original approach proposed by Kalka et. al. [20] to match SWIR to visible face images in variable challenging scenarios, including scenarios in which face images were captured behind different types of tinted glass. Experiments using our proposed face matching algorithm show that the use of randomly selected photometric normalization techniques (as proposed in [20]) 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 chapter is outlined as such. First, in section 2, the methodology behind the proposed face matching algorithm is described. Section 3 briefly discusses the databases

used and the results obtained from the approach while Section 4 provides a summary.

## 6.2 Methodology

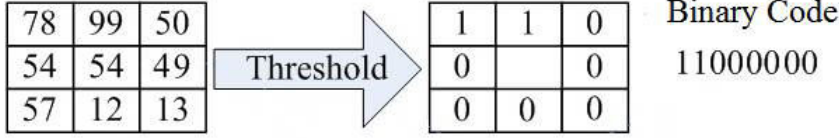
### 6.2.1 FR Matchers

In this work, both commercial and research software was employed to perform the face recognition experiments:

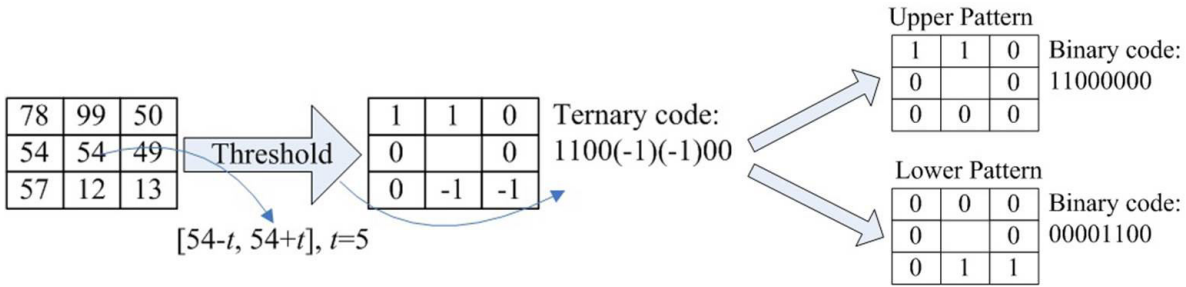
- Commercial software, more specifically Identity Tools G8 provided by L1 Systems
- Standard texture-based feature methods

Two different texture based feature extraction methods were used to test our algorithm, i.e. Local Binary Patterns (LBP) and Local Ternary Patterns (LTP). In the LBP operator, visually shown in Figure 6.1, patterns in an image are computed by thresholding 3x3 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  $LBP_{P,R}^{u^2}$  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. The binary pattern for pixels that are lying in a circle  $f_p$ ,  $P = 0, 1, \dots, P - 1$  with the center pixel  $f_c$ , is computed by the following:

$$f(n) = \begin{cases} 1, & \text{if } f_p - f_c \geq 0 \\ 0, & \text{if } f_p - f_c < 0 \end{cases} \quad (6.1)$$



**Figure 6.1:** Overview of the LBP operator



**Figure 6.2:** Overview of the LTP operator. Notice how the LTP operator splits the code into two separate LBP feature vectors.

After this, a binomial weight of  $2^P$  is assigned to each sign  $S(f_p - f_c)$  to compute the LBP code:

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c) 2^p \quad (6.2)$$

The advantage of using LBP is that they are 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. To overcome these limitations, local ternary patterns (LTP) are introduced, shown in Figure 6.2, and the quantization is performed as follows:

$$f(n) = \begin{cases} 1, & \text{if } f_p - f_c \geq t \\ 0, & \text{if } |f_p - f_c| \leq t \\ -1, & \text{if } f_p - f_c \leq -t \end{cases} \quad (6.3)$$

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.

### 6.2.2 Distance Metrics

After the texture based feature patterns are computed, two different distance metrics are used to obtain the final match score, namely the distance transform (DT) and the chi-squared ( $\chi^2$ ). The distance transform (defined as the distance or similarity metric from image X to image Y) is defined as follows:

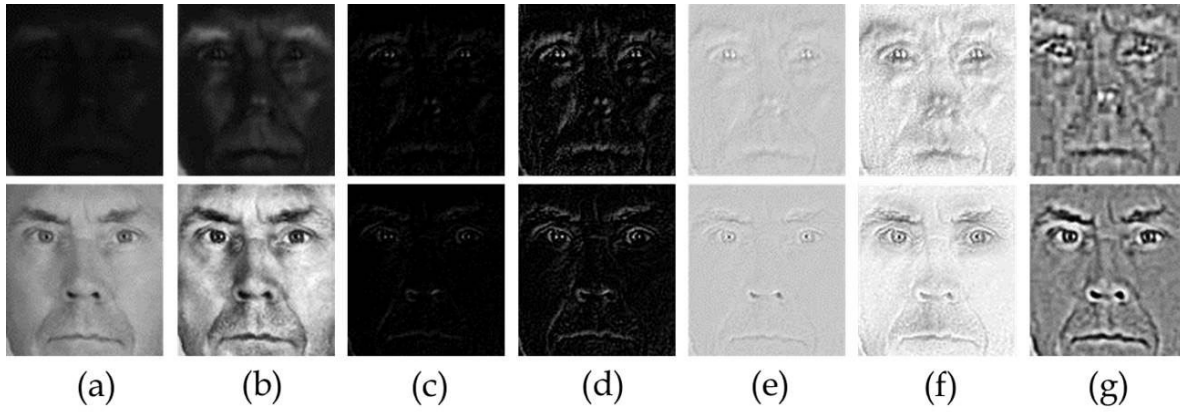
$$D(X, Y) = \sum_{Y(i,j)} w(d_x^{K_{Y(i,j)}}(i, j)) \quad (6.4)$$

where  $K_{Y(i,j)}$  is the code value of pixel  $(i, j)$  of image Y, and  $w$  is a user controlled penalty function. Additionally, the chi-squared distance is described as follows:

$$\chi^2(n, m) = \frac{1}{2} \sum_i^l \frac{h_n(k) - h_m(k)}{h_n(k) + h_m(k)} \quad (6.5)$$

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





**Figure 6.3:** Sample images from the visible gallery (bottom row) and SWIR probe (top row) after applying six different photometric normalization: (a) CLAHE (b) LBSSR (c) C-LBSSR (d) TBSSR (e) C-TBSSR (f) TT

sets respectively.

### 6.2.3 Photometric Normalization Empirical Study

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, as seen in Figure 6.3, in order to facilitate cross-spectral matching. More specifically, we employ the following techniques\*:

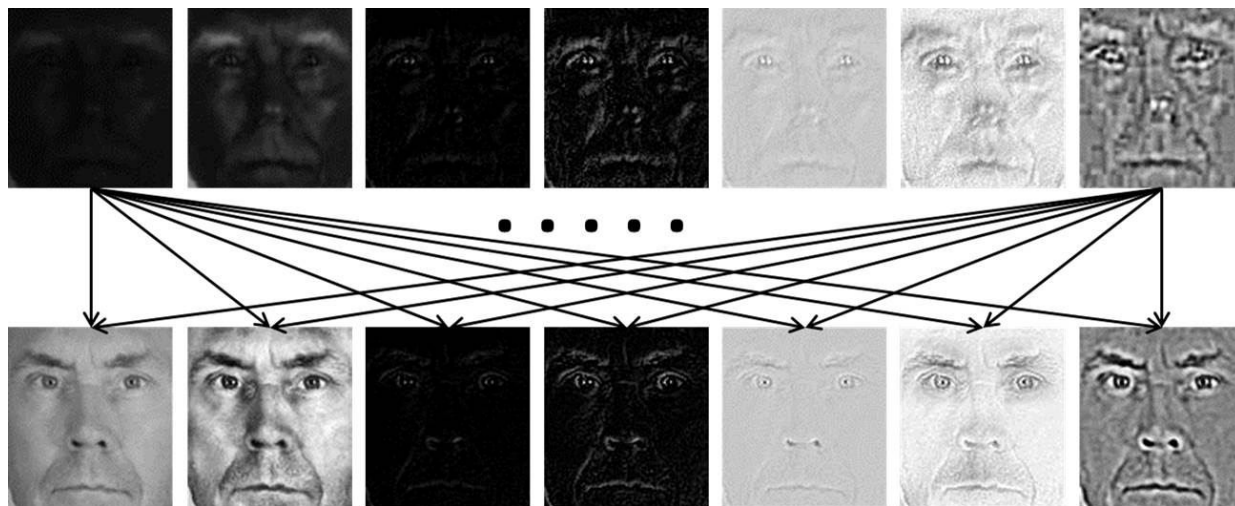
---

\*More details on the foundation of these photometric normalizations can be found in Chapter 2 Section 4.3

- Contrast Limited Adaptive Histogram Equalization (CLAHE) [82]
- Log Based Single-Scale Retinex (LBSSR) [83]
- LBSSR followed by CLAHE
- Tangent Based Single-Scale Retinex (TBSSR) [83]
- TBSSR followed by CLAHE (C-TBSSR)
- Tan and Triggs (TT) [84]

In this empirical study, the goal was to determine which combination of photometric normalization algorithms produces the best match scores between visible images and  $1550nm$  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 normalization used ( $n = 7$ ), 49 different photometric combinations are created per match (7 probe representation x 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 6.4.

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 operating characteristic (ROC) curve is computed for all 49 photometric normalizations. The ROC curve 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



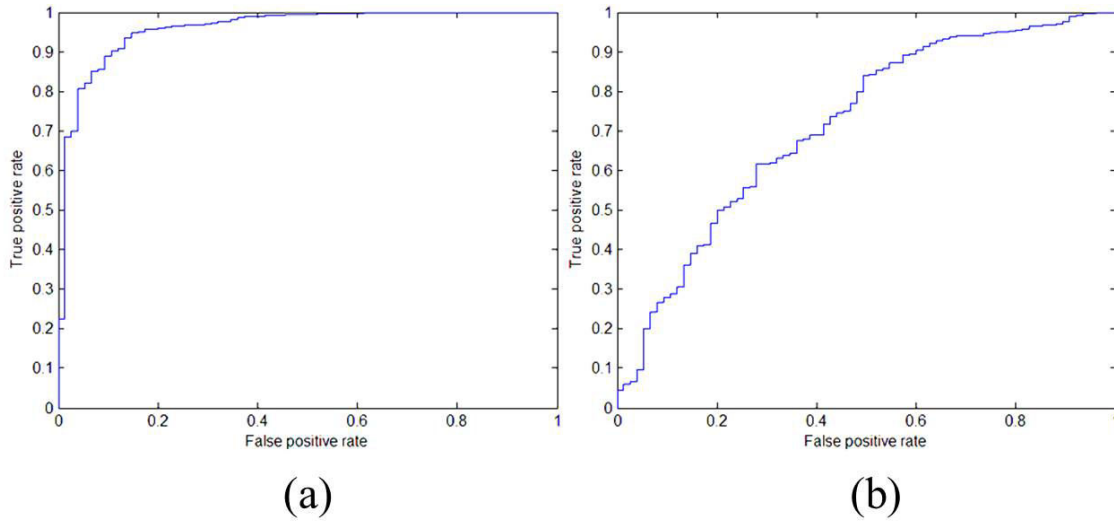
**Figure 6.4:** An overview of the cross-photometric empirical study done. Notice that each representation in the gallery set is matched against all representations in the probe set creating 49 combinations.

determine which photometric combination performs better than the others. Higher AUC's show combinations that have a wider gap between true positives and false positives, which, in then, results in higher performance.

After computing all AUC's for all 49 combinations, we can determine which photometric combinations result in higher performance. Sample ROC curves and the respective AUC's can be seen in Figure 6.5.

#### 6.2.4 Score Level Fusion

Since we known which photometric normalization combinations perform best and which combinations perform the worse, based on their respective AUC's, we can take advantage of multiple combinations to further increase the final match score. Simple fusion of all 49 combinations can be performed, as is performed in Kalka et al. [20]. However, this is not

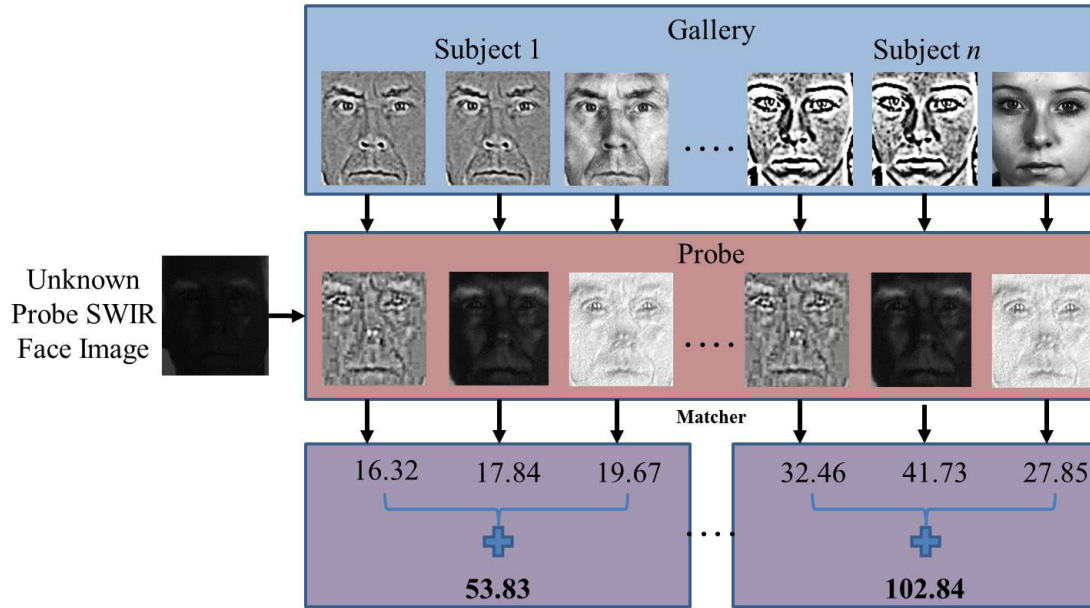


**Figure 6.5:** Sample ROC curves for two different photometric normalization combinations (a) Combination 44: 95.85% AUC and (b) Combination 2: 71.67% AUC.

feasible from a practical standpoint due to the computational resources it takes to apply 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 normalization 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 to be used. The final match score  $S$  for any probe image is computed by using the following formula:

$$S = \sum_{i=1}^3 m(Pt_i, Gt_i) \quad (6.6)$$

where  $Pt$  and  $Gt$  are the gallery and probe templates respectively and  $i$  represents the photometric normalization combination determined previously. Matching function  $m(Pt, Gt)$



**Figure 6.6:** An overview of the score level fusion scheme used to improve cross-spectral face recognition.

corresponds to the matching algorithms listed above, i.e. LBP and LTP. An overview of this process is illustrated in Figure 6.6.

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

- *Gallery:* TT - *Probe:* TT
- *Gallery:* TT - *Probe:* CLAHE
- *Gallery:* CLAHE - *Probe:* TBSSR

Another advantage to our approach is that instead of performing all 49 photometric normalization combinations as was proposed by Kalka et al. [20], 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. [20].

## 6.3 Evaluation

Two sets of experiments were conducted to demonstrate the efficiency of our algorithm. First, heterogeneous face recognition is performed on multiple Visible-SWIR databases and compared with Kalka et al. and the commercial software, G8. Secondly, a time efficiency (computational complexity) is performed to determine what the rate of increase is when compared to other methods.

### 6.3.1 Heterogeneous Face Recognition

In order to show the efficiency of our algorithms, we performed heterogeneous face recognition tests on seven different databases from the WVUMFD, as described in Chapter 3. The databases that were used are as follows:

- Ground Truth: No Glass
- Clear w/ 0% Film Tint Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior
- Clear w/ 0% Film Tint Glass Panel - 500mW 1550nm active illumination
- Clear w/ 80% Film Tint Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior
- Clear w/ 80% Film Tint Glass Panel - 500mW 1550nm active illumination

- Solarcool (2) Graylite Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior
- Solarcool (2) Graylite Glass Panel - 500mW 1550nm active illumination

140 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 (DT, $\chi^2$ ), a commercial face recognition algorithm (L1 Systems G8), and the original cross-photometric score level fusion approach described by Kalka et al. [20]. All experiments were performed on a 64-bit Windows 7 machine with 12GB of RAM running Intel Core i7 CPU @ 3.2Ghz using MATLAB R2012b. An overview of the results from this experiment can be seen in Table 6.1. All results shown are percentage of probe images that obtained a rank-1 identification.

	Face Matching Algorithm						
	LBP - $\chi^2$	LBP - DT	LBP - $\chi^2$	LTP - DT	L1's G8	Kalka et al [20]	Proposed Method
Ground Truth	62.14	80.14	70.71	86.43	81.43	35.00	<b>94.26</b>
0% Visible Lighting	38.57	50.00	47.86	53.57	56.43	19.29	<b>67.86</b>
0% Active SWIR Lighting	12.86	17.86	19.29	24.29	12.14	5.71	<b>35.71</b>
80% Visible Lighting	26.43	<b>40.82</b>	20.41	30.61	6.12	13.57	34.69
80% Active SWIR Lighting	7.14	12.14	8.57	13.57	4.29	5.00	<b>23.57</b>
Solarcool Visible Lighting	39.29	48.57	52.14	55.00	<b>69.29</b>	23.57	61.43
Solarcool Active SWIR Lighting	0.71	0.71	0.71	0.71	1.43	0.71	<b>4.29</b>

**Table 6.1:** Experimental results for the proposed eye detection and heterogeneous face recognition algorithm. Face recognition results show the rank-1 percentage

It can be seen that in most cases, the proposed heterogeneous score level fusion algorithm out performs all other algorithms, including Kalka et al. and the commercial software. In the

Clear w/ 80% Film Tint Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior database, the LBP-DT matcher performed the best, however, only out performing the proposed approach by  $< 6\%$ . Also, in the Solarcool (2) Graylite Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior database, the commercial software outperformed our proposed approach. Again, this only outperformed our approach by  $< 8\%$ . This shows that, even though the proposed approach is scenario dependent, it can outperform the previous photometric score level fusion and commercial software algorithms in most heterogeneous situations.

### 6.3.2 Time Efficiency

One of the main drawbacks to the algorithm proposed by Kalka et al. [20] is the length of time 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 the speed, as well as increase matching accuracy, our empirical study was performed to narrow the photometric normalization combinations down from 49 to 3. Although this helps speed up the process by approximately 18.5 times, its 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 6.2, where we can see the time it takes (in seconds) for a single probe to match a gallery images.

As we can see in Table 6.2, the proposed method, when using parallel processing, further



Algorithm	Kalka et al [20]	Proposed	Proposed w/ Parallel Processing
Avg. Time (sec)	12.059	0.650	0.207

**Table 6.2:** Results of the time efficiency test. All times are in seconds for a single probe to match a gallery face image.

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. [20]

## 6.4 Conclusion

In this chapter, 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 from the WVUMFD. We proposed an approach that enhances the capability of the original cross-photometric score level fusion proposed by Kalka et al. [20] 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 that 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% when using our ground truth data, which is about a 2.7 times improvement over the original algorithm [20] and a more than 7% 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 Clear w/ 80% Film Tint Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior database and the Solarcool (2) Graylite Glass Panel - Full interior ( $\sim 2600$  lux), 0 lux exterior database.







# Chapter 7

## Document to Live Facial Identification

### 7.1 Introduction

As discussed in Chapter 1, facial recognition has improved significantly for ideal cases such as visa and mugshot photographs [7]. However, much remains to be explored for non-ideal conditions [88]. Research has shown that the process of matching degraded facial photographs of a subjects ID documents against their high resolution counterparts (live subjects), also known as *document to live facial identification*, also fits into this category. The challenges associated with document to live facial identification can be grouped into the following three major categories [61,62]:

1. *Person-related factors* – Factors that are based on the variation of the individual’s facial appearance. Examples in this category include variations in pose, expression, and hairstyle. Aging also fits into this category because there is also a time lapse (that is often significant, i.e. up to a few years) between the comparison of the documented facial image to its higher resolution counterpart [89,90].

2. *Document-related factors* – Factors that are based on the variation of the type of document. Examples in this category include security watermarks embedded on facial images, variations in image quality and tonality across the face, and color cast of the photographs.
3. *Device-related factors* – Factors that are due to various limitations of the device. Examples in this category include limited device resolution, artifacts due to lighting, the type of image file format or compression used, and operator variability.

Document to live facial identification poses a practical problem for security officers, considering the person's identity document information can be potentially tampered, modified or stolen or even duplicated into another document that can be used for illegal or unauthorized purposes. Therefore, even if a state-of-the-art automated facial recognition (FR) system is used, such a challenging scenario as document to live face matching may not allow for the identification system to achieve the same level of performance that can be achieved in ideal conditions. As a result, the officer may decide to no longer rely on the FR system and try to manually match the query face image against a set of face images in a database (e.g. watch-list) by making visual comparisons. Such a process is prone to errors but, more importantly, prone to significant delays. Because of this, work is needed to design and develop a preprocessing methodology that is capable of restoring the degraded photographs from identity documents prior to comparing them against live (gallery) face images. In this regard, an investigation into several image restoration schemes in combination with academic and commercial face matchers is performed. Then, through an empirical evaluation study, the conditions that result in the highest identification rates are determined. For the purpose of this study, the West Virginia University (WVU) Identity Document Database, which is

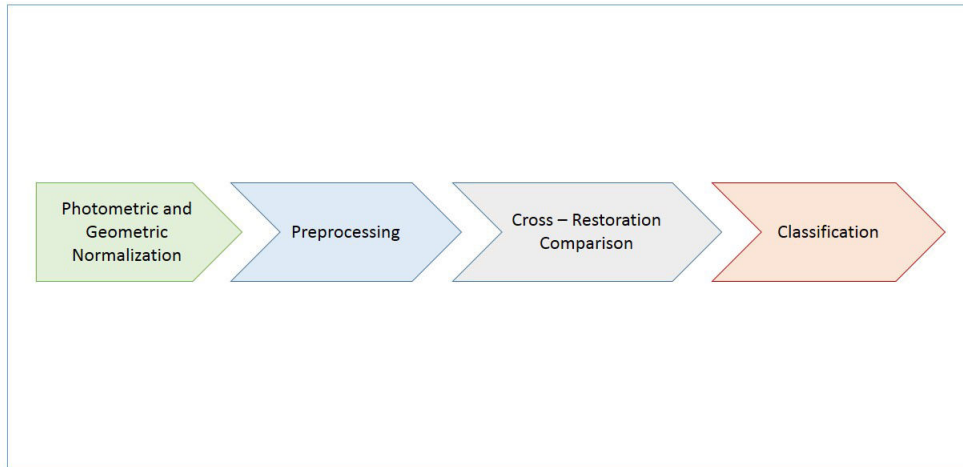
composed of 130 subjects coming from various countries, is used.

### 7.1.1 Analogy to Cross-Spectral FR

Similar to challenges with matching across the electromagnetic (EM) spectrum, document to live facial matching involves matching a low-quality facial image with its high resolution counterpart. Therefore, this process is dependent on the preprocessing element to improve image quality for better recognition. Also, because current automated facial recognition techniques work best for ideal situations, heterogeneous facial matching is used to evaluate the overall effectiveness in recognition performance. Heterogeneous facial matching has been developed for matching across the EM spectrum, which involves comparing the images from the visible spectrum to images acquired from different infrared spectrums, such as NIR, SWIR, MWIR and LWIR. Heterogeneous face recognition, proposed in [91], first was performed by matching NIR to visible images directly using random sub-space projections as well as using sparse representation classification. By using difference of Gaussian filtering with local encoding and feature normalization, the heterogeneous difference between the two spectral bands was alleviated.

In the case of document to live facial matching, heterogeneous facial matching is also adopted. However, instead of performing cross-spectral comparisons, *cross-restoration* comparisons are performed to evaluate identification performances. For cross-restoration recognition, comparisons between the original documented facial image, the restored facial images, and their ideal counterparts are performed. An advantage to heterogeneous facial matching, as illustrated in Figure 7.1, is that a better correlation between the visible output and the automated restoration is provided. Humans inherently have the ability to correlate the



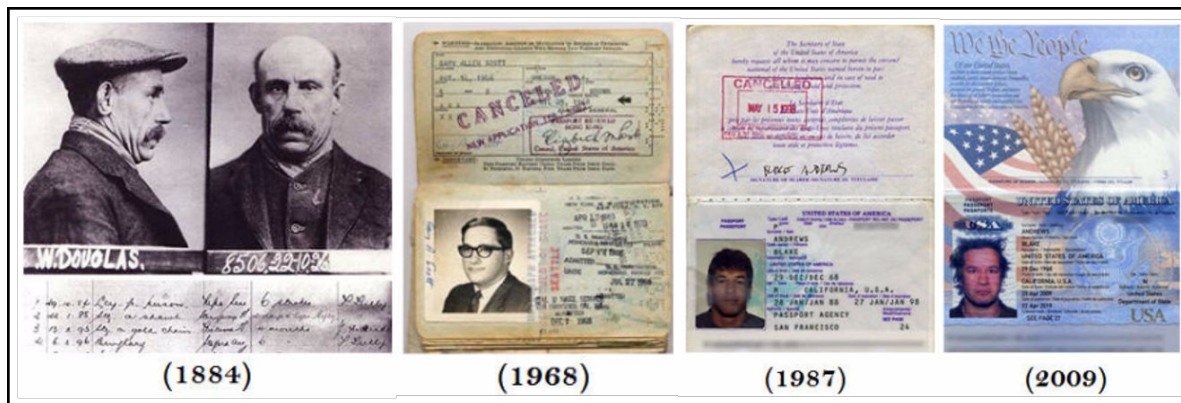


**Figure 7.1:** Flow chart representation illustrating the resemblance to matching across the electromagnetic (EM) spectrum where cross-restoration comparison is analogous to cross-spectral comparison throughout the EM domain.

improvement in image quality to better identify a person of interest, yet lack the speed to process the correlation efficiently (particularly when it comes to vast amounts of data). Therefore, the major challenge faced is that of translating these human capabilities to an algorithm in order to take advantage of a computers processing power. Conducting cross-restoration comparisons minimizes these issues while gaining a complete perspective of the restoration process.

### 7.1.2 Passport Standardization

Figure 7.2 demonstrates that identification documents have changed dramatically over the years. It is also important to acknowledge that recent developments have been made to meet society's contemporary needs. A set of recent ICAO standards have been incorporated for



**Figure 7.2:** A collection of passports ranging from 1884 to 2009. All are real passports and are available in the public domain [92].

*biometric passports* or *e-passports*. For these e-passports, critical information pertaining to the traveler is printed on the data page of the passport as well as stored in a radio-frequency identification (RFID) chip [92]. Additionally, face, fingerprint and/or iris templates of the legitimate owner are embedded as an additional layer of protection. However, not all passports are biometric friendly and there is a possibility that the information stored in the RFID chip could be compromised [61]. For these instances, facial recognition technology can be used as an alternative approach in order to confirm the traveler's identity. The limitations of facial recognition, in this context, are primarily due to document and device-related factors\* – particularly in the case of recent documents. In recent identification (ID) documents, security watermarks are observed across the face. Also, in some cases, a strong magenta cast can be observed on the photograph of the ID document. In both paradigms, the image quality can be severely impacted and the facial tonality can be diminished. Therefore, care must be taken to overcome these limitations of degraded photographs to improve image quality

---

\*Person-related factors are affected here due to the time lapse between the document and ideal images used in comparison

before face matchers are applied.

## 7.2 Pre-processing Techniques

This section provides the theoretical framework of the fundamental restorative building blocks used that include threshold-based (TB) denoising, total variation (TV) wavelet inpainting, and exemplar-based (EB) inpainting. Understanding these key building blocks provides the reader an analytical foundation of each independent strategy. Additionally, this understanding also helps compliment the experimental testing and observations of each of the key building blocks and their combination described in the following section.

### 7.2.1 Threshold-Based (TB) Denoising

Threshold-based (TB) denoising is the first building block that was incorporated into the restoration strategy to remove the noisy effects of various security markings. This generalized approach is done by considering the following major steps [62, 75]:

1. *Compute the Discrete Wavelet Transform* – First, the discrete wavelet transform (DWT) of the image is needed to convert the noisy image to the wavelet domain.
2. *Apply Thresholding Estimator* – Next, the thresholding estimator is needed to suppress the extraneous artifacts that correspond to the security watermark.
3. *Compute the Inverse Discrete Wavelet Transform* – Finally, the inverse discrete wavelet transform (IDWT) is used to reconstruct the resulting thresholded image.

In order for Step 2 to be the most effective, the thresholding estimator and value need to be selected in order to depict the overall denoising effectiveness. There are different estimators based on the threshold value calibration methods such as hard, soft or semi-soft thresholding. These settings are chosen based on the level of noisiness of the image where, for example, if an image is severely noisy then hard thresholding is chosen. Conversely, for images that have medium or rare levels of noisy behavior, semi-soft or soft thresholding levels are chosen, respectively. Each estimator removes unnecessary coefficients via the following equation

$$\hat{h} = \sum_{|\langle h, \psi_m \rangle| > T} \langle h, \psi_m \rangle \psi_m = \sum_m \Omega_T^q(\langle h, \psi_m \rangle) \psi_m, \quad (7.1)$$

where  $h$  is the noisy observation,  $\psi_m$  the mother wavelet function,  $m = (i, j)$  represents the scaling and translation, respectively, of the wavelet basis function,  $q$  the thresholding type, and  $T$  the threshold value. Given a chosen wavelet basis function  $\psi_m$ , equation (7.1) filters out the coefficients beyond the set threshold value  $T$  via the thresholding estimator  $\Omega_T^q$ . In terms of the input parameter  $x$ , the thresholding estimator  $\Omega_T^q$  is defined as

$$\Omega_T^H(x) = \begin{cases} x, & \text{for } |x| > T \\ 0, & \text{for } |x| \leq T, \end{cases} \quad (7.2)$$

$$\Omega_T^S(x) = \begin{cases} \text{sgn}(x) \cdot (|x| - T), & \text{for } |x| > T \\ 0, & \text{for } |x| \leq T, \end{cases} \quad (7.3)$$

and

$$\Omega_T^{SS}(x) = \begin{cases} 0, & \text{for } |x| \leq T \\ x, & \text{for } |x| > \mu T \\ \text{sgn}(x) \cdot \frac{|x-T|}{\mu-1}, & \text{otherwise,} \end{cases} \quad (7.4)$$

where  $\mu > 1$  and the superscripts  $H$ ,  $S$ , and  $SS$  denote hard, soft, and semi-soft thresholding, respectively. The advantage of applying TB nonlinear denoising, is that it is independent regardless of the type and condition of the acquired document.

It is important to note that image denoising using traditional (or decimated) orthogonal wavelets may exhibit visual artifacts due to the lack of translation invariance of the wavelet basis. Additionally, artifacts due to Gibbs phenomenon may occur in the neighborhood of discontinuities. To suppress such artifacts, the artifacts are “average out” via the translation dependence through “cycle spinning” defined by the following equation [65]:

$$\Theta_{TI}(h) = \frac{1}{|\Phi|} \cdot \sum_{\tau \in \Phi} \Theta(h_\tau)_{-\tau}, \quad (7.5)$$

where for  $\forall \tau \in \Phi$ ,  $\Theta_{TI}(h) = \Theta_{TI}(h_\tau)_{-\tau}$ . Here,  $\Phi$  is a lattice of  $\mathbb{R}^2$  for a two-dimensional image. This is called *cycle spinning* denoising. Equation (7.5) says that if there exists  $N$ -sample data, then pixel precision translation invariance is achieved by having  $N$  wavelet translation transforms (vectors) or  $|\Phi| = N$ . This results in  $O(N^2)$  operations. Therefore, similar to cycle spinning denoising, *thresholding-based translation invariant* denoising can be defined as:

$$\Theta_{TI}(h) = \frac{1}{|\Phi|} \cdot \sum_{m, \tau \in \Phi} \Omega_T^q(\langle h, (\psi_m)_\tau \rangle) (\psi_m)_\tau \quad (7.6)$$

The advantage of employing translational invariant wavelets is that using them improves the signal to noise ratio (SNR). Additionally, incorporating the averaging process, defined by equation (7.5), significantly reduces the oscillating artifacts [62]. Furthermore, it can be noted that additional SNR improvement can be achieved by properly selecting the threshold estimator  $T$ .

### 7.2.2 Total Variation Wavelet Inpainting

Although threshold-based denoising provides an approach that is both autonomous and adjustable, a key limitation is that traces of the security markings are still left in the image. To overcome these drawbacks, total variational (TV) wavelet inpainting in conjunction with threshold-based denoising was introduced [75]. The theoretical premise of TV wavelet inpainting is to first consider the following standard image model within the image domain  $\Omega$

$$v(x) = u(x) + n(x), \quad (7.7)$$

where  $u(x)$  represents the image not containing the watermark,  $n(x)$  represents the watermark itself, and  $v(x)$  represents the passport image. Next, let  $v(\alpha, x)$  and  $u(\beta, x)$  be the wavelet transforms given by

$$v(\alpha, x) = \sum_{j,k} \alpha_{j,k} \psi_{j,k}(x), \quad j \in \mathbb{Z}, k \in \mathbb{Z}^2 \quad (7.8)$$

and

$$u(\beta, x) = \sum_{j,k} \beta_{j,k} \psi_{j,k}(x), \quad j \in \mathbb{Z}, k \in \mathbb{Z}^2, \quad (7.9)$$

where  $\alpha = (\alpha_{j,k})$  are the stored wavelet coefficients in the image  $v$  and  $\beta = (\beta_{j,k})$  are the stored wavelet coefficients in the image  $u$ . Here,  $\beta$  are the *desired* coefficients and  $\alpha$  are the *damaged* coefficients where the goal is to recover as many desired coefficients while repairing the damaged coefficients. Considering the fact that additional noisy behavior can exist within the image, the following TV model is proposed [74]:

$$\min_{\beta_{j,k}} F(u, v) = \int_{\Omega} |\nabla_x u(\beta, x)| dx + \sum_{j,k} \lambda_{j,k} (\beta_{j,k} - \alpha_{j,k})^2 \quad (7.10)$$

subject to the constraint

$$\beta_{j,k} = \alpha_{j,k}, \quad (j, k) \in I, \quad (7.11)$$

where  $I$  represents the region of interest to be inpainted. Noting equation (7.10) the operator  $\nabla_x$  represents the spatial gradient ( $x \in \mathbb{R}^2$ ) and the parameter  $\lambda_{j,k}$  is a positive constant such that  $\lambda_{j,k} = 0$  in the inpainted region  $I$  otherwise it equals some positive constant to be chosen ( $\lambda > 0$ ). Minimizing equation (7.10) with respect to  $\beta = \beta_{j,k}$  achieves the following result

$$\frac{\partial F}{\partial \beta} = - \int_{\Omega} \nabla \cdot \left[ \frac{\nabla u}{|\nabla u|} \right] \psi_{j,k} dx + 2\lambda_{j,k} (\beta - \alpha), \quad (7.12)$$

where

$$\nabla \cdot \left[ \frac{\nabla u}{|\nabla u|} \right] \quad (7.13)$$

represents the curvature formula for the level lines of  $u$ . Taking into account the variational form, given by (7.12), TV wavelet inpainting provides a sophisticated process that accounts for variegated security markings; however, residual noisy effects are still eminent. Therefore, threshold-based denoising is still considered to remove the extraneous noise improving both image quality and recognition performance.

### 7.2.3 Exemplar-Based Inpainting

The effects of exemplar-based (EB) inpainting [93] as a potential preprocessing scheme was also investigated. The effectiveness of this method stems from the computation of patch priorities within the watermarked region  $\Omega$  of the image  $I$  in the pixel space. Given a patch  $\psi_p$  centered at the point  $p \in \partial\Omega$  where  $\partial\Omega$  represents the boundary of the region of interest, the *patch priority*  $P(p)$  is defined as the product between the *confidence term*  $C(p)$  and the *data term*  $D(p)$  given by [93]

$$P(p) = C(p)D(p) \quad (7.14)$$

where  $C(p)$  and  $D(p)$  are given by

$$C(p) = \frac{\sum_{q \in \psi_p \cap (I - \Omega)} C(q)}{|\psi_p|} \quad \text{and} \quad D(p) = \frac{|\nabla I_p^\perp \cdot \mathbf{n}_p|}{\alpha}. \quad (7.15)$$



It is important to note that at point  $p$  in equation (7.15)  $|\psi_p|$  is the area of the patch,  $\alpha$  is the normalization factor,  $\nabla I_p^\perp$  is the isophote representing both the direction and the intensity, and  $\mathbf{n}_p$  is the normal vector to the boundary  $\partial\Omega$ . By computing  $D(p)$ , the behavior of the isophotes within the targeted region  $I - \Omega$  are considered. Also, the measure of the reliable information surrounding the point  $p$ , defined by  $C(p)$ , is also estimated. From the patch priorities, texture and structure information is propagated by first finding the patch with the highest priority. Next, the *source exemplar* of the image is found by directly sampling the source region  $\psi_q$ , that is most similar to the patch region  $\psi_{\hat{p}}$ , given by the following equation

$$\psi_{\hat{q}} = \arg \min_{\psi_q \in (I - \Omega)} d(\psi_{\hat{p}}, \psi_q), \quad (7.16)$$

where  $d(\psi_{\hat{p}}, \psi_q)$  is the distance measure between the two patches  $\psi_{\hat{p}}$  and  $\psi_q$ . After computing the source exemplar, the value of each pixel to-be-filled is copied from the corresponding position within the estimated source region  $\psi_{\hat{q}}$ . Equations (7.14) - (7.16) pose an iterative process that interpolates the missing information within the security watermarked region.

### 7.3 Experimental Evaluation

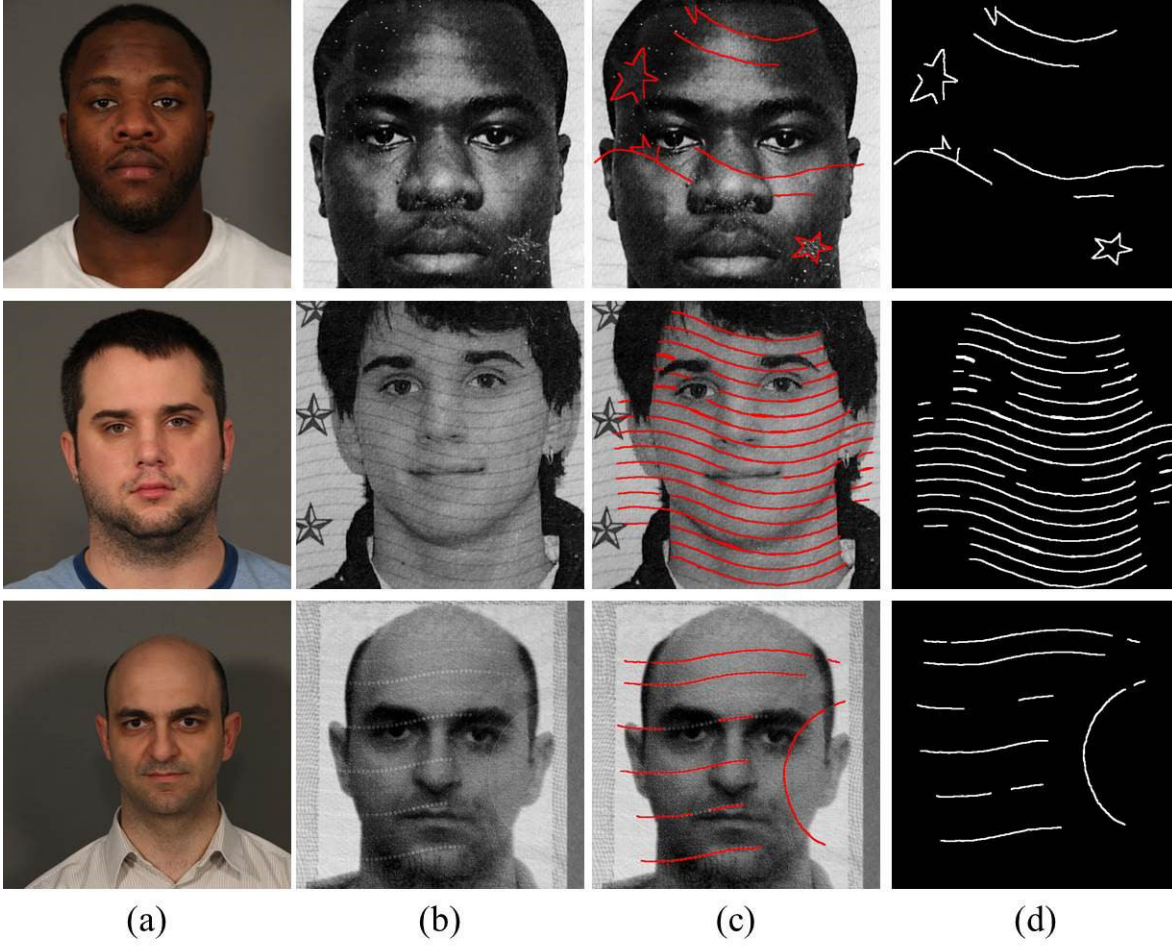
This section describes several evaluation procedures and methods to assess improvement in both image quality and recognition performance. Our experimental evaluations were performed on the *West Virginia University (WVU) Identity Document Database*, which is composed of 130 subjects, taken from various documents internationally. The types of documents composed in this dataset consist of passports, driver's licenses, and state identifications

from within the United States as well as various international regions such as the Middle East and European Union. Example images from this database can be seen in Figure 7.3. In preparation for empirical testing, manually annotated regions of interest (ROI) corresponding to the security watermark are gathered. Next, the preprocessing schema's described in 7.2, are applied. Furthermore, advanced preprocessing schemes, i.e. the combinations of TB denoising is independently coupled with TV wavelet inpainting and EB inpainting.

Each baseline and advanced preprocessing scheme is tested in terms of 1.) improvement in image quality and 2.) improvement in recognition performance. Understanding the improvement in image quality, via the Universal Image Quality (UIQ) metric [94], is important in terms of being able to evaluate and determine which preprocessing schemes are most robust. Next, investigating the improvement in recognition performance is needed to show which restoration method improves identification accuracy. The usage of academic algorithms, such as LBP and LTP texture-based approaches, are proven to be beneficial. This allows for a baseline performance metric. A commercial (state-of-the-art) algorithm is also used to show performance variations when employing the different image restorations approaches. Concurrently, the impact that mask selection (in terms of pixel width) has on identification performance is also studied. This is due to the fact that the aforementioned preprocessing schemes are heavily dependent on the level of mask annotation.

### 7.3.1 Evaluating Image Quality

Assessing improvement in image quality is performed using the UIQ metric, where the outputs of both restored image  $x$  and the ideal counterpart  $y$  governed are compared by the



**Figure 7.3:** Sample images used from the WVU Identity Document Database: a) ground truth gallery images, b) original document images, c) manually annotated passport masks in red, and d) the resulting masks marking the regions of interest (ROI).

following equation [94]:

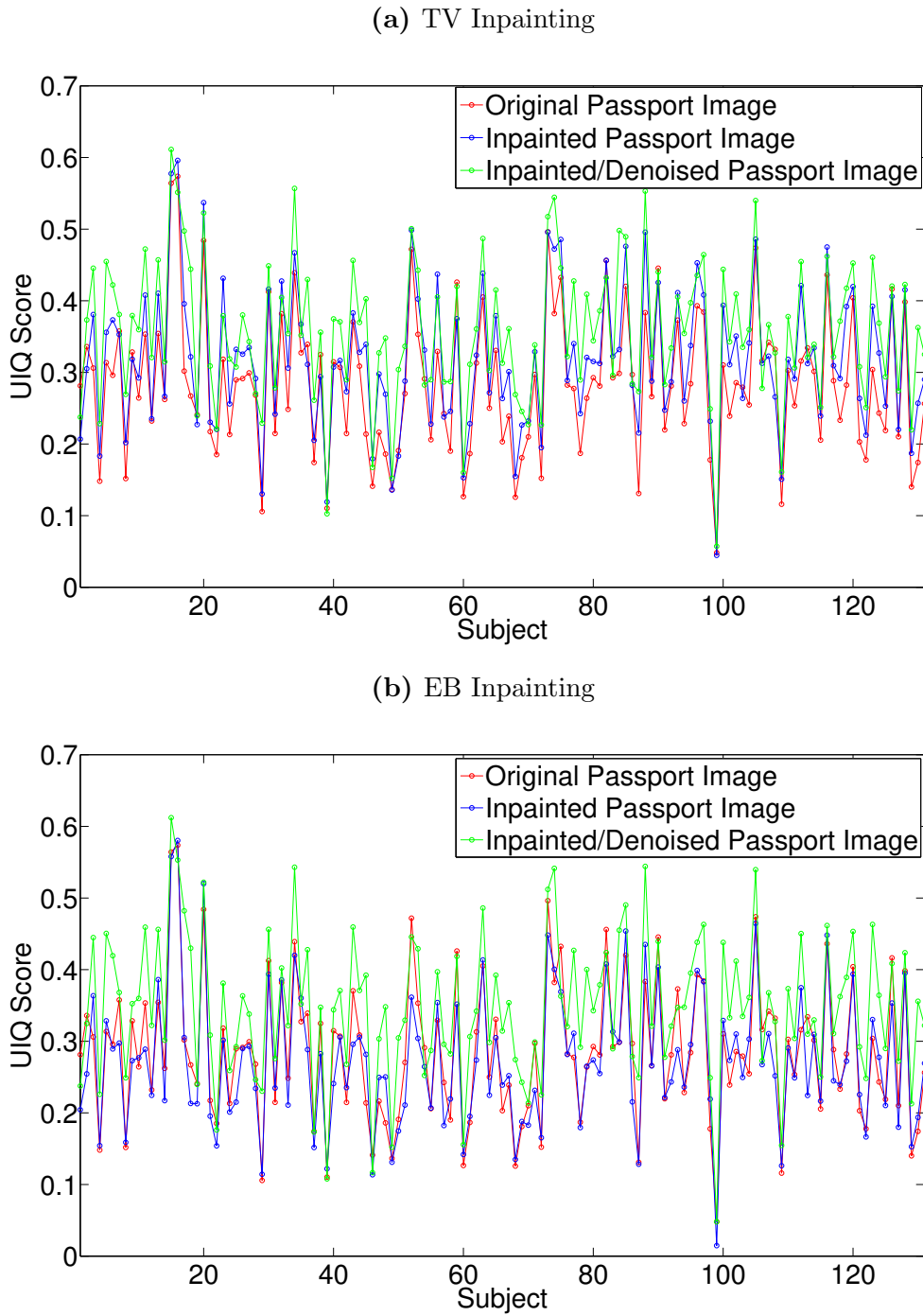
$$UIQ = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{\bar{x}^2 + \bar{y}^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 \sigma_y^2}, \quad (7.17)$$

where  $\bar{x}$  and  $\sigma_x^2$  are the sample mean and variance of the restored image,  $\bar{y}$  and  $\sigma_y^2$  of the ideal counterpart, and  $\sigma_{xy}$  is the covariance between the two. Using the UIQ metric, in this case,

evaluates the improvement in restoration from three main perspectives: loss of correlation, luminance distortion, and contrast distortion. Considering these factors together,  $UIQ \in [0, 1]$  such that as  $UIQ \rightarrow 1$  the restored image reaches maximum improvement. Figure 7.4 shows this trend where, for all 130 subjects, the comparisons between the previously mentioned restoration strategies are shown. Overall, it can be seen that using inpainting coupled with TB denoising shows a consistent trend as far as improvement in image quality – regardless of the type of inpainting method used. In other words, it is observed that using TV wavelet inpainting or EB inpainting in the inpainting/denoising combination, shows almost an identical trend in performance when evaluations are made on the entire dataset. From this assessment, intuitively one would expect that, when investigating rank-1 identification accuracy, there would be a similar trend in behavior.

### 7.3.2 Evaluating Texture Based Approaches

To show the how the applied restorations schemes fair in terms of improvements in identification, two academic texture based approaches to provide a baseline performance were used, namely local binary patterns (LBP) and local ternary patterns (LTP). The LBP image patterns are computed by thresholding 3x3 neighborhoods based on the value of the center pixel. Next, the resulting binary pattern is transformed to a decimal value where 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  $LBP_{P,R}^{u^2}$ , 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. The binary pattern for pixels that lie in a circle  $f_p$ ,  $P = 0, 1, \dots, P - 1$  with the center pixel  $f_c$ , is computed as



**Figure 7.4:** An illustration of the evaluation of the improvement in image quality where the UIQ metric is used to evaluate the restoration performance using a) TV wavelet and b) EB inpainting where each inpainting method was also coupled with threshold-based (TB) denoising. Note that the inpainting/denoising combination is the most effective regardless of the inpainting method that was employed.

$$f(n) = \begin{cases} 1, & \text{if } f_p - f_c \geq 0 \\ 0, & \text{if } f_p - f_c < 0. \end{cases} \quad (7.18)$$

From (7.18), a binomial weight of  $2^P$  is assigned to each sign  $S(f_p - f_c)$  to compute the respective LBP code given by

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c) 2^p. \quad (7.19)$$

The advantage of using LBP operators is that they are invariant to monotonic gray level transformations; however, one disadvantage is due to the noise sensitivity because thresholding is done by considering the center of the pixel region. Hence, local ternary patterns (LTP) are introduced where the quantization is performed as follows

$$f(n) = \begin{cases} 1, & \text{if } f_p - f_c \geq t \\ 0, & \text{if } |f_p - f_c| \leq t \\ -1, & \text{if } f_p - f_c \leq -t. \end{cases} \quad (7.20)$$

The resulting output 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.

### Distance Metrics

After the texture based feature patterns are computed, two different distance metrics are used to obtain the final match score, namely the distance transform (DT) and the chi-squared ( $\chi^2$ ). The distance transform (defined as the distance or similarity metric from image X to image Y) is given by:

$$DT(X, Y) = \sum_{Y(i,j)} w(d_x^{K_Y(i,j)}(i, j)), \quad (7.21)$$

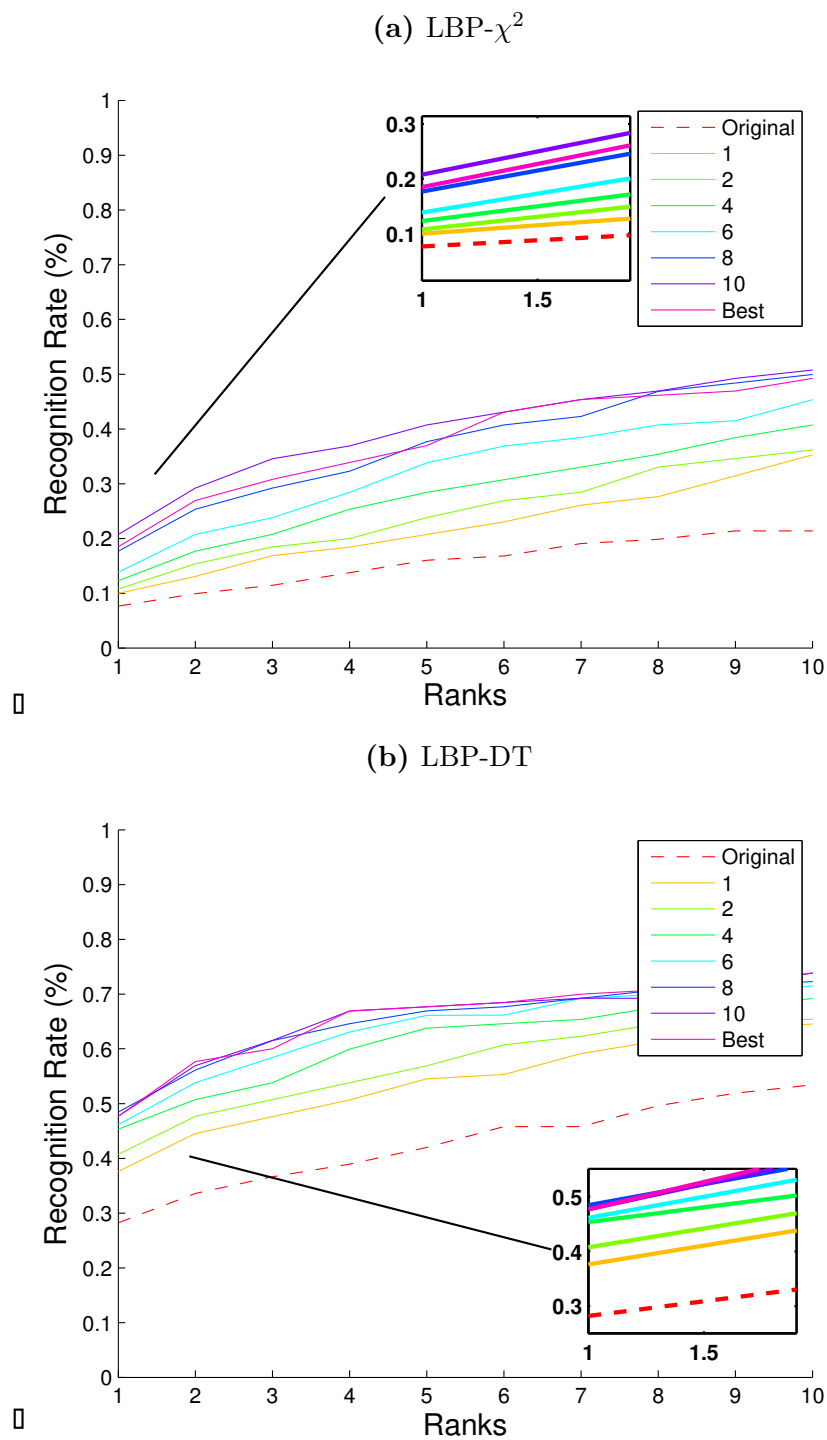
where  $K_Y(i,j)$  is the code value of pixel  $(i, j)$  of image Y and  $w$  is a user controlled penalty function. Additionally, the chi-squared distance is given by :

$$\chi^2(n, m) = \frac{1}{2} \sum_i^l \frac{h_n(k) - h_m(k)}{h_n(k) + h_m(k)} \quad (7.22)$$

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.

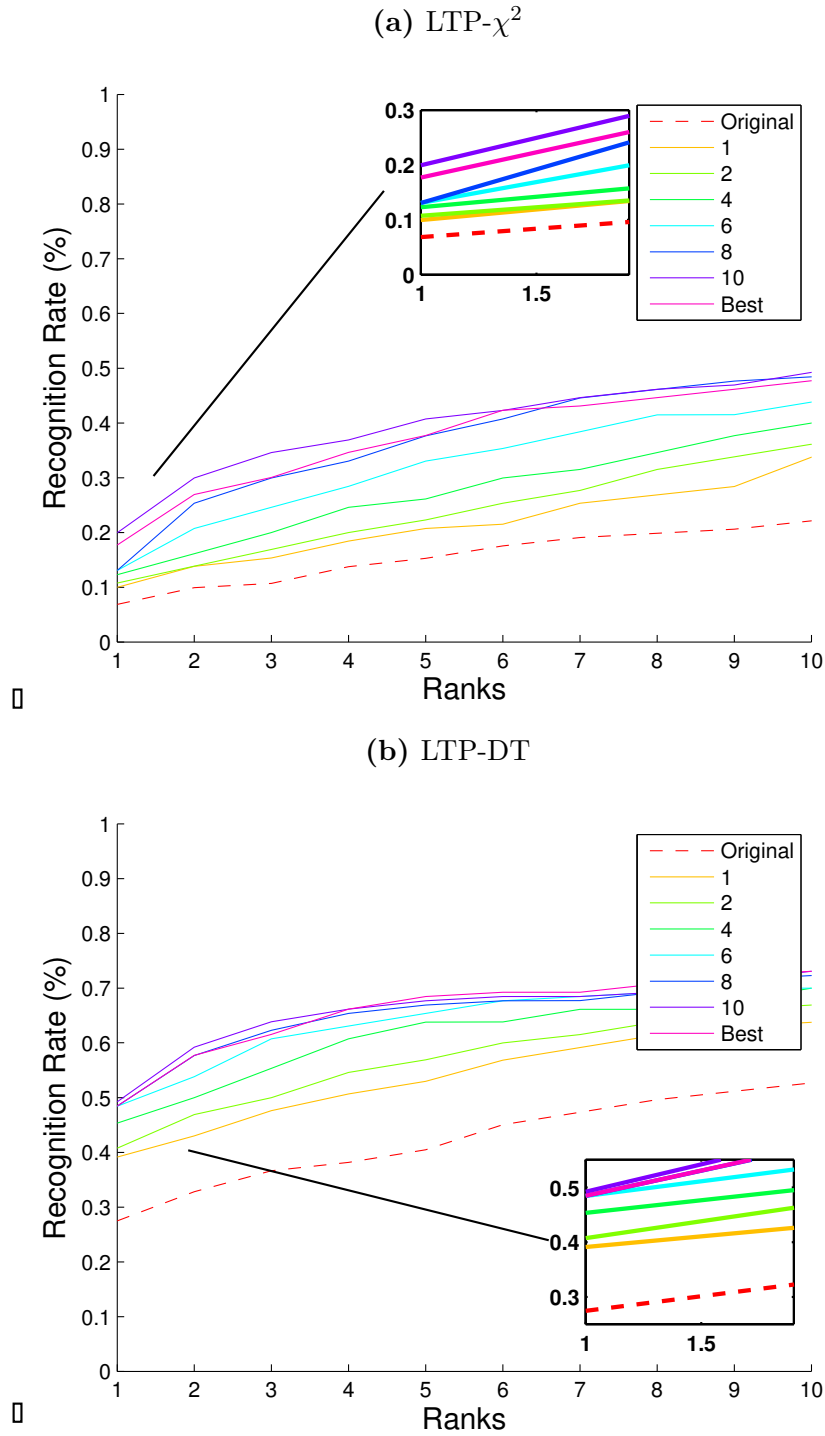
### 7.3.3 Assessing the Effects of Mask Dilation

The observations of the previous works [75] show that the pre-processing methodology depends heavily on mask annotation. Therefore, there is a need to explore and evaluate the impact of the mask generation in the context of facial recognition. For this aspect, the masks for each subject are demarcated to 1 pixel width and dilated to create different mask widths of 2, 4, 6, 8, and 10 pixel widths, respectively. Next, the image is interpolated using different preprocessing schemes. First, a focus on exploring the effects of TV wavelet inpainting

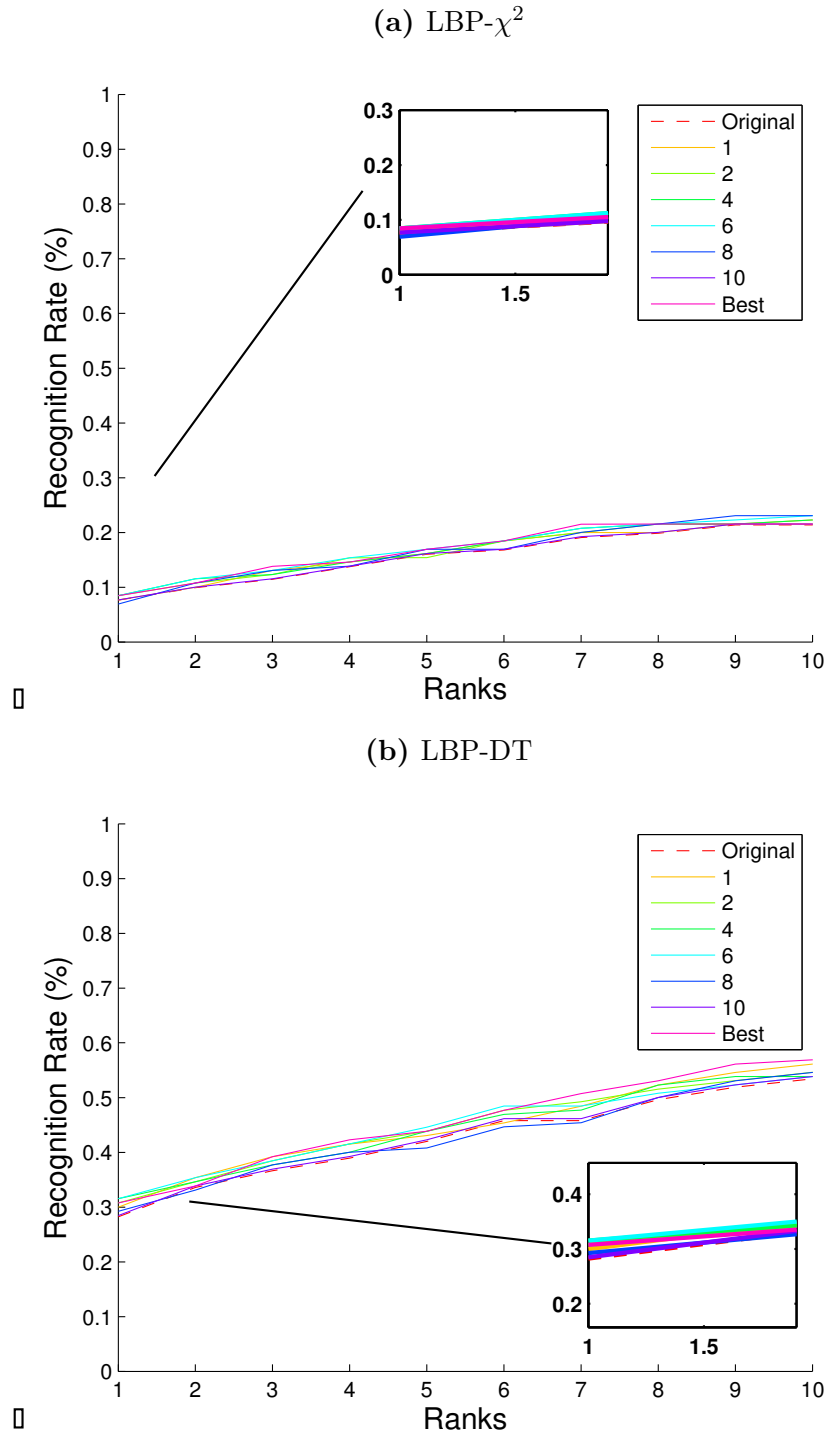


**Figure 7.5:** Identification results that solely use TV wavelet inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.

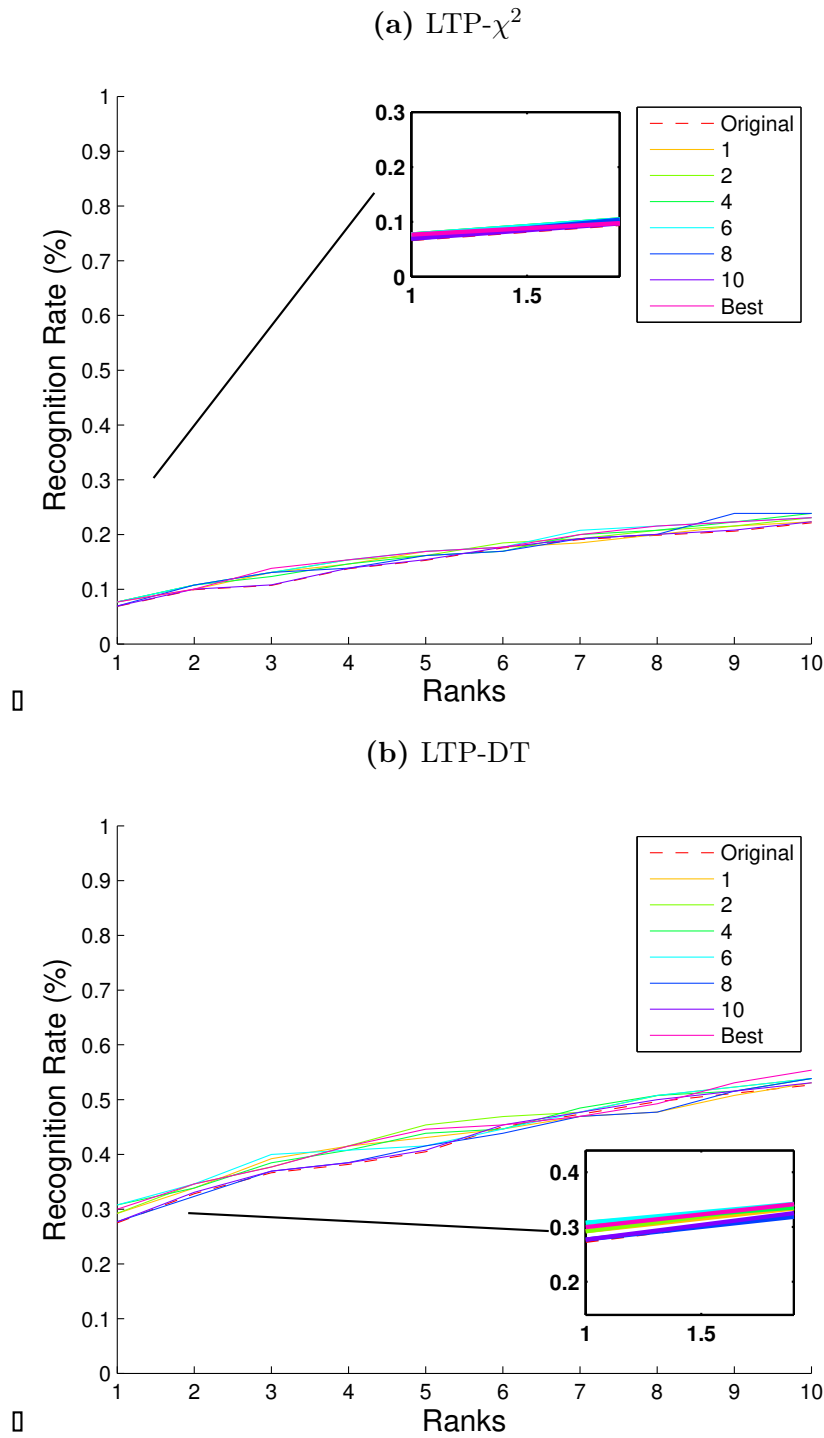




**Figure 7.6:** Identification results that solely use TV wavelet inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.



**Figure 7.7:** Identification results that solely use Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.



**Figure 7.8:** Identification results that solely use Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.

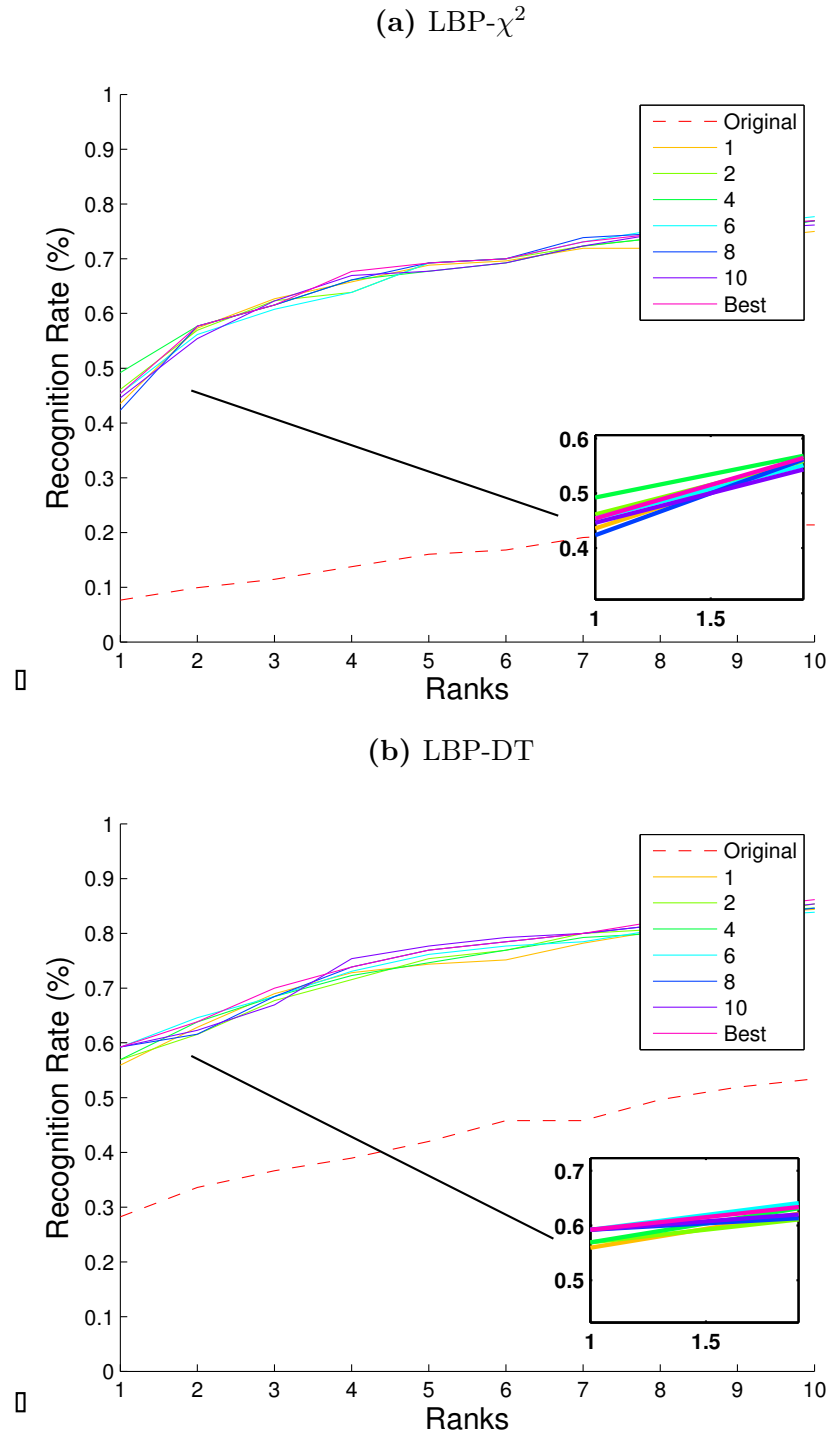
and EB inpainting is conducted. Texture based approaches are applied to show recognition performance considering it is observed that the ideal mask width may be distinctive for each documented image. Because of this, the recognition performance between original and best images are compared. Here, the original image is the raw passport image and the best image represents the restored image that corresponds to the optimum local mask dilation that is determined from the maximum UIQ metric. It is important to note that the mask dilation that produces the best image is not universal. For example, a best image could correspond to a mask dilation of 2 pixels and another could correspond to a mask dilation of 6 pixels. Applying these methods of comparison helps to observe the correlations between the improvement in image quality, optimum mask dilation, and recognition performance.

Comparing Figures 7.57.5 and 7.77.7, it is observe that applying TV wavelet inpainting shows more sensitivity in terms of mask generation than EB inpainting in terms of improvement in rank-1 accuracy. This is regardless of the texture based approach that was applied. For example, observing Figure 7.5 (a), a deviation of approximately 15% in terms of rank-1 performance when LBP with the  $\chi^2$  distance metric is seen. However, from Figure 7.7 (a), a much closer deviation in performance regardless of mask dilation when the same texture-based academic scheme is used for images that were preprocessed via EB inpainting is observed. Although applying the distance transform (DT) metric showed an improvement in rank-1 performance of approximately 20% when compared to the  $\chi^2$  metric, a similarity in behavior in terms of the impact that mask width has on identification performance can be seen. Furthermore, it can be noted that the recognition rates that correspond to the images are near the highest in terms of identification accuracy.

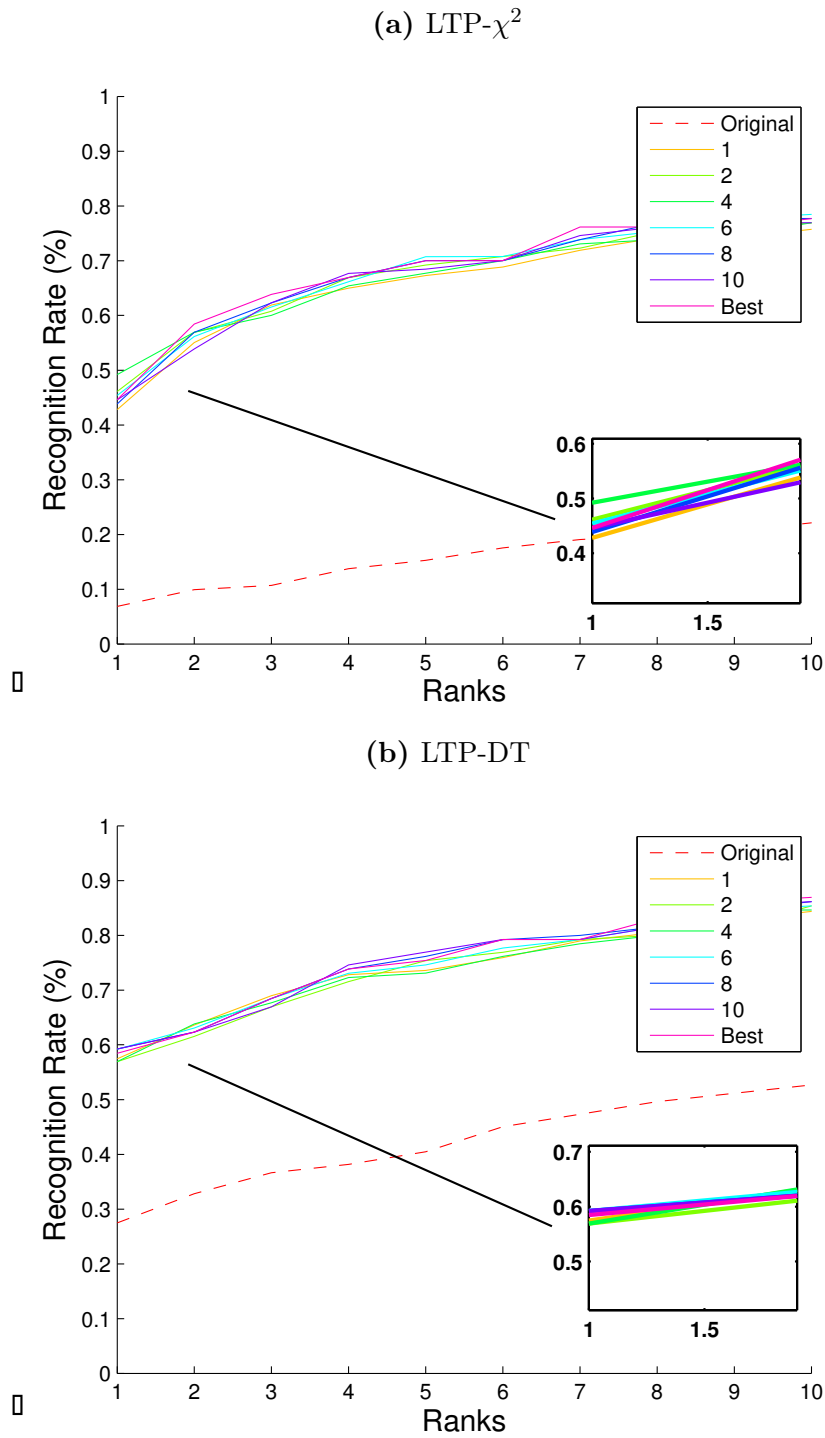
### 7.3.4 Assessing the Effects of Threshold-Based Nonlinear Denoising

It has been previously noted that, for documented facial images, there exist additional artifacts that affect the noisy behavior of the image, which can result from either document or device-related factors [61,62]. Furthermore, applying threshold-based (TB) denoising has shown to minimize this additional noisy behavior [75]. Therefore, care must be made to investigate the effects of coupling TB denoising with either inpainting method, while considering the effects of mask dilation mentioned in the previous section. We do this by first constructing two different types of training sets, where each set is denoised incrementally, varying the denoising level within the range of  $\sigma \in [2, 30]$ . Next, at each increment, the LTP-DT score is computed between the inpainted/denoised image and the ideal image and stored. It is important to note that, for each training set, the mean denoising level  $\bar{\sigma}$ , associated with the minimum LTP-DT score is also considered. This sample mean is used to denoise all of the respective documented facial images associated with the particular mask width.

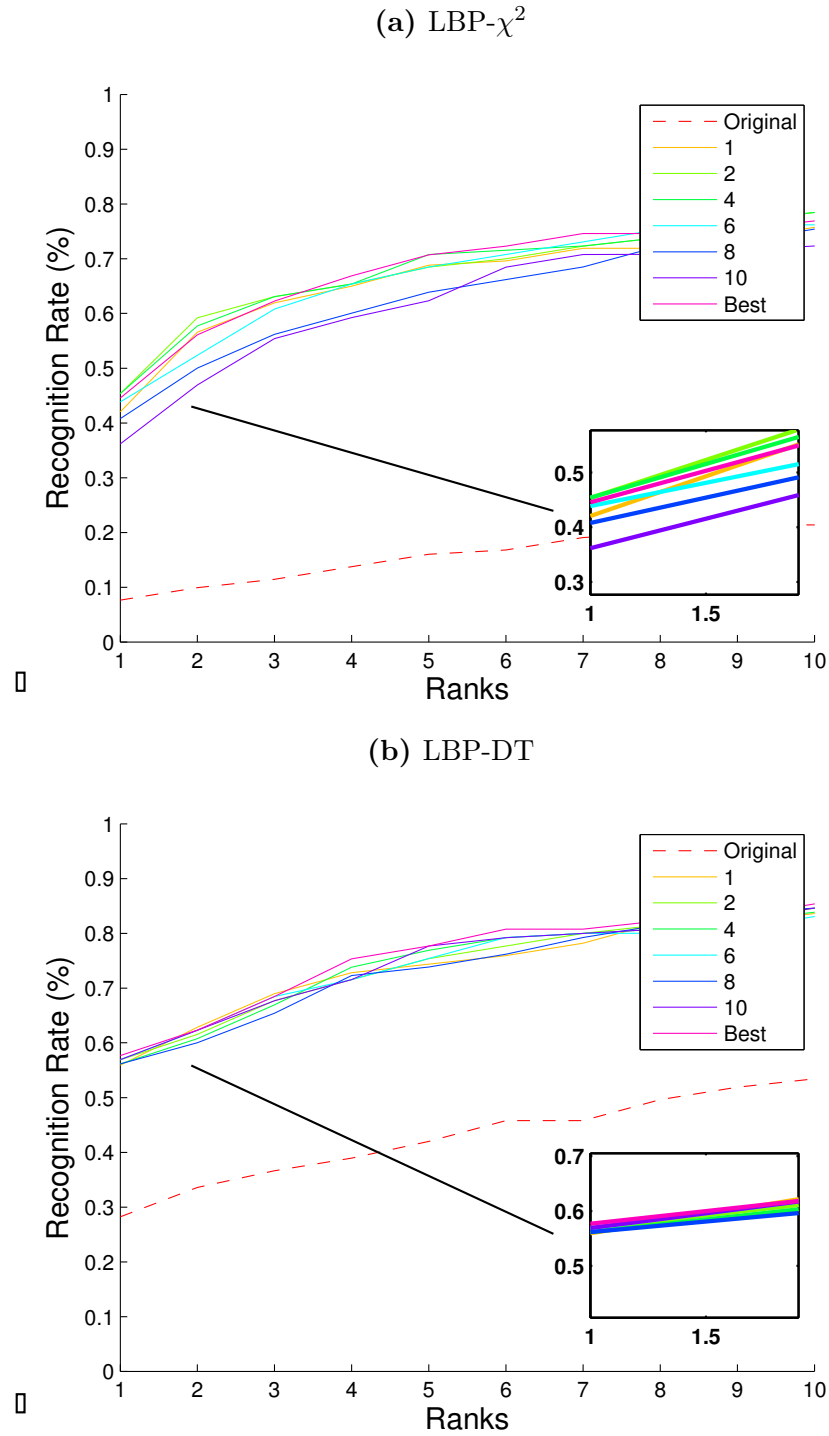
Examining Figures 7.97.10 and 7.117.12, it is seen that TB denoising showed dramatic improvement in rank-1 accuracy. This result was expected due to previous observations, i.e. when TV wavelet inpainting was applied [75]. However, closer inspection into the results show the trade-offs between using TV wavelet inpainting, EB inpainting, and mask dilation. Examining Figure 7.97.10, note that the coupling of TV wavelet inpainting with TB denoising showed less sensitivity in terms of mask annotation. In other words, it is observed that improvement in performance is independent of mask annotation when TV wavelet inpainting is used. However, from Figure 7.117.12, it is seen that there is more



**Figure 7.9:** Identification results that use denoising and Total Variation based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.

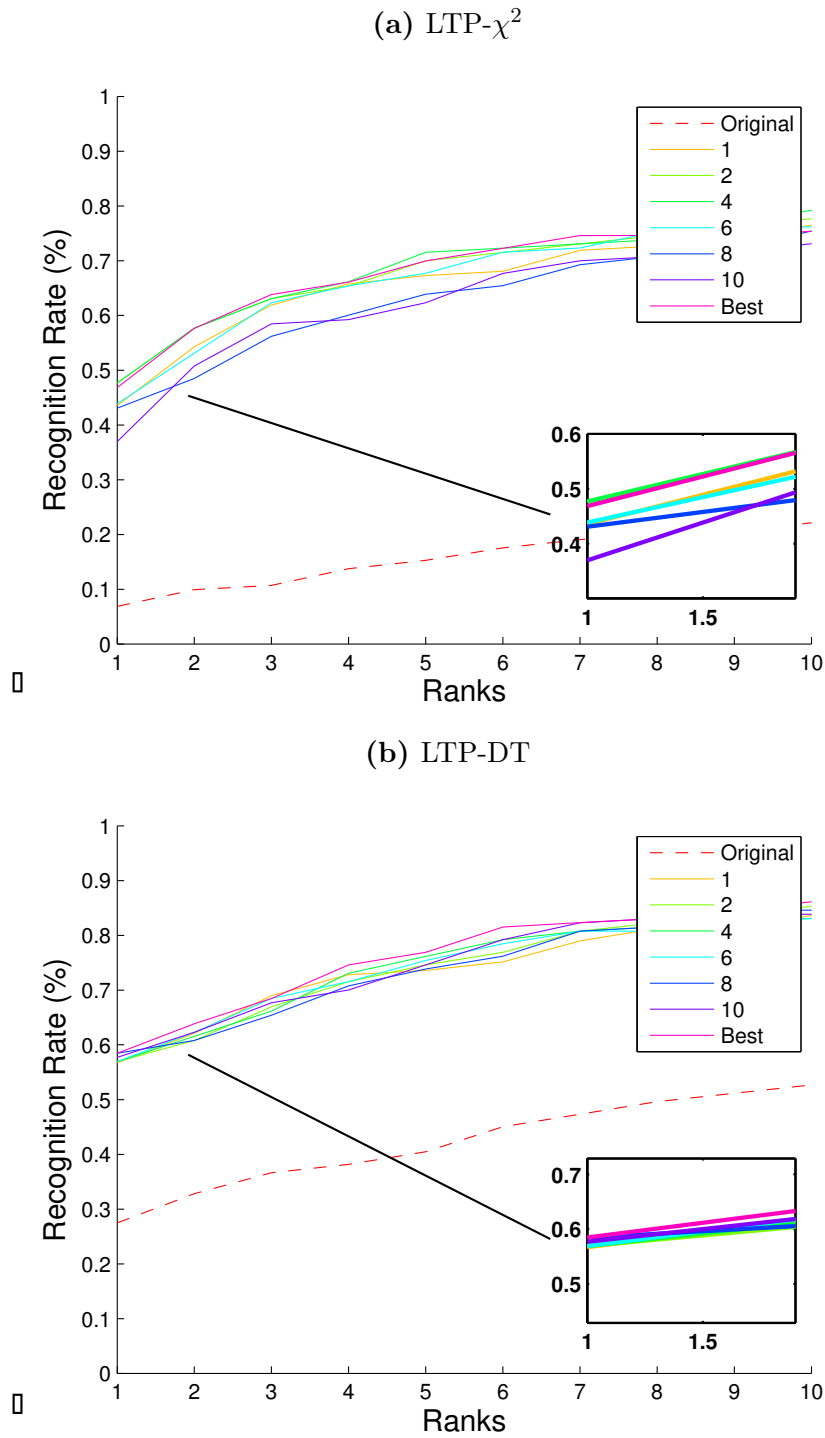


**Figure 7.10:** Identification results that use denoising and Total Variation based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.



**Figure 7.11:** Identification results that use denoising and Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LBP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.





**Figure 7.12:** Identification results that use denoising and Exemplar Based inpainting as the preprocessing scheme while comparing various mask widths (in pixels) to the original and best images. Here, LTP is used with different distance metrics of  $\chi^2$  and DT. Notice the sensitivity in this method when different mask dilations are chosen in both cases.

sensitivity in terms of mask selection when EB denoising is applied. Interestingly, it is observed that this variability is dependent on the distance metric applied. For example, comparing Figure 7.11 (a) to 7.11 (b), it is shown that when changing the distance metric from  $\chi^2$  to distance transform (DT), there is less sensitivity in terms of mask annotation.

### 7.3.5 Performance Assessment using a Commercial Matcher

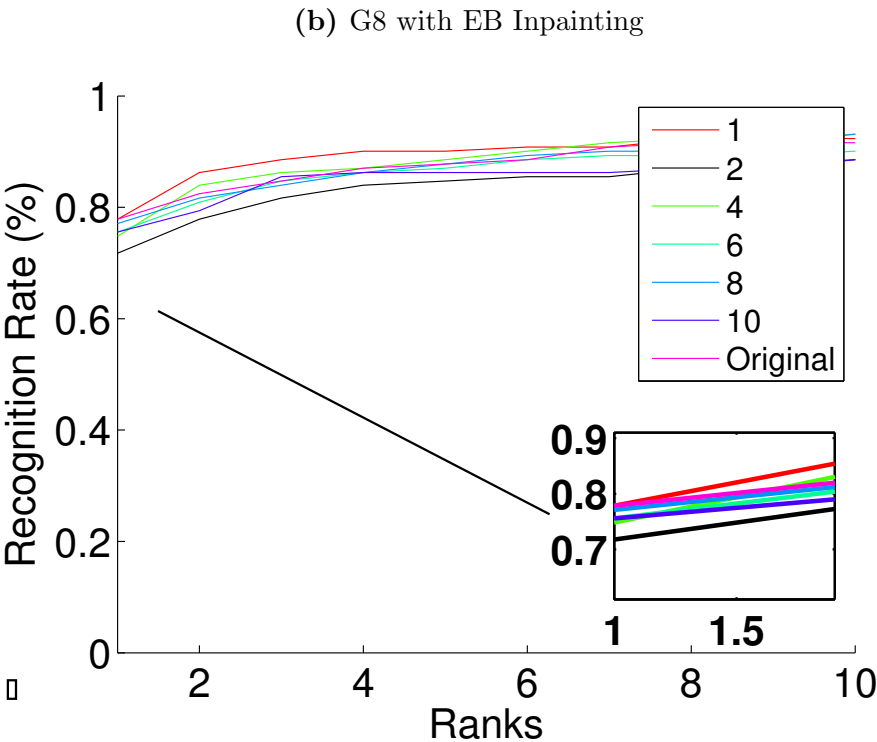
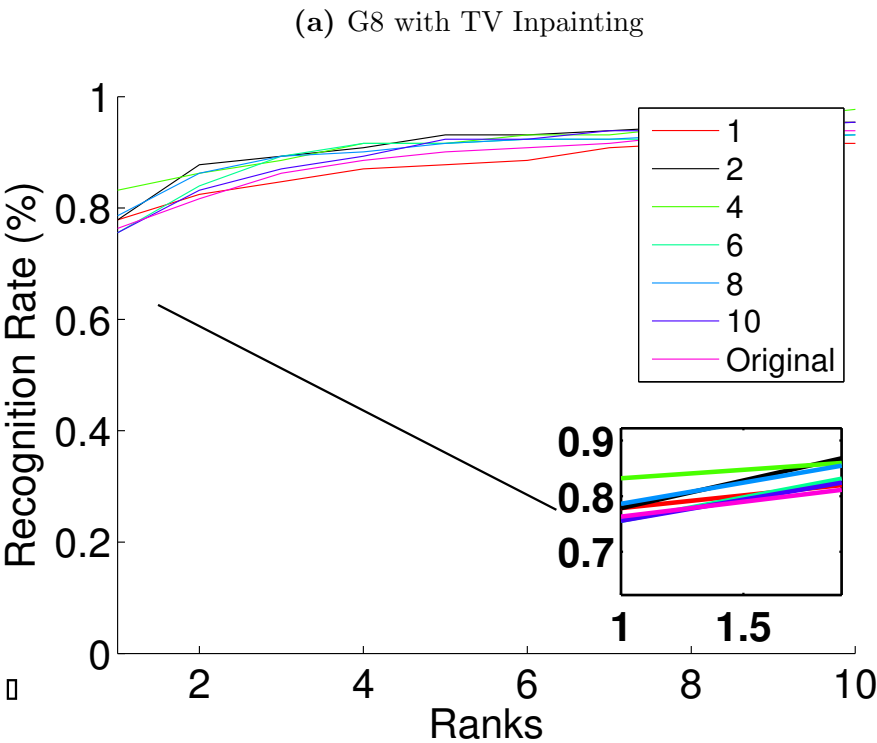
An investigation into the effects of either inpainting method from the perspective of a commercial matcher are conducted where the evaluation studies are performed using G8<sup>†</sup>. The purpose here is to determine whether TV wavelet inpainting or EB inpainting provides better preprocessing for improved recognition performance. Applying these techniques solely (i.e. without coupling TB denoising) provides a baseline understanding from a commercial point of view. From Figure 7.13, note the similarity in performance when either approach is applied. This is further shown in Table 7.1 where a slight improvement in performance of approximately 5% between the two methods is shown. Furthermore, note that this deviation becomes minimum as the rank order increases. These observations show that solely using either inpainting method is sufficient for improved performance.

## 7.4 Conclusion and Future Work

In this chapter, an investigation on the concept of document to live facial identification is performed, where different restoration strategies were applied to improve both image quality and identification performance. From an image quality standpoint, it is shown that the coupling effects of TB denoising with either inpainting method (i.e. TV wavelet inpainting or

---

<sup>†</sup>This algorithm was provided by L1 [www.11id.com](http://www.11id.com).



**Figure 7.13:** CMC curves illustrating commercial performance comparisons between TV wavelet and EB restoration strategies.

	TV Inpainting					
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Mask Width
<b>Original</b>	0.274	0.328	0.366	0.382	0.405	-
<b>Inpainted</b>	0.492	0.592	0.638	0.661	0.680	10
<b>Inpainted/Denoised</b>	0.592	0.631	0.685	0.731	0.746	6
<b>G8 Inpainted</b>	<b>0.832</b>	0.862	0.886	0.916	0.916	6
	EB Inpainting					
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Mask Width
<b>Original</b>	0.274	0.328	0.366	0.382	0.405	-
<b>Inpainted</b>	0.308	0.346	0.400	0.407	0.416	6
<b>Inpainted/Denoised</b>	0.585	0.638	0.685	0.746	0.770	Best
<b>G8 Inpainted</b>	<b>0.779</b>	0.863	0.885	0.901	0.901	1

**Table 7.1:** Top five ranks and which width mask generated the scores for both inpainting algorithms. Results are generated from the LTP-DTS texture based approach as well as the commercial algorithm G8.

EB inpainting) showed similar improvements in image quality according to the UIQ metric. From a facial recognition standpoint, several sensitivities and trade-offs that are dependent on the preprocessing method, mask dilation, and the texture based method used are noticed. Depending on the combination of the texture based method (LBP or LTP) with the distance metric ( $\chi^2$  or DT), it is noted that there are various sensitivities in terms of mask dilation when either TV wavelet inpainting or EB inpainting is used. However, these sensitivities are neutralized when coupled with TB denoising. It is speculated that this reduction in sensitivities is due to the cycle spinning behavior when TB denoising is applied. As a result, it is concluded that the inpainting-denoising combination is a prudent strategy considering inpainting mitigates the traces of various security watermarks, while TB denoising mitigates extraneous noisy artifacts. Furthermore, in terms of commercial recognition performance, it is observed that solely relying on total variation (TV) wavelet or exemplar-based (EB) inpainting is effective in achieving overall improvement in identification performance. Here,

it is hypothesize that because commercial algorithms, such as G8, already have a proprietary preprocessing scheme, incorporating either TV wavelet or EB inpainting methods are sufficient in improving rank-1 identification.

Although these pre-processing schemes do not increase the FR scores high enough to be considered statistically feasible for an automated system, this work provides an important step forward for operational use cases. Many operational FR systems typically have a “human in the loop” aspect in which, after identification is concluded, a face examiner visually examines multiple possible matches to the probe image with the hopes of being able to determine which potential face is the genuine match. By using the pre-processing techniques described above, the quality of the face image is increasing significantly. Eliminating the watermarks and creating a higher quality face not only increases the match scores of the document-live FR algorithms, it also increases the chances that a “human in the loop” examiner will be able to make a correct match. This achievement in document-live FR not only is statistically significant it is operationally relevant to everyday FR use cases. Also, it has been a global trend when issuing new passports to embed chips that contain an individuals biometric traits within them<sup>‡</sup>. Although this is not standard across the world, a majority of countries are moving toward the biometric passport, with a prediction of 80% of passports issued by 2015 containing biometric information. It is recognized that once biometric passports become standardized throughout the world, need for document-to-live FR may be moot. However, the proposed approach could still be considered viable in document to live historical forensics. For instance, if an old passport is found at a scene of a crime, the proposed method could be used to help identify the suspect in which the passport belongs

---

<sup>‡</sup><http://www.prweb.com/releases/ePassports/Biometrics/prweb3757714.htm>

to. By increasing the quality of the passport image using the proposed method, historical passport forensics could become significantly less challenging.

It is noted that applying TB denoising potentially discards important textural information from the facial image [62]. Therefore, a future direction in this area includes investigating the use of super-resolution algorithms that learn a priori the spatial distribution of the image gradient for frontal images of faces [95]. Additionally, it is noted that the current approach is also heavily dependent on manual mask annotation. This is limiting considering, as Figures 7.57.6 suggests, mask annotation produces additional sensitivities in terms of facial identification, where the most optimum performance might not completely be obtained. Therefore, another area of exploration would be to improve the mask selection, either automatically or semi-automatically, that optimizes both facial image quality and identification. Another area of exploration would be further investigation into the effects of preprocessing schemes from the standpoint of commercial algorithms. Doing so affords the opportunity to better understand commercial preprocessing and matching capabilities. Finally, an area that merits further investigation, is to extend these capabilities to improve image quality and face matching using very low quality images, where examples include images in closed-circuit televisions (CCTVs) and surveillance videos.









# Chapter 8

## Conclusions

### 8.1 Conclusions

As discussed in the introduction, FR trends are moving from traditional FR in the visible spectrum to higher bands in the EM spectrum due to the many advantages these wavelengths offer, such as covert capture at nighttime and robustness to obscurants such as fog and smoke. However, performing cross-spectral FR is a challenging problem due to the differences between the face representations. For example, in the SWIR band, specifically  $1550nm$ , all skin tones, regardless of race, appear black. Detecting facial landmarks and matching these types of images with their visible counterparts requires a unique approach that attempts to minimize the gap between the two representations. Common to most FR algorithms is the ability of normalizing the face to a standard position, in order for all facial features to be aligned across all subjects. One common way of performing this normalization is by using the centers of the subjects eyes to align on the same plane and resizing the face image so all images are the same resolution. Because FR trends are moving into higher bands of

the EM spectrum, it is important to have a fast and robust approach to obtaining these points. Through the application of eye detection, cross-spectral face recognition, and cross-representation face recognition, this work has proposed a complete FR system that efficiently and accurately detects the eyes in multiple bands of the EM spectrum as well as in non-ideal capture scenarios. After locating the eyes, the images are then normalized and used in the proposed cross-spectral FR algorithm, matching SWIR to Visible images, or Document to Live images.

To reiterate, this work provides five main contributions:

1. An extensive dataset was designed and collected that captured a large number of face images in the SWIR spectrum. Ground truth data, along with different challenging scenarios, including sunglasses and different tinted pieces of glass, were collected for the use of testing eye detection methods and cross-spectral face recognition algorithms.
2. A novel eye detection method was developed and described that incorporated face and eye localization using the normalized cross correlation coefficient and pupil detection using the proposed summation range filters. An evaluation on which pre-processing photometric normalization technique obtained the highest eye detection accuracy was also completed, while extensive testing on other available eye detection methods, including the state-of-the-art and commercial algorithm, showed that the proposed method performed best comparatively. It was also showed that the proposed algorithm was robust to image degradations such as image compression and spatial resolution. Finally, FR tests using multiple FR algorithms concluded that the eye locations produced from the proposed method resulted in higher recognition rates compared to other tested eye detection algorithms.

3. Multiple extensions to the baseline eye detection algorithm was described in order to detect eyes in multiple challenging non-ideal capture scenarios across different bands of the EM spectrum, including faces behind tinted glass, at long distances, with/without eye glasses, and extreme pose and illumination variation. By accurately classifying each challenging scenario, the appropriate method was able to be used in order to detect the eyes. It was shown that these improvements to the baseline algorithm not only increased detection rates, but outperformed all other tested eye detection methods. Finally, it was shown that the output of the proposed method extensions produced higher FR results in both identification and verification steps.
4. An extension to the work done by Kalka, et al. [20] was developed that quickly and accurately performed heterogeneous face matching algorithm that matched images in all bands of the SWIR spectrum to visible counterparts. By determining which photometric normalization combinations performed best, fusing these results showed a significant increase in cross-spectral matching performance compared to using all photometric combinations available. Not only did this increase the performance rate, it drastically decreased recognition time by over 50 times. It was also shown that the proposed method out-performed other academic and commercial face recognition algorithms on cross-spectral matching experiments.
5. A novel pre-processing method was developed in order to increase the performance of document to live face matching. A discussion on the analogy between document to live and cross-spectral matching was offered and multiple strategies was used in order to increase recognition performance. First, two different inpainting methods were used to help alleviate the noise that many face documents have due to watermarks.

Secondly a denoising algorithm was used to further eliminate unwanted noise and help increase recognition rates. It was shown that these pre-processing techniques significantly helped improve the document to live face recognition performance.

## 8.2 Future Work

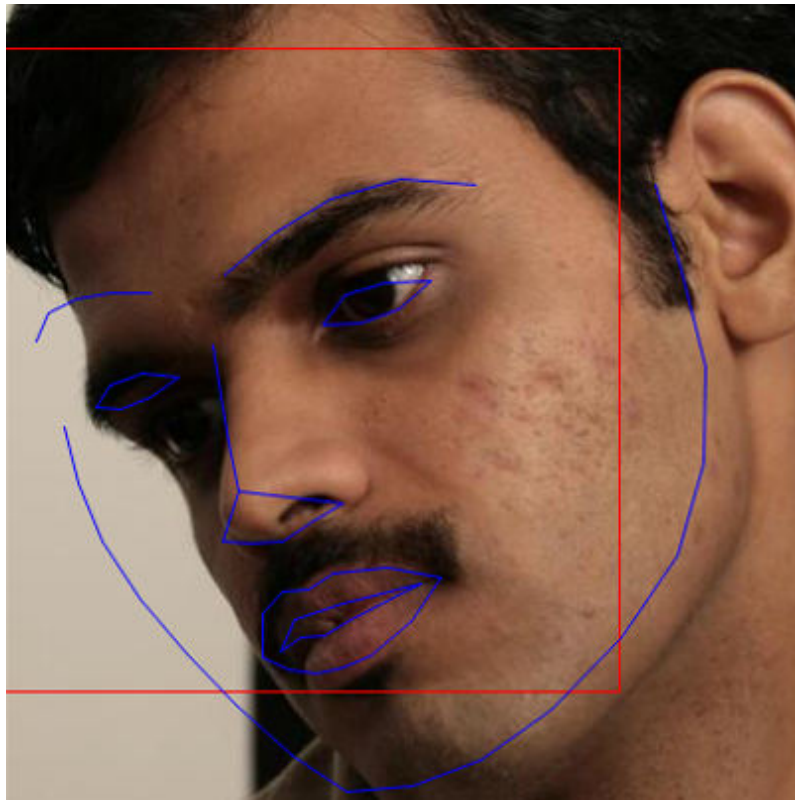
While the underpinnings of this work described provide several contributions to eye detection and cross spectral and document to live face matching, there are additional aspects to the methods described that will require future investigation moving forward.

During the time of this work, further research in the academic community was conducted on the use of facial landmarks to normalize a face for face recognition [96–100]. Active Appearance Models (AAM) and Active Shape Models (ASM) are widely used in this area in order to landmark, typically, 68 points across the face. Many algorithms, such as STASM\* and DLIB<sup>†</sup>, are available in the open source for use on any face image. Although these algorithms perform extremely well on constrained frontal poses, they tend to have difficulty when the image is in non-ideal conditions, as can be seen in Figure 8.1. Further experimentation also showed that these algorithms could not find any landmarks on SWIR images in the  $1550nm$  wavelength. It is understood that the use of just eyes for normalization may not be the optimal way of normalizing a face, but is merely a first stepping stone in advancing the normalization schemes in the SWIR spectrum and on highly unconstrained imagery. It is known that the trend for face image normalization is to use additional facial landmarks (other than just the eye locations) and that the addition of more landmarks may lead to

---

\*<http://www.milbo.users.sonic.net/stasm/>

<sup>†</sup>[dlib.net](http://dlib.net)



**Figure 8.1:** Facial landmarks detected on image with extreme pose and illumination variation. The open sourced algorithm DLIB was used. Notice the inaccuracy in landmark detection in this challenging image. Further testing determined that this algorithm could not detect any facial landmarks on SWIR face imagery.

higher FR rates. Because of this, it is the hope of the author that the techniques described above can be adapted to detect the nose, mouth, etc and, therefore, by adding additional points on those features, they can be used to help increase normalization accuracy. The described methods could also be adapted to allow these open sourced algorithms to perform better not only on unconstrained imagery, but on images in higher bands of the EM spectrum, such as SWIR.

Future work could also include algorithm optimization. Optimizing the algorithm in

terms of speed would be required in order to allow for real time eye detection or eye detection in video. Also, more challenging datasets should be explored, such as faces in the wild databases.

In terms of cross-spectral and document to live recognition, future research should be conducted on different bands of the electromagnetic spectrum, including the Mid Waved Infrared (MWIR) and Long Waved infrared (LWIR) spectra. Adapting the proposed algorithm to handle the different features of each respective band would be necessary in order to truly have an approach that works across the whole EM spectrum. Also, expanding the size of the WVUM database would be needed in order to test the scalability of the proposed approaches. In document to live recognition, further research should be conducted in having a more sophisticated method of watermark masking should be explored. Currently, all watermarks are manually annotated. However, in order to have an automated document to live FR system, this process would have to be automated or semi-automated (human in the loop). This would allow for the optimization of both facial image quality and identification. Also, extending the capability to improve the image quality and face matching using very low quality images, such as face images retrieved from close-circuit televisions (CCTV's) and surveillance videos.







# Bibliography

- [1] E. Mordini, S. Massari, “Body, biometrics and identity”, in *Bioethics*, 2008, Vol. 22, No. 9, pp. 488-498.
- [2] J. Ashbour, “Guide to Biometrics for Large-Scale Systems: Technological, Operational, and User-Related Factors”, in *Springer*, London, 2011.
- [3] M. Ao, D. Yi, Z. Lei, S.Z. Li, “Handbook of Remote Biometrics for Surveillance and Security”, in *Advances in Computer Vision and Pattern Recognition Series*, Springer, 2009.
- [4] A.K. Jain, P.J. Flynn, A. Ross, “Introduction to Biometrics” *Springer*, 2007.
- [5] A.K. Jain, C.D. Sarat, K. Nandakumar, “Soft biometric traits for personal recognition systems” in *Biometric Authentication*, Springer Berlin Heidelberg, 2004, pp. 731-738.
- [6] A.K. Jain, A. Ross, K. Nandakumar, “Introduction to Biometrics”, Springer Science and Business Media, 2011.
- [7] P. Grother, G. Quinn, P.J. Phillips, “Report on the Evaluation of 2D Still-Image Face Recognition Algorithms”, in *Nat’l Inst. of Standards and Technology* interagency/internal report (NISTIR) 7709, 2010.

- [8] A. Selinger, D.A. Scolinsky, "Face Recognition in the dark", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, June 2004.
- [9] N. Kalka, T. Bourlai, B. Cukic, L. Hornak, "Cross-spectral Face Recognition in Heterogeneous Environments: A Case Study on Matching Visible to Short-wave Infrared Imagery", in *International Joint Conference on Biometrics* (IEEE, IAPR), 2011.
- [10] N. Osia, T. Bourlai, "Holistic and Partial Face Recognition in the MWIR Band Using Manual and Automatic Detection of Face-based Features", in *IEEE Conf. on Technologies for Homeland Security*, Boston, USA, Nov. 2012.
- [11] T. Bourlai, A. Ross, C. Chen, and L. Hornak, "A Study on using Middle-Wave Infrared Images for Face Recognition", in *SPIE Biometric Technology for Human Identification IX*, Baltimore, USA, Apr. 2012.
- [12] H. Mendez, C. San Martín, J. Kittler, Y. Plasencia, E. García-Reyes, "Face recognition with LWIR imagery using local binary patterns", in *Advances in Biometrics*, Springer, Berlin Heidelberg 2009, pp. 327-336.
- [13] F. Nicolo, N. Schmid, "Long Range Cross-Spectral Face Recognition: Matching SWIR against Visible Light Images", in *IEEE Transaction of Information Forensics and Security*, 2012, pp. 1717-1726.
- [14] WVHTCF, "Tactical imager for night/day extended-range surveillance." accessed on Nov. 10, 2011. [Online]. Available: [http://www.wvhtf.org/departments/advanced\\_tech/projects/tinders.asp](http://www.wvhtf.org/departments/advanced_tech/projects/tinders.asp)

- [15] B. Klare, A.K. Jain, “Heterogeneous face recognition: Matching NIR to visible light images”, *Pattern Recognition (ICPR), 2010 20th International Conference on*, August, 2010, pp. 1513-1516
- [16] J. Ice, N. Narang, C. Whitelam, N. Kalka, L. Hornak, J. Dawson, and T. Bourlai, “SWIR Imaging for Facial Image Capture through Tinted Materials”, in *SPIE (Vol. 8353), Infrared Technology and Applications XXXVIII*, Baltimore, U.S.A., Apr. 2012.
- [17] T. Bourlai, B. Cukic, “Multi-Spectral Face Recognition: Identification of People in Difficult Environments”, in *IEEE International Conference on Intelligence and Security Informatics (ISI)*, Washington, U.S.A., June 2012.
- [18] P. Wang, G., M.B., Q. Ji, J. Wayman, “Automatic Eye Detection and Its Validation”, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, June 2005, pp. 164-171.
- [19] R. Valenti, T. Gevers, “Accurate Eye Center Location through Invariant Isocentric Patterns”, in *Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence*, Sept. 2012, pp. 1785-1798.
- [20] N. Kalka, T. Bourlai, B. Cukic, L. Hornak, “Cross-spectral Face Recognition in Heterogeneous Environments: A Case study on matching visible to Short-wave Infrared Imagery”, in *International Joint Conference on Biometrics IEEE/IAPR*, 2011, pp. 1-8.
- [21] M. Shafi, P. Chung, “A Hybrid Method for Eyes Detection in Facial Images”, in *Proc. World Academy of Science, Engineering and Technology*, August 2008.

- 
- [22] J. Dowdall, I. Pavlidis, and G. Bebis, "Face detection in the near-ir spectrum", in *Proc. Image and Vision Computing*, New Zealand, 2003, pp. 565-578.
  - [23] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, "An Eye Detection Algorithm using Pixel to Edge Information", in *Proc. International Symposium on Communications, Control and Signal Processing*, Marrakech, Morocco, March 2006.
  - [24] P. Viola, M. Jones, "Robust Real-Time Face Detection", in *International Journal of Computer Vision*, May, 2004, pp. 137-154.
  - [25] C. Whitelam, Z. Jafri, T. Bourlai, "Multispectral Eye Detection: A Preliminary Study", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Turkey, Istanbul August 2010, pp. 209-212
  - [26] T. Bourlai, C. Whitelam, I. Kakadiaris, "Pupil detection under lighting and pose variations in the visible and active infrared bands", in *Proc. International Workshop on Information Forensics and Security*, Foz do Iguacu, Brazil, 2011, pp. 1-6.
  - [27] Z. Zhu, Q. Ji, K. Fujimura, and Kuangchih Lee, "Combing Kalman Filtering and Mean Shift for Real Time Eye Tracking Under Active IR Illumination", in *Proc. International Conference on Pattern Recognition*, Quebec, Canada, 2002, pp. 318-324.
  - [28] T. Bourlai and Z. Jafri, "Eye detection in the Middle-Wave Infrared Spectrum: Towards Recognition in the Dark", in *Proc. International Workshop on Information Forensics and Security*, Foz do Iguacu, Brazil, 2011.
  - [29] J. Song, Z. Chi, J. Liu, "A Robust Eye Detection Method Using Combined Binary Edge and Intensity Information", in *Pattern Recognition*, June 2006, pp. 1110-1125.

- [30] W.M.K.W.M. Khairosfaizal, A.J. Nor'aini, "Eyes detection in facial images using Circular Hough Transform", in *Proc. 5th International Colloquium on Signal Processing & Its Applications*, March 2009, pp. 238-242.
- [31] R. Hsu, M. Abdel-Mottaleb, A. Jain, "Face Detection in Color Images", in *Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, pp. 696-706.
- [32] C. Chiang, W. Tai, M. Yang, Y. Huang, C. Huang, "A Novel Method for Detecting Lips, Eyes and Faces in Real Time", in *Real-Time Imaging*, 2003, pp. 277-287.
- [33] M. Shafi, P. Chung, "Eyes Extraction from Facial Images Using Edge Density", in *Proc. 7th IEEE International Conference on Cybernetic Intelligent Systems*, 2008.
- [34] K. Messar, J. Matas, J. Kittler, J. Luetttin, G. Maitre, "XM2VTSDB: The Extended M2VTS database" in *Second International Conference on Audio and Video-based Biometric Person Authentication*, March, 1999, pp. 965-966.
- [35] The BioID face database: [<http://www.bioid.com/downloads/facedb/facedatabase.html>]
- [36] M. Minear, D. Park, "A lifespan database of adult facial stimuli", in *Behavior Research Methods, Instruments, & Computers*, 2004, pp. 630-633.
- [37] H. Lu, W. Zhang, and D. Yang, "Eye detection based on rectangle features and pixel-pattern-based texture features", in *Proc International Symposium on Intelligent Signal Processing and Communication Systems*, Xiamen, China, 2007, pp. 746-749.
- [38] G. Feng, P. Yuen, "Multi-cues eye detection on gray intensity image", in *Pattern Recognition*, May 2001, pp. 1033-1046.

- [39] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 681-685, 2001.
- [40] B. Efraty, C. Huang, S.K. Shah, I.A. Kakadiaris, “Facial landmark detection in uncontrolled conditions”, in *International Joint Conference on Biometrics (IJCB)*, 11-13 Oct, 2011 pp.1-8.
- [41] X. Zhao, E. Dellandrea, L. Chen, I.A. Kakadiaris, “Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model”, in *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 2011, pp. 1417-1428.
- [42] P. Perakis, G. Passalis, T. Theoharis, I.A. Kakadiaris, “3D facial landmark detection under large yaw and expression variations” in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2013, pp. 1552-1564.
- [43] X. Yu, J. Huang, S. Zhang, W. Yan, D.N. Metaxas, “Pose-free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model”, in *International Conference on Computer Vision*, 2013.
- [44] X. Yu, Z. Lin, J. Brandt, D.N. Metaxas, “Consensus of Regression for Occlusion-Robust Facial Feature Localization”, in *Proc. European Conference on Computer Vision*, 2014, pp. 105-118.
- [45] Y. Yu, S. Zhang, F. Yang, D.N. Metaxas, “Localization of Multi-pose and Occluded Facial Features via Sparse Shape Representation”, in *Advances in Visual Computing: Springer*, 2013, pp. 486-945.

- [46] S. Li, L. Zhang, S. Liao, X. Zhu, R. Chu, M. Ao, R. He, “A Near-infrared Image Based Face Recognition System” in *FG*, April, 2006, pp. 455-460.
- [47] R. Jafri, R. Arabnia, “A Survey of Face Recognition Techniques”, in *Journal of Information Processing Systems*, June, 2009, pp. 41-68.
- [48] V. Vijayakumari, “Face Recognition Techniques: A Survey”, in *World Journal of Computer Applications and Technology*, 2013, pp. 41-50.
- [49] J.M. Pandaya, D. Rathod, J.J. Jadav, “A survey of face recognition approach”, in *International Journal of Engineering Research and Applications*, 2013, pp. 632-635.
- [50] T. Bourlai, N. Kalka, A. Ross, B. Cukic, L. Hornak, “Cross-spectral face verification in the short waved infrared (SWIR) band”, in *International Conference on Pattern Recognition (ICPR)*, August 2010.
- [51] J.Y. Zhu, W.S. Zheng, J.H. Lai, and S. Li, “Matching NIR face to VIS Face using Transduction”, in *IEEE Transactions on Information Forensics and Security*, 2014.
- [52] S. Li, D. Yi, Z. Lei, S. Liao, “The CASIA NIR-VIS 2.0 Face Database”, in *Proc. IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013, pp. 348-353.
- [53] M.A. Akhloufi, A. Bendada, J.C. Batsale, “Multispectral Face Recognition using non linear dimensionality reduction”, in *SPIE Defense, Security, and Sensing*, 2009.
- [54] H. Mendez, C. San Martin, J. Kittler, Y. Plasencia, E. Garcia-Reyes, “Face recognition with LWIR imagery using local binary patterns”, in *Advances in Biometrics*, Springer, 2009, pp. 327-336.



- 
- [55] N. Osia, T. Bourlai, “A spectral Independent Approach for Physiological and Geometric Based Face Recognition in the Visible, Middle-wave and Long-wave Infrared Bands”, in *Image and Vision Computing - Elsevier*, 2014.
- [56] T. Bourlai, “Applications of Passive Infrared Imaging to Forensic Facial Recognition”, in *InfraMation (Thermal Imaging Leading User’s Conference - Organized by FLIR)*, Loews Royal Pacific, 2013.
- [57] T. Bourlai, “Mid-wave IR Face Recognition Systems”, in *SPIE Newsroom Magazine - Defense & Security*, 2013, pp.1-3.
- [58] V. Starovoitov, D. Samal, B. Sankur, “Matching of faces in camera images and document photographs”, in *Int. Conf. on Acoustic, Speech, and Signal Processing IV*, pp. 2349-2352, 2000.
- [59] V. Starovoitov, D. Samal, D. Briliuk, “Three approaches for face recognition”, in *International Conference on Pattern Recognition and Image Analysis*, pp. 707-711, 2002.
- [60] N. Ramanathan, R. Chellappa, “Face verification across age progression” in *IEEE Transactions on Image Processing*, pp. 3349–3362, 2006.
- [61] T. Bourlai, A. Ross, A. Jain, “On matching digital face images against scanned passport photos”, in *First IEEE International Conference on Biometrics, Identity and Security (BIDS)*, 2009.
- [62] T. Bourlai, A. Ross, A. Jain, “Restoring degraded face images for matching faxed or scanned photos”, in *IEEE Transactions on Information Forensics and Security*, pp. 371-384, 2011.

- [63] Y. Li, J. Kittler, J. Matas, “Analysis of the lda-based matching schemes for face verification”, in *British Machine Vision Conference*, 2000.
- [64] S. Mohideen, S. Perumal, M. Sathik, “Image de-noising using discrete wavelet transform” in *International Journal of Computer Science and Network Security*, 2008.
- [65] R. Coifman, D. Donoho, “Translation-invariant de-noising”, in *In Wavelets and Statistics, Springer Lecture Notes in Statistics 103*, pp. 125–150, 1994.
- [66] M. Ferrara, A. Franco, D. Maio, D. Maltoni, “Face image conformance to iso/icao standards in machine readable travel documents”, in *IEEE Transactions on Information Forensics and Security*, pp. 1204–1213, 2012.
- [67] M. Bertalmio, A. Bertozzi, V. Caeselles, C. Ballester, “Image inpainting” in *Tech. rep., University of Minnesota*, 1999.
- [68] T. Chan, S. Kang, J. Shen, “Euler’s elastica and curvature based inpainting”, in *SIAM Journal of Applied Math*, pp. 564–592, 2002.
- [69] T. Chan, J. Shen, “Morphologically invariant pde inpaintings”, in *Tech. rep., UCLA*, 2001.
- [70] M. Bertalmio, A. Bertozzi, G. Shapiro, “Navier-stokes, fluid dynamics, and image inpainting and video inpainting” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
- [71] S. Kang, T. Chan, S. Soatto, “Inpainting from multiple view”, in *Tech. rep., UCLA CAM Report*, 2002.

- [72] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, J. Verdera, “Filling-in by joint interpolation of vector fields and grey levels”, in *IEEE Transactions on Image Processing*, pp. 1200–1211, 2001.
- [73] L. Demanet, B. Song, T. Chan, “Image inpainting by correspondence maps: A deterministic approach”, in *Tech. rep., UCLA CAM Report*, 2003.
- [74] T. Chan, J. Shen, H. Zhou, “Total variation wavelet inpainting”, in *SIAM Journal of Applied Mathematics*, 2006.
- [75] T. Bourlai, A. Clark, L. Best-Rowden, “Methodological insights on restoring face photos of multinational passports”, in *IEEE International Symposium on Technologies for Homeland Security (IEEE HST)*, 2013.
- [76] CEI 60825-1, *International Standard: Safety of Laser Products*, 2007.
- [77] M.P. Hansen, D.S. Malchow, “Overview of SWIR Detectors, Cameras, and Applications”, in *SPIE 6939*, 2008.
- [78] J. John, L. Zimmerman, P. Merkin, G. Borghs, C. Van Hoof, S. Nemeth, “Extended Backside-illuminated InGaAs on GaAs IR Detectors”, in *SPIE 4820*, 2003, pp. 453-459.
- [79] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection”, in *Proc. IEEE Computer Vision and Pattern Recognition, (CVPR)*, pp. 886-893, 2005.
- [80] G. Toderici, G. Evangelopoulos, T. Fang, I. Kakadiaris, “UHDB11 database for 3D-2D face recognition”, in *Image and Video Technology, Springer Berlin Heidelberg*, pp. 73-86, 2014.

- 
- [81] O. Jesorsky, K. J. Kirchbergand, R. Frischholz. “Robust face detection using the Hausdorff distance”, in *Audio and Video Biom. Pers. Auth.*, 1992, pp 90-95.
- [82] A.M. Reza, “Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement” in *Journal of VLSI signal processing systems for signal, image and video technology*, 2004, pp. 35-44.
- [83] D. Jobson, Z. Rahman, and G. Woodell. “Properties and performance of a center/surround retinex” in *IEEE Transactions on Image Processing*, March, 1997, pp. 451–462.
- [84] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions”, in *Transactions on Image Processing*, vol. 19, no. 6, 2010, pp. 1635-1650.
- [85] B. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, A. Jain, “Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A”, in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 1931-1939, 2015.
- [86] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, “Deepface: Closing the gap to human-level performance in face verification”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708, 2014.
- [87] Z. Rahman, D. Jobson, and G. Woodell, “Multi-scale retinex for color image enhancement”, in *Image Processing, 1996. Proceedings., International Conference on*, volume 3, Sept. 1996, pp. 1003-1006.

- 
- [88] S. Zhou, R. Chellappa, W. Zhao. “Unconstrained face recognition”, in *Springer*, 2006, pp. 1–8.
- [89] A. Albert, A. Sethuram, K. Ricanek, “Biometrics: Unique and Diverse Applications in Nature, Science, and Technology”, in *Implications of Adult Facial Aging Factors on Biometrics. Intech*, 2011.
- [90] K. Ricanek, K. Mahalingam, G. Mahalingam, A. Albert, R. Bruegge, “Age Factors in Biometric Processing”, in *Human Face Aging: A Perspective Analysis from Anthropology and Biometrics. The Institution of Engineering and Technology*, 2013.
- [91] B. Klare, A. Jain, “Heterogeneous face recognition: Matching nir to visible light images”, in *IEEE International Conference on Pattern Recognition (ICPR)*, pp. 1513-1516, 2010.
- [92] B. Andrews, “A brief history of us passport photographs”, *blakeandrews.blogspot.com/2009/05/brief-history-of-us-passport.html*, 2009.
- [93] A. Criminisi, P. Perez, K. Toyama, “Region filling and object removal by exemplar-based image inpainting”, in *IEEE Transactions on Image Processing*, pp. 1–13, 2004.
- [94] Z. Wang, A. Bovik, “A universal image quality index”, in *IEEE Signal Processing Letters*, pp. 81–84, 2002.
- [95] P. Hennings-Yeomans, S. Baker, B. Kumar, “Simultaneous super-resolution and feature extraction for recognition of low-resolution faces”, in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

- 
- [96] K. Seshadri, M. Savvides, “Robust modified active shape model for automatic facial landmark annotation of frontal faces”, in *IEEE 3rd International Conference on Biometrics: Theory, Applications and System*, pp. 1-8, 2009.
- [97] Y. Tong, X. Liu, F. Wheeler, P. Tu, “Automatic facial landmark labeling with minimal supervision”, in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 2097-2104, 2009.
- [98] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic, “A semi-automatic methodology for facial landmark annotation”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 896-903, 2013.
- [99] M. Pantic, L. Rothkrantz, “Automatic analysis of facial expressions: The state of the art”, in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pp. 1424-1445, 2000.
- [100] B. Fasel, J. Luettin, “Automatic facial expression analysis: a survey”, in *Pattern recognition*, pp. 259-275, 2003.



