2012

Homogeneous and Heterogeneous Face Recognition: Enhancing, Encoding and Matching for Practical Applications

Francesco Nicolo
West Virginia University

Homogeneous and Heterogeneous Face Recognition: Enhancing, Encoding and Matching for Practical Applications

by

Francesco Nicolo

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

Doctor of Philosophy
in
Electrical Engineering

Natalia A. Schmid, D.Sc., Chair
Erdogan Gunel, Ph.D.
Vinod K. Kulathumani, Ph.D.
Xin Li, Ph.D.
Matthew C. Valenti, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2012

Keywords: Face Recognition, Gabor Filters, Local Operators, Short Wave Infra Red, Visible, Quality Metrics, Multi-View

Abstract

Homogeneous and Heterogeneous Face Recognition: Enhancing, Encoding and Matching for Practical Applications

by

Francesco Nicolo

Face Recognition is the automatic processing of face images with the purpose to recognize individuals. Recognition task becomes especially challenging in surveillance applications, where images are acquired from a long range in the presence of difficult environments. Short Wave Infrared (SWIR) is an emerging imaging modality that is able to produce clear long range images in difficult environments or during night time. Despite the benefits of the SWIR technology, matching SWIR images against a gallery of visible images presents a challenge, since the photometric properties of the images in the two spectral bands are highly distinct.

In this dissertation, we describe a cross spectral matching method that encodes magnitude and phase of multi-spectral face images filtered with a bank of Gabor filters. The magnitude of filtered images is encoded with Simplified Weber Local Descriptor (SWLD) and Local Binary Pattern (LBP) operators. The phase is encoded with Generalized Local Binary Pattern (GLBP) operator. Encoded multi-spectral images are mapped into a histogram representation and cross matched by applying symmetric Kullback-Leibler distance. Performance of the developed algorithm is demonstrated on TINDERS database that contains long range SWIR and color images acquired at a distance of 2, 50, and 106 meters.

Apart from long acquisition range, other variations and distortions such as pose variation, motion and out of focus blur, and uneven illumination may be observed in multispectral face images. Recognition performance of the face recognition matcher can be greatly affected by these distortions. It is important, therefore, to ensure that matching is performed on high quality images. Poor quality images have to be either enhanced or discarded. This dissertation addresses the problem of selecting good quality samples.

The last chapters of the dissertation suggest a number of modifications applied to the cross spectral matching algorithm for matching low resolution color images in near-real time. We show that the method that encodes the magnitude of Gabor filtered images with the SWLD operator guarantees high recognition rates. The modified method (Gabor-SWLD) is adopted in a camera network set up where cameras acquire several views of the same individual. The designed algorithm and software are fully automated and optimized to perform recognition in near-real time. We evaluate the recognition performance and the processing time of the method on a small dataset collected at WVU.

Acknowledgements

I would like to express my greatest gratitude to my advisor, Dr. Natalia A. Schmid, who guided me during my graduate studies at WVU. This work wouldn't have been possible without her invaluable guidance and support. I am immensely grateful to her for teaching me how to perform high quality research and how to communicate the obtained results. I learned so much from her during my stay at WVU. I am also very thankful to her for helping me to correct the following manuscript.

I'm very thankful to Dr. Vinod Kulathumani who provided several suggestions, the network design and the multi-view face dataset employed in this work. I would like to thank Dr. Xin Li for his suggestions and excellent teaching efforts during the image processing course that I took at WVU. I would like to thank Dr. Matthew C. Valenti and Dr. Erdogan Gunel for accepting to be my committee members, taking the time to read my dissertation and providing me with helpful feedback. I would also like to thank Dr. Brian E. Lemoff of West Virginia High Technology Consortium Foundation for providing the SWIR datasets employed in this work.

I would like to thank all my lab colleagues and friends. My special thanks go to Jinyu Zuo that exchanged many ideas with me and provided great help during my years at WVU. I would like to also thank my good friends Xiaohan Chen and Nathan Kalka for their useful discussions and support. I want to thank Salvatore Talarico for his friendship.

I want to thank my uncle and aunt, my family in USA. I wouldn't be here without their support and help. I thank my girlfriend Syvale Lee for all her love and encouragement in difficult times. Finally, I would like to thank my parents, my brother and my sister for all their support and love.

Contents

List of Figures

List of Tables

Chapter 1

Overview and Objectives

1.1 Introduction

Biometrics is a science of automatically identifying individuals based on their unique physiological or behavioral characteristics [2]. Examples of physiological traits are face, iris, fingerprint, hand geometry and DNA. Behavioral features include keystroke dynamics, gait and voice. The work reported in this dissertation focuses on face biometrics. The dissertation presents a novel matching scheme for face recognition that is demonstrated in the range from the visible up to the Short Wave Infra Red (SWIR) spectral band. The designed method has application especially in surveillance scenarios (long range and/or variable distance from the camera) but is also demonstrated in a traditional setting (close distance from the camera).

1.2 A Face Recognition System

A typical face recognition system has two operation modes: (A) Enrollment and (B) Recognition. During the enrollment stage the individual usually stands at a close distance from the capturing camera. A block diagram of a face recognition system working in enrollment mode is shown in Fig. 1.1.

The enrollment process in Fig. 1.1 can be summarized as follows: (1) a video frame (image) is acquired; (2) face and additional landmarks are detected; (3) a vector of quality measures is generated for the biometric sample; (4) the acquired face sample is discarded

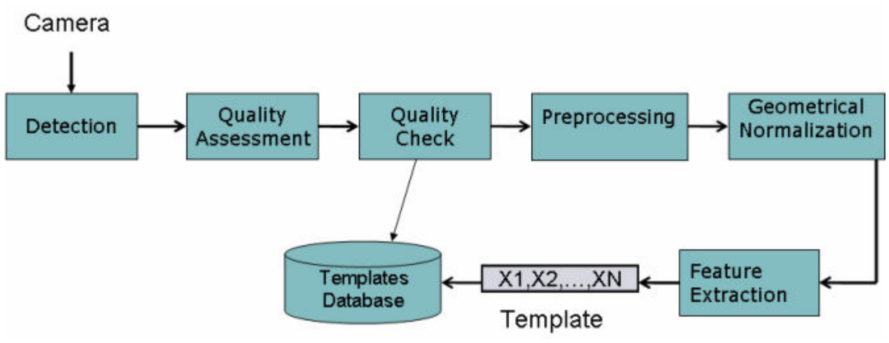


Figure 1.1: Block diagram of a face recognition system in enrollment mode (A).

or retained depending on its quality; (5) the face image is geometrically normalized to a canonical form; (6) the normalized face image is pre-processed (e.g. transformed to gray scale, convolved with a filter, etc.); (7) a feature vector is extracted from the processed face image (encoding stage); (8) the obtained feature vector is stored as a template in the database to form a gallery image in a gallery set and (optionally) the quality measures of the sample are stored together with the template.

When the face recognition system is working in recognition mode, the probe image can be acquired at a close or large distance and the individual may or may not be aware of being captured by a camera. The face recognition process is depicted in Fig. 1.2.

In Fig. 1.2, we note that the first seven steps (1)-(7) of the recognition process overlap with the steps of the enrollment procedure. The last blocks perform matching: (8) the probe feature vector is compared against a gallery set of templates previously stored in a database and matching scores are generated. The matching block may employ a distance (dissimilarity) measure as a matching score or alternatively adopts a similarity score; (9) a verification or identification result is produced depending on the operational mode.

Verification mode performs a 1 to 1 matching. A probe is claimed to belong to a specific entry in the gallery. Therefore, it is compared to only this entry. In this mode of operation, biometric systems compare the obtained matching score to a given threshold to produce a genuine or imposter decision.

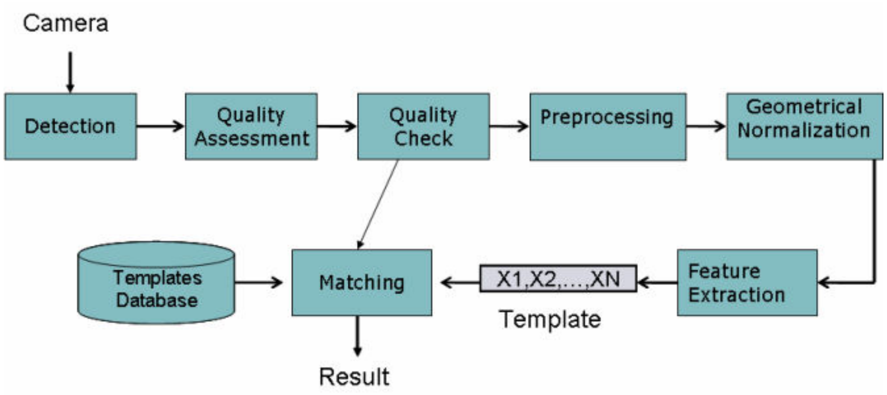


Figure 1.2: Block diagram of a face recognition system in testing mode (B).

In the identification case (1 to N matching) the face recognition system searches through the entire gallery set by comparing each gallery entry to a probe. In this case, the obtained matching scores can be sorted by value to define the best match. Sorting results in ranking.

The implementation in this work follows the schematics depicted in Fig. 1.1 and Fig. 1.2. However, the focus is placed on designing the quality check module, feature extraction and matching.

1.3 Dissertation Overview

Traditional face recognition systems are designed to process short range images acquired by a color camera. In this scenario (e.g. e-passport, access control, etc) state of the art face recognition algorithms are able to achieve high recognition rates (e.g. >99%). However, in a surveillance scenario where images are often acquired in the presence of difficult environments or at night time, the recognition rates of such algorithms may drop dramatically. To improve the recognition rates of face algorithms operating in difficult settings, other imaging modalities operating within a different range of the electromagnetic spectrum (Near InfraRed, Short Wave InfraRed, Medium/Long Wave InfraRed) have been introduced. In Chapter 2, we provide: (1) a brief overview of the traditional face recognition methods applied to visible data; (2) a summary of face recognition approaches that combine data from different spectral bands (multi-spectral imaging) and a review of methods that compare images across different spectral bands (cross-spectral matching).

Most of the described multi-spectral and cross-spectral algorithms are employed for surveillance purposes. In such applications, the recognition task can be challenging. Hence, when uncontrolled setting variations are present, it is very important to assess the quality of the acquired face sample; if very poor quality samples are adopted low recognition rates may be achieved.

A typical face recognition system includes a quality module that provides a set of measures (quality metrics) that describe different properties of the face images that are associated with the recognition performance of the biometric system. Hence, the quality metrics can be employed to discard poor quality samples according to a criteria or a method (see the quality check module in Fig. 1.1 and Fig. 1.2) that estimate the relationship of quality metrics with the matching performance. The quality metrics can also be adopted at the matching stage, where features of probe images are compared with the feature templates in the gallery. In this case the quality metrics can be employed to provide a confidence measure to the obtained matching result. The quality metrics can also be adopted simultaneously with the matching scores as input to an augmented classifier (see the matcher block in Fig. 1.1 and Fig. 1.2) to improve the recognition results. Chapter 3 introduces three methods that employ quality metrics to improve the recognition performance of a biometric matcher. The methods are demonstrated on FRGC 2.0 visible face dataset composed of face images of 202 subjects. The dataset is acquired in an uncontrolled illumination scenario and includes face expression variations (neutral and smiling).

In addition to non ideal settings described above, images may be also acquired under low lighting conditions or at night time. Surveillance cameras, in that case, usually employ Near InfraRed (NIR) imaging. Furthermore, if the probe images are long range images then the Medium/Long Wave Infra Red (MWIR or LWIR) bands are also adopted. Several face recognition methods employ an individual spectral band to perform recognition. However, this scenario does not often support modern face recognition applications. Surveillance cameras often switch between the visible and NIR band depending on the day or night environments. Since gallery and watch lists are traditionally composed of visible light face images, newly designed face recognition methods are expected to be successful in matching NIR data versus color images.

Apart from this case of cross-spectral comparison, attention has recently been turned to the previously unexplored part of the Short Wavelength IR spectrum ($1100nm - 1700nm$). Short Wave Infra Red (SWIR) modality provides clear images at a distance under difficult weather conditions or at night time and has several advantages with respect to NIR. A more detailed description of this technology is provided in Chapter 4. Despite the benefits of the SWIR technology, matching SWIR images against visible images is challenging, because of a photometric difference that exists between images of the same face class acquired at the two different spectral bands.

Chapter 4 introduces a novel encoding algorithm (feature extraction block in Fig. 1.1 and Fig. 1.2) that extracts and compare features between visible and SWIR images. The proposed method consists of a double stage mechanism that employs Gabor filters at the first stage and three multi-scale local operators at the second stage. The employed local operators are Simplified Weber Local Descriptor (SWLD), Local Binary Pattern (LBP), and Generalized Local Binary Pattern (GLBP). The multi-scale SWLD and LBP operators are employed to encode the magnitude responses and the GLBP to encode the phase responses of a small bank of Gabor filters. The proposed method is validated on two SWIR datasets of 48 subjects captured at three different operational distances (2m, 50m, 106m). The implemented algorithm outperforms Faceit G8, a state-of-art face recognition software, at the three distances.

In Chapter 5, the algorithm (introduced in Chapter 4) that employs a bank of Gabor filters and the three multi-scale local operators (SWLD, LBP, GLBP) is demonstrated on FRGC 2.0 dataset. The dataset contains face images of 100 subjects acquired in the visible spectral band. The employed images are collected at a close distance and in a controlled manner. It is shown that the adoption of a single local operator, particularly the SWLD operator applied to the magnitude response of Gabor filters, is sufficient to obtain high recognition rates in the visible spectral band. Furthermore, the obtained matching performance is compared with the performance of the commercial software Faceit G8.

In Chapter 6 we develop a simplified and also optimized version of the matching scheme demonstrated in Chapter 5. The proposed method encodes only the magnitude responses of Gabor filters by applying a single-scale SWLD operator. The encoding algorithm is employed for surveillance of individuals in a multi camera scenario. Particularly, a network of wireless cameras captures multiple pseudo-frontal views (± 20 degrees of out of plane rotation) of the same individual in the visible spectral band. The obtained images may have very low resolution, motion blur and pose variation. A quality based selection process is first employed to discard some of the images that are not suitable for recognition.

The network design and acquisition process follow the work by Kulathumani et al [3]. In our implementation, we further introduce an eye detection stage based on Haar wavelets and Adaboost classifiers [4] with respect to [3]. Since the face images are acquired at different distances from the camera, we geometrically normalize them according to a pool of standard resolutions by using the coordinates of detected eyes. This approach has the advantage to reduce the processing time when probe images have small resolutions. The images are transmitted to a fusion center and Gabor-filtered. The magnitude responses of filters are further encoded with an optimized SWLD operator. Finally, a matching score fusion strategy is adopted to process multiple views of the same individuals. All operational blocks in Fig. 1.2 are employed and described in the chapter. Matching ability of the method is evaluated on a dataset of 29 subjects collected at WVU. The matching results show that the proposed approach has similar or better matching performance compared to the performance of Faceit G8 and other methods based on local operators.

1.4 Summary of Contributions

The main contributions of this dissertation are as follows:

- Three non linear adaptive methods to enhance performance of biometric matchers by employing image quality metrics are developed. The three methods are applied to uncontrolled subset of the FRGC 2.0 face dataset.

- A novel encoding algorithm for cross-spectral comparison between Short Wave Infra Red (SWIR) and visible spectral bands is designed. A small bank of Gabor filters and three multi-scale local operators (SWLD, LBP, GLBP) are involved in encoding. The method is demonstrated on two datasets each composed of data from 48 subjects: PRE-TINDERS and TINDERS datasets. Both datasets include SWIR face images. PRE-TINDERS involves images acquired at a short distance (2 meters). TINDERS dataset includes long range images captured at two different operational distances (50 and 106 meters). The verification and identification performance of the algorithm are compared to the performance of the commercial software Faceit G8.

- The cross-spectral encoding method that adopts Gabor filters and the three multi-scale local operators (SWLD, LBP, GLBP) is further analyzed and applied to visible data. The experiments are carried out on the benchmark FRGC 2.0 dataset having images acquired at a short distance and under controlled settings. It is shown that adoption of an individual multi-scale operator is sufficient to achieve high recognition rates in such settings. Particularly, it is demonstrated that a method that adopts Gabor filters and SWLD provides high recognition rates. The recognition performance of the algorithm is compared to the performance of the commercial software Faceit G8.

- A fully automatic face recognition software that adopts Gabor filters and SWLD operators to encode face images is implemented. The software is designed to work with images of variable (mostly low) resolution acquired by a wireless camera network. It is also optimized to operate in near-real time. A multi-resolution template handling method is introduced, and a centralized fusion strategy is also applied to take advantage of the multiple views of the same subject in the network. The performance of the algorithm is evaluated on a dataset of 29 subjects collected at WVU. We further compare the proposed method with the commercial software Faceit G8 and other recognition schemes that adopt local operators.

Chapter 2

Literature Review

2.1 Face Recognition: Overview of Existing Methods

Face recognition has been an active research topic since the 1990s. The work of Turk and Pentland [5] on Principal Component Analysis (PCA) represents a milestone in the face recognition literature. The PCA or the Eigen-faces recognition method is introduced by Kirby and Sirovich [6] and demonstrated by Turk and Pentland on a face dataset of 16 face classes having 2500 images. The method is based on linear projections. A face image is represented in the feature space by a vector of weights that are obtained by multiplying a normalized face image treated as a vector by an empirical covariance matrix W. The matrix W is obtained by evaluating eigenvectors of the pseudo covariance matrix (scatter matrix) of the training images vectors.

The PCA method was extended to Linear Discriminant Analysis (LDA) or Fisher-faces [7] that, in the training stage (matrix estimation stage), takes into account of the intra-personal (or within-class) variation of the same individual due to pose, illumination and other factors. In this case, the linear projection matrix is produced by maximizing the ratio of the determinant of the between-class scatter matrix of the projected training samples to the determinant of the within-class scatter matrix of the projected training samples. The method often provides better recognition performance compared to PCA [7], however, it can only be applied in the case when multiple samples of the same subject are available for training.

Independent Component Analysis (ICA) is also a generalization of the PCA method [8]. ICA method requires only a single sample for training face class to create the linear projecting matrix. While the PCA algorithm minimizes the correlation (the second order statistic) between training face examples by adopting Singular Value Decomposition on a scatter matrix, the ICA algorithm attempts to make the training examples statistically independent by employing an objective function that involves higher (greater than two) order statistics (e.g. negentropy).

The encoding methods presented in the previous paragraphs are all known as linear methods. There are also a few nonlinear methods that are applied to face recognition. Some of these non linear algorithms for face recognition employ Neural Nets (e.g. Radial Basis Functions [9] or Feed Forward [10]) for the classification task. The basic idea is to reduce the dimensionality of the input images with a method (e.g. PCA) and then adopt the obtained vectors as input to a non linear network that outputs classification results. To employ raw images as input to nonlinear networks is not feasible because of the high dimensionality of images.

The work of Kim [11] employs a non linear extension of PCA called Kernel PCA (KPCA) for the face recognition problem. The method is similar to PCA, however in the eigen-decomposition problem each entry of the empirical covariance matrix is replaced with a Kernel function that can be represented as the inner product of two non linear functions (same mapping) applied to the correlating vectors (training images) pair. The non linear mapping projects the images into a higher dimensional space with respect to the input space such that a better separation of the training examples is obtained. The approach is demonstrated on the ORL face dataset composed of 40 subjects [12]. The results show that the KPCA provides lower Equal Error Rates (EERs) compared to EER of the PCA algorithm and to other non linear methods based on Neural Nets.

Elastic Bunch Graph Matching [13] is an algorithm that adopts Gabor filters [14] for face recognition. The encoding algorithm evaluates the filter responses of Gabor filters in small areas around fiducial points (e.g. eye, nose, mouth, etc) resulting in Gabor “jets”. The Gabor jets are organized in a graph form in which each edge corresponds to the distance between the considered points. The algorithm in the training stage builds a general face graph model in which each vertex is composed of a stacked group or a "bunch" of jets evaluated at a considered point for different subjects. When a new image has to be matched against one in the database, firstly a graph that maximizes a graph similarity function between the general graph model and a candidate model is found. The similarity function takes spatial (global and local) distortions into account and evaluates best jet similarities between the considered image jets and the model jets in each bunch. Once the distinctive model is generated, the obtained graph is matched against the individual graph models to find the highest similarity in the database (identification case).

The work of Lanitis et al. [15] adopts Active Shape Model (ASM) introduced by Cootes et al. [16] for the face recognition problem. The approach models the shape of a given set of facial landmarks. In the training stage landmarks (coordinates) of several faces are labeled to create a training set. The PCA approach is then applied to landmark point vectors to create a shape model. The models obtained on face probe images are used to warp the images to a mean face. The resulting images are shape-free, and an appearance based method (e.g. comparison of average gray-values on square patches) can now be applied to the obtained face images.

Edwards et al. [17] introduce Active Appearance Models (AAM) for face recognition. The method is a generalization of the ASM method. The ASM is adopted to find facial landmarks. Then the ASM method involves not only the shape model parameters but also a set of appearance features (e.g. gray-variation around the estimated landmarks). A concatenated feature vector of shape and appearance parameters is created and a PCA method is applied to this vector. At the matching stage, the AAM algorithm evaluates the distance between the probe model and comparing model iteratively, allowing for the model parameters (e.g. PCA vector coefficients) to vary within a certain range (evaluated at the training stage) to obtain the best fit within the current model. The method is demonstrated on a dataset of 25 subjects.

Ahonen et al. [18] [19] introduce the Local Binary Pattern (LBP) operator and apply it to the face images. The local operator is adopted to extract the local structure of face images, and the distribution of the local features are evaluated to create a template. The chi-square measure is adopted as a distance metric between two encoded images. In recent years, the LBP operator has become popular, and many improvements or variations have been proposed. For example, the work of Zhang [20] et al. improves LBP recognition performance by employing boosting methods and by reducing the dimension of features.

Zhang et al. [21] employ magnitude responses of a bank of 40 Gabor filters that are applied to face images. The magnitude responses are encoded with a single scale 8-bits LBP operator. The obtained local features are displayed in the form of a histogram. Intersection of histogram is adopted as a measure of similarity. The results are demonstrated on the FERET dataset [22].

Aly [23] makes use of the Shift Invariant Feature Transform (SIFT) to encode face images. SIFT has been introduced by David Lowes [24]. The algorithm in [23] is applied to AT&T and Yale databases and its' performance compared to PCA and LDA performances. The SIFT-based algorithm is shown to outperform the other two methods while employing a training set of smaller size.

The work of Tan and Trigs [25] combines LBP local features with local Gabor features. The obtained feature vectors are projected into a subspace by employing PCA, and then kernel discriminative common vectors (KDCVs) are adopted to extract the discriminative features. The algorithm is demonstrated on FRGC [26] and FERET datasets.

The work of Savvides et al. [27] adopts Minimum Average Correlation Energy (MACE) correlation filters for face recognition. For each individual in the database a correlation filter (template) is obtained by employing multiple training samples of the same face class and by enforcing constraints on the correlation function obtained from such samples. Different constraints (design) for the individual correlation filter are possible [28]. The probe image is multiplied in the Fourier domain with a template in database to obtain a cross-correlation function. The similarity measure is the Peak to Side-lobe Ratio (PSR) of the obtained cross-correlation function.

The work of Chan et al [29] employs a multi-scale LBP applied to raw face images, followed by a LDA approach to reduce the dimensionality of features. The algorithm is applied to the FERET and XM2VTS databases.

Zhang et al. [30] encode the phase response of a bank of Gabor filters applied to face images. The introduced method, called Histogram of Gabor Phase Patterns (HGPP), is demonstrated on FERET and CAS-PEAL [31] face datasets.

The Effective-Gabor Volume-LBP (E-GV-LBP) introduced by Lei et al. [32] encodes the magnitude responses of a bank of Gabor filters that are concatenated to form a 3D image (with XYZ axes). The volume is projected onto the orthogonal planes XY,YZ,XZ, and the LBP operator is applied to the obtained projections and transformed to a histogram form. The main drawback of the method is that it is computationally expensive. As a solution to the computational problem an "effective" formulation is proposed: the LBP operator is simultaneously applied to the three directions (XYZ). The method is applied to the FERET and FRGC 2.0 dataset.

The paper of Xie et al. [33] adopts the magnitude and phase of a bank of 40 Gabor filters to match images within the visible spectral band. In this work, the LXP and/or LBP operator has been adopted to encode magnitude and phase, followed by a block-based Fisher Linear Discriminant (BFLD) approach being applied to the FERET and FRGC 2.0 [26] dataset.

The work of Tan and Triggs [34] proposes two face recognition fusion (at the matching score level) schemes : (1) one that combines matching scores obtained by LBP and LTP operators applied to raw images and (2) a method that combines Kernel PCA, Gabor Features and LBP. The encoding schemes are demonstrated on three face datasets Extended Yale-B [35], CAS-PEAL-R1, and FRGC 2.0 under different illumination scenarios. The methods are designed to achieve illumination invariance and high recognition rates in the visible band.

2.2 Multi-Spectral and Cross-Spectral Face Recognition

The literature contains a number of works that perform face comparisons within the same spectral band, NIR (Near Infra-Red) or IR (Infra-Red). Several methods that combine diverse spectral image modalities at different levels (e.g. image, features, matching scores) and some works that address the cross-spectral comparison visible versus NIR are also available. However, a face recognition method that performs a cross-spectral comparison between visible and SWIR (Short Wave Infra-Red) face images has not been developed. Below, we review a few existing publications on the topic.

In the work of Chen et al. [36] a study of face recognition in the thermal IR and visible spectral bands is performed, by using PCA [37] and $Faceit\ G5$. It is shown that the performance of PCA in the visible spectral band is higher compared to the performance of PCA in the thermal IR spectral band, and that these data fused at the matching score level result in performance similar to the performance of the algorithm in visible band.

Pan et al. [38] demonstrate the effectiveness of a hyperspectral approach within the NIR spectral band. They collected a dataset of face images acquired at 31 narrow spectral bands in the range of 0.7μm-1.0μm. Then a spectral reflectance vector is evaluated in a few face (square) regions at the different wavelengths and is employed as a feature vector. The experiments performed on the hyperspectral dataset show that the adopted features are able to recognize individuals having different poses and facial expressions.

In their work, Kong et al. [39] perform fusion of NIR and thermal IR face images in the Discrete Wavelet Transform domain by employing images from the NIST/Equinox and the UTK-IRIS databases. They showed that, when the fused images are fed to the $Faceit$ recognition software, the resulting matching performance improves with respect to the case when the same face classes are compared within the same spectral band.

Li et al. [40] propose a method to compare face images within the NIR spectral band with different illumination scenarios. Their face matcher uses the LBP [18] operator to achieve illumination invariance and is applied to NIR images acquired at short distance (less than one meter).

Hariharan et al. [41] introduce an Empirical Mode Decomposition method to combine visible images with thermal IR. The images with the two different modalities are ordered in vectorial forms and an iterative mode decomposition algorithm is applied to obtain feature vectors. The obtained feature vectors from the two spectral bands are linearly combined. Then the algorithm uses the inverse transformation applied to the combined feature vector to obtained a fused image. The method is demonstrated to have better performance than wavelet-based fusion [39], PCA-based fusion and image averaging.

In their recent works Akhloufi and Bendada [42], [43] experiment with images from Equinox Database [44] (it includes visible, SWIR, MWIR, and LWIR images) and Laval University Multispectral Database [45](includes visible, NIR, MWIR, LWIR data). The first work [42] evaluates recognition performance within each spectral band by using a set of known face matching techniques.

In the second work [43] (performed on the same data) a classic Local Ternary Pattern (LTP) operator and a new Local Adaptive Ternary Pattern (LATP) operator are adopted to extract features from images. The authors involve multiresolution analysis in the "texture space" to fuse images from different spectral bands. They report that the fusion of face images from different spectral bands leads to improved recognition rates with respect to the case when images are matched within the same spectral band.

Lin et al. [46] introduce a Common Discriminant Feature Extraction (CDFE) method that brings images of different modalities (visible light (optical images and photographs), NIR and sketches) into a common feature space. It is shown that the proposed algorithm outperforms PCA and LDA (Liner Discriminant Analysis) [47], kernel PCA and kernel LDA [48], in the visible (optical)-NIR comparison and when visible (photo) images are matched against sketches.

The work of Yi et al. [49] proposes Canonical Correlation Analysis (CCA) to compare visible images (VIS) against NIR. Firstly, a dimensionality reduction method such as PCA or LDA is applied separately to VIS and NIR images (the training examples are within the same spectral band). Secondly, the CCA method obtains two unknown projection matrices W_{VIS}, W_{NIR} for VIS and NIR data. When VIS and NIR feature vectors are multiplied for their own projection matrices, the resulting correlation between the obtained projections is maximized. It is shown that the estimation of W_{VIS}, W_{NIR} is reduced to a Singular Value Decomposition problem. The method is demonstrated to improve the performance within the same spectral band and in the cross-spectral comparison when the PCA and LDA methods are adopted.

Liao et al. [50] apply a MB-LBP (Multi-scale Block Local Binary Patterns) descriptor to NIR and visible face images. The images in both spectral bands are preprocessed with Difference of Gaussian (DOG) filters, and then the MB-LBP operator is applied to the normalized images. A Gentle AdaBoost feature selection technique and a regularized LDA method are further evaluated and applied to a testing set of 52 subjects having images in both spectral bands. The implemented approach is shown to outperform CDFE [46], PCA-CCA and LDA-CCA [49] methods when visible images are compared to NIR.

Lei and Li [51] present a subspace learning framework named Coupled Spectral Regression (CSR) to match NIR versus visible (VIS) images. The algorithm first models the NIR and VIS separately and then learns two associated projections to project heterogeneous data to a common classification space. The algorithm is compared to PCA, LDA, CDFE [46], PCA-CCA and LDA-CCA [49] and others. The method is shown to be best under certain testing conditions.

In their paper Yi et al. [52] encode images captured at NIR and visible spectrum by adopting a Laplacian of Gaussian (LoG) filter. The filtered images are further converted into binary images that are locally partitioned into small patches. The method compares common patches (partial faces) between visible light images and NIR images. The experiments are performed on MBGC [53] data. The proposed method is compared to CDFE [46], PCA-CCA and LDA-CCA [49].

Wang et al. [54] adopt a Face Analogy (FA) method to compare VIS and NIR images. The method creates a correspondence between VIS images and NIR images by employing a local (pixel by pixel) normalization process that is applied to images acquired by different spectral modalities. The pixel-wise and scan-line estimated relationship is then employed to synthesize the VIS images from input NIR images. The images are manually aligned by employing 68 landmarks. The obtained VIS images are compared with the gallery images pixel by pixel. A LoG filter is also applied to further enhance the preprocessing method. The results show that the processing method provides a large increment of performance in the cross-spectral comparison of NIR vs VIS compared to the case when images are not preprocessed.

The work of Chen et al. [55] proposes a generative approach to compare NIR images to VIS images. A dictionary of patches from NIR and VIS spectral bands is created from training images; a correspondence between patches of the two vocabularies at the same face locations is created. In the testing stage LBP-based features are evaluated from NIR images, patch by patch, and the best correspondences in NIR vocabulary are found. The correspondent patches of the visible vocabulary are adopted to synthesize a VIS face image by employing a linear manifold mapping. The method ouperforms PCA-synthesis [56], LLM [57], LDA-CCA [49], and FA [54].

The work of Klare and Jain [58] employs a method based on LBP (Local Binary Patterns) and Histogram of Gradient (HOG) [59] features, followed by an improved LDA algorithm to reduce the dimensionality of feature vectors. This encoding strategy is applied to NIR and color images for their cross-spectral matching. As shown, the algorithm outperforms Cognitec's FaceVACS [60].

The recent method of Zhang et al. [61] synthesizes VIS images from NIR images. The methods divides training images from VIS and NIR modalities into small image patches that are randomly chosen to create two (VIS and NIR) overcomplete sparse vocabularies. The method firstly finds the best sparse approximation of the input NIR image by employing the NIR dictionary and L^1 minimization. Then the recovered projection coefficients are employed to create the VIS images by employing the VIS dictionary. A method that creates a correspondence between the VIS and NIR dictionary is proposed. It is shown that the method produces higher recognition rates compared to the case when input images are not processed or when the LLM method [57] is employed.

Chapter 3

Adaptive Quality Mappings

3.1 Introduction

In the previous chapter, we have described a set of multi-spectral and cross-spectral encoding algorithms that have application in a surveillance scenario. In this setup, where non ideal conditions such as pose variation, motion or out of focus blur, low light or uneven illumination and bad weather conditions may be present, the recognition task can be challenging. Hence, when uncontrolled setting variations are present, it is very important to assess the quality of the acquired face sample; if very poor quality samples are employed, only low recognition rates may be achieved.

The most common definition of biometric sample quality is at the image or signal level. A quality checking block is introduced into every biometric system to ensure that enrolled images/signals have sufficient quality to be further processed. All recent biometric systems extract a vector of quality measures. The components of a vector of quality measures, however, rarely have equal importance in terms of their relationship to the performance of the matcher. In practical applications (e.g. surveillance, authentication, etc.), it is necessary to keep a single biometric quality measure in order to decide if the biometric samples are suitable for further processing and matching.

Hence in a biometric system, a method that selects suitable quality metrics or that combines several quality measures into a single quality index has to be designed. The obtained quality index should retain a large amount (as much as possible) of the original information contained by the considered quality vector.

Quality metrics can also be employed at the matching stage; in this case quality vectors of a matching pair are employed simultaneously by the matcher to improve the recognition performance.

In this chapter, we propose three adaptive methods that employ quality metrics to improve the recognition performance of face matcher. The proposed methods are complementary to the face encoding process (e.g. scheme described in the next chapter).

In the next section, we review some works on the topic.

3.2 Related Work

Grother and Tabassi in their paper [62], relate the NFIQ (NIST Fingerprint Image Quality) to the matching performance of a fingerprint matcher. A Normalized Matching Score (NMS) measure is introduced and they show that quality of the fingerprint samples is related to NMS (especially genuine scores) in the fingerprint recognition problem.

Fierrez-Aguilar et al. [63] adopt a fingerprint quality metric (single index) to combine two fingerprint matchers. The matching scores produced by a ridge and minutiae-based matchers are normalized to a common range and linearly combined. The weights of the linear combination are proportional (fixed percents) to the quality of the fingerprint probe. The method is demonstrated to have better performance than the ones obtained with a sum rule [64].

The papers of Nandakumar et al. [65] [66] estimate the joint densities of quality and matching scores (genuine and imposter distributions). Then a likelihood ratio test between the estimated genuine and imposter distribution is adopted to evaluate the verification performance. In [65] the method is demonstrated on fingerprint and iris biometrics; for each modality a quality-based density is evaluated; hence a multi-modal distribution is obtained as a product of the individual density modalities. An improvement is obtained with respect to the case when the modalities are combined without the quality measures. In [66], the joint densities of single biometrics are described by a Gaussian Mixture Model (GMM). The model parameters are estimated with an Expectation Maximization (EM) algorithm.

The same method of [65] is adopted to obtain the verification performance. The method is demonstrated on several biometrics (fingerprint, iris, face, voice) and multiple matchers for the same biometric modalities are used.

The works of Kryszczuk and Drygajlo [67] [68], involve biometric sample quality at the matching stage by concatenating matching scores due to the original matcher and quality measures. These quality-based matchers are called Q-stack classifiers; the method is demonstrated on face and fingerprint biometrics by adopting Support Vector Machines (SVM), Bayes classifiers and Linear Discriminant-based (LDA) classifiers. In spite of the fundamental theory presented in these works, the obtained improvement of performance on individual face and fingerprint modalities are marginal.

Maurer and Baker [69] propose a quality-based fusion method that combines fingerprint and voice modalities. The approach is based on a Bayesian Belief Network (BBN) that models dependencies between matching scores, quality estimates, and global decision variables. These statistical dependencies are in the form of conditional distributions which are modeled as Gaussian, Gamma, Log-normal or Beta distributions. The method produces a recognition result by conditioning decision variables on quality as well as matching scores. The work reports an overall improvement of recognition performance when fingerprints and voice biometrics are fused.

The work of Fronthaler et al. [70] assigns different weights to the scores provided by multiple fingerprint classifiers (experts). The weights are estimated by employing an adaptive Bayes classifier that adopts the quality of fingerprint samples. The importance of the experts is dynamically assigned based on the quality of probe image. In the experiments, the combination of multiple matchers based on quality measure leads to noticeable improvements in performance.

Ozay et al. [71] generalize the NMS measure introduced by Grother and Tabassi [62] and suggest to relate quality and features of biometric face samples to a symmetric normalized match score. The approach employs a Bayesian network to estimate the relationship; the method is demonstrated on a face dataset of 40 subjects having 6 images per class. The work adopts a leave-one-subject-out cross-validation methodology to evaluate the performance. The results show an improvement of performance with respect to the case when only matching scores are adopted.

In this work, we describe three methods to improve the performance of a single matcher. The first method predicts a single QS (Quality Score) index to accept or discard a biometric sample for further processing; the second method evaluates a CS (Confidence Score) measure that trusts/discredits a matching score obtained from an image pair by employing quality metrics; the third method treats a vector of quality measures as a weak feature for the recognition task.

3.3 Methodology

This chapter makes use of non linear predictive tools for the design of quality enhanced matchers.

The proposed tools predict a set of quality measures and scores. In each of the three methods described below, the functional relationship between vectors of quality measures and the predicted (estimated) measures is not known and has to be modeled. The modeling problem is stated as a multivariate regression problem:

$$Y = f(X_1, \dots, X_K), \tag{3.1}$$

where f is a multivariate adaptive mapping, variable Y is the estimated output variable, i.e. that characterizes the overall quality, confidence in matching score, or quality enhanced decision, and $X_1, \dots, X_K$ is a vector of K input (predictive) variables, such as a vector of quality measures for a biometric sample or a concatenated vector of quality measures and matching scores. Since the true relationship between the input and output variables is not known, it is estimated using a set of labeled training data. The multivariate adaptive mapping $f(\,\cdot\,)$ can be implemented using a variety of multivariate functions such as Multivariate Adaptive Regression Splines (MARS), Radial Basis Function Networks (RBFN), Bayesian Networks (BN), Neural Networks (NN) and other mappings. The results reported here are obtained using a Feed Forward Neural Network (FFNN) (see Sec. 3.4.2).

3.3.1 Quality of Sample (QS)

Here we suggest a single quality index characterizing the overall quality of a biometric sample. Consider a set of biometric samples. Associated with these data are vectors of quality measures.

For example, for face biometrics, a matcher may be designed to employ the chi-square distance as Matching Score (MS), while the quality vector may be composed of 5 quality measures: (1) symmetry, (2) sharpness, (3) lighting, (4) resolution, (5) expression. These data can be used to design a single quality index by combining the entries of the vector. The main design requirement is that the values of the single quality index must be related to the performance of the original matcher. A description of the quality measures and the matching method adopted in our work is provided in Sec. 3.4.1.

We propose to use d-prime index as the combined quality index. This index is related to the relative entropy, which is asymptotically related to the performance of the matcher. Since the true distributions, means and variances of matching scores are not known, they are estimated using available labeled data. All labeled data are subdivided into two non-overlapping sets: training set and testing set. QS_A, the overall quality index of sample A, is defined as:

$$QS_A = \frac{|m(Imp.\ Scores)_A - m(Gen.\ Scores)_A|}{\sqrt{var(Imp.\ Scores)_A + var(Gen.\ Scores)_A}}, \tag{3.2}$$

where $m(\cdot)_A$ and $var(\cdot)_A$ are the sample mean and sample variance of genuine and imposter scores formed by involving the sample A.

Using the labeled training set, QS is estimated for every biometric sample. This requires that a set of genuine matching scores and a set of imposter matching scores involving the same biometric sample be formed. For unlabeled biometric samples, this task becomes almost impossible. However, having quality vectors associated with each biometric sample makes it possible to *predict* the QS of unlabeled data. The QS can be obtained as the output parameter of a nonlinear multivariate adaptive mapping applied to a vector of quality measures (input parameters). Let $\mathbf{Q}_A = [Q_{A,1}, \ldots, Q_{A,K}]^T$ be a vector of K quality measures characterizing a biometric sample A. The superscript T indicates the transpose operation. Let $f_{QS}(\ \cdot\ )$ be a nonlinear multivariate adaptive mapping that maps a vector of quality measure $\mathbf{Q}_A$ into the quality index QS_A. Let $\hat{f}_{QS}(\,\cdot\,)$ be its estimated version. Then QS_A is predicted as $QS_A = \hat{f}_{QS}(\mathbf{Q}_A)$. The predicted value of QS can then be used to decide if the underlying biometric sample should be retained or discarded to improve the performance of the original matcher.

3.3.2 Confidence in Scores (CS)

The second method evaluates the confidence level assigned to matching scores associated with a pair of biometric samples. The confidence in genuine and imposter scores (CS) is defined for a system that employs a distance measure as a Matching Score (MS):

$$CS_G = \begin{cases} 0, & MS_G < Q(MS_G)_x, \\ -\frac{MS_G - Q(MS_G)_x}{Q(MS_G)_y - Q(MS_G)_x}, & otherwise, \end{cases} \tag{3.3}$$

$$CS_I = \begin{cases} 0, & MS_I > Q(MS_I)_{1-x}, \\ -\frac{Q(MS_I)_{1-x} - MS_I}{Q(MS_I)_{1-x} - Q(MS_I)_{1-y}}, & otherwise, \end{cases} \tag{3.4}$$

where $Q(MS_G)_x$ and $Q(MS_I)_y$ are the quantile points at the quantile x and y for genuine and imposter scores, respectively. The levels of the quantiles were selected empirically. In our experiments, the values resulting in significantly improved verification performance are $x = 0.7$ and $y = 0.9$. Fig. 3.1 illustrates genuine and imposter distributions typical for a distance-based MS. The expressions for the CS are identical for matchers that adopt similarity scores with the only difference being that roles of the genuine and imposter scores are reversed (G subindex is replaced with I subindex and vice versa). In this work we provide examples that employ distances as matching scores.

The CS of biometric samples is predicted using a nonlinear adaptive mapping and vectors of quality measures (input parameters). The nonlinear adaptive mapping is trained using a set of labeled data in the form of vectors of quality measures and the corresponding CS values obtained using equations (3.3) and (3.4). At the testing stage, the CS value is predicted by employing only vectors of quality measures as input parameters. The obtained information is used to keep or discard the corresponding matching score in order to improve the performance of the original matcher. The procedure of predicting the CS of matching

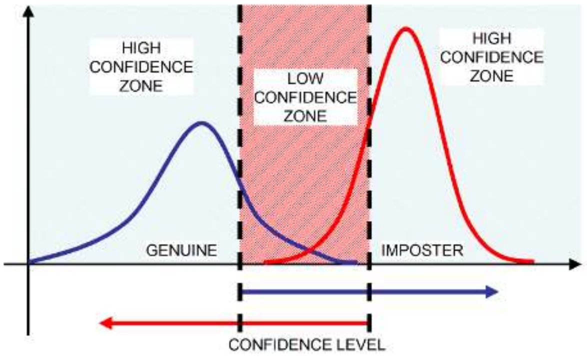


Figure 3.1: Illustration of the Confidence in Scores (CS).

scores between two biometric samples A and B is $CS_{AB} = \hat{f}_{CS}(\mathbf{Q}_A, \mathbf{Q}_B)$, where $\hat{f}_{CS}(\,\cdot\,)$ is a multivariate adaptive mapping (FFNN in our case) estimated using training data.

3.3.3 Quality Metrics and Template Features (QMTF)

The third method suggests to treat quality vectors as weak features that can be combined with biometric template features. Let $\mathbf{Q}_A$ and $\mathbf{C}_A$ be a vector of K quality measures and a template vector associated with a biometric sample A. Then the extended template of the sample A is the vector $\mathbf{F}_A = [\mathbf{Q}_A^T, \mathbf{C}_A^T]^T$.

When two biometric samples A and B are compared, the distance (or similarity) between $\mathbf{C}_A$ and $\mathbf{C}_B$ will be saved as a matching score MS_{AB}, while the quality vectors $\mathbf{Q}_A$ and $\mathbf{Q}_B$ will be treated as extra dimensions that may improve the performance of the original matcher, provided these dimensions contain sufficient discriminative information. As shown in Fig. 3.2, the input to the high dimensional classifier is a vector $[\mathbf{Q}_A^T, \mathbf{Q}_B^T, MS_{AB}]^T$. The output of the classifier is a decision made by the classifier. The decision is a binary valued variable corresponding to $\{Genuine, Imposter\}$. The high dimensional classifier is implemented using a nonlinear adaptive mapping. Denote $\eta_{A,B}$ as the output variable predicted using vectors of the quality measures and the matching score of the biometric samples A and B. Then the prediction procedure is described as $\eta_{A,B} = \hat{f}_{QST}\left(\mathbf{Q}_A^T, \mathbf{Q}_B^T, MS_{AB}\right)$, where $\hat{f}_{QST}(\,\cdot\,)$ is an estimated version of $f_{QST}(\,\cdot\,)$.

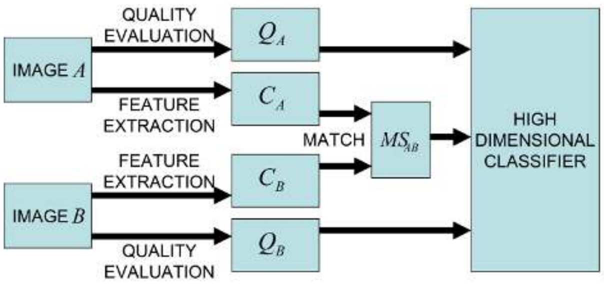


Figure 3.2: Combining Quality Metrics and Template Features ($QMTF$).

3.4 Experimental Results

3.4.1 Face Dataset and Face Quality Metrics

The experiments are performed on uncontrolled (varying illumination conditions) images of FRGC 2.0 face dataset [72]. The dataset is composed of 202 face classes having 10 images per class (5 have a neutral expression, 5 have a smiling expression). Examples of face images from FRGC 2.0 uncontrolled dataset are shown in Fig. 3.3.

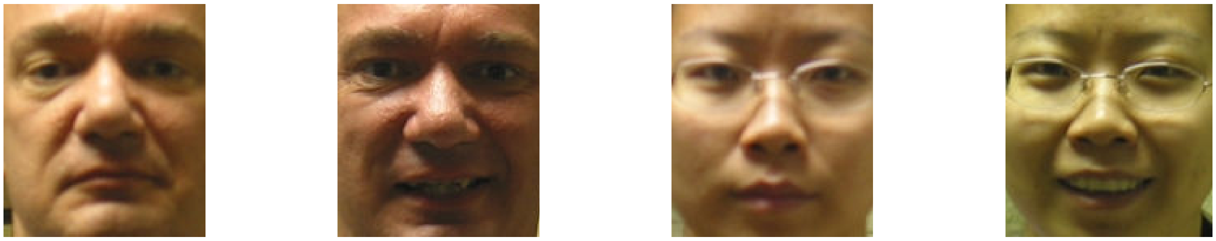


Figure 3.3: Sample images from FRGC 2.0 dataset with uncontrolled illumination conditions.

The face images are compared using a LBP matcher [18], resulting in a total of 9,090 genuine scores and 2,030,100 imposter scores. A chi-square distance is adopted as matching score (MS) [18]. To demonstrate our approach we employ 5 face quality metrics: (1) symmetry, (2) sharpness, (3) lighting, (4) resolution, (5) expression. The symmetry quality factor measures the amount of symmetry between left and right parts of the face, discriminating against both non central pose and uneven illumination along the face longitudinal direction [73]. To evaluate the symmetry measure, the normalized face image I is reflected along a vertical central axes located approximately at half width of the image; a reflected image I_R is obtained. Then both images are encoded in histogram form with a LBP operator using the same procedure adopted to produce the matching scores [18]. The symmetry measure is defined as the chi-square distance between histogram features $\hat{\mathbf{H}}$, $\hat{\mathbf{H}}_{\mathbf{R}}$ of images I, I_R:

$$S = \sum_{i=1}^{N} \frac{[\hat{H}(i) - \hat{H}_R(i)]^2}{[\hat{H}(i) + \hat{H}_R(i)]}, \tag{3.5}$$

where N is the length of the features vectors.

The sharpness metric evaluates the magnitude of the face image gradient [73]; given an image $I(x,y)$ of size $M \times N$ and let $G(x,y)$ be the gradient values at locations (x,y), the sharpness measure is defined as:

$$G = \sum_{x=1}^{M-2} \sum_{y=1}^{N-2} G(x,y). \tag{3.6}$$

The lighting measure evaluates the variation of luminance across different parts of the face. To assess the lighting metric, the face area is divided in K blocks of 16x16 pixels, and for each block the median pixel value M_i is calculated. The lighting factor is obtained as:

$$L = MAD\{M_1, M_2, ..., M_K\}, \tag{3.7}$$

where MAD is the Median Absolute Deviation estimator. The adopted metric is similar to the lighting quality measure implemented in [74] for the iris biometric.

The resolution factor indicates the original pixel distance between the eyes before geometrical normalization:

$$R = \sqrt{(X_L - X_R)^2 + (Y_L - Y_R)^2}, \tag{3.8}$$

where $(X_L, Y_L), (X_R, Y_R)$ are the coordinates of the left and right eyes in the original image space.

The quality measure evaluating the presence of expressions is a simple binary label (1,-1) that indicates if the subject has a neutral expression or a smiling expression (open mouth, for example).

The first three quality factors (symmetry, sharpness, lighting) are automatically evaluated (no training is required) while the resolution values and expression labels are obtained from the original metadata included in FRGC 2.0 dataset. These last two factors can also be automatically evaluated by applying an eye detector (e.g. [4]) for the resolution metric and by adopting an expression detection algorithm (e.g. [75]) for the expression quality factor, however it would require further portioning of FRGC data or involving other datasets in order to train the algorithms. For simplicity, we employ the original metadata that are available as a part of the face dataset. These assumptions don't influence the validity of the proposed methods.

3.4.2 Neural Network

The estimated nonlinear mappings are obtained by using Feed Forward Neural Network (FFNN), since they are able to describe any non linear relationship between the input and output parameters [76], [77], [78]. To establish topology of the network, we adopt a classic trial and error approach. We create a small network and we introduce new links and neurons until a desired value of the cost function (Mean Square Error) is reached on the training set. We validate the obtained design starting from a larger network and removing links and neurons until similar or better performance is obtained. The final design is achieved by trading off complexity and performance of the network. We adopt a single hidden layer neural net for the QS and CS methods and a two layer network for the $QMTF$ evaluation; all the hidden layers of the designed networks are composed of the sigmoid neurons (i.e. $1/(1+e^{-x})$) and the output layer is linear. An example of FFNN employed in the following experiments is depicted in Fig. 3.4.

For the QS method, 4 neurons are sufficient to describe the non linear relationship between input and output, instead for the CS method, 5 neurons are employed. During the training stage, an ensemble of neural net models are created by repeating the training process

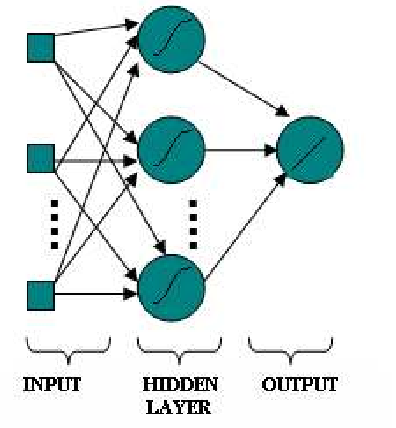


Figure 3.4: Example of a typical Feed Forward Neural Net employed in the experiments.

five times. Each model $\hat{f}_i(\,\cdot\,)$ is obtained by adopting a random subset (a fixed percent L) of available data for learning and remaining data ($V = 1 - L$ percent) for validation. For the QS the ratio L/V is 70/30 and for CS it is 80/20. The training process is stopped when the Mean Square Error (MSE) increases for six consecutive iterations on the validation set. In the testing stage, the output Y is obtained as a weighted sum of the model $\hat{f}_i(\,\cdot\,)$ applied to the input testing vector $\mathbf{X}$:

$$Y = \sum_{i=1}^{5} w_i \hat{f}_i(\mathbf{X}), \tag{3.9}$$

where the weights w_i are obtained as a correlation between training data and predicted training data. The predicted training data are determined by applying the non linear mapping $\hat{f}_i(\,\cdot\,)$ to the entire training set. The weights are normalized such that $w_1 + w_2 + ... + w_5 = 1$. Instead, for the $QMFT$ method, a single training procedure is employed for computational reasons. In this case a much larger amount of training data is adopted to train the network and to build a FFNN composed of two layers. The first hidden layer has 16 neurons and the second layer has 2 neurons. A ratio L/V of 60/40 is employed.

3.4.3 QS Evaluation

To form the training set, we randomly select 101 face classes from FRGC 2.0 dataset for a total of $1,010$ images (one half of the available data). The remaining $1,010$ face images are employed as the testing set. Quality vectors (quality factors 1 through 5) and QS of labeled data are adopted to obtain the ensemble of neural net models. The generated set of models are applied (see eq. 3.9) to the quality metrics of unlabeled images to predict QS. During performance evaluation, unlabeled images with a predicted QS value above a certain quantile are retained.

Fig. 3.5 shows three Receiver Operating Characteristic (ROC) curves parameterized by zero, 10% and 50% quantile levels. The ROC curve indicated as "original" is parameterized by zero quantile level, which means that no biometric samples were discarded. In the plot we have an increase of verification performance when we eliminate samples having (predicted) QS values lower than the considered quantile levels.

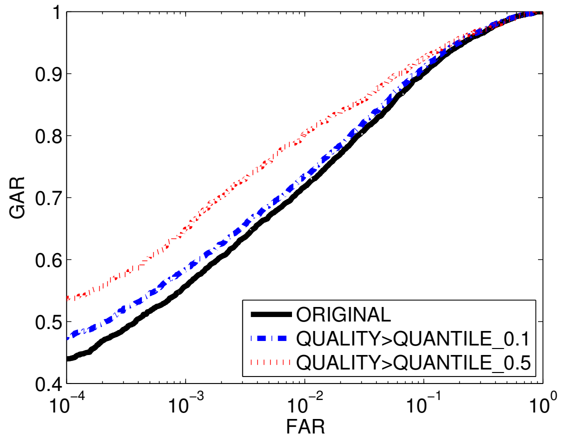


Figure 3.5: Performance improvement for face biometric achieved by selecting only images with high QS values.

Fig. 3.6 summarizes the results of ten trials. We note an improvement of performance as we discard samples with lower QS values. The Equal Error Rate (EER) improves on average by 16% (1.62% absolute decrement) at 50% quantile level. Verification rates evaluated at False Accept Rate (FAR) 10^{-3} reach an average relative increase of about 15% (8% absolute gain) at 50% quantile level. Fig. 3.6 also shows that other indexes of performance such as d-prime and verification rates measured at FAR 10^{-4} increase as low quality samples are removed.

3.4.4 CS Evaluation

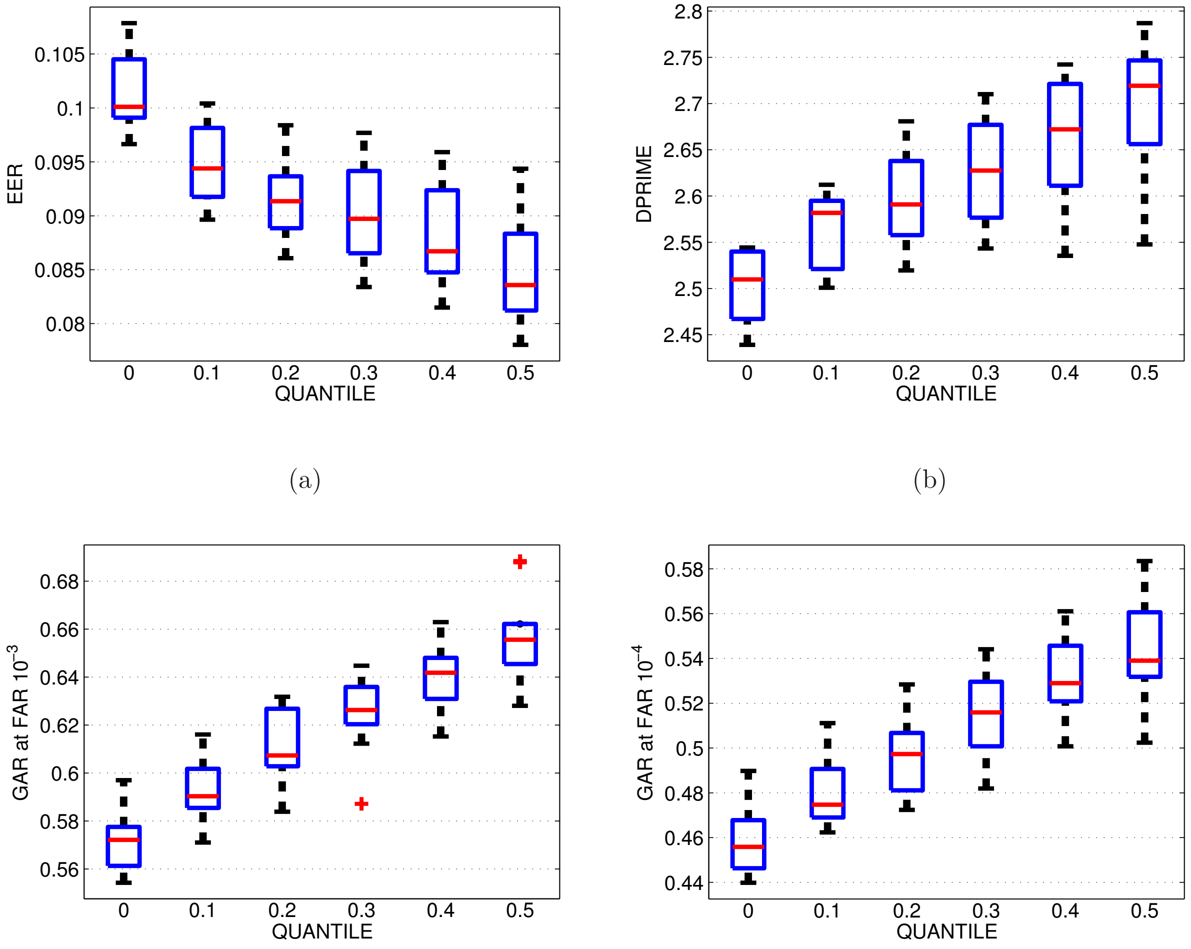
To assess the performance of the CS method, we randomly select a half of available data (images of 101 classes) to form a training set T. We employ all the $4,545$ possible genuine comparisons and a random subset of $4,545$ imposter scores from T. In this case, our experimental results suggest that the method requires a balance (in number) between genuine and imposter scores at the training stage. The matcher was designed to be symmetric with respect to quality vectors, that is, if $\mathbf{Q}_A$ and $\mathbf{Q}_B$ are two vectors of quality measures associated with face images A and B, training included both the pair $(\mathbf{Q}_A, \mathbf{Q}_B)$ and the pair $(\mathbf{Q}_B, \mathbf{Q}_A)$ and the associated matching score. Similarly to the QS case, the training process is repeated five times to obtain an ensemble of models. The obtained set of models are applied to the quality metrics of unlabeled images (other half of the data) to predict CS.

Fig. 3.7 shows three ROC curves: the original curve, the curve formed from face data with the predicted CS values exceeding 10% quantile and the curve formed from face data with the predicted CS values exceeding 50% quantile. The recognition performance improves when low confidence matching scores are discarded.

Fig. 3.8 summarizes the results of ten trials. We observe that EER improves on average by about 13.3% (1.4% absolute decrease) at 50% quantile level. Verification rates calculated at FAR 10^{-3} have an average relative increment of about 13% (7.3% absolute increase) at 50% quantile level. Fig. 3.8 also demonstrates that d-prime index and verification rates calculated at FAR 10^{-4} increase when low confidence matching scores are discarded.

3.4.5 QMTF Evaluation

To evaluate the performance of the $QMTF$ method, we randomly select a half of available data to obtain a training set T. We adopt all genuine $4,545$ comparisons and a random subset of $80,000$ imposter scores obtained from set T. Similarly to the CS case, we employ symmetric quality pairs to train the method, resulting in $9,090$ genuine and $160,000$ imposter



(a) (b)

(c) (d)

Figure 3.6: Performance improvement for face biometric obtained by selecting only images with a quality value larger than a certain quantile. The boxplots summarize the results obtained for 10 trials. Fig. 3.6(a) indicates the values of Equal Error Rate (EER); Fig. 3.6(b) illustrates d-prime values; Fig. 3.6(c) shows the verification rates (Genuine Accept Rates) evaluated at FAR equal to 10^{-3}; Fig. 3.6(d) displays the GARs evaluated at FAR equal to 10^{-4}.

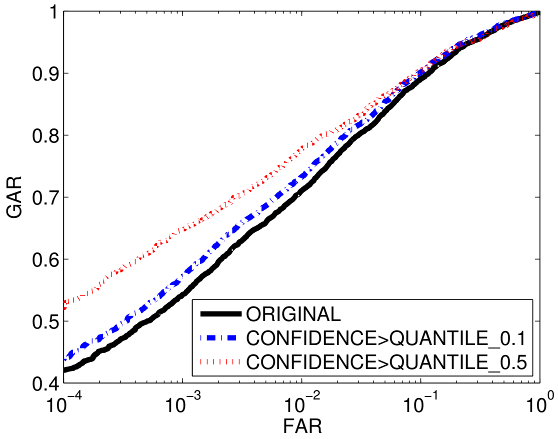


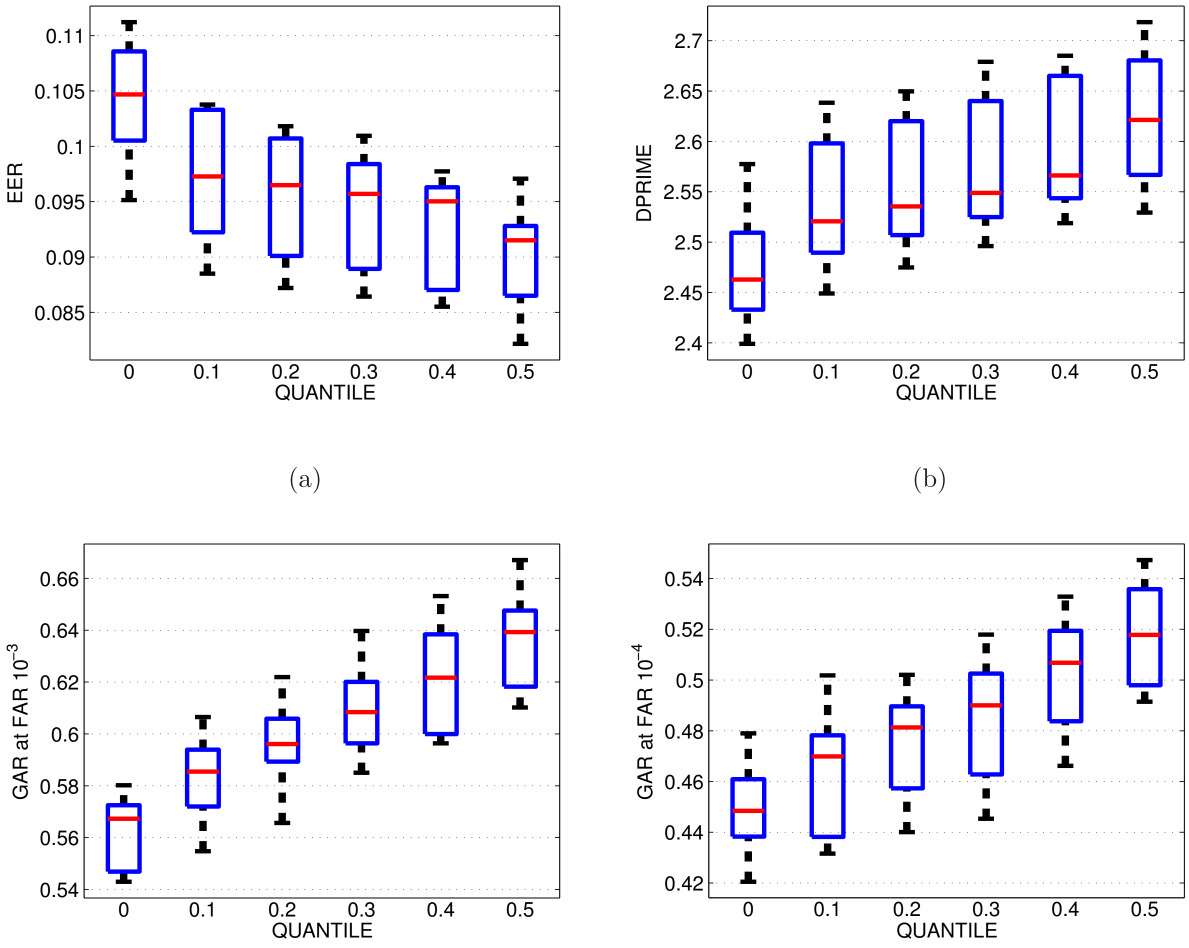
Figure 3.7: Performance improvement for face biometric achieved by selecting only images with high CS values.

training vectors. A typical ROC obtained by employing the $QMTF$ approach is displayed in Fig. 3.9; we note a substantial gain of performance when quality metrics are adopted jointly with matching scores to produce a verification result. The results of $QMTF$ evaluation for 10 trials are summarized in Table 3.1.

	Original EER (%)	New EER (%)	Original GAR (%) at FAR 10^{-3}	New GAR (%) at FAR 10^{-3}
Trial 1	10.43	8.25	55.39	66.04
Trial 2	10.18	7.81	58.28	65.62
Trial 3	11.05	8.66	54.47	62.58
Trial 4	9.53	7.29	58.67	66.97
Trial 5	9.36	7.27	60.17	68.26
Trial 6	9.77	8.21	57.34	67.73
Trial 7	10.85	8.91	59.35	64.74
Trial 8	10.64	8.16	55.05	62.60
Trial 9	10.61	8.14	56.35	64.59
Trial 10	10.47	7.78	53.38	61.18

Table 3.1: Matching performance for face biometric with/without quality metrics.

In columns 2 and 3 it shows the values of EER without/with quality metrics. In columns 4 and 5 the table illustrates the values of Genuine Accept Rate (GAR) evaluated at FAR 10^{-3}. We observe an average improvement of EER values of about 21.8% (2.24% absolute



(a)

(b)

(c)

(d)

Figure 3.8: Performance improvement for face biometric obtained by selecting only images with a confidence level higher than a certain quantile. The boxplots summarize the results obtained for 10 trials. Fig. 3.8(a) indicates the values of Equal Error Rate (EER); Fig. 3.8(b) illustrates d-prime values; Fig. 3.8(c) shows the verification rates (Genuine Accept Rates) evaluated at FAR equal to 10^{-3}; Fig. 3.8(d) displays the GARs evaluated at FAR equal to 10^{-4}.

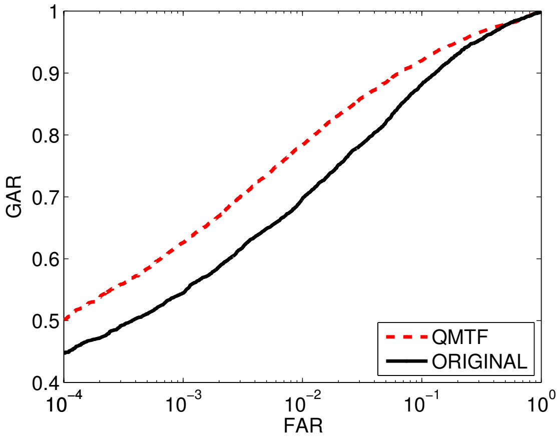


Figure 3.9: Performance improvement for face biometric achieved by combining Quality Metrics with Features Templates ($QMFT$).

decrease) and a 14.4% average improvement (8.19% absolute increment) of GAR values.

3.5 Conclusion

Three new methods for matching a single biometric using quality metrics are proposed. The methods are adaptive and use nonlinear mappings for making predictions on quality measures and corresponding verification scores. The reported experimental results illustrate the importance of predictive and selective integration of quality measures for face biometric authentication and show significant advantages compared to existing methods.

Chapter 4

Cross-Spectral Face Recognition: Visible vs. SWIR

4.1 Introduction

In the previous chapter we have shown that quality of images affects the recognition performance of a face recognition system operating under uncontrolled setting and we have proposed three different methods to improve the performance of the system by employing quality metrics. In the face recognition problem, not only the quality of the matching images influences the recognition performance of the system but also the spectrum at which they were acquired. Different spectral bands of electromagnetic spectrum such as visible, Near Infra Red (NIR) ($750nm - 1100nm$) and thermal Infra Red ($7 - 14\mu m$) have been used to collect images and videos of people for testing various face recognition algorithms. In majority of cases, these approaches are designed to perform face recognition within one specific spectral band. However, this scenario does not often support modern face recognition applications. Surveillance cameras, for example, often operate in both visible and NIR bands and switch between the bands depending on the night or day environments. Since gallery and watch lists are traditionally composed of visible light face images, newly designed face recognition methods are expected to be successful in matching NIR data versus color images. Apart from this case of cross-spectral comparison, attention has recently turned to previously unexplored part of the Short Wavelength IR spectrum ($1100nm - 1700nm$).

The SWIR band has several advantages over the NIR spectrum: SWIR imaging modality produces clear images in the presence of challenging atmospheric conditions such as rain, fog and mist; SWIR cameras produce high SNR images under low light conditions or at night time [79]; the SWIR wavelength is completely invisible to the human eye and is undetectable by silicon-based cameras [79]. This makes the modality suitable for covert long-range (up to 800m) applications [1]. On the other hand, the SWIR face images are photometrically distinct compared to color images.

An overview of the operational spectral ranges supporting different imaging modalities for face recognition is provided in Fig. 4.1.

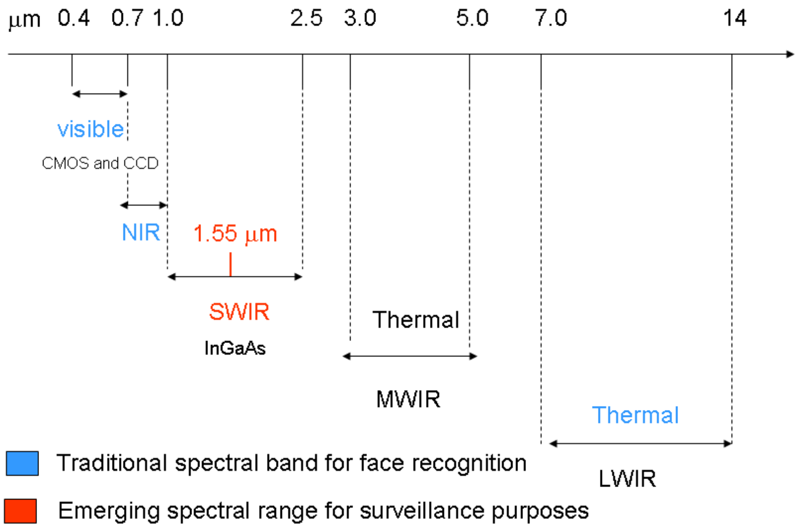

μm
0.4
0.7
1.0
2.5
3.0
5.0
7.0
14
visible
CMOS and CCD
NIR
1.55 μm
SWIR
InGaAs
Thermal
MWIR
Thermal
LWIR
Traditional spectral band for face recognition
Emerging spectral range for surveillance purposes

Figure 4.1: Traditional and emerging imaging modalities for face recognition and their places in the electromagnetic spectrum.

In this chapter, we introduce a novel cross-spectral encoding method to match visible images against SWIR images acquired at different distances. We adopt a small set of Gabor filters at the initial encoding stage. The filtered face images are then further encoded by an encoding scheme that involves three local operators to extract robust features across the two different spectral bands. These three operators are designed to encode both magnitude and phase of filtered images, resulting in a comprehensive encoding scheme. The obtained verification and identification results are compared against the performance of Faceit G8 [80], which is often viewed as a baseline and the state of the art algorithm for face recognition. We demonstrate that: (1) cross-spectral comparison of visible light face images versus images from the SWIR spectral band produces lower recognition rates compared to the case when images are matched within a spectral band (either visible or SWIR); (2) the proposed matching algorithm outperforms Faceit G8 at three different probe distances (2m, 50m, 106m); (3) overall cross-spectral recognition rates drop as the distance of probe SWIR images increases.

4.2 Preprocessing, Encoding and Matching

4.2.1 Face Normalization

We use the position of the eyes to normalize the face image to a canonical representation resulting in an image of a fixed size. Details about the adopted resolution are provided in Section 4.4. A similarity transformation S (rotation, scaling and translation) is applied to each image such that the eye locations are projected into fixed positions in the image space. The similarity matrix S in homogenous coordinates is given as:

$$S = \begin{pmatrix} s\cos(\theta) & -s\sin(\theta) & 0 \\ s\sin(\theta) & s\cos(\theta) & 0 \\ tx & ty & 1 \end{pmatrix},$$

where s, θ, tx, ty are unknown scale, rotation and translation parameters to be estimated.

The location of the eyes (4 scalar values in the image plane) are manually selected and are used to determine the defined parameters. The precise position of the eyes, however, are not critical for normalization in our approach, since we can easily accommodate for a displacement of some pixels. The position of the eyes can also be automatically determined, for example, with a Haar-based detector [4] trained on multi-spectral images. In this chapter, we focus on the encoder and matcher design rather than on a normalization scheme.

4.2.2 Preprocessing

Color images are transformed into gray scale images I_g by using a simple linear combination of the original RGB channels. Our experiments have shown that the outcomes of this linear combination are more robust compared to many other approaches which we have attempted when matching color images versus SWIR images encoded using Gabor filters. A SWIR image X (1550nm) is firstly preprocessed by using the log-transformation: $I_l = log(1 + X)$. Then, the image is further normalized as:

$$I_p = \frac{I_l - min\{I_l\}}{max\{I_l\} - min\{I_l\}} * 255, \tag{4.1}$$

where $min\{I_l\}$ and $max\{I_l\}$ are the minimum and maximum pixel values of the image I_l. This normalizes the histogram of the SWIR image to have intensities in the range [0,255]. The overall transformation redistributes the original darker pixels (most of the pixels in a SWIR image) over a much broader range and compresses the range of the original brighter pixels (a small amount of pixels). The resulting image is brighter and has a better contrast than the original. We note that the gray variation (trend) of the image pixels is preserved since the applied transformations are monotonic. Other preprocessing (more sophisticated) techniques that we have attempted to employ did not result in a noticeable improvement or increase of performance when applied in a cross-spectral verification scenario. Below we provide a brief description of the proposed methods and other previously developed methods that we have attempted for the cross-spectral preprocessing.

- HE (Histogram Equalization): The distribution of the pixel values of the entire image is equalized. The resulting cumulative distribution of the pixel values is a linear function of the image levels (e.g. $\{0, 1, ..., 255\}$).

- MRHM (Multi-Resolution Histogram Matching): This preprocessing method is inspired by the work of Efros [81] for texture synthesis and adapted for a completely different task. The visible image (converted to gray scale) and SWIR image are decomposed with a steerable pyramid decomposition [82]. The decomposition produces feature images at different resolutions and orientations (pyramid bands); we match distributions (histograms) of pyramid bands of the visible image with the ones of the SWIR image (or vice versa) [81]; we rebuilt the visible (or SWIR) image from the matched pyramid bands.

- CLAHE (Contrast Limited Adaptive Equalization) [83]: The Histogram Equalization is applied locally on square patches of 8×8 pixels. The maximum value of the local image histogram is clipped to a defined value. This limiting procedure avoids amplification of noise in the image.

- MSQI (Multi-Scale Self Quotient Image): This method is used in the literature [84] to reduce the effect of varying illumination that is observed in face images. The face image is smoothed by convolving it with Gaussian kernels. The original image is divided (pixel by pixel) by a linear combination of the obtained smoothed images. The produced image is re-normalized to the original range (e.g. [0,255] for a 8-bit image) using a non linear function (e.g. arctang or sigmoid).

- DOG (Difference of Gaussian): The DOG operator is a band-pass filter that enhances the edges of the images and removes noise [85]. The image is convolved with two different Gaussian kernels. The output image is obtained as a difference of the two convolved images.

- NNP (Neural Net Prediction): The approach is similar to our previous work demonstrated on the iris biometrics [86]. Normalized face images, visible (grayscale) and SWIR are the input and the output (or vice versa) of a Neural Network (NN). The input image (e.g. visible) is divided into small patches (3x3) that are reshaped as vectors. Each patch in a visible image predicts a single pixel in SWIR range. The mapping is learned by using patches from visible images as input parameters. The central pixels of corresponding (same location) patches from SWIR images are employed as output parameters. An extension of this method has also been implemented by adopting a Discrete Wavelet Transform (DWT). The prediction method is similar, however the input and target images are the wavelet bands of DWT applied to the visible and SWIR images. The prediction is carried out for each wavelet band and then the predicted bands are employed to rebuild a single (predicted) SWIR (or visible) image.

- PC (Phase Congruency): The method detects edges of the images based on the well-known Phase-Congruency principle [87]. A bank of log-Gabor filters is applied to the images to evaluate the local energy of the face images. The local maxima points of the energy function are taken as the feature image.
- CSD (Color Space Decomposition): The color image is decomposed to the three color channels and matching with SWIR images is performed by employing the singular color channels. The matching results of each color channel are combined at the matching score level by summing the individual contributions.

4.2.3 Filtering

A bank of 16 Gabor filters is applied to the cropped face images. In particular we use a set of filters at 2 scales and 8 orientations, resulting in a total of 16 filter responses. We adopt a much smaller number of filters with respect to other Gabor-based methods (e.g. [21] [30] [33]) that are applied to data at the visible spectral band. We note that such number of filters is sufficient to discriminate features in common between the color and SWIR images (especially when captured from a large distance). The experiments show that the adopted set of filters also works well when comparison is performed within the same spectral band.

The mathematical equation for the Gabor filter [13] is as follows:

$$G_{\alpha,\beta}(z) = \frac{\|k_{\alpha,\beta}\|}{\sigma^2} \exp\left[\frac{\|k_{\alpha,\beta}\|^2 \|z\|^2}{2\sigma^2}\right] [e^{ik_{\alpha,\beta}z} - e^{-\sigma^2/2}], \tag{4.2}$$

where σ^2 is the variance of the Gaussian kernel, $k_{\alpha,\beta}$ is the wave vector. The magnitude and phase of the wave vector determine the scale and orientation of the oscillatory term and z=(x,y). We adopt a Gabor kernel of size 32 × 32 pixels with the following wave vector:

$$k_{\alpha,\beta} = k_\beta e^{i\phi_\alpha}. \tag{4.3}$$

The adopted parameters for the complex vector are $k_\beta = \frac{\pi}{2^{\beta/2}}$ with $\beta = 3, 4$ and $\phi_\alpha = \frac{\pi}{\alpha} \times 8$ with $\alpha = 1, 2, .., 8$; the Gaussian kernel has standard deviation $\sigma = \pi$. These parameters are selected for all experiments described in this work.

A normalized and preprocessed face image $I(z)$ is convolved with the Gabor filter $G_{\alpha,\beta}(z)$ at orientation ϕ_α and scale k_β resulting in the filtered image: $Y_{\alpha,\beta}(z) = I(z) * G_{\alpha,\beta}(z)$, with " $*$ " denoting convolution. Denote by $\mathrm{Re}[Y_{\alpha,\beta}]$ the real part and by $\mathrm{Im}[Y_{\alpha,\beta}]$ the imaginary part of the filtered image. Then, the magnitude of the filtered images is $M_{\alpha,\beta} = \sqrt{\mathrm{Re}^2[Y_{\alpha,\beta}] + \mathrm{Im}^2[Y_{\alpha,\beta}]}$ and the phase is $P_{\alpha,\beta} = \arctan\{\mathrm{Im}[Y_{\alpha,\beta}]/\mathrm{Re}[Y_{\alpha,\beta}]\}$.

4.2.4 Encoding of Magnitude and Phase Response

In this work, we encode both magnitude and phase of Gabor responses and demonstrate that both of them are useful in a cross-spectral comparison. The recent paper of Xie et al. [33] adopts magnitude and phase of a bank of 40 Gabor filters to match images within the visible spectral band. In this work, the LXP and/or LBP operator has been adopted to encode magnitude and phase and was followed by a block-based Fisher Linear Discriminant (BFLD) approach applied to the visible datasets FERET [22] and FRGC 2.0. Here, we focus on a challenging problem where visible face images have to be matched against SWIR face images. We encode the magnitude and phase responses of 16 Gabor filters by employing three different multi-scale local operators. The encoded phase and magnitude are later combined at the feature level resulting in a robust representation for each face class. We do not involve any training data to achieve good recognition performance. The approach is applicable to the scenario where no SWIR data are available to train the recognition system. To encode the magnitude we use two distinct operators: Simplified Weber Local Descriptor (SWLD) and uniform Local Binary Pattern (LBP) operator. For encoding the phase we adopt a uniform Generalized Local Binary Pattern (GLBP) operator. The WLD is a new operator recently developed by Chen et al. [88]. It has been introduced to characterize textures in images and was illustrated on raw images. WLD has never been applied to Gabor-filtered images. This is one of the novelties claimed in our work. The WLD operator consists of two joint descriptors: a differential excitation operator and a gradient orientation descriptor. In this work we adopt only the differential excitation operator [88] to encode the magnitude filter response $M_{\alpha,\beta}$, resulting in a robust representation of face features.

The simplified version of the WLD operator is defined as:

$$SWLD_{l,r,n}(x) = \mathcal{Q}_l\left\{tan^{-1}\left[\sum_{i=0}^{n-1}\left(\frac{x_i - x}{x}\right)\right]\right\}, \tag{4.4}$$

where x_i are the $n = 12$ neighbors of x at radius $r = 1, 2$ and $\mathcal{Q}_l$ is a uniform quantizer with l quantization levels. In this work, we adopt $l = 135$ levels to discretize the output of the tan^{-1} function. An illustration of the application of SWLD operator to an intensity pattern is shown in Fig. 4.2.

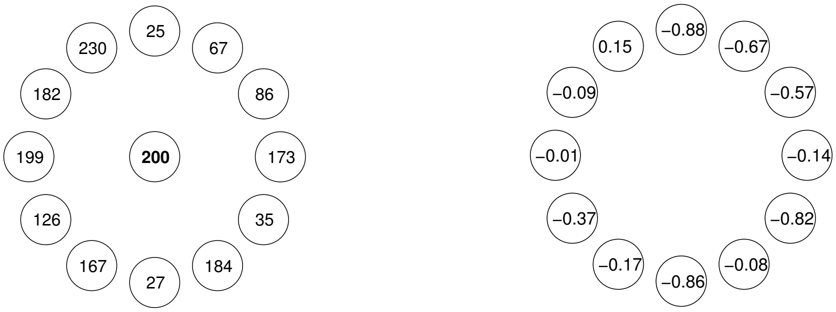


Figure 4.2: The figure shows an example of the application of SWLD operator to a generic intensity pattern having range [0, 255]. The left panel shows the original pattern; the right panel indicates the approximate contribution of each neighbor when SWLD is applied. The contributions are summed up and a non linear function is applied to the obtained sum. The obtained result is further quantized uniformly with a given number of levels.

The LBP operator is the other descriptor applied to the magnitude response $M_{\alpha,\beta}$. The main difference between its application in this work compared to all earlier applications (for example, see [21]) is that we consider the relationship among $n = 12$ neighbors at a radius of one and also two pixels:

$$LBP^{u}_{r,n}(x) = \mathcal{U}\left\{\sum_{i=0}^{n-1}\mathcal{I}(x_i - x)2^i\right\}, \tag{4.5}$$

where x_i are the neighbors of a value x at radius r, $\mathcal{U}$ is the uniform pattern mapping (see eq. 4.7) and $\mathcal{I}$ is defined as:

$$\mathcal{I}(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases} \tag{4.6}$$

Note that with the proposed encoding method we distinguish only uniform sequences. As shown in [89,90], the uniform pattern mapping reduces the number of binary features to be encoded, thereby preserving discriminative patterns. A binary pattern is defined being uniform if it contains at most two bitwise transitions from 0 to 1 or from 1 to 0 when the bit sequence is recorded circularly. For example, the sequence 011111111000 is a 12-bit uniform pattern whereas the sequence 010001011111 is not uniform. Denote by $\mathcal{U}\{d\}$ a binary decision made about the decimal number d, then $\mathcal{U}\{d\}$ is defined as:

$$\mathcal{U}\{d\} = \begin{cases} d, & \text{if } d_B \text{ is a uniform sequence} \\ N, & \text{otherwise} \end{cases} \tag{4.7}$$

where d_B is the value d expressed as power of 2 and N is the total number of uniform patterns formed using n bits. In our formulation, we work with $n = 12$ bit sequences which results in $N = 134$ uniform patterns. An example of the application of LBP operator to an intensity pattern is demonstrated in Fig. 4.3.

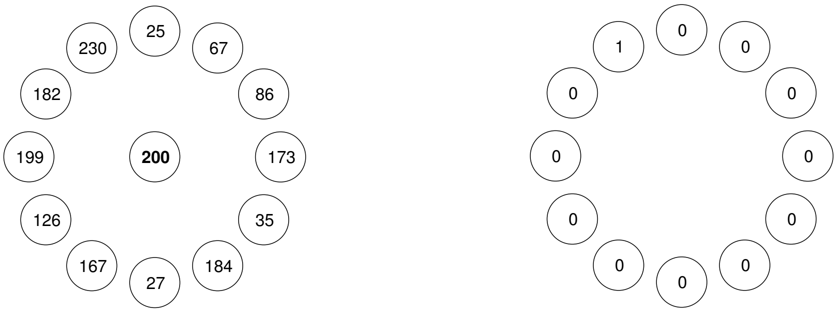


Figure 4.3: The plot displays an example of the application of LBP operator to a generic intensity pattern having range [0, 255]. The left panel shows the original pattern; the right panel indicates the binary sequence that is resulting from the application of LBP operator. If the sequence is uniform, then it is converted into a decimal number, otherwise it is replaced with a given number.

The SWLD and LBP operators are complementary operators in terms of the type of information that they encode [91]. The SWLD operator detects the edges and records their intensity values. The LBP detects orientation of the edges but does not encode the intensity values. We will show that these two operators combine well when encoding magnitude response of Gabor filters. The information that these two operators encode can be further aggregated with Gabor phase patterns leading to improved recognition performance.

To encode the Gabor phase response $P_{\alpha,\beta}$ (assumed to be defined on the interval $[0, 2\pi]$) we adopt a uniform generalized local binary pattern operator defined as:

$$GLBP^{u}_{r,n,t}(x) = \mathcal{U}\left\{\sum_{i=0}^{n-1} \mathcal{T}_t(x_i - x)2^i\right\}, \tag{4.8}$$

where x_i are the $n = 12$ neighbors of x at radius $r = 1, 2$. $\mathcal{U}$ is the uniform pattern mapping (see eq. 4.7) and $\mathcal{T}_t$ is a thresholding operator based on the threshold t. It is defined as:

$$\mathcal{T}_t(x) = \begin{cases} 1, & \mid x \mid \leq t \\ 0, & \text{otherwise.} \end{cases}$$

The defined GLBP operator is a generalization of the encoding method introduced in [92]. The main differences are as follows. The operator encodes only uniform binary sequences and similarly to the two other introduced encoding methods (SWLD, LBP), the operator considers the relationship among $n = 12$ neighbors at both radii $r = 1, 2$. The values for the thresholds were evaluated experimentally. It has been found that both $t = \pi/2$ and $t = \pi$ are good values when the operator is applied at radii $r = 1, 2$. We will further set $t = \pi/2$. An illustration of the application of GLBP operator to a phase pattern is displayed in Fig. 4.4.

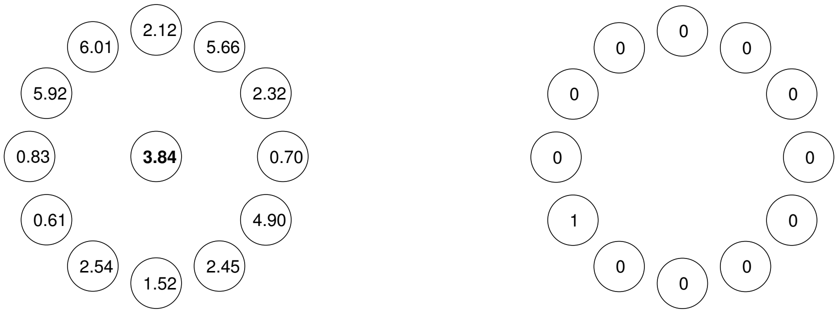


Figure 4.4: The figure demonstrates an example of the application of GLBP operator to a generic phase pattern having range $[0, 2\pi]$. The left panel displays the original pattern; the right panel indicates the binary sequence that is resulting from the application of GLBP operator. Differently from LBP operator, GLBP descriptor adopts a non zero threshold (e.g. $t = \pi$) to obtain the binary sequence. If the sequence is uniform, then it is converted to a decimal number, otherwise it is substituted with a given number.

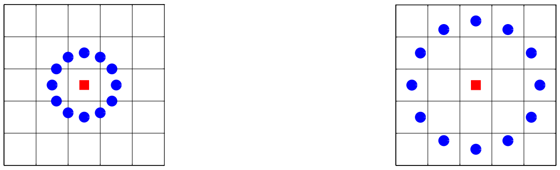


Figure 4.5: The figure indicates the support adopted by all three local operators. The application pixel is denoted with a square and its neighbors are indicated with circles. The left panel shows the operator with support parameters ($r = 1, n = 12$); the right panel displays the operator with support parameters ($r = 2, n = 12$).

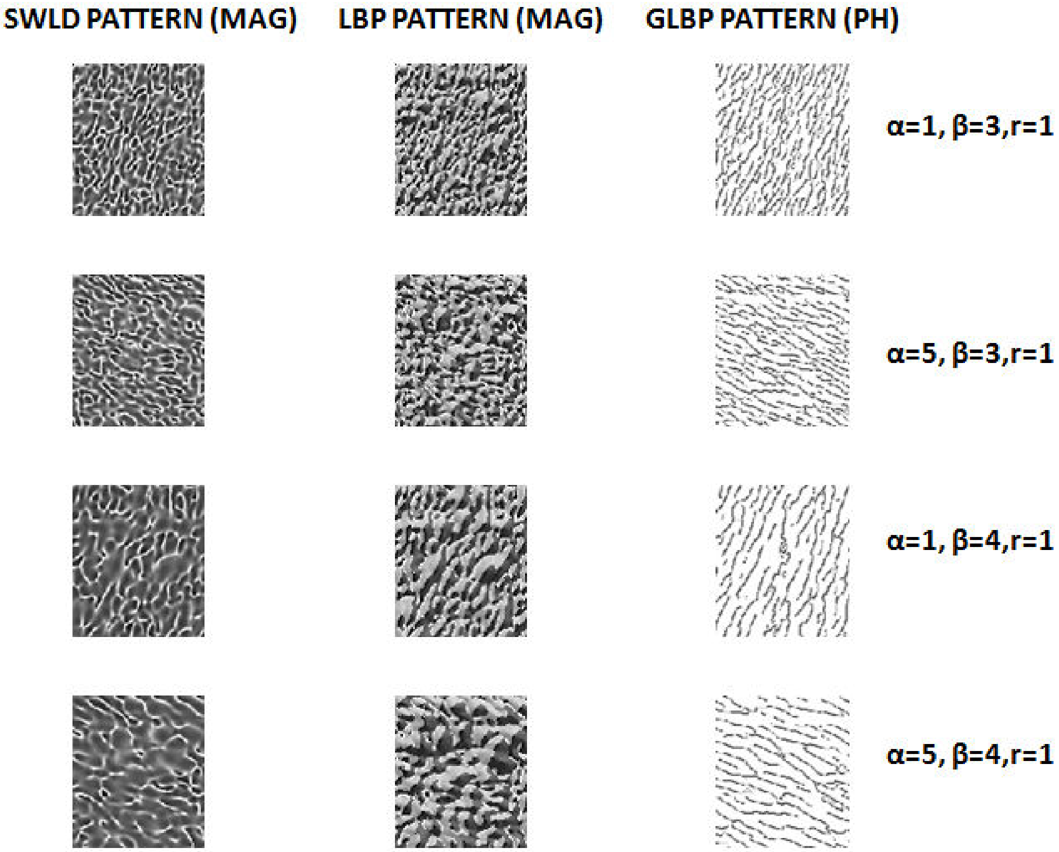


Figure 4.6: The figure shows the patterns from a 1550nm normalized face image produced by the Gabor filters and the three local operators (SWLD, LBP, GLBP). The images represent the encoded magnitude (MAG) and phase patterns (PH) for different orientations (α), scales (β) and radii (r). The original face image sample is shown in Fig. 4.9(b).

4.2.5 Local Features and Distance Measure

The block diagram of the proposed feature extraction method is displayed in Fig. 4.7.

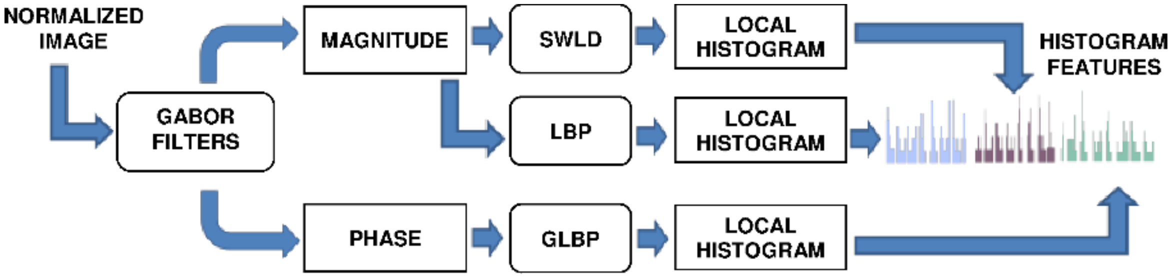


Figure 4.7: The figure displays the block diagram of the encoding scheme.

The preprocessed and normalized images are encoded with the bank of Gabor filters defined in Sec. 4.2.3. For each Gabor response we evaluate and separately store the magnitude and the phase response of individual filters. The magnitude response is further encoded using the SWLD and LBP operators. The phase response is encoded with the GLBP operator. Each encoded response is divided into non-overlapping blocks of size 8×8 resulting in N blocks. Blocks are displayed in the form of histograms containing 135 bins. Histograms are then concatenated to form a single vector.

The three vectors (one for each applied operator) are further concatenated to form a longer feature vector as shown in Fig. 4.7. The encoding process is repeated for each filter response of the Gabor bank, and vectors of features are stored in a matrix $\mathcal{H}$. Note that each of the three operators can also be treated as an independent encoder.

To compare feature vectors extracted from two images of two different (or the same) spectral bands we adopt a symmetric I-divergence distance [93]. Consider two images A and B. Denote by $\mathcal{H}_A$ and $\mathcal{H}_B$ the matrix of features extracted from images A and B, respectively. The symmetric I-divergence distance between these two matrices is defined as:

$$D(A,B) = \sum_{k=1}^{K} [H_A(k) - H_B(k)] \log \frac{H_A(k)}{H_B(k)}, \tag{4.9}$$

where K is the total length of the histogram features $H_A(k)$ (or $H_B(k)$) obtained by concatenating all rows of the matrix $\mathcal{H}_A$ (or $\mathcal{H}_B$).

Other distance or similarity metrics can also be employed to compare the histogram features. We have tested and compared three other well-know metrics: histogram intersection, euclidian distance, and chi-square distance.

The histogram intersection [21] similarity metric is defined as:

$$S(A,B) = \sum_{k=1}^{K} \min\left[H_A(k), H_B(k)\right], \tag{4.10}$$

The chi-square distance [18] metric has the following expression:

$$D(A,B) = \sum_{k=1}^{K} \frac{[H_A(k) - H_B(k)]^2}{[H_A(k) + H_B(k)]}, \tag{4.11}$$

Last, the euclidian distance metric is given by:

$$D(A,B) = \sqrt{\sum_{k=1}^{K} [H_A(k) - H_B(k)]^2} \tag{4.12}$$

The results of the comparison between these metrics are shown in Table 4.6 in Sec. 4.4.

4.3 Face Datasets

In our experiments we use two different collections of images acquired with the TINDERS (Tactical Imager for Night/Day Extended-Range Surveillance) system developed by Advanced Technologies Group, WVHTC Foundation [1]. The first set of data, called PRE-TINDERS, is composed of 48 frontal face classes (total of 384 images) at two wavelengths: visible spectrum and 1550 nm (SWIR spectrum). Four images per class are available for each spectral band: 2 images have neutral expression and 2 images depict the person talking (open mouth). All images in this dataset are acquired at a close distance from the camera (about 2 meters) in a single session. A light source at 1550nm is illuminating the face of the subjects for images captured in the SWIR spectral band. The original resolutions of the acquired images before normalization (see Fig. 4.8) are 640x512 (png format) for SWIR images and 1600x1200 (jpg format) for color images.

The second collection of data, called TINDERS, is composed of 48 frontal face classes having images at the same two spectral bands. The SWIR images are acquired at two long distances (50 and 106 meters). At both distances, 4/5 images per class are available: 2/3 have neutral expression and 2 have talking expression. A total of 478 images are available in the the SWIR band. The visible (color) images are collected at a short distance and in two sessions (3 images per session), and all of them have neutral expression, resulting in a total of 288 images. The original resolutions of images before normalization are 640x512 (png format) for SWIR images and 640x480 (jpg format) for color images.

Sample images from the PRE-TINDERS and TINDERS dataset are shown in Fig. 4.8 and Fig. 4.9, respectively.

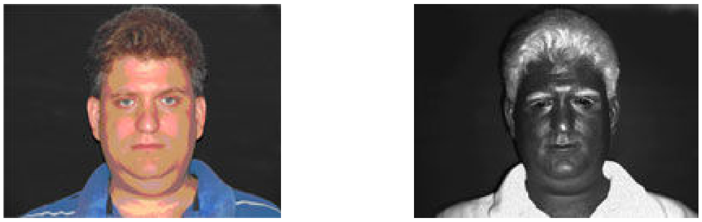


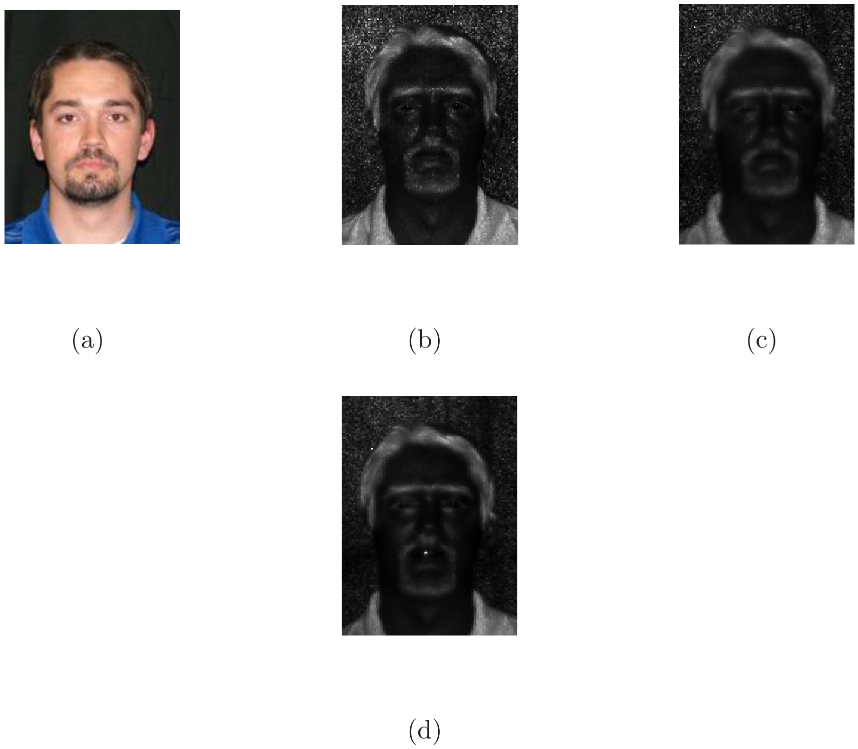
Figure 4.8: Sample images from PRE-TINDERS dataset from [1]: visible spectral band (left), SWIR (right). Face images are captured at a distance of 2 meters.

Note the difference of intensity distributions in these images. The human skin and eyes in the SWIR spectrum appear to be very dark because of the presence of moisture. Alternatively, the hairs are white because they turn to be highly reflective at those wavelengths.

4.4 Experimental Results

4.4.1 Short Range Data

In this section we summarize the experiments performed on PRE-TINDERS dataset. Each spectral band in the dataset is represented by 192 frontal face images (48 classes) having an expression variation. The face images acquired at short distance (2 meters) have high resolution and we geometrically normalize them to a standard resolution of 160 × 168. The experiments carried out on images acquired at close distance provide a baseline for performance comparison with respect to the case where the probe images are captured at large distances (50-106 meters).



(a) (b) (c)

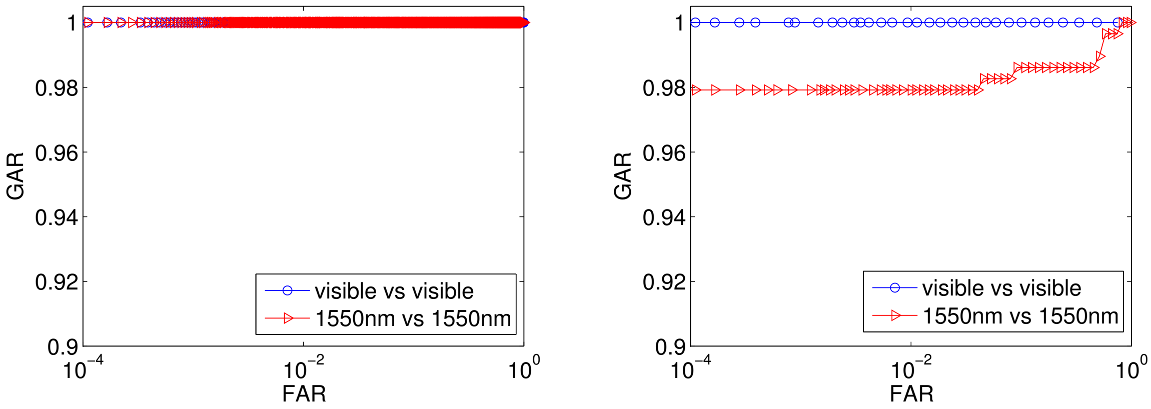
(d)

Figure 4.9: Sample images from TINDERS dataset: gallery visible image (a), neutral probe SWIR image at a distance of 50 meters (b), neutral probe SWIR image at a distance of 106 meters (c), talking probe SWIR image at a distance of 106 meters (d).

Comparison Within the Same Band

Our first experiment involves matching images within the same spectral band, that is, (1) visible versus visible, (2) 1550 nm versus 1550 nm. We compare the performance of the proposed matcher against the performance of Faceit G8.

Fig. 4.10(a) displays the Receiver Operating Characteristic (ROC) curves obtained with our matcher. Fig. 4.10(b) depicts the ROCs obtained using Faceit G8.



(a) (b)

Figure 4.10: Fig 4.10(a) shows ROC curves when images are matched within the same spectral band using the proposed method (SWLD+LBP+GLBP). Fig 4.10(b) displays ROC curves generated by Faceit G8 when images from the same spectral band are matched. Face images are captured at a short distance.

Note that both the proposed matcher (SWLD+LBP+GLBP) and Faceit G8 provide perfect recognition performance when applied to visible data. When face matching is performed within SWIR band (1550 nm), Faceit G8 provides recognition rate 98% (at False Accept Rate, FAR, ranging from 10^{-2} to 10^{-4}) compared to the proposed matcher which provides 100% recognition rate over the entire range of FARs. The matching results clearly indicate that the proposed algorithm performs well when matching is within the same spectral band.

Color vs 1550 nm at a short distance

In the following experiments we compare color face images against 1550nm probe images captured at a short distance (2 meters). Earlier in this chapter we indicated that images within the 1550 nm spectral band are nonlinearly transformed and normalized by following the equations in Sec. 4.2.2 prior to being encoded. Color images are transformed into gray scale images. We first evaluate the verification performance of matchers that adopt the single local operators to encode the Gabor responses and then we combine features of different local operators and compare the obtained performance to the one by Faceit G8. The matching results are summarized in Table 4.1.

METHOD	GAR (%) at FAR 10^{-1}	GAR (%) at FAR 10^{-4}	EER (%)	d-prime
GLBP (PH)	89.45	38.15	10.29	2.419
LBP (MAG)	97.40	62.96	4.69	3.109
SWLD (MAG)	95.57	65.49	7.16	2.821
LBP+SWLD (MAG)	99.09	72.79	3.48	3.269
FACEIT G8	93.49	71.94	7.71	1.433
SWLD+LBP+GLBP (MAG & PH)	98.96	76.30	2.99	3.272

Table 4.1: Verification rates (Genuine Accept Rates) at FAR 10^{-1} and FAR 10^{-4}, EER, d-prime index for single, combined local operators and Faceit G8 for cross-spectral comparison between visible images and 1550nm images captured at a short distance.

We note that the matcher based on phase patterns (PH) encoded by GLBP descriptor provides the lowest matching performance. The magnitude patterns (MAG) encoded by SWLD and LBP descriptors both provide significantly higher performance compared to GLBP. The LBP descriptor is the best when it operates at large FARs (e.g. 10^{-1}) whereas SWLD produces the best recognition rates at low FARs (e.g. 10^{-4}). Because of this behavior when the features produced by these two operators are combined, the achieved matching performance is increased with respect to the case when single operators are adopted to encoded magnitude responses. Finally, we combine the magnitude features with the phase features; as a result we observe that three performance indexes out of four have noticeable improvements, whereas the performance at FAR 10^{-1} is almost unchanged. This indicates that the overall best performance is obtained when we involve all three local operators to encode both the magnitude and the phase of Gabor filters. From these results we infer that the three adopted operators provide orthogonal information for cross-spectral face recognition. Therefore, the experiments on images acquired at large distances (next section) will provide further evidence.

In Fig. 4.11 we compare the verification performance (ROC) of Faceit G8 with the one achieved by the proposed approach when SWIR images are acquired at a short distance.

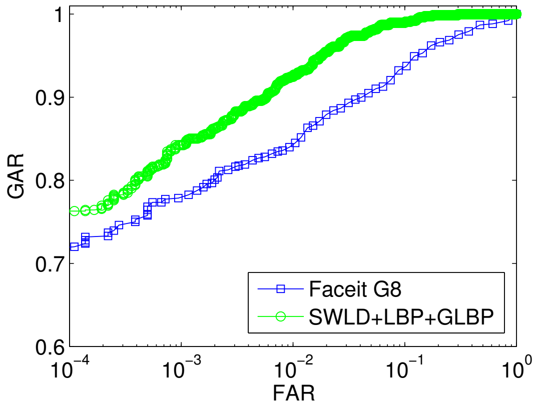


Figure 4.11: Cross-spectral comparison between visible images and 1550nm images acquired at a short distance. The plot shows ROCs for the proposed method (SWLD+LBP+GLBP) and Faceit G8.

Finally, in Table 4.2, we compare the identification performance (rank-1 recognition) for our matcher and Faceit G8.

METHOD	Rank-1 (%)
SWLD+LBP+GLBP	97.79
FACEIT G8	86.07

Table 4.2: Identification (rank-1 recognition) rates for the cross comparison between color images and images from 1550nm spectral band acquired at a short distance.

Tables 4.1, 4.2 and Fig.4.11 demonstrate that our approach achieves better verification and identification performance compared to Faceit G8 when probe images at 1550 nm acquired at a short distance are compared against the gallery composed of visible data. Overall, the gain in verification performance of our matcher with respect to Faceit G8 ranges from about 9% to 5% at FARs lower than 10^{-1}. The improvement in terms of identification performance achieves about 11.7%.

4.4.2 Long Range Data

In this section we evaluate performance of the proposed algorithm on SWIR data captured at a long distance from the camera (50 and 106 meters). We compare the performance of our matcher with the performance of Faceit G8. For long distances SWIR images, the eye detection module of Faceit G8 does not provide a good estimation of the eye location. Therefore, we manually detect positions of the eyes and supply them to the commercial recognition engine. The same coordinates are provided to our face normalization module. The implemented module normalizes color and long range SWIR images to a standard size of 112 × 120.

Color vs 1550 nm at a distance of 50 meters

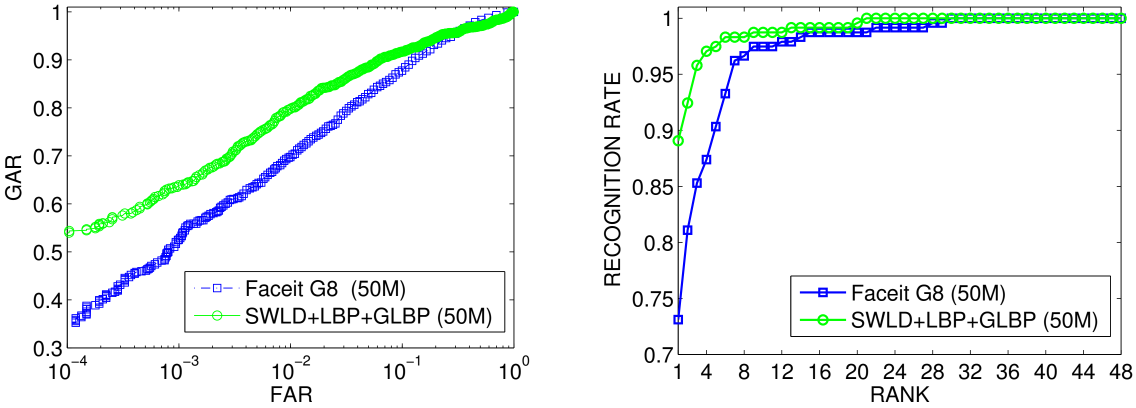
To evaluate the verification performance of our approach against Faceit G8 when SWIR images are captured at a distance of 50 meters, we match all available visible images against all available SWIR probe images. Similarly to the short distance case, we first calculate the verification performance of matchers that adopt single local operators applied to Gabor responses. We then assess performance of the matchers that combine features from the local operators and we compare them with the ones of Faceit G8. The results are summarized in Table 4.3.

METHOD	GAR (%) at FAR 10^{-1}	GAR (%) at FAR 10^{-4}	EER (%)	d-prime
GLBP (PH)	71.46	24.72	19.79	1.634
LBP (MAG)	85.05	36.24	12.60	2.277
SWLD (MAG)	86.48	24.30	12.11	2.228
LBP+SWLD (MAG)	88.48	44.29	11.00	2.420
FACEIT G8	88.10	35.57	11.21	1.390
SWLD+LBP+GLBP (MAG & PH)	91.70	53.82	8.70	2.619

Table 4.3: Verification rates (Genuine Accept Rates) at FAR 10^{-1} and FAR 10^{-4}, EER, d-prime index for single, combined local operators and Faceit G8 for cross-spectral comparison between visible images and 1550nm images captured at a distance of 50 meters.

The performance indexes in the table indicate that matchers with magnitude encoders overall provide better recognition performance compared to those with phase responses. The experiment outcome is analogous to the one obtained for the close distance SWIR data. In this case, the SWLD-based matcher provides a higher recognition rate at FAR 10^{-1} while the LBP operator provide a larger recognition at FAR 10^{-4}. Hence the two operators once more demonstrated complementary nature at different operational points. The combination of these two operators leads to an improvement of the performance indexes in the cross-spectral comparison. The matching performance indexes are further increased by aggregating the magnitude histograms with the phase histograms encoded by the GLBP operator. This strengthens the conclusions we reached for short range data. The best performance is obtained when we involve all three local operators that encode both magnitude and phase of Gabor filtered images.

In Fig. 4.12(a) we further compare ROC curves obtained with our method and those obtained with Faceit G8. The plots demonstrate that the proposed method outperforms the commercial engine, providing a gain in verification performance equal to 3.6% at FAR 10^{-1} and about 18.25% at FAR 10^{-4}.



(a) (b)

Figure 4.12: Cross-spectral comparison between visible images and 1550nm images acquired at a distance of 50 meters. The left panel shows ROCs for the proposed method (SWLD+LBP+GLBP) and Faceit G8. The right panel displays CMCs for the two matchers.

To evaluate the identification performance of our approach and Faceit G8 we match all SWIR probe images against a gallery built of a single visible image for each face class. The obtained Cumulative Match Curves (CMCs) are displayed in Fig. 4.12(b). Note the large gap of performance between two CMCs when the rank varies between 1 to 6. For the top match (rank-1) case, the gain in performance is about 15.97%. The values of the identification rates of the proposed matcher and commercial software are shown in Table 4.4.

METHOD	Rank-1 %
SWLD+LBP+GLBP	89.08
FACEIT G8	73.11

Table 4.4: Identification (rank-1 recognition) rates for the cross-spectral comparison of color images versus 1550nm images captured at a distance of 50 meters.

Color vs 1550 nm at a distance of 106 meters

To assess the verification performance of our matcher and Faceit G8 on images acquired at 106 meters, we compare all available visible images against all available SWIR probe images. We follow the same experimental protocol we employed for data collected at a close range and at 50 meters. The obtained verification performance is shown in Table 4.5.

METHOD	GAR (%) at FAR 10^{-1}	GAR (%) at FAR 10^{-4}	EER (%)	d-prime
GLBP (PH)	54.97	9.93	31.50	0.926
LBP (MAG)	80.66	18.16	15.04	2.013
SWLD (MAG)	77.60	20.35	16.62	1.875
LBP+SWLD (MAG)	83.68	21.91	14.18	2.115
FACEIT G8	82.64	22.19	14.17	1.187
SWLD+LBP+GLBP (MAG & PH)	84.03	31.70	13.83	2.073

Table 4.5: Verification rates (Genuine Accept Rates) at FAR 10^{-1} and FAR 10^{-4}, EER, d-prime index for single, combined local operators and Faceit G8 for cross-spectral comparison between visible images and 1550nm images acquired at a distance of 106 meters.

Again, the overall results indicate that matchers employing magnitude responses provide better recognition performance compared to matches with encoded phase. LBP operator provides better performance at FAR 10^{-1}, while the SWLD is better at FAR 10^{-4}. When SWLD and LBP-based matchers are combined we obtain further improvement in recognition performance compared to Faceit G8 at FAR 10^{-1}, but Faceit G8 shows better performance at FAR 10^{-4}. However, when we combine features from all three local operators the proposed method outperforms Faceit G8.

The verification performance of our approach compared to Faceit G8 is displayed in Fig. 4.13(a). We note that the performance improvement ranges from about 1.4% at FAR 10^{-1} to 9.5% at FAR 10^{-4}.

Table 4.6 summarizes the verification performance of the proposed method by employing different distance/similarity metrics. The obtained results indicate that I-divergence metric produces the lowest EER (Equal Error Rate) and the largest d-prime (separability index) under this challenging scenario.

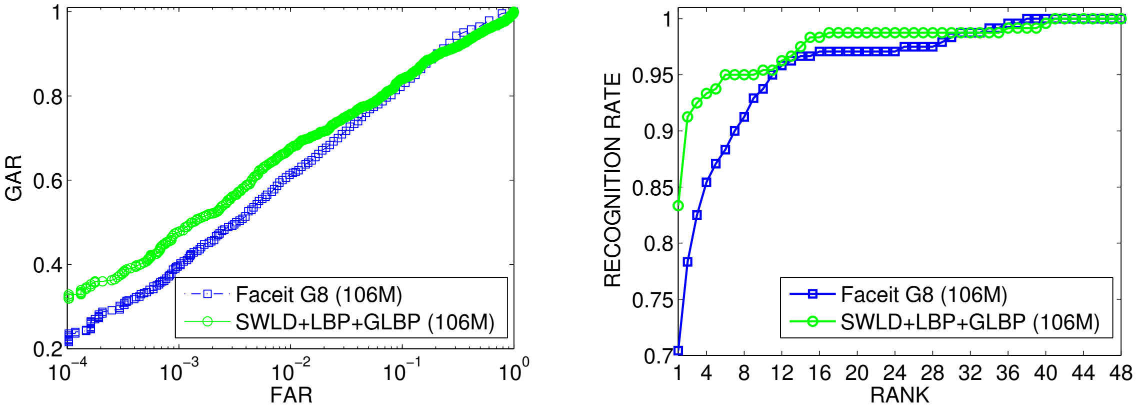
METRIC	EER (%)	d-prime
Euclidian Distance	31.61	0.879
Histogram Intersection	16.41	1.855
Chi-Square Distance	14.68	1.977
I-divergence	13.83	2.073

Table 4.6: EER and d-prime indexes for the proposed method followed by a matcher adopting Euclidian distance, Histogram intersection, Chi-square distance or I-divergence. The results are for cross-spectral comparison between visible images and 1550nm images acquired at a distance of 106 meters.

Similarly to the 50 meter case, we also assess the identification performance. All SWIR probe images are matched against a gallery having a single visible image for each face class in the dataset. The obtained Cumulative Match Curves (CMCs) are shown in Fig. 4.13(b). The figure demonstrates that a large separation exists between the CMC of Faceit G8 and our approach when the rank values are between 1 to 8. From Table 4.7 we conclude that the gain in identification performance of our approach with respect to Faceit G8 for top match (rank-1) is about 12.9%.

METHOD	Rank-1 (%)
SWLD+LBP+GLBP	83.33
FACEIT G8	70.42

Table 4.7: Identification (rank-1 recognition) rates for the cross-spectral comparison of visible images versus 1550nm images acquired at a distance of 106 meters.



(a)

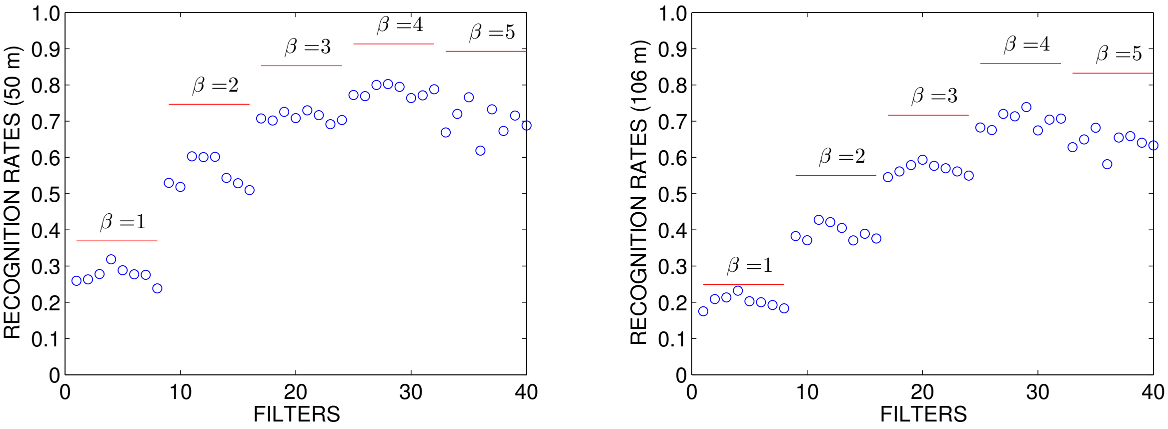
(b)

Figure 4.13: Cross-spectral comparison between visible images and 1550nm images captured at a distance of 106 meters. The left panel shows ROCs for the proposed method (SWLD+LBP+GLBP) and Faceit G8. The right panel displays CMCs for the two matchers.

4.4.3 Recognition Performance vs. Parameters of Gabor Filters

In this section we relate the long-range recognition performance to the parameters of the Gabor filters employed for the recognition task. We show that the adopted number of Gabor filters and the chosen scale parameters provide optimized performance in matching SWIR vs. visible face images.

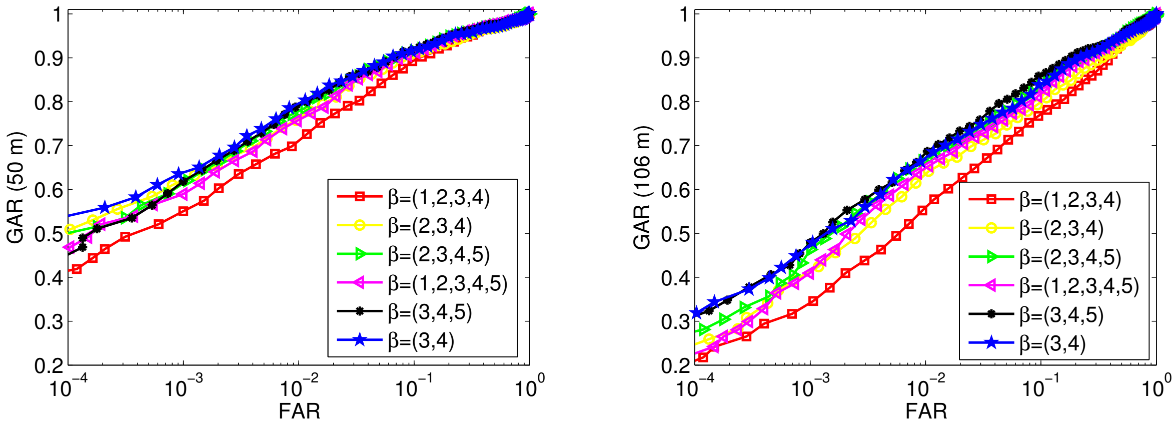
In Sec. 4.2.3 we introduced and described a bank of Gabor filters employed for our experiments. Bank of Gabor filters for face recognition were introduced in the work of of Wiskott et al. [13]. Wiskott et al. suggest to use a bank of 40 filters having 5 scales and 8 orientations [13]. We implemented the bank of filters described in [13] and applied it for cross matching SWIR and visible face images. Our experiments show that when local descriptors are adopted to encode Gabor responses, only certain scales of such a bank of Gabor filters provide good recognition rates. In this section, we evaluate the performance of a matcher that employs only an individual Gabor filter or a bank of filters parameterized by a different number and combination of scale parameters. The (SWLD, LBP, GLBP) operators are applied to filter responses by following the procedure described in Sec. 4.2. The performance of individual filters are assessed in the cross-spectral comparison at 50 and 106 meters. In Fig. 4.14(a), 4.14(b) the recognition rates evaluated at FAR 10^{-1} for individual filters are indicated with circles. Each circle corresponds to the performance of a filter at a given scale and orientation and each line corresponds to the performance of the combined filters (at the feature level) for 8 different orientations evaluated at a single scale. The filters are ordered from the smallest scale ($\beta = 1$) to largest scale ($\beta = 5$). We note that filters at the same scale have performance within a certain range, and no particular orientation of the filters provides the best recognition result across all filter scales. When filters at different orientations and having the same scale are combined, a noticeable gain in performance is obtained. The best scale in terms of performance is $\beta = 4$. The performance drops as β varies in the order $\beta = 3, 5, 2, 1$.



(a) (b)

Figure 4.14: The figure displays the verification performance (GAR) at FAR 10^{-1}. The circles indicates the verification performance for a given orientation and scale β. The lines indicate the performance when all filters (8 orientations) at given scale β are combined. On top of the circles and lines the corresponding scale parameter is also shown. The left panel indicates the 50 meters case; the right panel demonstrates the 106 meters case.

Fig. 4.15(a), 4.15(b) show verification performance (ROC curves) for different banks of filters parameterized by different numbers and combinations of scales (all orientations are employed). When all filters from all different scales are involved ($\beta = (1, 2, 3, 4, 5)$) we obtain one of the worst ROCs at both distances (50m, 106m). Fig. 4.15(a) shows this. When we employ the combination of filters with parameters $\beta = (3, 4)$, we obtain the best recognition performance at 50 m. Fig. 4.15(b) indicates that both combination $\beta = (3, 4)$, and the combination $\beta = (3, 4, 5)$ have similar performance at 106 m (combination $\beta = (3, 4)$ boosts the tail of the ROC and combination $\beta = (3,4,5)$ boosts the main portion of the ROC). The plots demonstrate that the two combinations are the best among other combinations at a given distance. We also note that the processing time for $\beta = (3,4)$ is about 66% of the processing time for $\beta = (3,4,5)$. Hence, we conclude that the best choice in terms of recognition performance and processing time at both long range distances is $\beta = (3,4)$.



(a) (b)

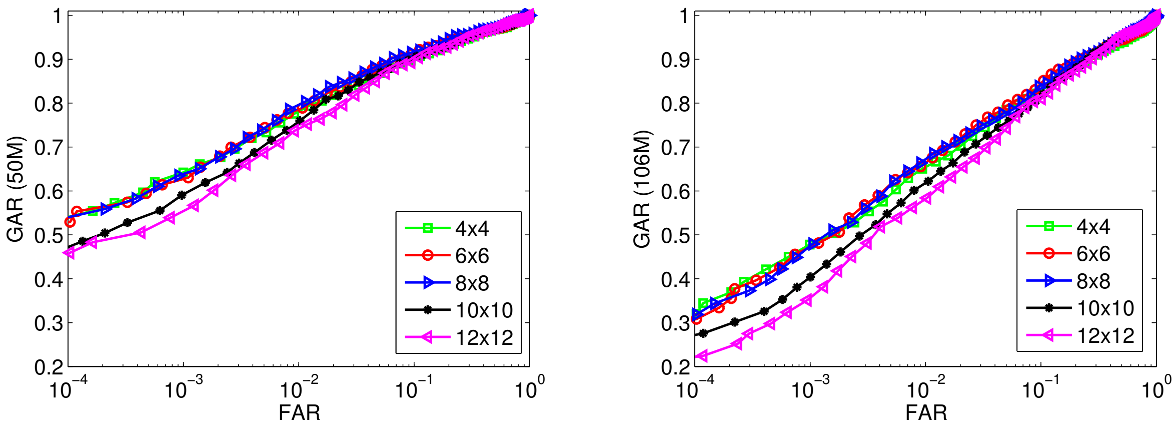
Figure 4.15: ROC curves for different combinations of scale parameter β in the encoding process. The left panel is for the case of 50 meters; the right panel is for the case of 106 meters.

4.4.4 Effect of Partitioning of Gabor Patterns

As described in Sec. 4.2.5, the encoded magnitude and phase of Gabor filtered images, are partitioned into square patches for further processing. In this section, we evaluate long-range recognition performance for different partitioning of the encoded patterns. We consider five patch sizes $P = \{(4 \times 4), (6 \times 6), (8 \times 8), (10 \times 10), (12 \times 12)\}$ pixels. We choose to work with even partition sizes because the size of the normalized images is also even.

Fig. 4.16(a), 4.16(b) show the verification performance for different patch sizes for the two operational distances (50 m, 106 m).

The plots show that at both distances the patch sizes $(4 \times 4), (6 \times 6), (8 \times 8)$ provide about the same verification performance (the three ROCs overlap), however the processing time for (4×4) and (6×6) is about 4 and 1.5 times the time employed for patch size (8×8).



(a)

(b)

Figure 4.16: ROC curves for different partition size of the Gabor patterns. The left panel is for the case of 50 meters; the right panel is for the case of 106 meters.

Furthermore, the performance of the matcher decreases when the patch size is larger than (8×8) (i.e. $(10 \times 10), (12 \times 12)$). Therefore, we conclude that the patch size (8×8) is the best partition size for the Gabor patterns.

4.4.5 Another Method of Comparison Between the Proposed Method and Faceit G8

The matching scores of our recognition approach and the scores of Faceit G8 engine can be combined to improve recognition performance of the designed cross-spectral face recognition system. We do not seek for an optimal fusion scheme.

The main purpose of applying a match score fusion is two-fold: (1) to emphasize that the proposed recognition method and Faceit G8 are orthogonal in terms of extracted features; (2) to demonstrate that the best performance, in a fusion scenario, is obtained when our method is given a larger weight than Faceit G8.

Different rules are presented in the literature for combining matching scores or classifiers [64, 94]. We employ a weighted sum rule which is a simple and effective rule [95]. Since Faceit G8 employs a similarity measure and our approach adopts a distance measure, we first transform similarity and distance measures into a common space by means of transformation below:

$$MS_{G8}' = 1 - \frac{MS_{G8} - min\{MS_{G8}\}}{max\{MS_{G8}\} - min\{MS_{G8}\}} \tag{4.13}$$

$$MS_{PM}' = \frac{MS_{PM} - min\{MS_{PM}\}}{max\{MS_{PM}\} - min\{MS_{PM}\}} \tag{4.14}$$

where MS_{G8} indicates the matching score produced by Faceit G8 and MS_{PM} is the matching score of the Proposed Method (PM). We have adopted a min-max normalization scheme which is well known in the literature [96]. It produces scores in the range [0,1]. We estimate the minimum and maximum values of matching scores by employing a representative sample (about 1/3) of the available data.

The weighted sum rule states that the combined matching scores is given by:

$$MS = w_1 \times MS_{G8}' + w_2 \times MS_{PM}', \tag{4.15}$$

where ”×” stands for multiplication.

The values of w_1 and w_2 are usually chosen to be a function of a performance metric (e.g. d-prime, Equal Error Rate, a single point on ROC curve, etc) from a training set. Here, we use the entire set to estimate the weights. The weights satisfy a normalization constraint: $w_1 + w_2 = 1$. We will show that to obtain the best improvement of matching performance we have to choose $w_1 > w_2$. Furthermore, since the combination of the scores leads to a substantial performance improvement, we may conclude that two recognition methods (ours and Faceit G8) employ different encoding methods (note that Faceit G8 algorithm is unknown). Indeed, in the literature it is shown that classifiers that combine well employ uncorrelated or negatively correlated features [94].

To achieve the best improvement of recognition performance, we maximize the following gain function:

$$G(w_1) = \int_0^1 [GAR_{FUS}(R, w_1) - GAR_{PM}(R)]\, dR \tag{4.16}$$

where GAR_{PM} are the values of Genuine Accept Rate obtained with the proposed method, parameterized by R (False Accept Rate), and GAR_{FUS} are the values of GAR obtained with fusion parameterized by w_1 and R. The expression geometrically indicates the area delimited by Receiver Operating Characteristic (ROC) curve obtained with fusion and ROC curve obtained with the proposed method. The maximum gain value G^* is achieved for an optimal weight w_1^*. The optimal weight w_2^* is obtained by enforcing the normalization constraint: $w_2^* = 1 - w_1^*$.

Fig. 4.17 shows that the approximate (up to 2 decimal digits) maximum values of gain G^* is obtained for $w_1^*(50) = 0.61$ (and $w_2^*(50) = 0.39$) at a distance of 50 meters and for $w_1^*(106) = 0.60$ (and $w_2^*(106) = 0.40$) at a distance of 106 meters.

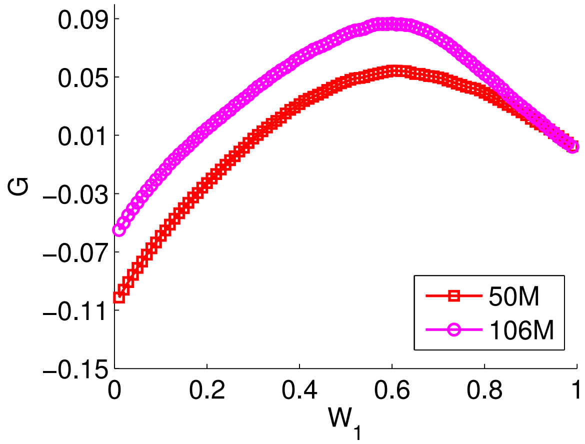


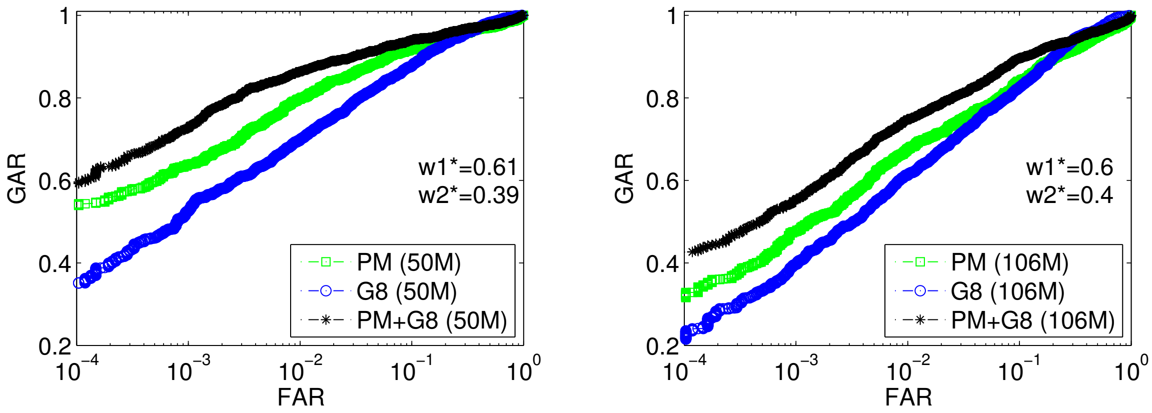
Figure 4.17: The value of the gain function G parameterized by the value of w_1 when the Proposed Method (PM) is combined with Faceit G8. The maximum gain is obtained for the approximate optimal weight $w_1^*(50) = 0.61$ (and $w_2^*(50) = 0.39$) at a distance of 50 meters and $w_1^*(106) = 0.60$ (and $w_2^*(106) = 0.40$) at a distance of 106 meters.

Fig. 4.18(a), 4.18(b) demonstrate performance improvement due to combining the two classifiers. Note the improvement over the entire range of FARs. The values of the optimal weights are proportional to the recognition performance of the individual methods and the largest gain of performance is obtained when the weight of our recognition scheme is 20% more than the weight of Faceit G8.

4.4.6 Recognition Rates vs. Range

In this section we summarize verification and identification results for cross-spectral matching and plot them as a function of the acquisition distance.

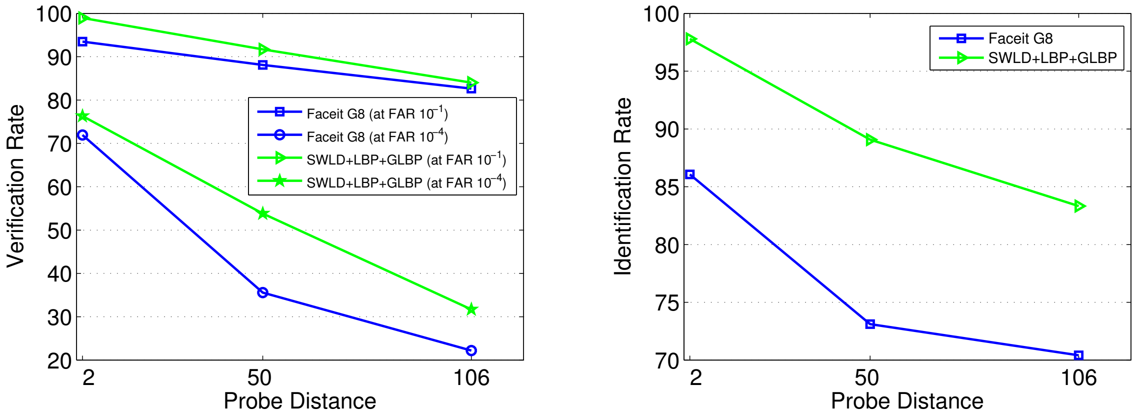
The plots 4.19(a), 4.19(b) indicate that the proposed method outperforms Faceit G8 in



(a) (b)

Figure 4.18: Cross-spectral comparison between visible images and 1550nm images captured at 50 and 106 meters. The panels display ROCs for the Proposed Method (PM), Faceit G8 (G8), and their combination (PM+G8) when we employ the optimal weights $w_1^*(50) = 0.61$ and $w_2^*(50) = 0.39$ for the 50 meters case (left panel) and weights $w_1^*(106) = 0.60$ and $w_2^*(106) = 0.40$ for the 106 meters case (right panel).

both scenarios at all operational distances and the overall recognition rates decrease as the capture distance of the probe images increases.



(a) (b)

Figure 4.19: Recognition rates of the proposed encoding scheme (SWLD+LBP+GLBP) and Faceit G8 as a function of probe distance. The left panel shows verification rates at FAR equal to 10^{-1} and FAR equal to 10^{-4}. The right panel displays identification (rank-1 recognition) rates.

4.5 Conclusion

In this chapter, we have described a novel method to compare a probe set of face images acquired in the SWIR spectral band against a gallery of visible light images. The implemented scheme involves a small set of Gabor filters and three local operators. An operator known as SWLD was applied to encode the magnitude of the Gabor filter responses, which is claimed here as a new application of SWLD operator. This operator was jointly employed in the Gabor domain with a uniform 12-bit LBP operator to encode magnitude of the Gabor response and was further complemented with a uniform GLBP descriptor to encode the phase response. The method is demonstrated on two datasets with SWIR images acquired at three different distances (2, 50, and 106 meters). The proposed approach is shown to outperform in terms of cross-spectral (SWIR vs. visible) verification and identification the commercial face recognition engine Faceit G8.

Chapter 5

Application of SWLD, LBP, GLBP to Visible Data

5.1 Introduction

In Chapter 4, we introduced a face recognition algorithm for cross-spectral matching. The designed algorithm employs Gabor filters at the first stage and SWLD, LBP, GLBP operators at the second stage. We have demonstrated that the encoding scheme is effective when applied to match SWIR vs. visible light data. In this chapter, the developed method for cross-spectral recognition is tested under a traditional scenario where the images are collected in the visible spectral band in a controlled manner and at a short distance (e.g. a few meters). We analyze the performance of the multi-operator scheme and also the performance of the matchers that employ individual local operators to encode Gabor-filtered images. We demonstrate that the adoption of a single local operator is sufficient when a controlled scenario is employed for data collection. Particularly, we show that an encoding method based on Gabor filters and the Simplified Weber Local descriptor produces high recognition rates. The recognition performance of this algorithm is compared against the performance of the commercial software Faceit G8.

5.2 Database Description

For the experiments we employ 100 face classes from the FRGC 2.0 dataset [26]. The dataset is composed of 1000 high-resolution images (10 images per class) captured in a controlled scenario (studio setting) having two different illumination setups. This is a scenario used to collect data for security applications (e.g. US VISA program), in which face images are captured at a short distance (e.g. 1-2 meters) with a silicon-based camera. The collection of samples is composed of frontal face images; half of the images have a neutral pose and the remaining half have a smiling expression. The average pixel distance between the eyes in the original face images is about 250 pixels. Examples of face images from the FRGC 2.0 dataset are shown in Fig. 5.1.

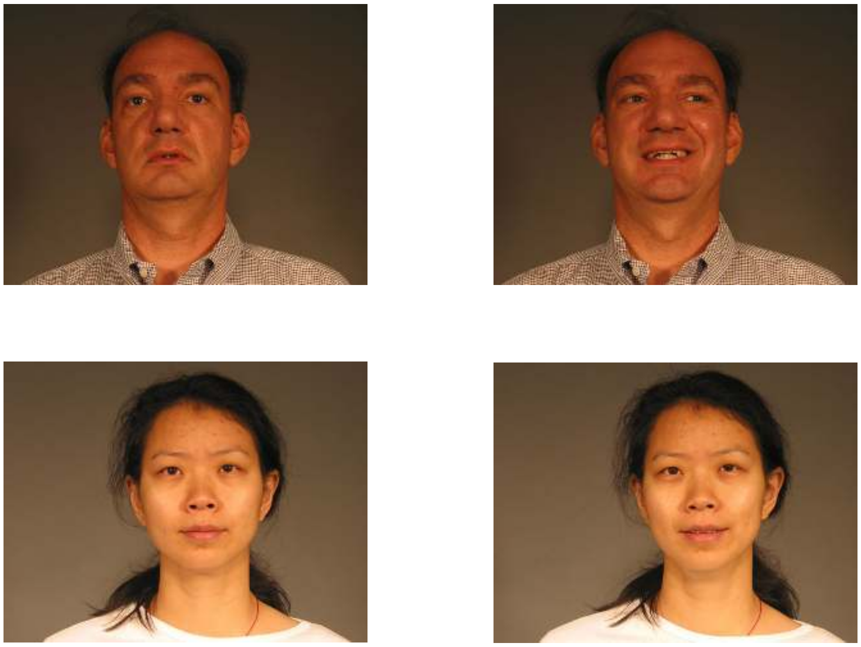


Figure 5.1: Sample images from FRGC 2.0 dataset acquired in controlled settings.

5.3 Experimental Results

5.3.1 Preprocessing

Prior to encoding and matching, the color images are geometrically normalized. They are brought into a canonical form by adopting the coordinate of the eyes as landmarks and by following the procedure described in Chapter 4 (see Sec. 4.2.1). The eye coordinate values are obtained from the original metadata included in FRGC 2.0 dataset. The images are normalized to have a resolution of 128x144 pixels. The adopted resolution is sufficient for the recognition task and makes the computation feasible. The color images are transformed into gray scale by employing a linear combination of the color channels. No other preprocessing is applied to the face images.

5.3.2 Encoding

To evaluate the verification performance we adopt the cross-spectral encoding scheme introduced in Chapter 4. The normalized visible images are filtered with the Gabor bank of filters defined in eq. 4.2. The employed parameters for the complex vector are $k_\beta = \frac{\pi}{2^{\beta/2}}$ with $\beta = 3, 4$ and $\phi_\alpha = \frac{\pi}{\alpha} \times 8$ with $\alpha = 1, 2, .., 8$. The adopted Gaussian kernel has the standard deviation $\sigma = \pi$. A total of sixteen filters is used. The three multi-scale local operators (SWLD, LBP, GLBP) defined in Sec. 4.2.4 are applied to the Gabor-filtered images. We partition the encoded images in square blocks of size 8×8 and generate histograms with 135 bins per block. We follow the procedure described in Sec. 4.2.5 to create features vectors. The symmetric I-divergence distance is adopted as matching score. We report the performance of the complete scheme that simultaneously employs all three local operators to process the filtered images. We also analyze the performance of the three encoding methods that employ the individual local operators (SWLD, LBP, GLBP) to encode the filtered images. For comparison purposes, we also process the FRGC 2.0 dataset with the Faceit G8.

5.3.3 Recognition Results

In this section, we evaluate: (1) the verification performance of the complete encoding scheme described in Sec. 4.2; (2) the performance of the three encoding schemes that adopt the individual multi-scale local operators to encode the Gabor-filtered face images; (3) the performance of the commercial software Faceit G8. The verification performances are assessed by matching all available images in the dataset, obtaining a total of 449,500 matching scores. The results of the comparison (ROC curves) are shown in Fig. 5.2.

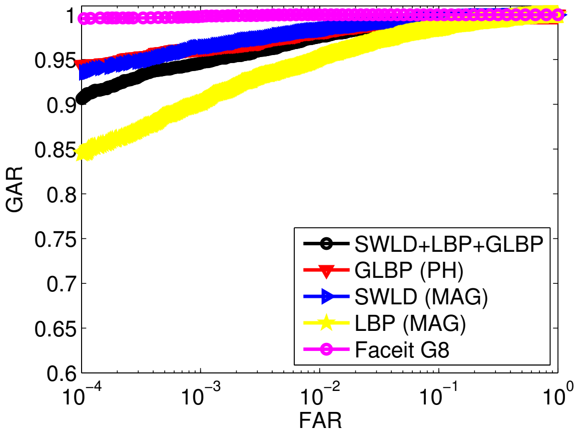


Figure 5.2: Verification performance evaluated on FRGC 2.0 dataset captured in a controlled scenario. The plot shows ROCs for matchers employing individual local operators (SWLD, LBP, GLBP), combined local operators, and Faceit G8.

Fig. 5.2 indicates that in the visible spectral band, both the multi-scale Simplified Weber Local descriptor (SWLD) applied to the magnitude response of Gabor filters and the multi-scale Generalized Local Binary Pattern (GLBP) applied to the phase response of Gabor filters (GLBP) produce very high recognition rates. The multi-scale Local Binary Pattern (LBP) operator applied to the magnitude responses of Gabor filters instead provides relatively low recognition rates. The ROC curves of SWLD and GLBP methods shown in Fig. 5.2 overlap at many different operational points. Table 5.1 indicates that there is no significant difference in recognition rates between these two methods. However, in Chapter 4, we have shown that the GLBP method provides the lowest performance when used to match SWIR against visible data. Chapter 6 will show also that a GLBP based approach (that encodes the phase response of Gabor filters) produces the lowest recognition rates amongst all the other methods based on local operators when applied under an uncontrolled scenario in the visible spectral band.

We also note, while in a cross-spectral comparison the combination of the three operators (SWLD+LBP+GLBP) at the feature level with the given parameters provides an effective encoding scheme, however, this is not true when visible data are matched. The results show that in the visible spectral band a scheme that employs a single local operator can be sufficient to obtain high recognition rates.

Fig. 5.2 further shows that the commercial software Faceit G8 overall provides the best performance among all the tested methods. However at some operational points (FAR values), we note an overlap with the ROCs obtained with SWLD and GLBP methods. Hence we provide a more detailed analysis by summarizing the data in Table 5.1.

METHOD	EER (%)	GAR (%) at FAR 10^{-1}	GAR (%) at FAR 10^{-2}	GAR (%) at FAR 10^{-3}	d-prime
LBP (MAG)	2.91	98.51	97.59	90.20	3.17
GLBP (PH)	1.42	99.67	98.28	96.31	3.45
SWLD (MAG)	1.39	99.54	98.30	96.41	3.33
SWLD+LBP+GLBP (MAG & PH)	1.82	99.27	95.27	94.61	3.32
FACEIT G8	0.13	100	100	99.80	3.61

Table 5.1: Verification rates (GAR) at FAR 10^{-1} and FAR 10^{-3}, EER, d-prime index for individual local operators (SWLD, LBP, GLBP), combined local operators and Faceit G8, for FRGC 2.0 visible dataset in a controlled scenario.

The table indicates that Faceit G8 provides perfect recognition performance in the range of FAR between 10^{-2} and 1. The recognition rates of the SWLD method, evaluated at FAR values 10^{-1} and 10^{-2}, are only 0.4% and 1.7% lower than the corresponding Faceit G8 performance. A more noticeable difference of performance with Faceit G8 is observed at lower FAR rates (e.g. 2.4 % at FAR 10^{-3}). The FRGC 2.0 dataset is a well known benchmark dataset, and the commercial software may have been previously trained on this dataset. To the best of our knowledge, no published face recognition method reports 100% recognition (evaluated at FAR 10^{-1}) on this dataset (controlled scenario). Instead, the implemented encoding schemes based on local operators have not employed any image from the dataset for training purpose. Since the recognition rates obtained with SWLD method are already quite large (e.g. GAR at FAR 10^{-1} is greater than 99.5%) we do not attempt to modify the implemented face matcher to obtain better performance on the FRGC 2.0 dataset (the improvement would be marginal). Instead, we optimize the SWLD method to obtain good performance in a more challenging scenario (surveillance in a camera network) and compare it against the performance of the commercial software and other methods based on local descriptors.

5.4 Conclusion

In this chapter, we have shown that in the visible spectral band, a recognition scheme that employs Gabor filters and an individual local operator can be sufficient for the face recognition problem in a traditional setting. In particular, we have shown that the SWLD matcher provides high recognition rates when face images have large resolution and are captured at a close distance. The obtained results motivate the work described in Chapter 6. In Chapter 6, we will show that the adoption of magnitude responses from Gabor filters followed by the application of the Simplified Weber Local Descriptors is a good recognition scheme when used in an uncontrolled scenario with multiple views of the same subject captured in the visible spectral band. The magnitude response of Gabor filters changes smoothly for adjacent face locations. The Gabor phase response, instead, changes quickly across different face locations. In uncontrolled settings, where a displacement exists between probe and gallery images, magnitude responses of Gabor filters are robust features. In Chapter 6, we will demonstrate that the magnitude-based Gabor features provide higher recognition rates than phase-based Gabor features in an uncontrolled setting. Chapter 6 will also show that the optimization of the matcher design leads to low processing time when applied to low resolution images.

Chapter 6

Face Recognition in a Wireless Camera Network

6.1 Introduction

In the previous chapter we have demonstrated that an encoding method based on Gabor filters and Simplified Weber Local Descriptor (SWLD) achieves high recognition rates in a scenario where subjects are controlled and visible images are compared. In this chapter we propose an optimized and simplified version of the same algorithm to recognize individuals in a surveillance scenario. In particular, we describe a fully automatic face recognition system that detects and recognizes individuals in near real time. The system uses low-power embedded cameras to acquire video frames of subjects in an uncontrolled environment and opportunistically extracts frontal face images in real time. Images may be strongly blurred, have low resolution or large pose variability. To select images of acceptable quality, a quality based selection process is employed to discard some of the images that are not suitable for recognition. Then, the face images are geometrically normalized according to a pool of four standard resolutions, by using coordinates of detected eyes. The images are transmitted to a fusion center which has a multi-resolution template gallery set. Then we employ a simplified and an optimized version of the encoding algorithm described in Chapter 5 that adopts Gabor filters and the SWLD operator. The method extracts features from the normalized probe face images. At the fusion center the comparison between gallery images and probe images acquired by a wireless network of seven embedded cameras is performed. A score fusion strategy is adopted to produce a single matching score. The performance of the proposed algorithm is compared to the commercial face recognition engine Faceit G8 and other methods based on local descriptors. The experiments show that the overall system is able to provide similar or better recognition performance than the commercial engine with a shorter computational time, especially with low resolution face images.

6.2 Related Work

The work of B. Lovell [97] proposed an enhanced version of PCA (Principal Component Analysis) and FLD (Fishers Linear Discriminant) to extract features in a surveillance scenario. The approach first obtains optimal weights for pairwise features and then builds a feature chain in order to determine the weights for all features. The proposed approach is low-complexity by design and the algorithm performance is evaluated on CMU PIE [98], FERET [22], and Asian Face databases [99]. The results show that the method performed well compared to other benchmark appearance-based methods.

The recent approach of D. Kisku et al. [100] adopts Gabor filters to encode face at multi-views (frontal and non-frontal) and further encode filtered data with PCA and a generalization of LDA (Linear Discriminant Analysis). The obtained feature vectors are linearly combined with weights estimated using a training set. The classification is performed with three different classifiers, namely, K-Nearest Neighbor, SVM with linear kernel and SVM with RBF. The algorithm is designed to work off-line and it is tested on the UMIST [101] dataset having controlled and good quality images (200 pixels of resolution, no blur).

The recent GE (General Electric) system proposed by F. Wheeler et al. [102] consists of a single network node with a pair of PTZ cameras to track, detect and recognize an individual at a distance of 15-20m. The active vision system employs adaptive background modeling to detect walking subjects and performs face detection adopting Pittsburgh Pattern Recognition FT-SDK [103]. The recognition is performed frontally with the Cognitec FaceVACS [60] software. The system has to firstly track the individual in the field and then zoom in on the targeted face. The adopted face recognition algorithm requires at least 30 pixels between the eyes to perform recognition.

E. Kokiopoulou et al. [104] address the problem of distributed classification of multiple views of the same object that are collected in an ad hoc camera network (without a central node). An average consensus algorithm is implemented for estimating the unknown object class by adopting an objective function. At the initial stage the objective function is locally estimated based on observation of the single node; as the algorithm progresses all observations are gradually taken into account in the estimation of the objective function. The method is illustrated on the UMIST dataset [101]; the obtained results show that if the training set is sufficiently large, the consensus classification decision is equivalent to the decision of a centralized system that has access to all observations.

The work of Banduml et al. [105] designs a network of four cameras to solve the problem of person re-identification. The system during the training stage produces a set of face-dependent classifiers by acquiring several video frames for each subject walking in the network. The method employs features based on Modified Census Transformation (MCT) to detect face images. The obtained face images have different pose, resolution and level of blur. Hence, the face images are geometrically normalized and partitioned in local patches of size 8×8 pixels and Discrete Fourier Transform (DCT) coefficients are evaluated for each partition to obtain face features. Hence, a Support Vector Machine (SVM) classifier for each individual is created. In the testing stage, the system detects an individual in the scene and tracks it to acquires several images. The network applies the available set of classifiers to identify the walking subject in the training set. It is shown that the system, without any human supervision, achieves a verification rate of 60% on a dataset of 10 subjects and obtains a verification rate of 86% after relevant feedback by an operator.

6.3 Network Description

The face images are captured by a network of wireless embedded cameras. A Logitech 9000 series camera [106], a Beagleboard [107] (with an OMAP processor), and a wireless card are assembled to form the individual embedded camera unit. Each embedded camera performs basic pre-processing operations such as background subtraction to detect an individual in a given scene. This is followed by a Haar cascade classifier-based face detection [4, 108] and an image selection process that discards very poor quality (heavy motion blur, poor resolution and bad pose) samples. Only frontal face images likely to yield high recognition accuracy are transmitted to a base station for face recognition. Using this setup, we perform an experiment with 29 individuals walking through a network of 7 embedded cameras. The adopted network design has been introduced in the work by Kulathumani et al [3]. In [3] the commercial software Faceit G6 is adopted to verify the achievable recognition performance by using a multi-camera network and shows how multiple cameras improve the chance of acquiring a suitable face image for recognition. In this chapter, instead, we show how face recognition can be implemented in near-real time by encoding the images with a Simplified Weber Descriptor, especially when the resolution of images are small. Its recognition accuracy is comparable to those of Faceit G8.

The network setup is illustrated in Fig. 6.1.

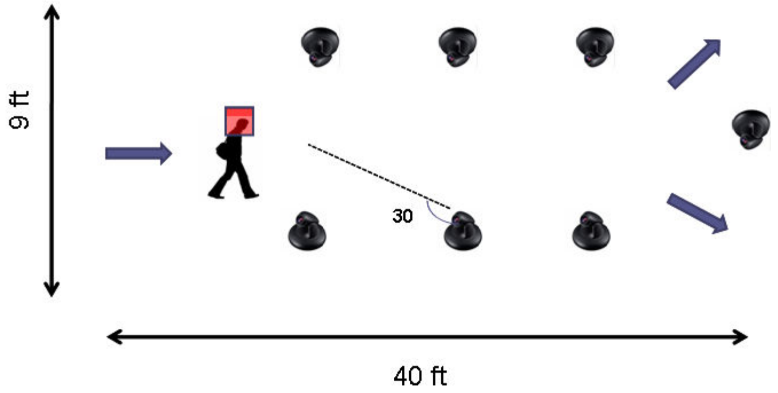


Figure 6.1: Illustration of the network setup. A face sample is captured in a network of 7 embedded cameras

In the next section we describe the face detection process that follows the work by Kulathumani et al. [3] (an additional eye detection step has been introduced in our implementation with respect to [3]).

6.4 Face Detection

We employ the Viola-Jones [4] face detector based on Haar wavelets and Adaboost classifiers. The face detector is designed to detect only frontal images (out of plane rotation up ±20 degrees with respect to the frontal view) and to extract the face area. Sometimes, the face detector produces false positives for detected faces by including side views as well as non-facial images. In order to reduce these false positives, we perform an additional step of eye detection after detecting the face. The eye detector algorithm is also based on the Viola-Jones approach. The eye detection step is essential to geometrically normalize the face to a canonical form. More details are provided in Sec. 6.5.1.

The detected face images are further analyzed for the presence of low resolution and motion blur. The Discrete Cosine Transformation (DCT) is employed to estimate and detect motion blur. DCT detects low and high frequency components, and the degree of motion blur can be estimated from the number of high frequency components [109]. We assess the presence of motion blur by observing the DCT coefficients of the top 10% high frequency components. Alternatively, the measure of blur defined by Eq. 3.6 could also be employed. During the process of image selection, we apply a predetermined threshold on the resolution of a face image and the blur metric to discard poor quality frames. These thresholds are empirically found based on the off-line analysis of the impact of different quality images on the matching performance. We empirically determine a necessary minimum resolution of size 28×28 pixels for a detected face before it is transmitted to the base station.

At the end of this phase, a total of 223 face images are obtained for all the 29 subjects with a minimum of 5 and a maximum of 12 images per subject. These face images are used to test the performance of our recognition technique and compare it with other algorithms. Approximately 90% of the bandwidth is saved on each transmitted frame by extracting only the face portion of the image. It is further reduced by the transmission of a small subset of frames.

6.5 Face Recognition

In this section, we introduce a simplified and optimized version of the face recognition algorithm described in Chapter 5. An optimized eye detection scheme and multi-resolution handling method for face templates is also described.

6.5.1 Preprocessing and Normalization

Detected face images are brought to a canonical form prior to being encoded. We employ the same (similarity) transformation described in Sec. 4.2.1 to project eye locations to two fixed points in the image space. The selected points correspond to the eye coordinates of a single face image taken as a reference.

The positions of the eyes are automatically detected with Haar features and an AdaBoost classifier [4] using the OpenCV library [108]. In our implementation, to reduce the number of false alarms, we enforce a few geometrical constraints: detection areas for left and right eyes, maximum vertical distance and minimum horizontal distance between candidate landmarks. In a surveillance scenario, images are usually captured in an unconstrained environment. The distance to the camera is unknown, and various distortions are observed in the images.

In particular, the acquired face images may have variable resolution (the number of pixels representing the face in the image). Hence, before we detect the eye coordinates we downsample high resolution face images and upsample low resolution face images to deal with face images at variable scale. This strategy is adopted only for the purpose to improve the eye detection algorithm.

At the stage of feature extraction during the enrollment, instead of defining a standard resolution for resampling the normalized images, we create feature templates at different image resolutions (a predefined set). This solution avoids excessive upsampling to a given resolution when captured images are extremely small (e.g. 28×28 pixels). This is beneficial for recognition methods that employ local features, since upsampling may produce aliasing effects that can destroy the local structure of the face images.

The adopted strategy increases processing time at the enrollment stage, which is usually done off-line, but has the advantage to decrease the average processing time at the testing stage. Particularly, low resolution images will be processed much faster compared to high resolution images.

A detailed summary is provided in Sec. 6.6.1 and Sec. 6.6.3.

6.5.2 Feature Extraction

The proposed feature extraction method is based on an algorithm that employs Gabor filters at the first stage and Simplified Weber Local Descriptor (SWLD) at the second stage [110]. In Chapter 5, the algorithm was demonstrated on a controlled scenario. In this chapter we demonstrate the algorithm in an uncontrolled scenario where multiple views of the same individual are available. Differently from the previous chapter, we employ a different design for the Gabor filters and the local descriptor to obtain good recognition performance and a short processing time.

To extract face features, firstly, the normalized images are convolved with a bank of 32 Gabor Kernels given by:

$$G_{\alpha,\beta}(z) = \frac{\|k_{\alpha,\beta}\|}{\sigma^2} \exp\left[\frac{\|k_{\alpha,\beta}\|^2 \|z\|^2}{2\sigma^2}\right] [e^{ik_{\alpha,\beta}z} - e^{-\sigma^2/2}], \tag{6.1}$$

where σ^2 is the variance of the Gaussian kernel, $k_{\alpha,\beta}$ is the wave vector and $z = (x, y)$. We adopt a 32 × 32 pixel Gabor kernel having the following wave vector:

$$k_{\alpha,\beta} = k_\beta e^{i\phi_\alpha}. \tag{6.2}$$

The adopted parameters for the complex vector are $k_\beta = \frac{\pi}{2^{\beta/2}}$ with $\beta = 1, 2, 3, 4$ and $\phi_\alpha = \frac{\pi}{\alpha} \times 8$ with $\alpha = 1, 2, .., 8$. The Gaussian kernel has standard deviation $\sigma = \pi$. We note that with respect to the case described in Chapters 4 and 5, where we match larger resolutions images (i.e. equal or larger than 110 × 110 pixels), we employ a filter bank that also adopts small scale filters ($\beta = 1, 2$). From the experiments, we noted that micro-features captured by the small scale filters are also useful. Furthermore, smaller scale filters are appropriate because the Gabor kernel size is comparable to the size of the normalized images. A smaller kernel size (e.g. 16 × 16 pixels) has also been tested but the experiments have shown that the adoption of a smaller kernel size didn't improve the performance of the matcher on the tested dataset. Hence the kernel size 32 × 32 has been employed, resulting in a fast convolution operation between filters and face images.

After filtering images with the Gabor kernels, each Gabor magnitude response is encoded with the following single-scale SWLD descriptor:

$$SWLD_{l,r,8}(x) = \mathcal{Q}_l \left\{ tan^{-1} \left[\sum_{i=0}^{7} \left(\frac{x_i - x}{x} \right) \right] \right\}, \tag{6.3}$$

where x_i are the neighbors of x at radius $r = 1$ and $\mathcal{Q}_l$ is a uniform quantizer with l quantization levels.

The obtained operator is a robust edge detector that preserves the intensity values along the edges. With respect to the implementation in Chapters 4 and 5, the support of the local operator is reduced at the first radius and at 8 neighbors. These parameters appear to be a good choice for low resolution images. In the following experiments we adopt $l = 64$ levels to discretize the output of the tan^{-1} function.

Finally, all obtained patterns are partitioned into nonoverlapping regions of size 8 × 8 pixels, and a 64 bin histogram of feature values is assembled for each block. Individual histograms are concatenated, resulting in a template vector. We note that the overall feature extraction process does not require any kind of training.

To compare two images A and B, we adopt the symmetric I-divergence distance as a matching score:

$$d(A, B) = \sum_{k=1}^{K} (H_A(k) - H_B(k)) \log \frac{H_A(k)}{H_B(k)}, \tag{6.4}$$

where K is the length of the feature vectors $H_A(k)$ and $H_B(k)$ obtained from images A and B. The feature vectors are normalized such that matching pairs at different resolutions produce distance scores within the same range.

In the remainder of this chapter the described method will be called Gabor-SWLD. Note that we explicitly indicate (with Gabor prefix) an encoding method that has a Gabor-filtering stage to distinguish it from a method that does not employ the Gabor filters. In the following section, we compare the performance of Gabor-SWLD approach with (1) the performance of the commercial software Faceit G8 [80], (2) the performance of a matcher based on the approach introduced in [21] by encoding the magnitude responses of a bank of Gabor filters with a 8-bit Local Binary Pattern (LBP) operator, (3) the performance of a matcher similar to [92] by encoding the phase responses of a bank of Gabor filters with a 8-bit Local Difference Pattern (LDP) operator, (4) the performance of a method that follows [19] that applies a 8-bit LBP operator directly to raw images. We show that our approach achieves better verification performance than Gabor-LBP [21], Gabor-LDP [92], LBP [19] and has comparable or better verification performance to the Faceit G8 software. We also demonstrate that our method can be extremely fast when applied to low resolution images (e.g. images at a large distance from the camera) compared to Faceit G8 that has a processing time independent of the resolution.

6.6 Experimental Results

6.6.1 Dataset and Preprocessing

During enrollment, 5 images per subject are collected at close distance from a single camera to obtain a total of 145 gallery images. Also as described in Sec. 6.4, we obtain 223 probe images with an average of about 7 images per face class using our multi-camera network based face detection experiments. The captured probe images have an out of plane rotation up to ±20 degrees with respect to the frontal view. They also have variable resolution and different amount of blur involved due to the limited capture volume of cameras in the network. The eye detector is applied to the gallery images (at the enrollment stage) obtaining a total of 136 detected gallery images and to probe images (at the testing stage) resulting in a total of 216 probe images. The eye detection rate can be improved by adding an eyeglasses detector. However, to keep the processing time low we adopt a single Haar cascade trading off complexity versus performance. The normalized probe images are partitioned into 7 resolutions: 28×28, 38×38, 47×47, 56×56, 75×75, 103×103, and 122×122 pixels. The distribution of probe images grouped by their original resolution is shown in Fig. 6.2. Note that most of the probe images have low resolution.

As explained in Sec. 6.5.1, we predefine a set of resolutions to rescale the probe images

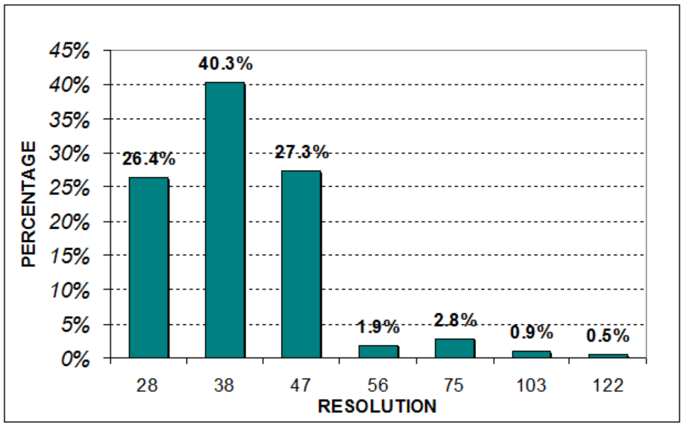


Figure 6.2: Distribution of probe images grouped by resolution.

prior to encoding and matching. All high resolution probe images (6% of all probe images) are scaled to 64 pixels; images with resolution 47, 38, and 28 pixels are mapped to 48, 40, and 32 pixels by using a bicubic interpolation. In general, once the sets of standard resolution have been defined, the incoming probe images can be rescaled to the closest size in this set.

In the following three subsections, we analyze the performance of our recognition system. We first evaluate the recognition performance in computing the receiver operating characteristics by considering all the probe images that are collected for the 29 subjects Next, we evaluate the processing time required to encode and match face images. Finally, we evaluate the recognition performance after employing a score fusion strategy by which, for each subject, the probe image that yields the best score with respect to each gallery image is selected and the receiver operating characteristics are computed based on these selected probe images.

6.6.2 Performance Using Data From a Single View

The performance of the method is compared to the performance of the commercial software Faceit G8, Gabor-LBP approach, Gabor-LDP method and LBP matcher. Faceit G8 employs a similarity measure as matching score, instead Gabor-SWLD (64 levels), Gabor-LBP (8 bits), Gabor-LDP (8 bits) and LBP (8 bits) employ a symmetric I-divergence distance. Other measures of distance such as Chi-Square (Eq. 4.11) or Histogram Intersection (Eq. 4.10), can be adopted. Based on our experiments, the I-divergence distance performs reliably compared to other metrics. Furthermore all methods that employ Gabor filters (including our approach) are using the same bank of Gabor filters for a fair comparison.

In the following experiments the set P of 216 probe images is matched against the gallery set G composed of 136 images. Each probe image is individually submitted for verification. The resulting Receiver Operating Characteristic (ROC) curves are shown in Fig. 6.3

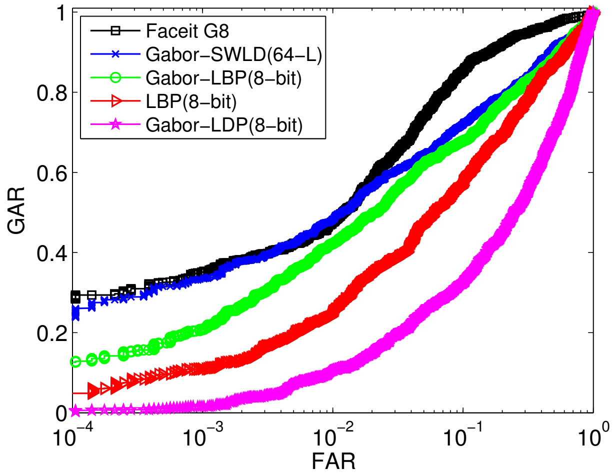


Figure 6.3: ROCs characterizing performance of Faceit G8, our method (Gabor-SWLD), Gabor-LBP, LBP, and Gabor-LDP matchers for a single camera.

The results in Fig. 6.3 show that Gabor-LDP provides the worst performance. The LBP approach outperforms the Gabor-LDP method, however the Gabor-LBP approach provides better performance compared to LBP. Finally, the Gabor-SWLD approach (the proposed method) outperforms Gabor-LBP. The experiments show that the Gabor-LDP that employs phase responses of Gabor filters is not a suitable scheme for an uncontrolled scenario and for matching within the visible spectral band. The Gabor-LDP provides poor performance because the Gabor phase response changes rapidly across various locations within a face image. In uncontrolled settings and at low resolutions, when a misalignment exists between gallery and probe images, the phase features of the two images can be extremely different at the same locations. Instead, the magnitude response of Gabor filters varies slowly for adjacent face locations. As a consequence, it is less sensitive to misalignments or pose differences that exist between the comparing images. We also note that the Gabor-LDP [92] is a particular version of the Gabor-GLBP (as explained in Chapter 4) method. Hence, we have demonstrated that a method (Gabor-SWLD) that employs the magnitude response of Gabor filters works best with respect to other methods that employ local descriptors in an uncontrolled scenario. We further note that the Faceit G8 software provides the best performance for the operational points in the range 0.05 to 1 of False Accept Rate (FAR), however its performance is similar to the performance of the Gabor-SWLD matcher for the operational points in the range 10^{-4} to 0.05 of FAR. The cumulative results show that all four matchers provide low verification rates ($< 50\%$) at FAR below 10^{-2}. The complete summary of performance is provided in Table 6.1.

MATCHER	dprime	EER (%)	GAR (%) at FAR 10^{-2}	GAR (%) at FAR 10^{-3}
Gabor-SWLD	1.5795	20.81	48.32	33.38
Faceit G8	1.1251	11.91	47.63	35.21
Gabor-LBP	1.4498	21.52	42.21	20.81
LBP	1.1245	25.67	25.35	10.95
Gabor-LDP	0.5534	38.56	10.65	1.58

Table 6.1: d-prime, Equal Error Rate (EER), and Genuine Accept Rate (GAR) evaluated at FAR 10^{-2} and FAR 10^{-3} for Faceit G8, our method (Gabor-SWLD), Gabor-LBP, LBP, and Gabor-LDP matchers using a single camera.

6.6.3 Processing Time

In this section we compare the processing time of the proposed face recognition method against the processing time of Faceit G8. Broadly speaking, the processing time can be divided into two parts: (1) creation of a probe template from the detected face image which we denote as operation CT, and (2) matching the probe template against gallery templates, which we denote as operation $Match$. In our approach, the operation CT is composed of eye detection, normalization of face image into a canonical form using eye positions, selection of appropriate resolution s from a predefined set S, rescaling the face image to resolution s and finally extracting features from the face image. For the Faceit G8 software the corresponding operations are performed by function $CreateTemplate$.

In our approach, the operation CT for a probe image of resolution 64 pixels (the largest resolution supported by the set S) requires only 92.6% of the processing time employed by Faceit G8 and the $Match$ function at such resolution takes 80% of the matching time of Faceit G8, when performed on the same machine. Furthermore, the processing time of our method reduces dramatically as the resolution decreases. Conversely, the processing time of Faceit G8 remains constant and is independent of the resolution of probe images. In particular, at the lowest resolution of 32 pixels the operation CT for our approach is approximately 4 times faster compared to Faceit G8, and the $Match$ function at such resolution for our method is 5 times faster than Faceit G8. Table 6.2 summarizes the average processing time employed by Faceit G8 and (our) Gabor-SWLD method approach for both CT and $Match$ operations. The processing time has been evaluated on a PC equipped with Intel Pentium® 4 (2004) single core CPU operating at 3.20GHz and 1GB of RAM.

Image Size	CT		$Match$	
	Faceit G8 (msec)	Gabor-SWLD (msec)	Faceit G8 (msec)	Gabor-SWLD (msec)
64×64	730	676	100	80
48×48	730	246	100	46
40×40	730	195	100	29
32×32	730	177	100	19

Table 6.2: The table displays processing time (in milliseconds) for CT and $Match$ operations. The processing time is shown as a function of image size for Faceit G8 and Gabor-SWLD.

The results of the comparison are also displayed in Fig. 6.4.

6.6.4 Performance Using Data From Multiple Views

In this section we evaluate verification performance of the proposed encoding and matching algorithm (Gabor-SWLD) for the case when images are collected by multiple cameras. The performance of our method is compared to the performance of Faceit G8, the Gabor-LBP approach and the LBP matcher. Face images are captured by a camera network with an average detection rate of 7 probe images per walking subject, when the subject is in the field of view of the network.

Prior to performing experiments, we group the set P of 216 probe images (see its descrip-

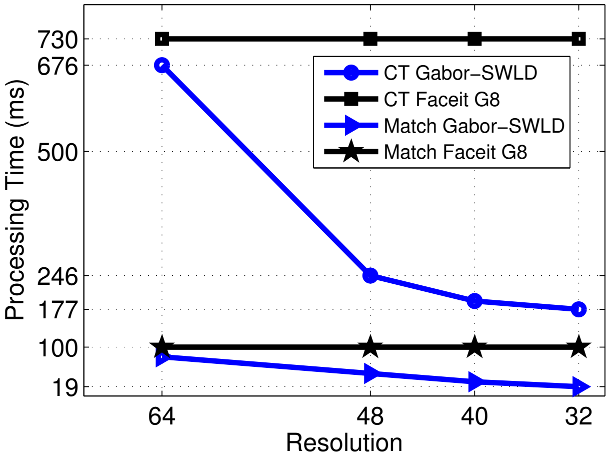


Figure 6.4: The graph displays the processing time (in milliseconds) for *CT* and *Match* operations of the Gabor-SWLD method and Faceit G8 software as a function of image resolution.

tion in Sec. 6.6.1) into 29 groups (the number of groups is equal to the number of walking subjects $i = 1, 2, \ldots, 29$) with an average number of 7 images per class. Let Q_i, $i = 1, \ldots, 29$ be the i-th subset with the images of the i-th class. At the fusion center, we compare each group Q_i of probe images against all images in the gallery G. Each group Q_i contains a different number of probe images (the minimum number is 4 and the maximum number is 10 images). The images have different resolutions. The position of the head and the level of motion blur also vary from image to image. The data collection was repeated a number of times for each subject walking in the field of view of the network. In the following experiments, we assume that any group of images Q_i is a set of images acquired from an equivalent larger camera network (10 cameras). We also assume that cameras were not always able (e.g. because of pose or occlusion) to detect the face of the subject walking in the network. The average number of probe images per each individual is equal to 7 face images.

To compare each group Q_i of probe images against the images in the gallery G, we employ a score fusion strategy. In particular, we adopt the max or min rule [64] depending on the tested matcher. At the fusion center, k_i images of the current group Q_i are compared against the first image of the gallery G and the lowest matching score among the k_i is selected. This strategy is applied to Gabor-SWLD, Gabor-LBP, LBP and, Gabor-LDP matchers, since the

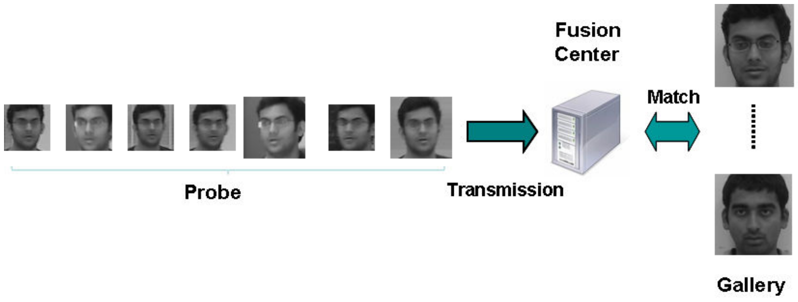


Figure 6.5: Matching process at the fusion center. An average of 7 face images having different resolutions are transmitted to the fusion center. Individual images are compared with the gallery in the database. A fusion strategy is employed to combine the matching results.

output of these matchers is a distance score. When Faceit G8 is employed, the adopted fusion rule selects the maximum of the scores, since the output of Faceit G8 is a similarity score. This normalization process is applied to all comparisons performed against remaining gallery images. The ROC curves characterizing the verification performance of the multiview face recognition method are shown in Fig. 6.6.

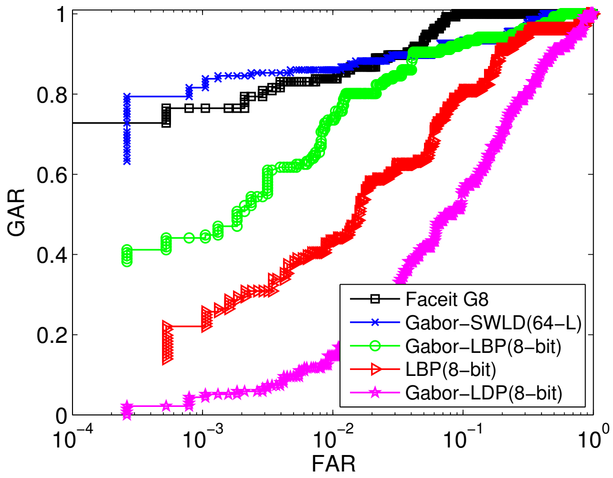


Figure 6.6: Verification performance for Faceit G8, our method (Gabor-SWLD), Gabor-LBP, LBP, and Gabor-LDP matchers using a camera network. The network detects an average of 7 probe images per walking subject.

The results demonstrate that the involvement of multiview data and a score fusion strategy leads to considerable performance improvement compared to the results in Fig. 6.3.

This is especially noticeable for the Gabor-SWLD approach and the Faceit G8 matcher. Note that our matcher performs best in the FAR range between 10^{-3} and 10^{-2}, while Faceit G8 outperforms our method in the range of FAR between 10^{-1} and 1.

MATCHER	dprime	EER (%)	GAR (%) at FAR 10^{-2}	GAR (%) at FAR 10^{-3}
Gabor-SWLD	3.054	7.35	86.03	80.51
Faceit G8	2.097	5.19	83.82	76.46
Gabor-LBP	2.719	8.09	73.53	44.12
LBP	1.942	15.51	43.38	18.12
Gabor-LDP	1.268	24.25	14.71	3.31

Table 6.3: d-prime, Equal Error Rate (EER), and Genuine Accept Rate (GAR) evaluated at FAR 10^{-2} and FAR 10^{-3} for Faceit G8, our method (Gabor-SWLD), Gabor-LBP, LBP, and Gabor-LDP matchers. The performance is shown for a multiview camera network employing a min or max rule for fusion.

Table 6.3 presents a comparison of the four matching methods based on a few operating points. The table shows that the Equal Error Rate (EER) of Faceit G8 is the lowest, however, our approach provides the best separability index (d-prime) and verification rates (Genuine Accept Rate) at FAR 10^{-2} and at FAR 10^{-3}.

We have also tested a strategy that averages the available matching scores (since the number of views k_i is variable for each user walking in the network) but in this case the fusion method produces a much lower improvement of recognition performance for all four methods based on local descriptors as shown in Table 6.4. The averaging method seems to be sensitive to few genuine comparisons that may generate high distance scores. For example, suppose that one of the views of the walking subject is very challenging for the matcher (e.g. has a large displacement in both pitch and yaw view angle with respect to the gallery image). Suppose also that the other views of the same subject are more suitable for the matcher. In this case, if we average the contribution of the challenging view with the more favorable views, the fusion result can be greatly affected from that single view that produces an undesired matching score. If we instead select the most favorable matching scores, we are going to provide much larger improvement for all the four methods based on local descriptor as shown in Table 6.3. Furthermore, the minimum rule works best in our system because the probe images are minimally preprocessed and are extremely different in terms of pose, resolution, blur, etc. On the other hand, by employing a larger processing time with respect to the proposed method, the commercial software tries to be optimal for every single comparison and works best when the scores are averaged as shown in Table 6.4.

In conclusion, we have shown that our method provides the best trade off between computational time and recognition performance amongst all considered methods by employing a minimum fusion rule.

MATCHER	dprime	EER (%)	GAR (%) at FAR 10^{-2}	GAR (%) at FAR 10^{-3}
Gabor-SWLD	2.426	8.10	72.79	56.25
Faceit G8	2.176	2.18	91.18	81.25
Gabor-LBP	2.425	11.76	61.40	40.07
LBP	1.856	18.37	48.53	27.21
Gabor-LDP	0.979	30.88	11.76	7.35

Table 6.4: d-prime, Equal Error Rate (EER), and Genuine Accept Rate (GAR) evaluated at FAR 10^{-2} and FAR 10^{-3} for Faceit G8, our method (Gabor-SWLD), Gabor-LBP, LBP, and Gabor-LDP matchers using a multiview camera network employing an average rule for fusion.

6.7 Conclusion

In this chapter, we have demonstrated the use of a wireless camera network to perform near real-time face recognition that yields high accuracy and robustness. We have employed a portable camera network composed of off-the-shelf components in which individual units perform background subtraction and face detection, and transmit only relevant face images to a fusion center for recognition. The adopted network design follows the work of Kulathumani et al. (an eye detection scheme is added in our implementation) [3]. The technique relies on opportunistically acquiring suitable face images as a subject walks through the multi-camera network.

We have designed an efficient face recognition algorithm employing Gabor filters and Simplified Weber operator. A multi-resolution handling method for the face features is also introduced. We evaluated the performance of this system for face recognition using a dataset of 29 subjects collected using the portable camera network. Our networked face recognition system is able to achieve comparable recognition performance to the commercial Faceit G8 software while reducing the required processing time.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this dissertation we developed a set of face recognition schemes to compare face images acquired across spectra: (visible vs SWIR).

Chapter 1 introduced the concept of face recognition and briefly described the contributions of the dissertation.

Chapter 2 provided an overview of the state of the art in the field of face recognition. It further described face matchers and preprocessing approaches that are adopted to compare face images within or across different spectral bands of the electromagnetic spectrum.

Chapter 3 presented three non linear methods that employ image quality metrics to enhance recognition performance of a face matcher. The first method selected biometric samples at the image level and decided if the input image is suitable for further processing and matching. The second method accepted or discarded a matching result by providing a confidence measure in the obtained comparison. The third method treated quality vectors of face images as weak features that are employed simultaneously with the matching scores in the matching stage. The proposed schemes estimated a non linear relationship between the input image quality metrics and a well defined quality index or score. We estimated the relationship by employing Feed Forward Neural Net. Labeled data were adopted to train the network. We applied the three methods to a face dataset of 202 face classes (a total of 2020 images from FRGC 2.0 dataset) acquired in the presence of different illumination and expression scenarios. The results showed that when the proposed methods were employed, a significant improvement of performance was obtained.

In Chapter 4, we designed a robust matching method to compare a gallery of face images acquired in the visible (VIS) spectral band against a probe image acquired in the Short Wave InfraRed (SWIR). The problem is very challenging, since the SWIR images are photometrically different from the VIS images. Furthermore, the probe images were acquired at large distances (up to 106 meters). The proposed method consists of a double stage mechanism that employs Gabor filters at the first stage and three multi-scale local operators at the second stage. The employed local operators are Simplified Weber Local Descriptor (SWLD), Local Binary Pattern (LBP) and Generalized Local Binary Pattern (GLBP). The multi-scale SWLD and LBP operators were employed to encode the magnitude responses and the GLBP to encode the phase responses of a small bank of Gabor filters. All the obtained encoded patterns were partitioned in square patches. The distributions of features were locally displayed in a histogram form. All the obtained histograms were concatenated to build a long feature vector. We employed a symmetric I-divergence measure to compare feature vectors.

The implemented method was demonstrated on two different datasets having 48 face classes (a total of 1150 face images): PRE-TINDERS and TINDERS. The probe SWIR images were collected at three operational distances: 2, 50 and 106 meters. We evaluated the verification and identification performance of the implemented method and we compared them with the performance of Faceit G8, which is often considered state of the art for facial recognition. We showed that: (1) the VIS-SWIR comparison provides lower recognition rates with respect to the case when images in the same spectral band (VIS or SWIR) are compared; (2) the implemented matching method outperforms in terms of verification and identification performance the commercial software in the VIS-SWIR comparison at three operational distances; (3) the proposed encoding method provides best recognition performance when a relatively small number of Gabor filters is employed; (4) overall face recognition rates of all considered methods decrease as the distance of probe SWIR images increases.

In Chapter 5, we tested the matching scheme introduced in Chapter 4 on visible face data from FRGC 2.0 dataset. The method, that employs Gabor filters and the three multi-scale local operators (SWLD, LBP, GLBP), was applied to 1000 face images acquired in a controlled scenario. We also adopted the three recognition methods that employed the individual local operators (SWLD, LBP, GLBP) to evaluate the recognition performance. We showed that adoption of a single local operator is sufficient to obtain high recognition rates (up to 99.6%) when tested on visible data. Particularly, it was demonstrated that the SWLD operator applied to the magnitude response of Gabor filters is a good scheme when employed in a controlled scenario. We further compared the performance of the implemented method with the commercial software Faceit G8.

In Chapter 6, we presented a fully automatic face recognition software that detects and recognizes individuals in an uncontrolled scenario. The matching algorithm employs a simplified and optimized version of the scheme that encode magnitude response of a bank of Gabor filters with SWLD operator. We demonstrated the implemented algorithm in a wireless camera network that acquires visible light images in near real time and sends the detected face images to a fusion center that performs recognition. To acquire the face images we followed the network design of the work by Kulathumani et al. [3]. In our implementation, we further introduced an eye detection stage based on Haar wavelets and Adaboost classifiers [4] with respect to [3]. The faces of the individuals, walking in the network, were captured at different distances from the cameras and have low resolutions (minimum 28 × 28 pixels). A multi-resolution handling method for feature vectors was introduced; specifically, the obtained face images were normalized to a standard resolution chosen from a pool of four resolutions. The normalization process was performed by employing the eye coordinates. This approach has the advantage to reduce the processing time when probe images have small resolutions. Particularly, low resolution images are processed much faster compared to high resolution images. This solution also avoids excessive up-sampling to a given resolution when captured images have low resolution. We evaluated the performance of the implemented system on a dataset of 29 subjects collected at WVU. In the experiments we evaluated: (1) the verification performance (ROC curves) by considering all the probe images that were collected for the 29 subjects; (2) the performance after employing a score fusion strategy that selected for each walking subject the probe image that yielded the best score with respect to each gallery image. We showed that in both scenarios, the recognition performance of the proposed method was similar or better than the one of the commercial software Faceit G8 and of other approaches that employ local descriptors. Finally, we showed that the implemented matching method is four times faster than Faceit G8 at the lowest resolution (28 × 28) supported by the system.

7.2 Future Work

The implemented matching methods were designed to work only on near frontal images (up to 20 degree out-of-plane rotation with respect to the frontal view), however the methods can be extended to non frontal images in different ways. For example the non frontal images can be deformed to a standard frontal view by employing an Active Shape Model [15], which requires training images to define a shape model. The method would find the best shape model for different given poses. Once the model is found, the image could be deformed to a frontal view. After this normalization process, we could employ the designed matching methods without modifying the implementation.

Alternatively, a possible approach to non frontal recognition is to employ a series of classifiers that have gallery images at different view angles. In the training stage, templates at different view angles could be created for the same subject. Particularly, the implemented matching methods could be employed on images having different pose, by detecting a few reference landmarks in the face (eyes, nose, mouth, etc.) and by employing a normalization scheme. In the testing stage, we would estimate the pose of the probe image and extract features from the image. A face classifier would be selected based on the estimated pose. Then, the chosen classifier would be employed for matching. A precise estimation of the pose (e.g. yaw angle) is not necessary if several templates having different view angles are present in the database. When the probe images would be submitted for recognition, we could employ a small subset of classifiers trained in a certain angular range that includes the estimated pose. The subset of classifiers would produce a set of matching scores. Finally, we could select the best matching score across different views and subjects.

The main advantage of the implemented encoding methods is that they don't require training or labeled data to produce a recognition result. Initially, we have designed the Gabor filters to have certain parameters by employing only few face images. Then, we employed three local operators (SWLD, LBP, GLBP) to encode the magnitude and phase response of Gabor filters applied to face data. Hence, we have shown that the proposed cross-spectral matching scheme outperforms a state of the art recognition software (Faceit G8) when SWIR images are compared to visible light data. Although the proposed method provides larger recognition rates than Faceit G8, we could further improve the recognition performance if a certain amount of SWIR and visible data would be available to train the face matcher. For example, we could assign larger weights to more informative features during the matching process. The feature vectors of encoded images could be divided into well defined segments. Then, we could evaluate a recognition performance index (e.g. Equal Error Rate or others) on the given training data by employing feature vectors that are composed only of an individual segment. A set of performance indexes would be obtained by adopting all available segments for recognition. Hence, the weights could be chosen to be proportional to the obtained indexes. Then, we would employ the complete (all segments) feature vectors to test a probe image. During the testing stage, the obtained weights would assign different importance to each segment. The described method could be applied for both verification and identification scenarios, without making any assumptions about the specific face class that is being tested. Other well-known boosting methods (e.g. Adaboost) could also be applied to increase the recognition performance of the implemented matching method.

In this work, we implemented a fully automatic system to compare low resolution images in a camera network. We demonstrated the approach by employing only visible data. However, the implemented sofware could be employed with few modifications (face and eye detector only) in a camera network that acquires SWIR probe images.

Appendix A

A Method for Ranking Quality Metrics

In this appendix, we briefly summarize a method for evaluating and ranking various quality metrics applied to biometric images, based on their ability to predict recognition performance of a biometric recognition system.

A.1 Method Description

Consider a dataset Θ of N biometrics samples. A quality assessment algorithm Ψ accepts as input a biometric sample (signal or image) A and produces, in general, a vector $\mathbf{Q_A} = \Psi(\theta_A)$; similarly a distinct sample B is mapped to $\mathbf{Q_B} = \Psi(\theta_B)$. Consider also a matching algorithm that takes the samples pair (A, B) and generates a matching score MS_{AB}, we theorize there exists a non linear function f that relates the quality measures $\mathbf{Q_A}$ and $\mathbf{Q_B}$ to the obtained matching score MS_{AB}.

Formally we have that:

$$MS_{AB} = f(\mathbf{Q_A}, \mathbf{Q_B}). \tag{A.1}$$

The function f is unknown and must be estimated. In general since a physical model describing such relationship cannot be established, we then appeal to a ”black box” approach. Such an approach does not assume any particular physical model or mathematical relation-

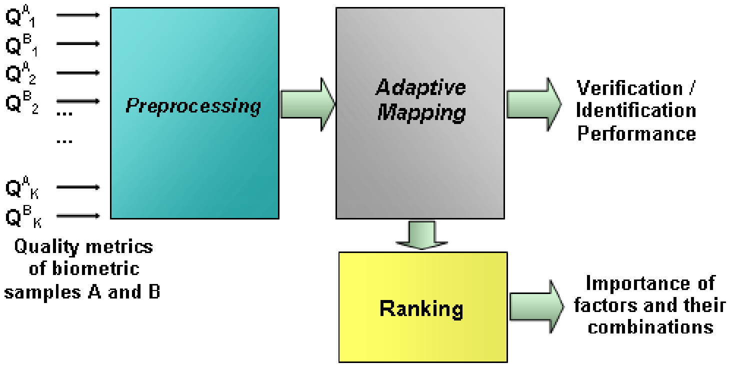

Q^A_1
Q^B_1
Q^A_2
Q^B_2
...
...
Q^A_K
Q^B_K
Quality metrics of biometric samples A and B
Preprocessing
Adaptive Mapping
Verification / Identification Performance
Ranking
Importance of factors and their combinations

Figure A.1: General block diagram of proposed method for given matching algorithm and quality metrics

ship between input and output. Hence, we can estimate the function f using a Feed Forward Neural Network (FFNN) or other non linear regression methods (e.g. MARS, RBF, others). We also distinguish the case when samples A, B belong to same individual (genuine pair) and the case when A, B belong to distinct individuals (imposter pair). Then, for each of these two cases a different mapping f exists. This method is deployed as a Matlab software program divided into three operational modules. The preprocessing module (1) that has as input quality scores pairs and outputs relative quality metrics or augmented quality metrics. The adaptive mapping module (2) that estimates the unknown relationship between quality metrics and matching scores. The ranking module (3) that provides the best combination of quality measures. Fig. A.1 provides a general block diagram that summarizes the modules' functions and interconnections.

A.1.1 Preprocessing

The preprocessing block introduces the idea that quality metrics of a pair of biometric samples A, B may also be described from a single relative quality vector $\mathbf{Q_{AB}}$ rather than a pair $\mathbf{Q_A}$, $\mathbf{Q_B}$. Let K be the dimensionality of a quality vector $\mathbf{Q_A}$ of sample A and $\mathbf{Q_B}$ of a sample B; each component of such a vector represents a distinct and well defined quality measurement. Then the preprocessing block provides three different operational modes:

1. Output of the preprocessing block is the vector $[\mathbf{Q_A}, \mathbf{Q_B}]$ obtained by concatenating

vector pairs $\mathbf{Q_A}, \mathbf{Q_B}$. The dimensionality of the output vector is 2K.

2. Output of the preprocessing block is a relative quality vector $\mathbf{Q_{AB}}$ of dimensionality K. The i-th component $\mathbf{Q_{AB}}$ is a function of the i-th component of quality metrics $\mathbf{Q_A}$ and i-th component quality metrics $\mathbf{Q_B}$.

3. Output of the preprocessing block is the vector $[\mathbf{Q_A}, \mathbf{Q_B}, \mathbf{Q_{AB}}]$ obtained by concatenating vector triplets $\mathbf{Q_A}, \mathbf{Q_B}, \mathbf{Q_{AB}}$. The dimensionality of the output vector is 3K.

In mode (2) a predefined transformation g relates quality metrics of pair samples A, B having the same vector index. For simplicity of implementation, we apply the same function g to each component of the considered pair. If $\mathbf{Q}_A = [Q_{A,1}, \ldots, Q_{A,K}]^T$ and $\mathbf{Q}_B = [Q_{B,1}, \ldots, Q_{B,K}]^T$, then formally we have:

$$\mathbf{Q_{AB}} = g(\mathbf{Q_A}, \mathbf{Q_B}) = [g(Q_{A,1}, Q_{B,1}), \ldots, g(Q_{A,K}, Q_{B,K})]^T. \tag{A.2}$$

The choice of g is heuristic: different candidate functions are tested and applied to same and different biometric datasets and quality metrics. We have explored functions such as $min(\mathbf{Q_A}, \mathbf{Q_B}), |\mathbf{Q_A} - \mathbf{Q_B}|, \mathbf{Q_A} + \mathbf{Q_B}, \mathbf{Q_A} \times \mathbf{Q_B}$ and compositions of these functions with other non linear mappings. Others functions can be tested but we have found that a non linear function of $|\mathbf{Q_A} - \mathbf{Q_B}|$ consistently improves prediction performance. In particular we have identified the two non linear functions of $|\mathbf{Q_A} - \mathbf{Q_B}|$:

$$\mathbf{Q_{AB}} = \tanh\left[\pi \frac{|\mathbf{Q_A} - \mathbf{Q_B}|}{\mathbf{Q}_{\max} - \mathbf{Q}_{\min}}\right] \tag{A.3}$$

$$\mathbf{Q_{AB}} = \frac{2}{1 + \exp\left[-\alpha \frac{|\mathbf{Q_A} - \mathbf{Q_B}|}{\mathbf{Q}_{\max} - \mathbf{Q}_{\min}}\right]} - \frac{1}{2} \tag{A.4}$$

where $\mathbf{Q}_{\max} - \mathbf{Q}_{\min}$ is a vector composed of ranges of individual quality metrics. Since quality measurements are assumed to be positive or null, we have that all components of $\frac{|\mathbf{Q_A} - \mathbf{Q_B}|}{\mathbf{Q}_{\max} - \mathbf{Q}_{\min}}$ are between 0 and 1. The final relative quality measurement $\mathbf{Q_{AB}}$, in both cases, is also in the interval [0,1].

For all three considered modes, the output of the preprocessing block will feed the adaptive mapping module.

A.1.2 Adaptive Mapping Module

The adaptive mapping module mathematically provides a solution to a multivariate regression problem. A general regression model can be described as

$$\hat{Y} = f(\mathbf{X}) \tag{A.5}$$

where $\hat{Y}$ is an estimation of the unknown output Y, in our case the matching score MS_{AB}, and $\mathbf{X}$ is a vector of predictive variables, such as pairs of quality metrics for biometric samples A and B.

We employ Feed Forward Neural Network (FFNN) to estimate $\hat{Y}$. To establish topology of the network we adopt a classic trial and error approach. From the analysis on different biometric data we have concluded that one hidden layer is sufficient to describe the non linear relationship between quality scores and matching scores. The number of necessary neurons to characterize the mapping depends on specific biometric data and is also different for genuine and imposter cases. During the training stage, an ensemble of neural net models are created by repeating the training process H times. Each model is obtained by adopting a random subset (a fixed percent L) of available data for learning and the remaining data ($V = 1 - L$ percent) for validation. We employ a ratio L/V equal to 60/40 . The final non linear model is chosen as a weighted average of H individual functions applied to testing data. This procedure improves validation results for all considered biometric data. The parameter k is optimized repeating the procedure several times and the H averaging weights are chosen as follows:

$$w_i = \frac{R_i^2}{\sum\limits_{j=1}^{H} R_j^2} \tag{A.6}$$

where R_i^2 is the squared correlation between predicted values and actual values in the training set and $\sum\limits_{i=1}^{k} w_i = 1$.

A.1.3 Performance Measures

We introduce three measures to evaluate goodness of fit between the measured matching scores and predicted matching scores. In particular the considered criteria are: squared Pearson's correlation (R^2), Mean Square Error (MSE) and F statistics (F). The employed measures are quite intuitive and easy to calculate; they are used only as relative measures (for ranking purposes) rather than absolute indication of fit (other measures can be employed in that case). Consider the i-th measured matching score y_i in the testing set and the corresponding predicted value $\hat{y}_i$, then the squared Pearson's correlation is defined as:

$$R^2 = \left(\frac{\sum_{i=1}^{N} (y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{(N-1) s_y s_{\hat{y}}} \right)^2 \tag{A.7}$$

where N is number of matching scores, $\overline{y}$ and $\overline{\hat{y}}$ are respectively sample mean of measured and predicted matching scores, s_y and $s_{\hat{y}}$ are sample variance of measured and predicted matching scores. Statistically, R^2 represents the part (real number between 0 and 1) of variance in the matching score variability that can be explained by the predictive model.

Mean Square Error is a classical measure of performance defined as:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{A.8}$$

As a last measurement, we introduce the F-statistic defined as the ratio between the mean square error of the linear model and mean square error of its prediction:

$$F = \frac{\sum\limits_{i=1}^{N} (\hat{y}_i^L - \bar{\hat{y}}^L)^2}{\sum\limits_{i=1}^{N} (\hat{y}_i^L - \hat{y}_i)^2/(N-2)} \tag{A.9}$$

where $\bar{\hat{y}}^L$ is mean value of all points $\hat{y}_i^L$ of regression line in the scatter plot $(y_i, \hat{y}_i)$;

A.1.4 Ranking of Quality Factors

The ranking block described in this appendix, given a specified number of quality metrics, performs an exhaustive search of best possible combination of quality measures based on the statistical measures introduced in the previous section. If K is the number of quality metrics involved, the possible number of combinations is $2^K - 1$. If K is a large number, a more efficient algorithm (e.g. tree algorithm) can be adopted to rank combination of quality measures.

A.2 Example of Ranking Table

In the experiments we adopt the commercial face recognition software FaceIt G6 [80]. The software deployed in a SDK (Software Development Kit) form provides three operational modules. The enrollment module, matching module and quality module. The enrollment module creates templates from images and stores it in the mass memory. The matching module for each considered biometric sample pair produces a non negative similarity score. The quality module instead provides ten quality metrics and an overall quality measure that ranges from 0 to 10. Each metric is characterized by a real number with two decimal places.

The quality factors included are:

1. Darkness: Used to determine if there are any underexposed parts of the subject's face.

2. Brightness: Used to determine if there are any overexposed parts of the subject's face.

3. Exposure: Measures the degree of overexposure and underexposure on the subject's face. 0-10 are quality measures, with 10 indicating the highest quality.

4. Focus: Used to determine if the image is out of focus.

5. Resolution: Used to determine whether the head is sufficiently large. With a live subject, the software should be able to obtain a perfect 10.

6. Cropping: Used to ensure that the entire head is safely within the picture. If the head is touching the boundary or partly outside the boundary (cropped), the result is 0. The score improves as the subject's head moves away from the boundary.

7. Glare: Used to determine if there is adverse glare, usually in the subject's glasses. The possible values are between 0 and 10. A 10 indicates that there was no glare found,

meaning this is a higher quality image for face recognition. 0 indicates that there was glare found in the image.

8. Faceness: Measures the likelihood that a face is unobstructed and suitable for face recognition. An obscured face will have a low quality score and will likely degrade face recognition performance. Non-faces will give a low Faceness score. 0-10 are quality measures, with 10 indicating the highest quality.

9. Contrast: Measures the contrast in the face region of the image. A poor contrast value indicates that the face may lack detail from too little or too much contrast in the image. 0-10 are quality measures, with 10 indicating the highest quality.

10. Texture: Measures the effective resolution of the subject face for use with high-resolution face recognition algorithms. Low resolution, poor focus, or over-compression will lower the score. 0-10 are quality measures, with 10 indicating the highest quality.

11. Overall: Overall quality factors obtained as minimum of a selected combination of previous quality factors. Overall quality score close to 0 indicates bad quality image and a quality score close to 10 indicates a very good quality image.

For the following example we employ a WVU Face dataset composed of 1,745 face images for a total of 270 biometric classes. We apply the Faceit G6 matching algorithm to each image pair resulting in 6,074 genuine comparisons and 1,515,566 imposter scores. We consider all genuine scores and a representative subset of 109,000 imposter scores to evaluate the prediction performance of the face quality metric. Then we apply the Faceit G6 quality module to produce eleven quality scores for each image. A random sample composed of half of the genuine scores is chosen as training set and remaining half as testing set. In the imposter case, we use 9,000 scores for training purpose and remaining 100,000 scores for testing. Table A.1 reports the prediction performance of a single hidden layer FFNN composed of 10 neurons applied to genuine testing data when the third preprocessing scheme is employed (best scenario). Shown in Table A.1 are the performance measures obtained for all individual factors and some combinations of interest. Table A.1 indicates the best individual factors are 3 (Exposure) and 5 (Resolution); the table also shows that the best predictive combination is obtained when factors 1 through 10 are involved. Fig.A.2 instead shows the scatter plot of predicted imposter scores versus measured imposter scores for the combination 1 through 9 which is the best in the imposter case. In Fig. A.2, we observe that the scores are compacted in the space around the origin and the range of predicted values is very limited with respect to the measured matching scores, indicating that the plots for the imposter case are not very informative. We note, instead the scatter plot in Fig. A.3, relating predicted genuine scores to actual genuine scores, can be employed to evaluate goodness of fit for a given combination of quality factors.

Factors	R^2	MSE	F
1	0.201596	505.888	766.33
2	0.111198	563.291	379.71
3	0.274761	457.519	1149.83
4	0.104965	566.480	355.93
5	0.229737	489.123	905.21
6	0.000514	678.816	1.56
7	0.002136	630.910	6.50
8	0.136530	550.393	479.89
9	0.050384	601.671	161.03
10	0.025500	616.821	79.42
11	0.193296	513.334	727.22
1,2	0.316748	429.473	1406.99
1,3,8	0.395692	381.079	1987.28
3,4,5,6,7,8	0.526292	300.131	3371.91
1,2,3,4,5,6,7,8,9	0.553939	281.349	3769.00
1,2,3,4,5,6,7,8,9,10	0.554554	283.189	3778.40

Table A.1: Prediction performance for genuine face matching scores.

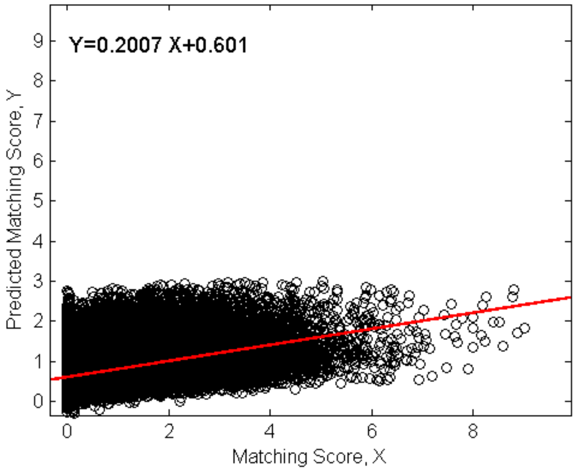


Figure A.2: Scatter plot of predicted face matching scores versus actual matching scores in the imposter case. The results are obtained with a single hidden layer FFNN with 12 neurons and involving the best combination of quality factors for the imposter case (1 through 9). Equation of regression line is also provided.

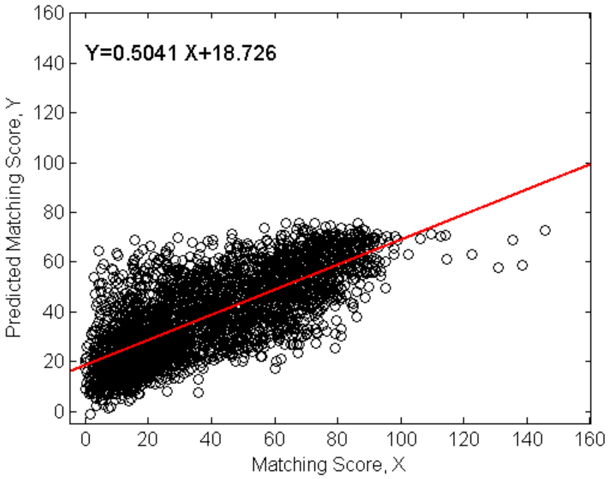


Figure A.3: Scatter plot of predicted face matching scores versus actual matching scores in the genuine case. The results are obtained with a single hidden layer FFNN with 10 neurons and involving the best combination of quality factors for the genuine case (1 through 10). Equation of regression line is also provided.

References

[1] West Virginia High Technology Consortium Foundation, "Tactical imager for night/day extended-range surveillance," `http://www.wvhtf.org/departments/advanced_tech/projects/tinders.asp`, accessed on Nov. 10, 2011.

[2] A.K. Jain, R.M. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[3] V. Kulathumani, S. Parupati, A. Ross, and R. Jillela, "Collaborative face recognition using a network of embedded cameras," in *Distributed Video Sensor Networks*, pp. 373–389. Springer, 2011.

[4] P.A. Viola and M.J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.

[5] M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[6] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103–108, Jan. 1990.

[7] P.N. Belhumeur, J. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," in *Proc. of European Conference on Computer Vision*. 1996, vol. 1064 of *Lecture Notes in Computer Science*, pp. 45–58, Springer.

[8] M. Bartlett and T. Sejnowski., "Independent components of face images: A representation for face recognition," in *Proc. of the 4th Annual Joint Symposium on Neural Computation*, May 1997.

[9] M.J. Er, S. Wu, J. Lu, and H.L. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 697–710, May 2002.

[10] R. Bhati, "Face recognition system using multi layer feed forward neural networks and principal component analysis with variable learning rate," in *Proc. of IEEE International Conference on Communication Control and Computing Technologies*, Oct. 2010, pp. 719–724.

[11] K.I. Kim, K. Jung, and H.J. Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 40–42, Feb. 2002.

[12] AT&T Laboratories Cambridge, "AT&T face database," `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`, accessed on Nov. 10, 2011.

[13] L. Wiskott, J. Fellous, N. Krger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779, 1997.

[14] J.G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A*, vol. 2, no. 7, pp. 1160–1169, July 1985.

[15] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, July 1997.

[16] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, Jan. 1995.

[17] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face recognition using active appearance models," in *Proc. of European Conference on Computer Vision*. 1998, vol. 1407 of *Lecture Notes in Computer Science*, pp. 581–595, Springer.

[18] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. of European Conference on Computer Vision*. 2004, vol. 3021 of *Lecture Notes in Computer Science*, pp. 469–481, Springer.

[19] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[20] G. Zhang, X. Huang, S. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (LBP)-based face recognition," in *Advances in Biometric Person Authentication*, S. Li, J. Lai, T. Tan, G. Feng, and Y. Wang, Eds., vol. 3338 of *Lecture Notes in Computer Science*, pp. 179–186. Springer, 2005.

[21] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. of IEEE International Conference on Computer Vision*, 2005, vol. 1, pp. 786–791.

[22] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[23] M. Aly, "Face recognition using SIFT features," `http://www.vision.caltech.edu/malaa/publications/aly06face.pdf`, accessed on Nov. 10, 2011.

[24] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, pp. 91–110, Nov. 2004.

[25] X. Tan and B. Triggs, "Fusing gabor and LBP feature sets for kernel-based face recognition," in *Proc. of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2007, AMFG'07, pp. 235–249.

[26] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 947–954.

[27] M. Savvides and B.V.K.V. Kumar, "Efficient design of advanced correlation filters for robust distortion-tolerant face recognition," in *Proc. of the IEEE Conference on Advanced Video and Signal Based Surveillance*, 2003, pp. 45–52.

[28] P. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and horner efficiency," *Optics Letters*, vol. 16, no. 11, pp. 829–831, June 1991.

[29] C. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," in *Advances in Biometrics*, S. Lee and S. Li, Eds., vol. 4642 of *Lecture Notes in Computer Science*, pp. 809–818. Springer, 2007.

[30] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 57–68, 2007.

[31] B. Cao, S. Shan, X. Zhang, and W. Gao, "Baseline evaluations on the CAS-PEAL-R1 face database," in *Proc. of Advances in Biometric Person Authentication, 5th Chinese Conference on Biometric Recognition*, S.Z. Li, J. Lai, T. Tan, G. Feng, and Y. Wang, Eds. 2004, vol. 3338 of *Lecture Notes in Computer Science*, pp. 370–378, Springer.

[32] Z. Lei, S. Liao, M. Pietikäinen, and S.Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 247–256, 2011.

[33] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of gabor magnitude and phase for face recognition," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1349–1361, 2010.

[34] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, June 2010.

[35] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, June 2001.

[36] X. Chen, P.J. Flynn, and K.W. Bowyer, "IR and visible light face recognition," 2005, vol. 99, pp. 332–358.

[37] M. Turk and P. Alex, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[38] Z. Pan, G. Healey, M. Prasad, and B.J. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003.

[39] S.G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B.R. Abidi, A. Koschan, M. Yi, and M.A. Abidi, "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," *International Journal of Computer Vision*, vol. 72, no. 2, pp. 215–233, 2007.

[40] S.Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 627–639, 2007.

[41] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, and M. Abidi, "Fusion of visible and infrared images using empirical mode decomposition to improve face recognition," in *Proc. of IEEE International Conference on Image Processing*, Oct. 2006, pp. 2049–2052.

[42] M. Akhloufi and A.H. Bendada, "Multispectral infrared face recognition: a comparative study," in *Proc. of Quantitative InfraRed Thermography*, July 2010, number 3.

[43] M. Akhloufi and A.H. Bendada, "A new fusion framework for multispectral IR face recognition in the texture space," in *Proc. of Quantitative InfraRed Thermography*, July 2010, number 2.

[44] Equinox Corporation, "Human identification at a distance (HID)," `http://www.equinoxsensors.com/products/HID.html`, accessed on Nov. 10, 2011.

[45] M.A. Akhloufi, A. Bendada, and J.C. Batsale, "Multispectral face recognition using non linear dimensionality reduction," in *Proc. of SPIE Visual Information Processing XVIII*, 2009, vol. 7341.

[46] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. of European Conference on Computer Vision*. 2006, vol. 3954 of *Lecture Notes in Computer Science*, pp. 13–26, Springer.

[47] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[48] J. Yang, A.F. Frangi, J. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[49] D. Yi, R. Liu, R. Chu, Z. Lei, and S.Z. Li, "Face matching between near infrared and visible light images," in *Proc. of IAPR/IEEE International Conference on Biometrics*, 2007, pp. 523–530.

[50] S. Liao, D. Yi, Z. Lei, R. Qin, and S.Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Proc. of IAPR/IEEE International Conference on Biometrics*, 2009, pp. 209–218.

[51] Z. Lei and S.Z. Li, "Coupled spectral regression for matching heterogeneous faces," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1123–1128.

[52] D. Yi, S. Liao, Z. Lei, J. Sang, and S.Z. Li, "Partial face matching between near infrared and visual images in MBGC portal challenge," in *Proc. of IAPR/IEEE International Conference on Biometrics*, 2009, pp. 733–742.

[53] National Institute of Standards and Technology, "Multiple biometric grand challenge (MBGC) 2008," `http://www.nist.gov/itl/iad/ig/mbgc.cfm`, accessed on Nov. 10, 2011.

[54] R. Wang, J. Yang, D. Yi, and S. Li, "An analysis-by-synthesis method for heterogeneous face biometrics," in *Advances in Biometrics*, vol. 5558 of *Lecture Notes in Computer Science*, pp. 319–326. Springer, 2009.

[55] J. Chen, D. Yi, J. Yang, G. Zhao, S.Z. Li, and M. Pietikäinen, "Learning mappings for face synthesis from near infrared to visual light images," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 156–163.

[56] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proc. of IEEE International Conference on Computer Vision*, Oct. 2003, vol. 1, pp. 687–694.

[57] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2005, vol. 1, pp. 1005–1010.

[58] B. Klare and A.K. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," in *Proc. of International Conference on Pattern Recognition*, 2010, pp. 1513–1516.

[59] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 886–893.

[60] Cognitec Systems, "FaceVACS software developer kit," `http://www.cognitec-systems.de`, accessed on Nov. 10, 2011.

[61] Z. Zhang, Y. Wang, and Z. Zhang, "Face synthesis from near-infrared to visual light via sparse representation," in *Proc. of International Joint Conference on Biometrics*, Oct. 2011, pp. 1–6.

[62] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 531–543, Apr. 2007.

[63] J. Fiérrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A.K. Jain, "Incorporating image quality in multi-algorithm fingerprint verification," in *Proc. of International Conference on Biometrics*, 2006, pp. 213–220.

[64] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[65] K. Nandakumar, Y. Chen, A.K. Jain, and S. Dass, "Quality-based score level fusion in multibiometric systems," in *Proc. of International Conference on Pattern Recognition*, Aug. 2006, vol. 4, pp. 473–476.

[66] K. Nandakumar, Y. Chen, S. C. Dass, and A. Jain, "Likelihood ratio-based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 342–347, 2008.

[67] K. Kryszczuk and A. Drygajlo, "Improving classification with class-independent quality measures: *Q-stack* in face verification," in *Proc. of International Conference on Biometrics*, 2007, pp. 1124–1133.

[68] K. Kryszczuk and A. Drygajlo, "Improving biometric verification with class-independent quality information," *Signal Processing, IET*, vol. 3, no. 4, pp. 310–321, July 2009.

[69] D.E. Maurer and J.P. Baker, "Fusing multimodal biometrics with quality estimates via a bayesian belief network," *Pattern Recognition*, vol. 41, no. 3, pp. 821–832, 2008.

[70] H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Fingerprint image-quality estimation and its application to multialgorithm verification," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 331–338, 2008.

[71] N. Ozay, Y. Tong, F.W. Wheeler, and X. Liu, "Improving face recognition with a quality-based probabilistic framework," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2009, pp. 134–141.

[72] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 947–954.

[73] X. Gao, S.Z. Li, R. Liu, and P. Zhang, "Standardization of face image sample quality," in *Proc. of International Conference on Biometrics*, 2007, pp. 242–251.

[74] J. Zuo and N.A. Schmid, "Global and local quality measures for NIR iris video," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami, FL, June 2009, pp. 1–6.

[75] C. Shan, S. Gong, and P.W. McOwan, "Robust facial expression recognition using local binary patterns," in *Proc. of IEEE International Conference on Image Processing*, 2005, pp. 370–373.

[76] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, July 1989.

[77] E.K. Blum and L.K. Li, "Approximation theory and feedforward networks," *Neural Networks*, vol. 4, pp. 511–515, July 1991.

[78] V. Kůrková, "Kolmogorov's theorem and multilayer neural networks," *Neural Networks*, vol. 5, pp. 501–506, Mar. 1992.

[79] GoodRich Corporation, "Surveillance using SWIR night vision cameras," `http://www.sensorsinc.com/facilitysecurity.html`, accessed on Nov. 10, 2011.

[80] L-1 Identity Solutions, Inc., "Faceit SDK," `http://www.l1id.com/pages/101-faceit-sdk`, accessed on Nov. 10, 2011.

[81] A.A. Efros and T.K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. of IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1033–1038.

[82] E.P. Simoncelli and W.T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *Proc. of IEEE International Conference on Image Processing*, 1995, vol. 3, pp. 444–447.

[83] K. Zuiderveld, *Contrast Limited Adaptive Histogram Equalization*, pp. 474–485, Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[84] H. Wang, S.Z. Li, and Y. Wang, "Face recognition under varying lighting conditions using self quotient image," in *Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 819–824.

[85] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. of the Royal Society of London. Series B, Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[86] J. Zuo, F. Nicolo, and N.A. Schmid, "Cross spectral iris matching based on predictive image mapping," in *Proc. of IEEE International Conference on Biometrics: Theory Applications and Systems*, Sep. 2010, pp. 1–5.

[87] P. Kovesi, "Phase congruency detects corners and edges," in *Proc. of the Austrialian Pattern Recognition Society Conference: DICTA*, 2003, pp. 309–318.

[88] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.

[89] M. Topi, O. Timo, P. Matti, and S. Maricor, "Robust texture classification by subsets of local binary patterns," in *Proc. of International Conference on Pattern Recognition*, 2000, vol. 3, pp. 935–938.

[90] T. Ojala, M. Pietikäinen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, July 2002.

[91] J. Chen, G. Zhao, and M. Pietikäinen, "An improved local descriptor and threshold learning for unsupervised dynamic texture segmentation," in *Proc. of IEEE International Conference on Computer Vision Workshops*, 2009, pp. 460–467.

[92] Y. Guo and Z. Xu, "Local gabor phase difference pattern for face recognition," in *Proc. of International Conference on Pattern Recognition*, 2008, pp. 1–4.

[93] I. Csiszar, "I-divergence geometry of probability distributions and minimization problems," *The Annals of Probability*, vol. 3, no. 1, pp. 146–158, 1975.

[94] A. Ross, A.K. Jain, and J. Qian, "Information fusion in biometrics," in *Proc. of International Conference on Audio- and Video-Based Biometric Person Authentication*, 2001, AVBPA '01, pp. 354–359.

[95] R. Cappelli, D. Maio, and D. Maltoni, "Combining fingerprint classifiers," in *Proc. of the First International Workshop on Multiple Classifier Systems*. 2000, MCS '00, pp. 351–361, Springer.

[96] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, pp. 2270–2285, Dec. 2005.

[97] B.C. Lovell, S. Chen, A. Bigdeli, E. Berglund, and C. Sanderson, "On intelligent surveillance systems and face recognition for mass transport security," in *International Conference on Control, Automation, Robotics and Vision*, Dec. 2008, pp. 713–718.

[98] R. Gross, "CMU PIE Database," `http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261`, accessed on Feb. 10, 2012.

[99] Xi'an Jiaotong University, "Oriental Face Database," `http://www.aiar.xjtu.edu.cn/groups/face/English/HomePage.htm`, accessed on Feb. 10, 2012.

[100] D.R. Kisku, H. Mehrotra, P. Gupta, and J.K. Sing, "Robust multi-camera view face recognition," in *Computing Research Repository*, 2010.

[101] D.B. Graham and N.M. Allinson, *Characterizing Virtual Eigensignatures for General Purpose Face Recognition*, vol. 163 of *Face Recognition: From Theory to to Applications, NATO ASI Series F, Computer and Systems Sciences*, pp. 446–456, F. Fogelman-Soulie, and T. S. Huang, 1998.

[102] F.W. Wheeler, R.L. Weiss, and P.H. Tu, "Face recognition at a distance system for surveillance applications," in *Proc. of IEEE International Conference on Biometrics: Theory Applications and Systems*, Sep. 2010, pp. 1–8.

[103] PittPat, "PittPatt SDK v5.2.2 Documentation," `http://www.pittpatt.com/documentation/`, accessed on Feb. 10, 2012.

[104] E. Kokiopoulou and P. Frossard, "Distributed classification of multiple observation sets by consensus," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 104–114, Jan. 2011.

[105] M. Banduml, K. Bernardin, M. Fischer, H.K. Ekenel, and R. Stiefelhagen, "Multi-pose face recognition for person retrieval in camera networks," in *Proc. of IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010, pp. 441–447.

[106] Logitech USA, "Logitech QuickCam Pro 9000," `http://www.logitech.com/en-us/webcam_communications/webcams/devices/6333`, accessed on Nov. 10, 2011.

[107] BeagleBoard, "Beagle Board System reference Manual Revision C 3.0 2009," `http://www.beagleboard.org`, accessed on Nov. 10, 2011.

[108] Sourceforge, "OpenCV library," `http://sourceforge.net/projects/opencvlibrary`, accessed on Nov. 10, 2011.

[109] U. Park, A. Jain, and A. Ross, "Face Recognition in Video: Adaptive Fusion of Multiple Matchers," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[110] F. Nicolo and N. Schmid, "A method for robust multispectral face recognition," in *Proc. of Image Analysis and Recognition*. 2011, vol. 6754 of *Lecture Notes in Computer Science*, pp. 180–190, Springer.